Making the Black Box Transparent: State of the Art in Explainable Machine Learning for Structural Design and Assessment

Mohsen Zaker Esteghamati[1], Jingcheng Wang[2], Xiaowei Wang[3], Stephanie G. Paal[4], Jacob Murphy[5], Ali Namin[6], Abdullahi Salman[7], Adam Ado Sabari[8], Muhammad Ahsan Ibrar[9], Ram Mazumder[10], Yue Li[11], Abdollah Shafieezadeh[12]

[1] Assistant Professor, Department of Civil and Environmental Engineering, Utah State University, Logan, UT,84322, Email: mohsen.zaker@usu.edu **(Corresponding Author,** *Address: 4110 Old Main Hill, Logan, UT, 84322***)**
[2] Associate Professor, College of Civil Engineering, Fuzhou University, Fuzhou 350108, China, Email: jingchengwang@fzu.edu.cn
[3] Associate Professor, Department of Bridge Engineering, Tongji University, Shanghai 200092, China, Email: xiaoweiwang@tongji.edu.cn
[4] Williams Brothers Construction Company Associate Professor, Zachry Department of Civil & Environmental Engineering. Texas A&M University, College Station, TX, 77840, Email: spaal@civil.tamu.edu
[5]Graduate Student, Zachry Department of Civil & Environmental Engineering. Texas A&M University, College Station, TX, 77840, Email: jacob70920@tamu.edu
[6] Graduate Student Department of Civil and Environmental Engineering, Utah State University, Logan, UT, 84322, Email: ali.namin@usu.edu
[7] Associate Professor, Department of Civil and Environmental Engineering, University of Alabama in Huntsville, Huntsville, AL 35899. Email: ams0098@uah.edu
[8] Graduate Student, Department of Civil and Environmental Engineering, University of Alabama in Huntsville, Huntsville, AL 35899. Email: aas0041@uah.edu
[9] Graduate Student, Department of Civil and Environmental Engineering, University of Alabama in Huntsville, Huntsville, AL 35899. Email: mi0025@uah.edu
[10] Assistant Professor, Department of Civil Engineering & Construction, Georgia Southern University, Statesboro, GA 30458. Email: rmazumder@georgiasouthern.edu
[11] Leonard Case Jr. Professor in Engineering, Department of Civil and Environmental Engineering, Case Western Reserve University, Cleveland, OH 44106. Email: yxl1566@case.edu
[12] Lichtenstein Professor, Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH 43210. Email: shafieezadeh.1@osu.edu

Abstract

Machine learning (ML)-based solutions have gained traction in various structural engineering applications, from structural design to assessment and monitoring. Nevertheless, the black-box nature of advanced ML models, and the resultant limited interpretation and transparency, are among the primary barriers to their broader adoption and implementation in the field. eXplainable ML (XML) is an interdisciplinary field that improves the understanding of the ML model performance. Despite the potential of XML to increase the accessibility of ML, the scattered available literature and the lack of a domain-specific holistic review of XML have created a significant knowledge gap for its application in structural engineering. Therefore, this paper presents a targeted review of the XML definition, nomenclature and taxonomy, frequently used algorithms, and domain-specific literature. Additionally, three case studies are presented to illustrate different classes of XML algorithms and their implementation on diverse structural engineering problems at the component-, structure-, and inventory levels, providing insights into how these techniques can provide engineering-oriented interpretations that enhance understanding of the studied problems.

Keywords*: Explainable Machine learning; Artificial intelligence; Structural Engineering; Design and Assessment*

1. Introduction

The application of machine learning (ML) in structural engineering has accelerated in recent years, enabling advances in design and assessment of structures (Kareem 2020; Salehi and Burgueño 2018; [3–6]. More recently, researchers have begun to explore the use of e**X**plainable **M**achine **L**earning (**XML**) to tackle the opacity of complex ML models. XML provides a means to understand and interpret the output of ML systems, fostering a new form of human-ML collaboration [7,8]. The growing emphasis on causal inference [9], scientific discovery [10], and algorithmic fairness [11] further motivates the use of XML. Additionally, the emergence of legal requirements for ML-based systems, such as the "right to explanation", suggests that XML may soon become a necessary component of ML-based solutions in engineering applications [12].

Among various XML concepts, *model transparency* is a key principle that supports users' understanding and trust [7,13]. In this context, transparency refers to the extent to which a model's inner workings can be explained and communicated [8], and is typically characterized along three different dimensions: simulatability (i.e., simple enough to be simulated by humans), decomposability (i.e., the ability to disintegrate a model into parts that can be explained), and algorithmic transparency [14]. While providing source code, dataset, and training procedure contributes to transparency, XML methods go further by providing direct user-oriented explanations, especially for those lacking technical expertise [15]. These explanations can then be delivered through various formats, including text, numerical output, visual representations, tabular data, speech, expressive motions and lights, or mixed forms [16].

XML offers transformative potential for meaningful breakthroughs in structural design and assessment by improving ML transparency, thereby overcoming key barriers to its broader adoption. However, despite the availability of various surveys on ML applications in the field [5,17–19], there is currently no literature that maps XML concepts, nomenclature, and algorithms to the domain-specific requirements of structural engineering. Therefore, this study, first, reviews XML concepts, classification taxonomies, and common algorithms. Next, through a targeted literature review and illustrative examples, the accrued knowledge is adapted and applied to topics in structural engineering, with the goal of increasing access to ML in the field through XML.

The paper is organized as follows: Section 2 presents a review of XML literature from related fields, providing a comprehensive overview of existing definitions and taxonomies, common algorithms, and their classification. Section 3 introduces the XML literature on structural engineering design and assessment at three levels: component, structure, and inventory. Section 4 provides three illustrative examples that cover different XML techniques applied at different scales. Lastly, Sections 5 and 6 conclude the paper by outlining future directions for research and the integration of XML in structural engineering, as well as a summary of the discussed material. The compiled information and illustrative examples aim to provide a curated resource for advanced practitioners and academics interested in XML, enabling them to identify and apply best practices in implementing these methods for their intended applications in the design and assessment of the built environment.

2. A Review of XML Definition, Taxonomies, and Techniques

2.1. Definition and nomenclature

XML does not have a single universally accepted definition due to its evolving nature and the diversity of methods in this field. For example, Grunning [33] defines XML as models that "*enable human users to understand, appropriately trust, and effectively manage the emerging generation of AI partners*". In contrast, Arrieta et al. state "*given an audience, an explainable AI is one that produces details or reasons to make its functioning clear or easy to understand*" [20]. Even the definition of the term "explainability" varies across the literature. For example, Hu et al. [34] define an *explanation* as a group of statements that clarifies or provides a reason for an action or belief, and categorize explanations by their purpose, type, and presentation. Doshi-Velez and Kim [22] defined explainability as "*The currency in which we exchange beliefs*". As shown in Table 1, numerous concepts are associated with XML, including transparency, understandability, comprehensibility, explicability, interpretability, and explainability [20–22,24,29–32]. In particular, *explainability* and *interpretability* are two terms that are sometimes used interchangeably, although some authors distinguish between them [10,22,24,35]. Rudin [36] defines *interpretability as* designing a model to be inherently transparent, whereas *explainability* provides post-hoc explanations for a black-box model. Accordingly, interpretability relates to the model's internal structure, whereas explainability is associated with how users understand its outputs [37]. Lipton [38] frames this distinction as interpretability addresses "*How does the model work*?" whereas explainability asks "*What else can the model tell me*?".

Developing fully interpretable ML models is challenging, as interpretability requires the ability to synthesize the model into understandable components and the transparency of the underlying algorithms [39]. Given the trade-off between model interpretability and flexibility, black-box models are often preferred for capturing complex patterns. Therefore, this paper primarily focuses on post-hoc explainability applicable across a wide range of ML models.

Table 1. Common nomenclature related to XML

Term	References	Definition and notes
Explainability	[20,21]	Explanation as an interface between AI and human; explanations should be comprehensible and accurate proxy of how model works; often is considered post-hoc
Interpretability	[22,23]	Degree which a human can understand the model and its purpose; often is considered an intrinsic property of the model
Transparency	[23,24]	The model by itself is understandable (does not require explanations)
Explicability	[21]	The properties of AI model are inspectable
Correctability /Corrigibility	[25–27]	AI model can be adapted by human agent in a manner that ensure its correctness; associated with AI robustness
Comprehensibility	[21,26,28]	The ability of the model to represent its learning in a human understandable fashion, focus on user-side cognitive load
fidelity	[29]	Accuracy of the explanations to. represent the original model behavior locally
Faithfulness	[30]	The degree to which explanations represent the original model internal reasoning; separates real from fake explanations
Decomposability	[24]	Ability to understand and inspect each component of the model
Controllability	[31]	The ability of user to steer, interact and refine the AI output
trustworthiness	[32]	the degree to which explanations improve user trust, a complex concept that combine understandability, fidelity and robustness

2.2. Explainability techniques classification

From a broad perspective, XML can be categorized as either ante-hoc or post-hoc methods [37]. Ante-hoc (i.e., “white-box”) models achieve transparency by design, such as linear regression or decision tree-based models. In contrast, post-hoc methods ("black-box" models) create approximations of the model’s prediction process to explain how model reaches its outcome [16]. Minh et al. [8] further refined this classification into pre-modeling explainability, interpretable models, and post-modeling explainability.

The post-hoc XML techniques are often categorized into two main groups: model-agnostic and model-specific methods. Model-agnostic approaches focus solely on the input-output behavior of the model [34], and can be applied to a variety of different ML algorithms. Model-agnostic techniques include explanation by simplification, feature relevance explanation, local explanation, architecture modification, explanation by example, and visual explanation [16]. On the other hand, model-specific approaches leverage the internal structure of models, such as gradients and activation [40]. Some authors suggest that post-hoc explainability can also be categorized as local or global approaches. Local explanations seek to explain individual predictions and are generally relevant to ML algorithms where the decision affects a person’s “right to explanation” ([34,41]). These methods offer advantages due to their higher clarity and lower computing cost [42]. In contrast, global models focus on the overall behavior of the ML model, and might be more computationally demanding [43].

2.3. Existing XML taxonomies and relevance to structural engineering

Different taxonomies are proposed to classify XML techniques based on scope, usage, or methodology. Figure 1 shows several common taxonomies, namely functionality-based, conceptual, result-based, and mixed taxonomies [44], whereas Table 2 compares the advantages and disadvantages of each presented taxonomy.

The functionality-based taxonomy [20,45] focuses on how XML techniques extract information from a model and includes local perturbation, leveraging structure, meta-explanation, architecture modification, and examples. Local perturbations evaluate how small input changes affect response to identify the most influential input, often using gradients. Leveraging structure-based methods use model architecture, particularly in neural networks, to generate explanations by encoding the model's internal connections and layers into the explanation process. For example, the layer-wise relevance propagation technique [46] tracks information flow through the network layers from the output to the input. Lastly, meta-explanations identify simple user-centric patterns by examining and synthesizing multiple model explanations [47]. For example, spectral relevance analysis [48] clusters individual heatmaps to analyze the model predictive capabilities.

Results-based taxonomies [49] classify XML based on the nature of their output, typically derived from functioning-based categories. The three main categories are feature importance, surrogate models, and example-based methods. Feature importance methods quantify the influence of each input on model predictions, often by calculating the average marginal contribution of the feature across all possible combinations [50]. Surrogate models are simplified models that approximate the behavior of more complex models, such as Local Interpretable Model-agnostic Explanation (LIME) [29,51]. Examples-based methods use specific instances to demonstrate how the model works, including counterfactuals (how different inputs would change the outcome), adversarial examples (how small changes can cause

Table 2. Comparison of different existing taxonomies for XML

Taxonomy	Strength	Weakness
Functionality-based	Insights on internal mechanisms of XML techniques; easy to extend	Focus on inner-workings may prevent broader view on model output when used in isolation
Result-based	Organized by output, user-friendly (e.g. practitioners can selecte XML technique based on the form of explaination needed)	Lacks details on how models work, potential to miss on methodological nuances on relationship between different techniques
Conceptual	Structured through high-level dimensions, extensible dimensions	Emphasis technical categories, may not align with non-technical stakeholders interest, overwhelming when expanded
Mixed	Integration of other taxonomies, breadth and comprehensiveness, modularity to adopt other categories specific to a certain domain	Oversimplified overlaps by assuming all categories equally weighted, subjectivity on exclusion of dimensions

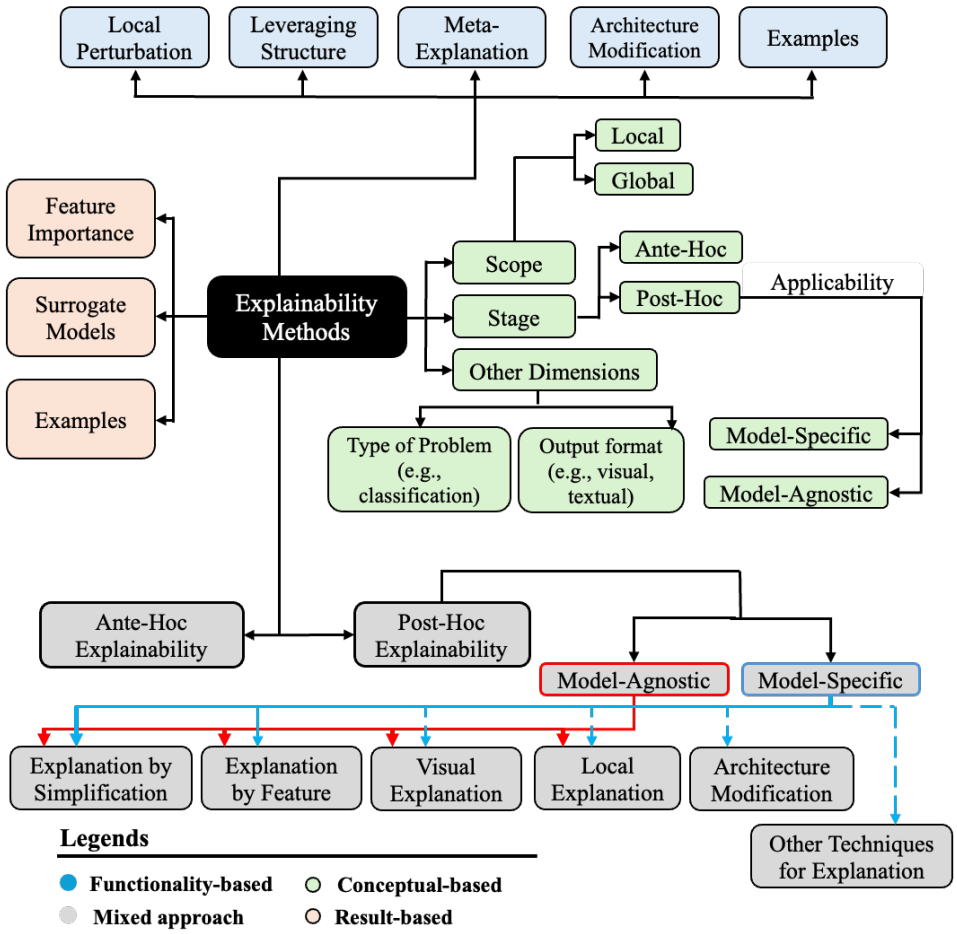

Figure 1. Classifying XML techniques using different taxonomies

incorrect predictions), and influential instances (which training data points had the most impact on the model's behavior). Concept-based taxonomies utilize various dimensions to construct concepts, including stage, applicability, and scope, for classifying XML approaches [40,52]. It is worth noting that additional dimensions exist in the literature for categorizing XML techniques [53,54]. In the context of structural

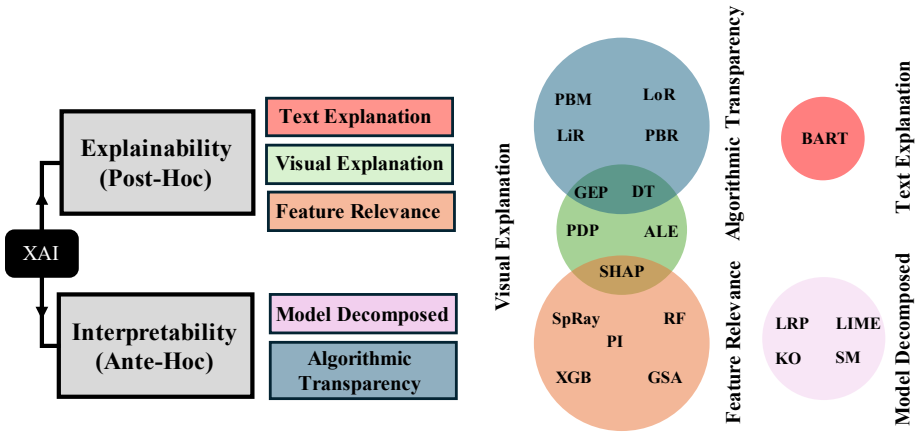

Figure 2. Common XML techniques

engineering, no effort has been made to develop such a domain-specific taxonomy. A mixed taxonomy (borrowing elements from other taxonomies; similar to work by Angelov et al. [41], Belle and Papantonis [14], and Minh et al.[8], among others) may better serve the diverse requirements of the field by adopting multiple dimensions, including ML-related (e.g., mechanism or context), engineering-related (physical interpretations or computational cost), and stakeholder-related (e.g., regulatory relevance) dimensions. In contrast, a result-based taxonomy offers the most straightforward adoption as it is intuitive (e.g., structural engineers are familiar with surrogate models and feature importance analysis) and requires less technical cost to implement. Developing such a taxonomy is a crucial step toward a more standardized use of XML in the field.

2.4. Description of common XML techniques

As discussed in the previous section, ML can be explained through two main avenues: post-hoc and ante-hoc (Figure 2). This section describes several widely used XML techniques in each category.

Three common post-hoc explainability methods are Shapley additive explanations (SHAP) [55], partial dependence plots (PDP) [56], and accumulated local effects (ALE) [57]. SHAP draws on game theory to assign each feature an importance value by treating inputs as players in a coalition game, where the model output is the shared payoff [41]. PDP and ALE visualize how individual inputs influence model predictions, capturing both linear and nonlinear relationships. While PDP measures the marginal effect of a feature on the average output, it ignores feature dependencies. ALE addresses this by using conditional distributions and averaging local differences, making it more robust to correlated inputs. Figure 3 illustrates the inner workings of these algorithms.

Two common methods of achieving XML through the form of the model (i.e., ante-hoc or white-box models) are genetic programming (GP) [58] and genetic expression programming (GEP) [59]. Both algorithms produce a model that is represented as a simple mathematical equation, allowing for the relationships between features to be easily explained when compared to more complex ML techniques such as artificial neural networks.

Physics-informed ML (PIML) models represent a form of ML models that possess elements of both ante-hoc and post-hoc XMLs: PIMLs embed physical laws directly into the ML architecture, which can be viewed as some level of built-in (algorithmic) explainability; however, they still leverage deep neural

networks that are complex and require post-hoc XML techniques to be fully understood. Common structural engineering applications, such as infrastructure performance assessment and reliability analysis, often rely

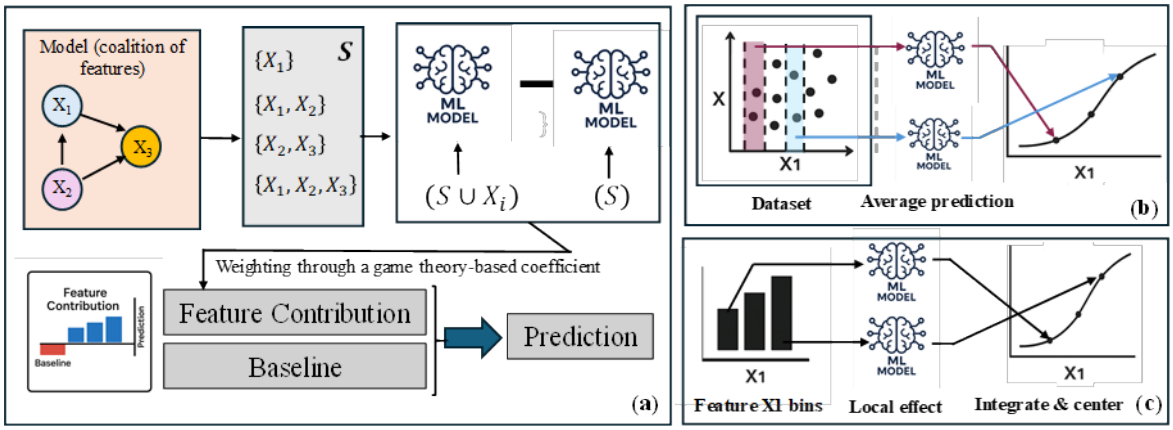

Figure 3. Schematics of how three widely used post-hoc XML techniques work: (a) Shapley explainers, (b) partial dependence plots, and (c) accumulated local effect

Table 3. Representative studies on using Physics-based ML techniques used in structural engineering

References	Application	Model Trained	Performance Evaluation Technique
[60]	Damage Classification	Regression Trees, LDA, KNN, DT, RF	Confusion Matrix
[61,62]	Structural Performance Prediction	DNN, LSTM	Mean Absolute Error, Root Mean Squared Error, Relative Error
[63,64]	Structural Health Monitoring	Gaussian Mixture Model, Physics-guided Neural Network	Type I, Type II Errors, Variance Variation
[65,66]	Pipelines Condition Assessment and Failure Analysis	KNN, DT, RF, NB, AdaBoost, LGBoost, XGBoost, CatBoost	Correlation Matrix, Confusion Matrix
[67]	Data Generation	RT, GAN	ROC, Confusion Matrix

on traditional physics-based models. However, these conventional physics-based models face computational challenges when dealing with complex and large-scale engineering problems [68,69]. PIML models integrate physical laws into ML models, enhancing the accuracy and explainability of predictions while overcoming computational challenges [68,70]. While ML models typically require input, output, and validation data for training, physics-based simulations can generate or augment these datasets [67]. Recently, generative adversarial networks (GANs) have gained traction in structural engineering for producing realistic synthetic data, particularly for failure analysis and structural health monitoring [67,71]. Table 3 shows various physics-based ML models, their applications, and performance evaluation techniques in structural engineering problems in recent years.

The explainability of PIML models depicts how physical laws are embedded into ML models and how information is transformed and exchanged within their algorithms. Since PIML models feed mechanical properties into the nodes/layers of the ML architecture and ground them on physical laws and principles, they are more interpretable than other ML architectures. Furthermore, these models and ML algorithms can be performed simultaneously, resulting in hybrid physics-informed ML models with higher accuracy [72].

Researchers also use knowledge-driven facts to interpret ML models with higher accuracy, referred to as Knowledge Ontology [73]. This process helps address confusion when a single symbol/character/icon is used for multiple semantic relationships [74]. Other techniques include the visualization of latent spaces, which analyze the internal relationships of models; physics-based regularization, which evaluates physical constraints during the learning process; gradient-based methods; and cross-validation with physics-informed metrics.

3. XML application in structural engineering

XML has emerged as a powerful tool in structural engineering, enhancing prediction, assessment, and optimization across material, component, structural, and regional levels. To explore the state-of-the-art and practical applications of XML in structural engineering, this section reviews its implementation at the material and component levels, the structural level, and the regional level.

3.1. XML for component-level and material-level assessment

This section reviews the application of XML in component-level assessment, including behavior prediction, damage assessment, and design optimization, as well as material-level assessment, focusing on concrete, steel, and their associated bonds. Table 4 provides representative publications using XML for this domain. A general finding is that SHAP has been the most widely used XML technique for both component- and material-level assessment, covering almost 100% among the representative publications. PDP and ALE are employed in relatively few investigations (occupying about 13%) to reveal how strengths of material or components are affected by design parameters. Significant efforts have been made in component-level XML assessment, to explain the influence of material and structural design parameters on failure modes and component capacity properties. The examined capacity properties include plastic hinge lengths, compressive and shear strength, drift and curvature-based limit states of various types of components such as columns, walls, slabs, beams, joints, and pile foundations.

For columns, XML has been widely applied to classify failure modes and predict capacity properties. Mangalathu et al. [60] used XML to classify failure modes of reinforced concrete (RC) columns, while Wakjira et al. [75] predicted the plastic hinge length of rectangular RC columns. Wang et al. [76] used XML to predict the drift capacity of precast concrete columns, and Mansouri et al. [77] developed an XML model to estimate the shear strength of circular concrete-filled steel tubular columns. Moreover, XML models have been used to explain the compressive strengths or axial load-carrying capacity of various column types [78–83]. For walls, XML has also been applied to primarily predict shear strength and failure modes. Feng et al. [84] developed an XML model for shear strength prediction of squat RC walls, and Mangalathu et al.[60] used XML to classify failure modes of RC shear walls. Chen et al. [85] used XML for seismic damage evaluation of RC walls based on surface crack images, and Zhang et al. [86] further extended XML applications to predict the shear strength of RC flanged shear walls. For slabs, XML has been used to estimate punching shear strength, which is a critical factor in slab design. Wu and Zhou [87] applied XML to predict the punching shear strength of two-way reinforced concrete slabs, while Mangalathu et al. [88] focused on flat slabs without transverse reinforcement. For beams, Wakjira et al. [89] used XML to predict the shear strength of RC beams, while Zhang et al. [90] applied XML to estimate the flexural capacity of FRP-strengthened RC beams. Additionally, SHAP was used to explain the feature behavior of five ML models for the prediction of T-beam shear strength. These models included decision trees, which are themselves an XML model, along with random forest, GBDT, LightGBM, and XGBoost [91]. Through the use of SHAP, a wide range of ML techniques can become explainable, allowing for a

more widespread acceptance of these techniques in practice and in research. GP has also been used to model deep beam shear [92], steel fiber reinforced concrete beams [93], and beams without stirrups [94].

Table 4. Representative publications focusing on XML applied to structural material and component-level ML studies.

References	Studied component/material	Target Variable	Best ML model employed[a]	Explainability technique(s)
Structural Material-level Assessment				
[95]	Concrete	Compressive strength	LightGBM	SHAP
[96]	Concrete with manufactured sand	Compressive and tensile strength	XGBoost	SHAP
[62]	Steel fiber-reinforced recycled aggregate concrete	Compressive and splitting tensile strength	GBRT and AdaBoost	SHAP, PDP
[97]	Concrete	Creep behavior	LGBM	SHAP
[98]	Corroded reinforced concrete	Bond strength of corroded reinforcement	Combination of KNN, MLP, RF and XGBoost	SHAP
[99]	Ultra-high-performance concrete and normal concrete interface	Bond strength	CatBoost	SHAP
Ke et al. 2023b	CFRP-steel epoxy-bonded interfaces	CFRP-steel interface bond strength	CatBoost	SHAP
[101]	Steel	Fatigue strength	Hybrid model combining XGBoost and LightGBM	SHAP
Kulasooriya et al. 2023	Basalt-fiber reinforced concrete (BFRC)	Compressive, flexural, and tensile strength	Decision Tree, Gradient Boosting Tree, LightGBM	SHAP, LIME
Structural Component-level Assessment				
[60]	RC column	Failure mode classification	RF classifier	SHAP
[75]	RC column	Capacity prediction (drift, axial strength)	Stacking ensemble model	SHAP
[76]	Precast concrete columns	Drift capacity	XGBoost	SHAP
[102]	Bridge piers	Seismic demand and fragility	ANN	SHAP
[77]	Circular concrete-filled steel tubes	Shear strength	XGBoost	SHAP
[103]	Timber column	Fire resistance	RF	SHAP, PDP
[84]	Wall	Shear strength	XGBoost	SHAP
[86]	Wall	Shear strength	XGBoost	SHAP
[104]	Wall	Failure modes	XGBoost	PDP, ALE, SHAP
[86]	Wall	Damage and crack pattern	NN, RF, DT	SHAP

[105]	Wall	The ration of Shear wall area-to-floor area	GBDT	SHAP
[87]	Slab	Punching shear strength	PSO-SVR	SHAP
[88]	Slab	Punching shear strength	XGBoost	SHAP
[89]	Beam	Shear and flexural strength	XGBoost	SHAP
[90]	Beam	Flexural capacity	GBDT	SHAP
[106]	Connection	Failure mode	XGBoost	SHAP
[107]	Connection	Shear capacity	GBM	SHAP
[108]	Connection	Fatigue failure of butt-welded connections	XGBoost	SHAP

For joints, Gao and Lin [106] developed an XML model to predict the failure mode of beam-column joints, while Ye et al. 2024 used XML to estimate the shear capacity of UHPC joints. In addition, Barun et al. [108] developed an XML model to predict fatigue failure in small-scale butt-welded joints. For pile foundations, XML has been used to predict load-bearing capacity and seismic energy dissipation mechanisms. Karakaş et al. [109] applied XML to predict the ultimate bearing capacity of piles, while Wang et al. [110] used XML to assess seismic energy dissipation mechanisms in scoured bridge pile-group foundations. Lastly, XML has used for ML beyond conventional structural components, for example, to provide explainability to a model for wind turbine blade damage [111].

A limited number of component-level studies focused on optimizing component design. Reinforcement layouts in RC columns and beams have been optimized using XML to ensure both performance and cost-effectiveness. For example, Cakiroglu et al. [112] applied XML to optimize the pile diameter for cantilever soldier pile retaining walls, while Iqbal et al. [113] used XML to study the effects of rebar size and volume fraction of glass fibers on the tensile strength retention of GFRP rebars in alkaline environments.

For material-level assessment, XML has been applied to predict the compressive strength, elastic modulus, and bond strength of various types of concrete. Jia et al. [95] and Lyngdoh et al. [96] used XML to predict the compressive strength of concrete, while Zhang et al. [62] applied XML to estimate the compressive and splitting tensile strength of steel fiber-reinforced recycled aggregate concrete. Liang et al. [97] developed an XML model to predict the creep behavior of concrete, and Hu et al. [99] used XML to estimate the bond strength between UHPC and normal-strength concrete. SHAP was further used to provide explainability for eight machine learning models including random forest, support vector machine, and artificial neural networks for concrete strength predictions [114]. Through the use of SHAP the authors were able to determine not only the impacts of the features on the final prediction but also on each other. Focusing on ante-hoc explainability, GEP has been used to predict the compressive strength of high-strength concrete [115], bagasse ash-based concrete [116], and geopolymer concrete [117]. In these studies, the GEP algorithm was utilized to generate a model that is highly explainable, allowing for further analysis and commentary on the relationships between the features. Beyond concrete, XML has also been used to predict the bond strength and fatigue life of steel. Ke et al. [100] developed an XML model to predict the bond strength of CFRP-steel interfaces, while Yan et al. [101] applied XML to predict the fatigue strength of steel. Ge et al. [98] used XML to predict the bond strength of corroded reinforced concrete following high-temperature exposure.

3.2. XML for structure-level design

The application of XML techniques in building design and assessment has gained significant attention over the past five years (Table 5), particularly within the framework of performance-based earthquake engineering (PBEE) for explaining the roles of structural and seismic intensity parameters in seismic response, capacity, fragility, loss, and resilience. SHAP plays the dominant role among all the XML

Table 5. Representative publications focusing on XML applied to structural-level ML studies

Publication	Structure	Target Variable	Best ML model employed	Explainability technique(s)
[118]	Masonry-infilled frames	Natural vibration period of masonry infill concrete frame structures	XGBoost	SA, SHAP, ICE
[119]	Masonry-infilled frames	Fundamental period of vibration of masonry infilled RC frames	NN	PDP, ICE, ALE, SHAP
[120]	Masonry-infilled frames	Fundamental time period of masonry-infilled RC frames	XGBoost	SHAP
[121]	Masonry-infilled frames	Natural period of vibration	RF	SHAP, LIME, PDP
[122]	Steel moment frames	Seismic damage states	RF	SHAP
[123]	Steel moment frames	Fragility function parameters	ANN	SHAP
[124]	RC frame buildings	Seismic fragility curve parameters	XGBoost	SHAP
[125]	RC frame buildings	Compressive membrane/arch action resistance, Progressive collapse resistance-displacement curve	DCN	SHAP
[126]	Seismically deficient RC frame	Damage index	LGBM	SHAP, PDP, LIME, ALE
[127]	High-rise building with interference from a low-rise building	Wind pressure coefficients	XGBoost	SHAP
[128]	Bridges	Pier drift ratios, Bearing displacements, Deck displacements	XGBoost with Bayesian optimization	SHAP
[129]	Bridges	Damage classification	ANN, MLP and RF	SHAP

[130]	Bridges	Repair cost of earthquake-damaged bridges	Extra Trees ensemble	SHAP, Feature dependence plots
[111]	wind turbine blade	Mahalanobis distance-based damage index	XGBoost	SHAP
[131]	pipelines	Failure modes of corroded pipelines	XGBoost	SHAP
[132]	Steel Moment-Resisting Frames	Resilience index	XGBoost	SHAP
[133]	Bare frames, and frames with infills	Maximum inter-story drift ratio	GPR	SHAP
[134]	RC buildings	Maximum Inter-story Drift Ratio (MIDR)	LightGBM	Contributions Plots, PDP, Shapley values, SHAP
[135]	Steel moment frames	Peak inter-story drift ratio (PIDR), Peak floor acceleration (PFA)	GPR	SHAP
[136]	Cross-Laminated Timber (CLT) buildings	Maximum Inter-story Drift Ratio , Maximum Roof Drift Ratio	RF	SHAP
[137]	Steel frames equipped with shape memory alloys (SMAs)	Amplification coefficient for energy modification factor	XGBoost	PDP, SHAP
[138]	High-rise buildings with outrigger systems	Maximum inter-story drift, Maximum top acceleration, Maximum top displacement	DNNs	PDP, SHAP
[139]	Construction projects	Cost impacts of nonconformances	RF algorithm with Gravitational Search Algorithm	SHAP
[140]	RC and masonry buildings	Seismic vulnerability index	ANN	SHAP
[141]	Buildings	Overall damage grade, Building collapse grade and Building leaning grade	XGBoost with SMOTE	SHAP
[142]	Two-story damped steel frame	Maximum inter-story displacement	Aggregation model combining XGBoost and SWT	SHAP
[143]	High-rise buildings	Unit construction cost	BP-SVM combined algorithm	SHAP, PDP, and ICE
[144]	High-rise buildings	Aerodynamic force coefficients, and Wind-induced dynamic response coefficients	XGBoost	SHAP

[145]	Residential buildings	Fire damage area and Fatality/Casualty prediction	XGBoost	SHAP
[146]	Buildings	Property loss rate due	IBTEM	SHAP, PDP
[147]	Offshore concrete bridge	Life-cycle seismic performance	LSTM	LSTM-Shapley framework
[148]	Buildings	Seismic performance metric	XGBoost	SHAP, LIME, PDP, ICE
[149]	Tunnel	Tunnel damage state classification, Damage indicators, and Seismic fragility prediction	XGBoost	SHAP, and Feature importance score analysis
[150]	RC Bridges	Post-earthquake vertical load-carrying capacity	ANN	SHAP,PDP

techniques (occupying over 95%), while about 25% studies adopt ICE, PDP, or LIME for structural assessment and analyses of frames and buildings.

The studies include: (1) Identifying Critical Parameters: using techniques such as SHAP, ICE, LIME, or PDP to identify critical structure-related parameters that significantly affect natural (fundamental) periods of masonry-infilled frames [118–121]; (2) Explaining Seismic Response Influences: leveraging XML to interpret how seismic responses (e.g., maximum inter-story drifts and floor acceleration) of buildings are effected by structural and seismic loading features [133–138,142]; (3) Seismic Capacity and Damage Analysis: utilizing SHAP to investigate the impact of various features on seismic capacity and damage states of steel moment frames [122]; (4) **Fragility**: employing SHAP to identify key parameters in predicting the seismic fragility of steel moment-, reinforced concrete-, and masonry-frame buildings [123,124,140,141]; (5) **Loss Estimation:** applying PDP to explore the effects of structural-modeling-related parameters on seismic loss of steel moment-frame buildings [151–154]; and (6) **Resilience Assessment:** using SHAP to analyze the importance, effects, and interactions of design parameters in evaluating seismic resilience of buildings [132]. Beyond earthquake applications, XML techniques have also been adopted to explain the developed ML models for assessing construction quality, predicting costs [139,143], estimating progressive collapse resistance [125], evaluating aerodynamic interference effects between tall and short buildings [127,144], and fire loss prediction [145,146].

For bridges, a critical transportation infrastructure, XML techniques have been implemented to understand how design features affect behaviors of bridges under operational conditions or against natural hazards, especially within the PBEE framework. Applications include (1) **Seismic Response Prediction:** Evaluation of seismic responses in terms of the column drift, bearing deformation and superstructure acceleration incorporating time-dependent material deterioration and multiple sources of uncertainty [128,147,155]; (2) **Damage State Classification:** classification of seismic damage state using numerical models considering aging corrosion effect or using the post-earthquake damage data [129,156]; (3) **Seismic Fragility Assessment:** application of XML techniques to assess the fragility of rocking column-supported bridges [157]; and (4) **Repair Cost Predictions:** prediction of post-earthquake repair cost of damaged bridges [130]. Despite these advancements, applications of XML in seismic resilience assessment for

bridges remain unexplored as of the time of this manuscript's preparation. In addition to earthquakes, XML have also been applied to bridge health monitoring. using XML techniques and structural health monitoring (SHM) data, researchers have predicted and interpreted load-deformation correlation and cumulative displacement of long-span suspension bridges subjected to temperature, wind and vehicle loads [158–160].

Besides building and bridge structures, XML techniques have been extended to various other areas in structural engineering, including nuclear power plants [161], wind turbines [111], pipelines [131], road embankments [162], tunnels [149], deep beam shear estimation [163], and gridshell structures [164].

3.4. XML For regional-level assessment

ML vulnerability and risk assessment models are increasingly being developed for infrastructure systems and buildings and bridges at regional scales. These models address some of the limitations of traditional analytical and empirical approaches by reducing computational costs for regional seismic loss estimation, handling numerous input variables and complex spatial dependencies in infrastructure assessments, and capturing intricate interrelationships among physical, social, and economic factors necessary for community resilience modeling [165,166]. Since ML vulnerability models for infrastructure systems and regional assessments are intended to support mitigation and recovery planning, it is crucial to understand the predictions provided by these models. However, the lack of explainability and interpretability in ML models hinders such understanding. To address this issue, efforts have been made to interpret and explain ML models for regional and system-level assessments. The most common explainability techniques used in the literature for infrastructure systems and regional assessments are PDP, ALE, and SHAP, as summarized in Table 6.

Several studies have employed PDP and ALE to explain ML models for buildings, bridges, and infrastructure systems. For example, [167] used PDP to explain an ML model developed to predict climate-induced disaster-related damages, focusing on property damage caused by wind hazards. The study considered attributes related to hazard, climate, land cover, social, housing, demographic and economic data. Various ML techniques were evaluated, and random forest-based regression tree was found to be the best performing model. PDP was successfully used to explain the complex interrelationships between input variables and predicted property damage. The impact of each individual variable and pairs of variables was studied. Johnson et al. [166] used PDPs as a tool to depict spatial network vulnerabilities to disruptions. A GBM model was trained to analyze disruptions in a freight network. The PDPs were then used to study how disruptions at different locations impacted network efficiency while controlling for other input variables and spatial dependencies. In a similar approach, [168] used ALE as a tool to examine the relationship between road network attributes and groundwater quality parameters, aiming to identify road transportation factors influencing water quality.

SHAP is the most common technique for explaining ML models for infrastructure systems and building and bridge inventories. [169] used SHAP to explain the results of an extreme gradient boosting model developed to predict and analyze electricity, gas, and water consumption in large-scale buildings in New York City. Similarly, Chen and Zhang [170] and Roeslin et al. [171] also employed SHAP to explain ML models for regional seismic risk and damage assessment.

Mangalathu et al. [104] compared the use of PDP, ALE, and SHAP to explain ML models for regional-scale bridge damage assessment in California. Seven ML models were evaluated and XGBoost was found to be the best performing model. XGBoost's importance factor was also considered as an explainability technique. The authors compared the four methods in terms of their ability to rank input variables, identify trends, predict influence on output, explain specific predictions, and explain how the model makes decision.

XGBoost importance factor can only rank input variables and identify any correlations between them. PDP and ALE offer similar insights, but PDP cannot capture correlations between input variables. SHAP is the only technique that can explain the risk predictions and identify the factors that contribute to the predictions. The authors concluded that no single technique could fully explain the ML models and recommended the use of more than one technique to adequately understand and explain ML models.

Table 6. Representative publications focusing on XML applied to regional-level ML studies

Ref	Studied infrastructure	ML models employed	Explainability technique(s)
[104]	Regional damage assessment of bridges in California.	7 models (Ridge, SVR, KNN, DT, RF, AdaBoost, XGBoost)	ALE, PDP, SHAP
[168]	Impact of road transport systems on groundwater quality.	XGBoost	ALE
[166]	Freight network vulnerabilities (intermodal road and rail freight transportation network).	GBM	PDP
[169]	Interrelationships between electricity, gas, and water consumption in large-scale buildings.	XGBoost	SHAP
[167]	Predict wind-related property damages based on disaster, landcover, climate, social, economic, housing, and demographic-related records.	Classic regression trees and regression trees ensemble techniques (i.e., bagging, random forest, and boosting)	PDP
[172]	Assess regional seismic resilience considering geological, building, and social dimensions.	Dempster-Shafer evidence theory and cluster analysis	Global sensitivity analysis
[170]	An automated machine learning (AutoML) framework to predict the casualty rate and direct economic loss induced by earthquakes.	SVM, kNN, RF, XGBoost, CatBoost, and LightGBM	SHAP
[171]	Building damage prediction model using empirical regional data.	Logistic regression, SVM, DT, and RF	SHAP
[173]	Improving the seismic performance of a regional group of small-to-medium span highway bridges.	ANN	SHAP
[174]	Seismic repair cost rate prediction for residential buildings due to seismic damage.	Three distance-based regression models, a Decision Tree Regression model and two Ensemble regression models (RF and Gradient Boosting Regression).	SHAP

4. Illustrative Examples

This section presents three case studies that apply different XML techniques to various structural engineering problems at three different scales: component-, structure-, and inventory-level to showcase the potential of these approaches over a diverse set of problems. The first example uses an ante-hoc XML technique based on genetic expression programming, whereas the other two models use post-hoc explainability based on SHAP, permutation importance, and accumulated local effects.

4.1. XML for component level assessment: Ultimate shear prediction of deep reinforced concrete beams with knowledge transfer from slender RC beams

The first illustrative example investigates the application of XML to the problem of modeling deep beam shear strength through the use of genetic expression programming (GEP) [175] and transfer learning (TL) [176,177].

4.1.1. Dataset and Background

In structural engineering, collecting high-quality experimental data is often constrained by cost, time, and practical feasibility. To address this issue, this study explores data-scarce modeling through XML, with an emphasis on model explainability, which is crucial for engineering judgment and design decision-making. Transfer learning is used here to improve predictions on a small "target" dataset by leveraging knowledge from a larger, related "source" dataset. TL is a more complex task than traditional ML as it involves multiple datasets and mechanisms for transferring knowledge from the source to the target in a beneficial manner. Genetic expressions programming (GEP) is used to address the issue of explainability as it is a white-box model where the relationship between features used to make predictions are plainly observable.

The dataset utilized in this study consists of 1,849 data points spanning a wide range of deep and slender reinforced concrete beams without shear reinforcement [178]. This dataset consists of material properties of the beam's constituent materials as well as the geometric properties of the beam itself. The data was collected over the span of 59 years from 1948 to 2007. The beams contained within the dataset are both high- and low-strength concrete, square and circular cross sections, and tested under various loading configurations. The slender beams, those having an a/d ratio of more than 2.5, were taken as the source dataset to pretrain the model and the deep beams, those having a/d ratio less than 2.5, were used as the target dataset and used to fine tune the model. The dataset features can be seen in Table 7. The data is log scaled before being used as input for training and prediction.

4.1.2. Model Development

To accommodate differences between source and target datasets, the 2M2P algorithm (Figure 4) is employed. This method is designed for both homogeneous (same features) and heterogeneous (different features) transfer scenarios. To address the feature space mismatch zero vectors are added for missing features, creating a shared input space. A GEP population is first trained on the source dataset. This trained population is then evaluated on the target data, and the best-performing individuals are selected. These models undergo a forced mutation, a step that introduces small random changes to adapt the model to the target data while preserving learned knowledge. The mutated population is then fine-tuned on the target training set, and the best model is selected based on validation performance. Table 8 provides the hyperparameters utilized in model training.

In this application the source data consists of slender beam shear strength tests, while the target data consists of deep beam shear strength tests. To mimic real-world data scarcity, only 6, 9, and 15 deep beam data points are used. Furthermore, certain features are intentionally removed to simulate heterogeneous transfer conditions. Three models are compared: A source-only GEP model trained only on the slender beam data, a baseline GEP model trained only on the few deep beam data points, and the 2M2P model incorporating transfer learning.

The results can be seen in Figure 4. When only 6 deep beam points are available, the source model performs best. However, at 9 and 15 data points, the 2M2P model consistently outperforms both the source and baseline models, demonstrating stable performance across 10 randomized runs. The best-

Table 7. RC Beam Dataset Features

Parameter	Description	Unit	Mean	Standard Deviation	Minimum	Maximum
b_w	Web Width	mm	213.01	212.94	21.00	3000.00
b	Flange Width	mm	256.76	230.21	21.00	3000.00
h	Depth	mm	364.35	254.05	51.00	3140.00
d	Effective Depth	mm	320.25	237.77	41.00	3000.00
amv	Distance from point of inflection to maximum moment location	mm	953.22	823.70	80.00	9000.00
astm	Distance from center of applied load to the center of support	mm	1000.97	845.97	80.00	9000.00
a/d	Span to depth ratio	-	3.20	1.82	0.25	15.06
bear	Length of bearing pole	mm	107.13	75.15	0.00	6000.00
ρ_w	Horizontal reinforcement ratio	%	2.24	1.52	0.10	9.50
f'_c	Concrete compressive strength	MPa	34.86	18.34	6.10	127.50
ag	Maximum aggregate size	mm	18.48	6.96	1.00	50.00
f_y	Steel yield strength	MPa	462.37	172.14	267.00	1779.00
mpvd	Derived parameter	-	138.62	156.01	0.00	2768.30
v_u	Ultimate shear strength	MPa	129.59	153.21	1.90	1575.00

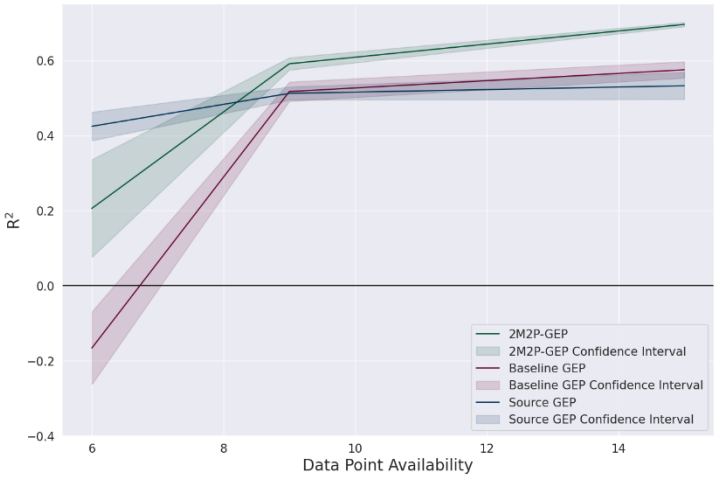

Figure 4. Data availability analysis

Table 8. Hyperparameters used for training ML of RC Beam example

Parameter	Description	Values
Head length	Number of operators and features in the head of each gene	3,**5**,8
Genes	Number of genes that comprise the chromosome	3,**4**,5,6
Population size	Number of chromosomes/expression trees in the population	36, **50**
Batches	Number of populations isolated from each other and evaluated	**50**
Source generations	Number of generations the source model is allowed to execute	**200**
Baseline generations	Number of generations the target model is allowed to execute	**400**
Transfer generations	Number of generations the transfer model is allowed to execute	**200**
Tournament size	Number of chromosomes/expression trees competing with each other for reproduction	3,**5**,10,30
Champions	Number of highest scoring chromosomes/expression trees copied from previous generation	**3**
Linking function	Function that links the chromosomes	addition, **multiplication**
Fitness function	Function utilized to evaluate the fitness of the chromosomes/expression trees	**MAE**
Transfer Function	Function used to select individuals for transfer	**SSIP,** WPE
Operators	Mathematical operations used in the chromosomes/expression trees	**addition, subtraction, multiplication, division**

performing model generated by the 2M2P algorithm for the 15 data point availability can be seen below. Equation 1 illustrates the high level of interpretability as the equation is simply the product of three terms consisting of variables that represent values commonly collected during beam experiments.

$$V_u = \left(b_w * (-astm + d) + d * \sqrt{f_c'}\right) * \left(-\frac{a}{d} + b + f_c' * (b_w - 5)\right) \\ (b + f_c' * (b_w - 5) - h) \quad (1)$$

4.1.2. Explainability

The explainability of this XML approach is rooted in the architecture of the GEP model. As a white-box method, GEP provides full transparency into the steps that link input features to the model's predictions. The model comprises two interconnected components: the genotype and the phenotype (Ferreira 2001). The genotype, which is shaped during the training and mutation processes, is a linear string of characters that includes both operators, such as mathematical functions like addition or multiplication, and terminals, which represent constants or dataset variables. Additionally, the genotype maintains a modular organization

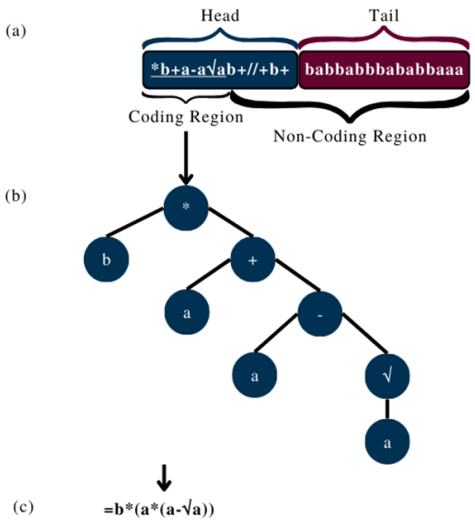

Figure 5. GEP model form (a) the genotype string (b) the phenotype expression tree (c) the

in which the "head" section contains the active, expressed portion of the string, while the "tail" consists of unused elements reserved for potential structural expansion. During the prediction phase, this string is decoded by the GEP algorithm and expressed as a mathematical equation, known as the phenotype, which is used in prediction and evaluation. This transformation reveals the internal logic of the model in a form that is both readable and interpretable. The resulting equation can be viewed directly, as shown in Figure 5(c), or visualized as an expression tree, as illustrated in Figure 5(b), offering multiple layers of interpretability. Because the final model is a concise mathematical expression composed of familiar engineering variables, its structure can be readily understood by domain experts. This design ensures that the model remains both flexible and interpretable, making it particularly well-suited for use in civil engineering applications where understanding the relationship between variables is essential.

4.2. XML for structure-level assessment: Evaluating the impact of structural parameters and seismic intensity measures on seismic responses of multi-span highway girder bridges

The second illustrative example explores the capability of XML in interpreting a traditional Neural Network (NN) trained to predict longitudinal seismic responses of regular multi-span highway girder bridges. The database was developed using finite element modeling accounting for various parameters of the bridges.

4.2.1. Database and model features

The examined bridges are ordinary highway bridges among the national highway systems in China. They have a typical span length of 30 m to adapt to industrial prefabricated superstructures, which are supported by reinforced concrete (RC) bridge bents and abutments through spherical steel bearings. One of the bents has longitudinally fixed bearings while others have sliding bearings. Each bent comprises a cap beam, two extended pile-shafts (i.e., Type II with capacity-protected pile). For this reason, the bottom of columns is fixed in the three-dimensional finite element models established in OpenSees [179], where nonlinear fiber-section-based beam-column elements and bilinear force-displacement constitutive models are assigned to columns and sliding bearings, respectively, to capture their nonlinear behavior under earthquakes. For conciseness, detailed modeling techniques can be referred to [180].

To develop a sound database, five parameters of bridges were considered as uniformly distributed random variables and selected as features (input variables) of the NN model: column diameter ($D \sim$ U[1.4m,

2.5m]), column axial load ratio ($\alpha \sim$ U[0.1, 0.3]), column height ($H \sim$ U[10m, 25m]), column longitudinal reinforcement ratio ($\rho_l \sim$ U[0.6%, 2.5%]), and number of span ($N_s \sim$ U[2, 5]). Note that the assumed uniform distributions aim to cover a wide range of engineering scenarios as well as to assure a balanced database for the expediency of model training. Accordingly, 720 bridges were sampled using Latin hypercube sampling, and randomly paired to a selected 720 ground motions (Wang et al. 2021; Baker et al. 2011) for a total number of 720 nonlinear time-history analyses. The peak column drift (γ_p) is adopted as the engineering demand parameter for seismic performance assessment, and thereby chosen as the label (output variable) of the NN model. On the other hand, multiple seismic intensity features are often required to achieve high-confidence ML modeling of structures [182]. Accordingly, seven seismic intensity measures were also selected as features of the NN model: peak ground acceleration (*PGA*), peak ground velocity (*PGV*), peak ground displacement (*PGD*), spectral acceleration at a period of 1 second (*Sa*10), spectral velocity at a period of 1 second (*Sv*10), Housner intensity (*HI*), and the state-of-the-art average spectral acceleration (*AvgSa*) with proper lower and upper period bonds for highway bridges (Feng et al., 2024). Finally, a comprehensive dataset consisting of 720 data for 12 features (including five structural parameters and seven seismic intensity measures) and the label (peak column drift) was established and can be accessed via Zonedo (https://doi.org/10.5281/zenodo.14211631).

4.2.2. Model development

The dataset was normalized using the Z-score method and then randomly grouped into a training set and a test set by 80% and 20%, respectively. Next, a traditional NN model was trained using the training data and an open-source Scikit-learn package in Python. Except for hidden layer sizes, other hyperparameters of the NN model were taken as the optional values provided by Scikit-learn. To reduce model complexity (avoiding overfitting) and minimize the computational demand for hyperparameter optimization, both the number of hidden layer N and the number of neurons in each hidden layer Q were restricted to values less than 20. Using the grid search method, N and Q were tuned to be N = 11 and Q = 6 as the optimal NN hyperparameters. This optimization process took approximately 1,000 seconds on a personal computer with Intel Core i7 12700K @ 4.9 GHz processor and 32 GB DDR4 RAM.

Figure 6 shows the performance of the NN model on training and testing set. In generally, NN reasonably estimated peak column drifts with a R-squared value larger than 0.90. The average errors of NN model are

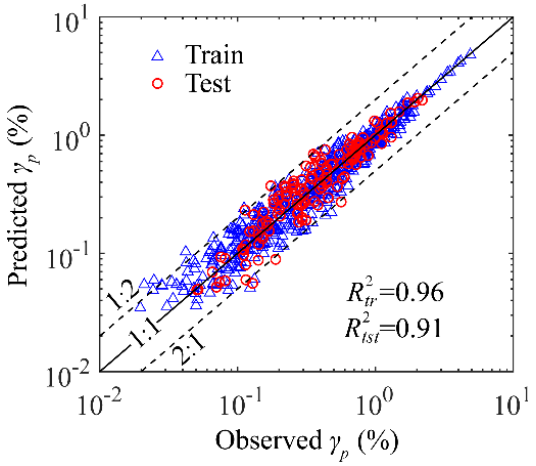

Figure 6. NN performance on training and testing set to predict peak column drifts

22.9 % and 23.7% on average for the training set and test set, respectively, falling into the typical error range from 15% to 25% for the applications of traditional ML in structural seismic response predictions[184,185]. For large drift responses (exceeding 1.0%) of primary concern, the prediction error decreases significantly to 9.6%. Therefore, the prediction accuracy of the ML model is generally acceptable.

4.2.3. Explainability techniques

Two XML techniques, i.e., SHAP and permutation importance [186], were selected to explain the NN model. Figure 7(a) displays the ranking of feature importance, which is represented by the mean absolute value of feature SHAP values (mean(|SHAP value|)) over all the samples (Mangalathu et al. 2020). Note that SHAP values presented in this example quantify the impact of features on the Z-score normalized peak column drifts. A larger mean (|SHAP value|) indicates a higher importance. Overall, the seismic intensity measures play a more critical role than the structure-related ones in predicting peak column drifts. In particular, the state-of-the-art *AvgSa* is the most important feature with the largest mean(|SHAP value|) of about 0.4, which indicates that a loss of *AvgSa* as the feature may lead to a change of peak column drift as large as approximately 40% of its standard deviation on average. Therefore, it is reasonable to adopt *AvgSa* as an optimal intensity measure for probabilistic seismic demand modeling of the investigated bridges in terms of peak column drifts for future studies. Figure 7(b) summarizes SHAP results over all the samples, where features are ranked in descending order of importance (i.e., mean(|SHAP|)) from top to bottom. Each point in the plot represents the feature SHAP value in an instance and is colored based on the feature value from blue (low) to red (high). When red dots cluster on the right (positive SHAP values) and blue dots cluster on the left (negative SHAP values), it indicates a positive correlation between the feature and the model output. Conversely, it indicates a negative correlation. It can be found from Figure 7(b) that red dots of seismic intensity-based features (particularly *AvgSa*, *PGD*, *PGV*, *HI*), number of spans (N_s), and column axial load ratio (α) are mostly distributed on the right side of the figure with positive SHAP values. This implies a larger value of these features would trigger a larger peak column drift. By contrast, red dots of the column height (H) are mostly negative, indicating that a higher column could lead to a smaller column response. Compared to the above features significant impact on peak column drifts. Specifically, the reinforcement ratio mainly influences the cases where the yielding of steel reinforcement occurs. The column diameter affects both the mass (i.e., the section area multiplied by the concrete strength and axial load ratio) and stiffness of structures, eventually exerting a limited impact on seismic responses.

To understand the stability of explainable results across different XML approaches for the examined bridges, Figure 7(c) shows the ranking of permutation-based feature importance, which is obtained by permuting the feature values and comparing the model loss with the permuted feature values to that with the original feature values [186]. A large decrease in model accuracy with the permuted feature values indicates a high importance. Note that each feature is permuted 1,000 times to obtain the statistical properties of permutation importance. In this box chart, three vertical lines of boxes from the left to right represent the 25% quantile, median, and 75% quantile of permutation importance, respectively. The feature importance ranking is determined based on the median permutation importance, while the confidence level of such a ranking can be evaluated by examining the variation of permutation importance. In general, the permutation-based importance ranking is quite close to the SHAP-based one in terms of the top three and last three influential features. Form Figure 7(c), *AvgSa* is again identified as the most influential feature, i.e., permuting values of *AvgSa* would lead to most decrease in accuracy score of the NN model. By contrast, disrupting the sequence of column reinforcement ratio (ρ_l) hardly lowers the accuracy of the model, indicating a negligible feature impact. It is worth noting that the span number (N_s) exhibits a larger

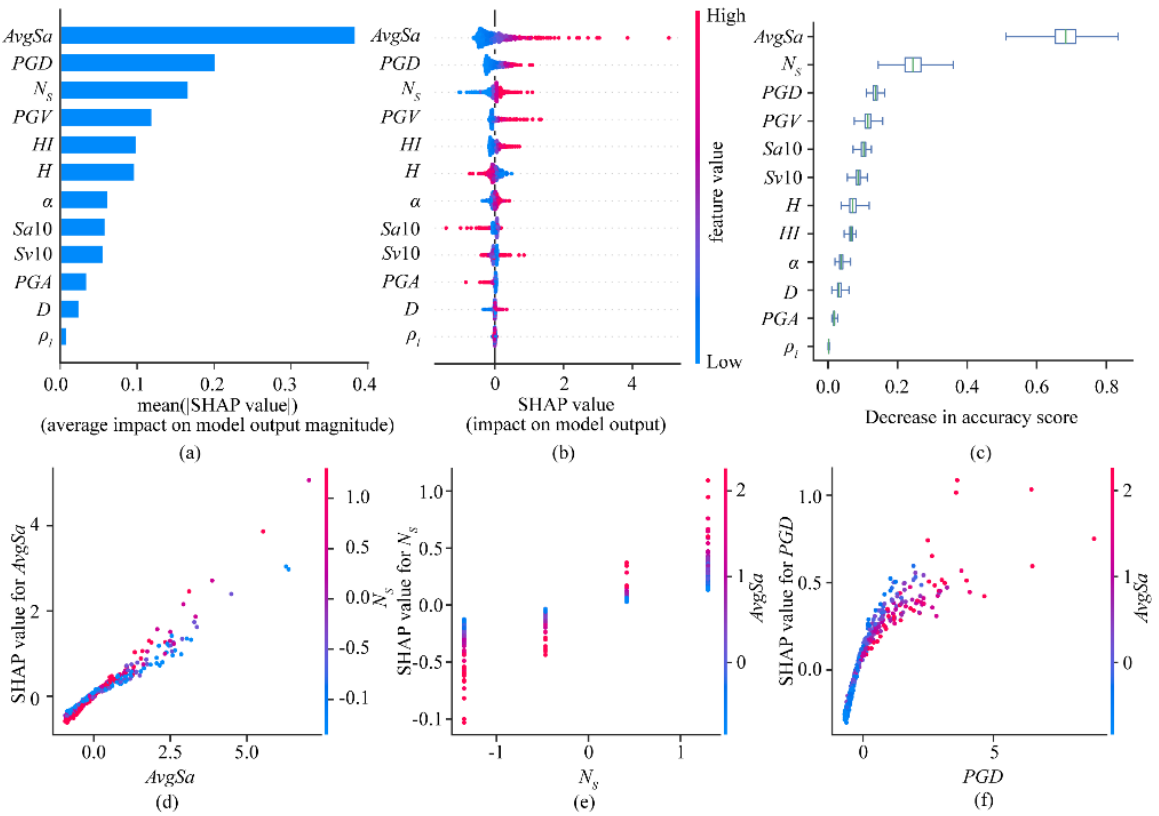

Figure 7. Interpretation results for bridge seismic response predictions: (a) ranking of SHAP-based feature importance, (b) summary of SHAP values, (c) ranking of permutation-based feature importance, SHAP dependence plots for (d) AvgSa, (e) Ns, and (f) PGD.

permutation importance than *PGD* (see Figure 7(c)), whereas *PGD* is identified as more influential by SHAP (see Figure 7(b)). These results imply that applying various XML techniques helps to comprehensively explain the ML models. Nevertheless, the general trend of variable relative importance is quite stable for the examined case. Additionally, the permutation-based feature importance ranking exhibits relatively high credibility and stability, supported by minimal overlaps in the box plots of adjacent features.

To explore feature interactions, Figures 7 (d)-(f) present the SHAP dependence plots for top three influential features (i.e., *AvgSa*, N_s, and *PGD*). The horizontal and vertical axes represent the feature value and SHAP value, respectively, of the first feature. The color of point indicates the value of the second feature, with blue indicating low values and red indicating high values. From Figure 7(d), the dispersion of SHAP value for a given *AvgSa* value is relatively low, which means the effect of *AvgSa* on column drift estimates is less affected by other features. By contrast, *AvgSa* significantly influences the SHAP values of N_s and *PGD* (Figures 7(e) and (f)), which means the contributions of N_s and *PGD* on column drift predictions depend on the average spectral acceleration of ground motions. The results of feature interactions further highlight the finding that *AvgSa* exhibits an importance much higher than that of other features.

4.3. XML for inventory-level assessment: Evaluating the impact of design and geometry parameters on annualized loss of a steel building inventory

The third illustrative example applies XML to understand an ML model trained to predict the seismic losses of a building inventory of 621 special moment-resisting frames.

4.3.1. Database and model features

The building inventory was generated through an automated seismic design tool, and covers a wide range of geometries (from 1- to 20-story buildings) with varying footprints and designs [188]. The building had a typical story height of 13 ft, with the first story height between 13 to 26ft. All buildings are assumed to be located at Los Angeles, California. Two-dimensional finite element models were generated in OpenSees, where frame elements were modeled using elastic elements with plastic hinges at both ends. The plastic hinges followed the Ibarra-Medina-Krawinkler model with bilinear hysteretic [189].

The buildings were subjected to 240 ground motions from 12 past events in California with a magnitude between 6 to 7, and peak ground acceleration of 0.03g to 0.61g [190]. The maximum drift and acceleration responses of all floors were estimated through dynamic nonlinear history analysis and post-processed to derive building-level fragility function and loss values. A component-based approach based on FEMA P-58 [191] was used for loss functions, where floor-level responses were used to quantify damage to structural and nonstructural components. It should be noted that the underlying data and assumptions in this component-based approach influence the model's performance, and would affect model explainability as different geometric and design parameters may be better correlated to a specific loss assessment methodology [153,192]. The damage information was then translated into the repair cost of individual components and was aggregated to derive building repair cost at the building level for different shaking intensities. Lastly, the conditional distribution of losses over shaking intensities was integrated over site's hazard to derive expected annual loss (EAL)[193,194].

4.3.2. Model development

A training workflow comprising wrapper-based feature selection, hyperparameter tuning, and model evaluation was used to train several ML algorithms to estimate annual seismic loss in terms of repair costs based on building geometry and design parameters [195]. These "design-oriented" models aim to use only information available during the design process (early or detailed), providing opportunities for an ML-based performance design through inverse problem formulation [196]. A *k*-fold cross-validation approach was used to tune hyperparameters and evaluate models, where models are developed based on *k*-1 sets and tested on the kth set for *k* iterations.

An ensemble learning-based model, random forest (RF), was selected to estimate EALs for different buildings. RF uses the result of a number of weak learners based on bagging (parallel learning). RF decorrelated constituent weak learners (i.e., decision stumps) by training each stump on bootstrapped sample of training data and separate feature set.

Figure 8 shows the performance of the RF model in predicting EAL for both the training and test sets. For the test set, the RF model achieved an R^2 value of 0.90, 0.99, 0.99, and 0.99 for structural, nonstructural acceleration-sensitive, nonstructural drift-sensitive, and total losses. Overall, the model exhibits higher accuracy for nonstructural drift- and acceleration-sensitive losses than structural losses. Since nonstructural losses largely govern the total loss, the model's performance for total loss achieves similar accuracy as nonstructural loss predictions.

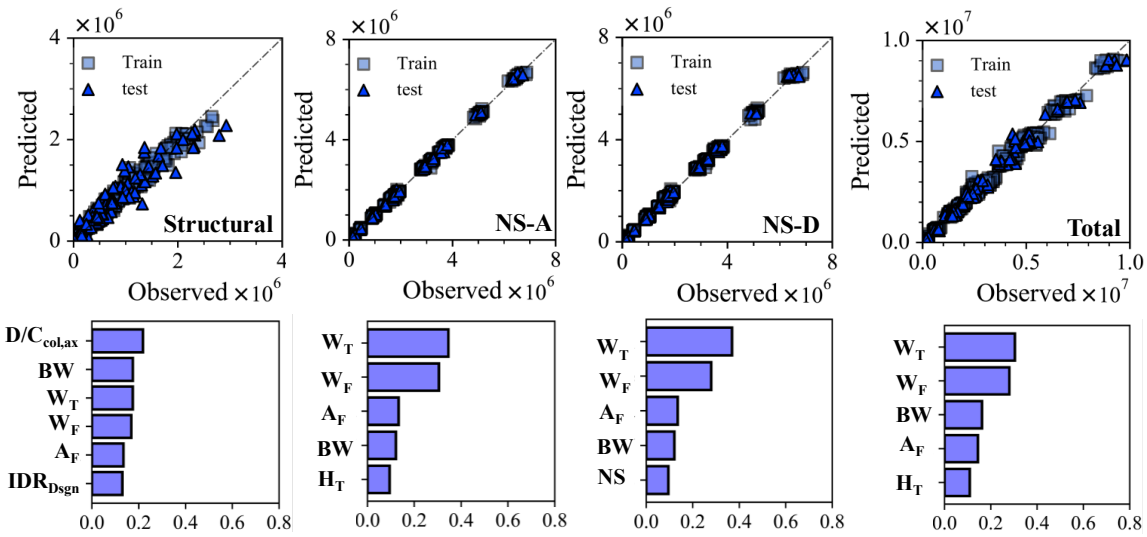

Figure 8. Top: RF performance on training and testing set to predict structural, nonstructural acceleration- sensitive, nonstructural drift-sensitive, demolition, and total expected annual loss. **Bottom:** Feature importance of each model

Figure 8 also includes feature importance plots highlighting that weight-related parameters, such as total (W_T) and frame weights (W_F), as well as floor area (A_F), are influential across all assembly types. Furthermore, parameters critical for predicting nonstructural losses— including total and frame weight, bay width (BW), A_F, and building height (H_T) —also drive the total loss prediction. In contrast, structural losses are primarily influenced by two design-specific parameters: the demand-to-capacity ratio of column ($D/C_{col,Ax}$) and design drift (IDR_{Dsgn}). This observation aligns with engineering principles, as nonstructural losses are governed by the quantity and characteristics of components, which are directly calculated based on building geometry in FEMA P-58 method for structures with similar systems and occupancy in a given location.

4.3.3. Explainability techniques

ALE was selected to explain the RF models. Figure 9 shows a clear linear relationship between structural losses and bay width and floor area. In contrast, the relationships between structural losses and design drift and the column axial demand-to-capacity ratio are nonlinear and increasing. Structural losses are particularly sensitive to design drift values around 0.25% and demand-to-capacity ratios that are about 15%. Additionally, the relationship between frame weight and structural losses is nonlinear, with losses being lower than average at both extremes of the frame weight spectrum.

The most influential parameters are identical for nonstructural acceleration-sensitive losses and total losses. This observation is intuitive, as acceleration-sensitive nonstructural losses dominate total losses. Both floor area and bay width demonstrate a linear increase in relation to these loss types, while the relationships between weight-related parameters (frame and total weight) and losses are nonlinear. Specifically, the relationship is increasing for nonstructural acceleration-sensitive losses, but for total losses, it peaks at mid-range frame weights and then decreases, suggesting that frames with intermediate weights incur higher-than-average total losses. Overall, weight-related parameters, building height, and floor area significantly impact all loss types, with losses generally increasing as these parameters increase.

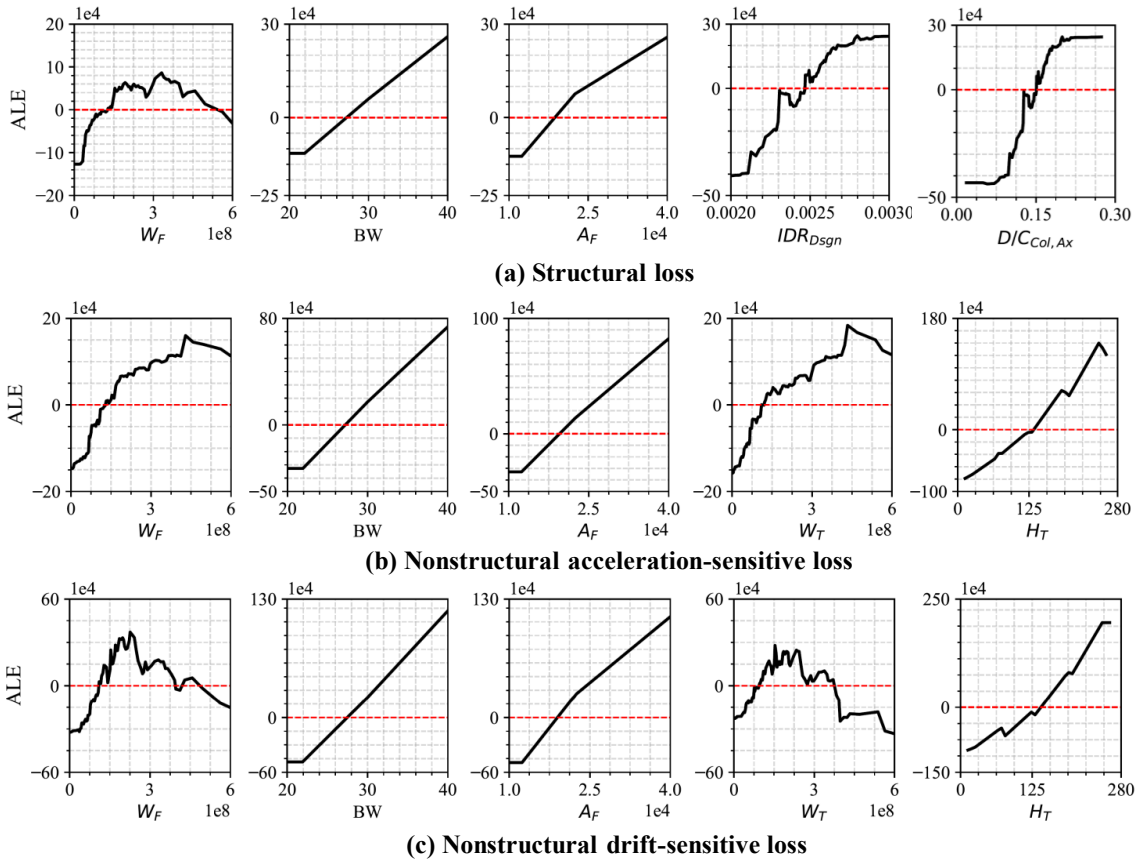

Figure 9. ALE results for (a) structural, (b) nonstructural acceleration-sensitive, and (c) total loss.

5. Challenges and Future Directions for XML in Structural Engineering

1. Current research efforts are directed at post-hoc explainability, with limited attention given to ante-hoc explainability (i.e., models with transparent inner workings). Future research should aim to develop and apply interpretable models, such as decision trees, to combine explainability with improved accuracy, thereby further enhancing their acceptance in structural engineering practice.
2. Most studies only use a single XML method, typically SHAP, for explaining ML models, with few comprehensive comparisons of different XML techniques for structure-level assessment. To advance the field, future research should prioritize systematic comparisons of various XML methods to better understand their strengths and limitations, and conduct benchmark studies to compare the performance of various XML techniques.
3. The current literature of XML for structural design and assessment against natural hazards emphasizes capacity, response/demand, and loss assessment, while resilience assessment remains largely unexplored. Future research should also apply and benchmark XML techniques across a broader spectrum of structural and geotechnical applications, particularly in areas like wind, fire, and soil-structure interaction, evaluations of recovery processes, and decision-making in multi-hazard scenarios.

4. Future research should also explore more computationally efficient XML algorithms, such as sampling-based approximations or surrogate modeling, and evaluate trade-offs between fidelity and computational feasibility.
5. XML techniques are often not intended to provide causal explanations. While non-causal explanations are still useful, understanding causal relationships would benefit many engineering problems, enabling more generalizable ML solutions [197–199]. Future efforts can focus on integrating causal inference methods with XML frameworks to communicate causal relationships. For example, investigation of the distinction between correlation-based explanations and causal reasoning; The reason that causal explanations are important to provide insights in engineering design and risk mitigation; and case study of applications that causal XML integration is helpful to understand the effects of design on fragility.
6. Most existing XML studies are based on the ML models trained on relatively small or synthetic datasets. However, the performance of XML techniques relies on the predictive performance of ML models. In this regard, the quality of training data and the robustness/generalization of ML models have been a concern for XML-based structure-level assessment. Compared to material-level and component-level data, structure-level experimental data is significantly less documented due to high costs, limiting its availability for ML-based assessments. While monitoring data is more abundant, challenges arise from the limited occurrence of extreme hazard events and restricted data accessibility, which can impact the reliability of ML models. To enhance the robustness and generalization of ML models for structural assessment, future efforts should focus on developing extensive, high-quality databases. Ideally, the databases incorporate experimental, field monitoring, and post-event reconnaissance data.

6. Conclusion

This paper reviews XML definitions, classifications, and techniques, and provides three structural-engineering-specific examples at different scales to illustrate how XML can connect the purely data-driven nature of ML to engineering-oriented insights. Despite the lack of a universal definition for XML, the literature often distinguishes between "interpretability" and "explainability". In this context, interpretable models are designed to be inherently understandable, whereas explainability refers to post-hoc explanations of a complex model through different local and global techniques. Different taxonomies exist to classify XML techniques, including functionality-based, conceptual, result-based, and mixed taxonomies; however, none of these taxonomies have been tailored to the domain-specific needs of structural design and assessment. A targeted literature review of structural engineering literature demonstrates the dominance of post-hoc explainability through global model-agnostic methods, particularly SHAP, across various applications.

Applying XML techniques to three case studies highlighted their potential to infer engineering-oriented insights beyond establishing a predictive or classification model. The component-level case study demonstrated the feasibility of utilizing ante-hoc XML to provide more accurate and robust predictive models. By generating highly accurate equations based on existing datasets, regardless of the size, it is possible to define the relationship between key input parameters and performance. Applying post-hoc explainability to structure- and inventory-level problems demonstrated its capability to identify critical features, as well as the type of relationship they may have with the target variable, across black-box models such as random forests and neural networks. For example, feature importance-based explanations suggested

that seismic intensity measures (particularly AvgSa) were more important than structure-related parameters for peak column drift. Additionally, increasing *AvgSa*, PGD, PGV, number of spans, and column axial load ratio increases columns' peak drift responses, whereas column diameter and reinforcement ratio had a negligible impact. Lastly, applying ALE to the steel inventory problem revealed a linear relationship between structural losses and bay width and floor area, as well as a nonlinear increasing trend between the same loss and the column axial demand-to-capacity ratio and design drift. Lastly, weight-related design parameters, building height, and floor area affect all types of losses.

The paper identifies several future research directions. Despite the overall similarity between the outcomes of different post-hoc explainability techniques, there exist instances where various techniques rank a specific feature differently, suggesting the need for benchmarked comparison across diverse XML techniques to understand their limitations. While XML has been, to some extent, applied for seismic assessment, its application to other hazards or multi-hazard decision-making remains largely underexplored. Lastly, the computational cost and scalability are another critical consideration for the practical adoption of XML techniques in structural engineering, particularly for complex ML models such as large-scale infrastructure inventories**.**

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] Kareem A. Emerging frontiers in wind engineering: Computing, stochastics, machine learning and beyond. Journal of Wind Engineering and Industrial Aerodynamics 2020;206:104320. https://doi.org/10.1016/j.jweia.2020.104320.

[2] Salehi H, Burgueño R. Emerging artificial intelligence methods in structural engineering. Engineering Structures 2018;171:170–89. https://doi.org/10.1016/j.engstruct.2018.05.084.

[3] Kareem A. Emerging frontiers in wind engineering: Computing, stochastics, machine learning and beyond. Journal of Wind Engineering and Industrial Aerodynamics 2020;206:104320. https://doi.org/10.1016/j.jweia.2020.104320.

[4] Salehi H, Burgueño R. Emerging artificial intelligence methods in structural engineering. Engineering Structures 2018;171:170–89. https://doi.org/10.1016/j.engstruct.2018.05.084.

[5] Wang X, Mazumder RK, Salarieh B, Salman AM, Shafieezadeh A, Li Y. Machine Learning for Risk and Resilience Assessment in Structural Engineering: Progress and Future Trends. Journal of Structural Engineering 2022;148:03122003. https://doi.org/10.1061/(ASCE)ST.1943-541X.0003392.

[6] Zaker Esteghamati M, Bean B, Burton H, Naser MZ. Beyond development: Challenges in deploying machine learning models for structural engineering applications. Journal of Structural Engineering 2024. https://doi.org/10.1061/JSENDH/STENG-13301.

[7] Dwivedi R, Dave D, Naik H, Singhal S, Omer R, Patel P, et al. Explainable AI (XAI): Core Ideas, Techniques, and Solutions. ACM Comput Surv 2023;55:194:1-194:33. https://doi.org/10.1145/3561048.

[8] Minh D, Wang HX, Li YF, Nguyen TN. Explainable artificial intelligence: a comprehensive review. Artif Intell Rev 2022;55:3503–68. https://doi.org/10.1007/s10462-021-10088-y.

[9] Nogueira AR, Pugnana A, Ruggieri S, Pedreschi D, Gama J. Methods and tools for causal discovery and causal inference. WIREs Data Mining and Knowledge Discovery 2022;12:e1449. https://doi.org/10.1002/widm.1449.

[10] Marcinkevičs R, Vogt JE. Interpretable and explainable machine learning: A methods-centric overview with concrete examples. WIREs Data Mining and Knowledge Discovery 2023;13:e1493. https://doi.org/10.1002/widm.1493.
[11] Dignum V. Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Cham: Springer International Publishing; 2019. https://doi.org/10.1007/978-3-030-30371-6.
[12] Lisboa PJG, Saralajew S, Vellido A, Fernández-Domenech R, Villmann T. The coming of age of interpretable and explainable machine learning models. Neurocomputing 2023;535:25–39. https://doi.org/10.1016/j.neucom.2023.02.040.
[13] Lepri B, Oliver N, Letouzé E, Pentland A, Vinck P. Fair, Transparent, and Accountable Algorithmic Decision-making Processes. Philos Technol 2018;31:611–27. https://doi.org/10.1007/s13347-017-0279-x.
[14] Belle V, Papantonis I. Principles and Practice of Explainable Machine Learning. Front Big Data 2021;4. https://doi.org/10.3389/fdata.2021.688969.
[15] Bhatt U, Xiang A, Sharma S, Weller A, Taly A, Jia Y, et al. Explainable machine learning in deployment. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, New York, NY, USA: Association for Computing Machinery; 2020, p. 648–57. https://doi.org/10.1145/3351095.3375624.
[16] Brdnik S, Šumak B. Current Trends, Challenges and Techniques in XAI Field; A Tertiary Study of XAI Research. 2024 47th MIPRO ICT and Electronics Convention (MIPRO), 2024, p. 2032–8. https://doi.org/10.1109/MIPRO60963.2024.10569528.
[17] Málaga-Chuquitaype C. Machine Learning in Structural Design: An Opinionated Review. Frontiers in Built Environment 2022;8.
[18] Sun H, Burton HV, Huang H. Machine learning applications for building structural design and performance assessment: State-of-the-art review. Journal of Building Engineering 2021;33:101816. https://doi.org/10.1016/j.jobe.2020.101816.
[19] Thai H-T. Machine learning for structural engineering: A state-of-the-art review. Structures 2022;38:448–91. https://doi.org/10.1016/j.istruc.2022.02.003.
[20] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 2020;58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012.
[21] Bruckert S, Finzel B, Schmid U. The Next Generation of Medical Decision Support: A Roadmap Toward Transparent Expert Companions. Front Artif Intell 2020;3:507973. https://doi.org/10.3389/frai.2020.507973.
[22] Doshi-Velez F, Kim B. Towards A Rigorous Science of Interpretable Machine Learning 2017. https://doi.org/10.48550/arXiv.1702.08608.
[23] Páez A. The Pragmatic Turn in Explainable Artificial Intelligence (XAI). Minds & Machines 2019;29:441–59. https://doi.org/10.1007/s11023-019-09502-w.
[24] Lipton ZC. The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue 2018;16:31–57. https://doi.org/10.1145/3236386.3241340.
[25] Kulesza T, Burnett M, Wong W-K, Stumpf S. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. Proceedings of the 20th International Conference on Intelligent User Interfaces, Atlanta Georgia USA: ACM; 2015, p. 126–37. https://doi.org/10.1145/2678025.2701399.
[26] Schmid U, Finzel B. Mutual Explanations for Cooperative Decision Making in Medicine. Künstl Intell 2020;34:227–33. https://doi.org/10.1007/s13218-020-00633-2.
[27] Teso S, Kersting K. Explanatory Interactive Machine Learning. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, New York, NY, USA: Association for Computing Machinery; 2019, p. 239–45. https://doi.org/10.1145/3306618.3314293.
[28] Subhash V, Chen Z, Havasi M, Pan W, Doshi-Velez F. What makes a good explanation?: A harmonized view of properties of explanations, 2022.

[29] Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA: Association for Computing Machinery; 2016, p. 1135–44. https://doi.org/10.1145/2939672.2939778.
[30] Jacovi A, Goldberg Y. Towards Faithfully Interpretable NLP Systems: How should we define and evaluate faithfulness? 2020. https://doi.org/10.48550/arXiv.2004.03685.
[31] Pentenrieder A, Hahn P, Schaffrath S, Krieger B, Brzoska S, Peters R, et al. Designing Explainable and Controllable Artificial Intelligence Systems Together: Inclusive Participation Formats for Software-Based Working Routines in Industry. In: Shajek A, Hartmann EA, editors. New Digital Work: Digital Sovereignty at the Workplace, Cham: Springer International Publishing; 2023, p. 135–48. https://doi.org/10.1007/978-3-031-26490-0_8.
[32] Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. Journal of Biomedical Informatics 2021;113:103655.
[33] Gunning D. Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), Nd Web 2017;2:1.
[34] Hu ZF, Kuflik T, Mocanu IG, Najafian S, Shulner Tal A. Recent Studies of XAI - Review. Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, New York, NY, USA: Association for Computing Machinery; 2021, p. 421–31. https://doi.org/10.1145/3450614.3463354.
[35] Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining Explanations: An Overview of Interpretability of Machine Learning. 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), 2018, p. 80–9. https://doi.org/10.1109/DSAA.2018.00018.
[36] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1:206–15. https://doi.org/10.1038/s42256-019-0048-x.
[37] Retzlaff CO, Angerschmid A, Saranti A, Schneeberger D, Röttger R, Müller H, et al. Post-hoc vs ante-hoc explanations: xAI design guidelines for data scientists. Cognitive Systems Research 2024;86:101243. https://doi.org/10.1016/j.cogsys.2024.101243.
[38] Lipton ZC. The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue 2018;16:31–57. https://doi.org/10.1145/3236386.3241340.
[39] Peng X, Li Y, Tsang IW, Zhu H, Lv J, Zhou JT. XAI Beyond Classification: Interpretable Neural Clustering. Journal of Machine Learning Research 2022;23:1–28.
[40] Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A Survey of Methods for Explaining Black Box Models. ACM Comput Surv 2018;51:93:1-93:42. https://doi.org/10.1145/3236009.
[41] Angelov PP, Soares EA, Jiang R, Arnold NI, Atkinson PM. Explainable artificial intelligence: an analytical review. WIREs Data Mining and Knowledge Discovery 2021;11:e1424. https://doi.org/10.1002/widm.1424.
[42] Liang Y, Li S, Yan C, Li M, Jiang C. Explaining the black-box model: A survey of local interpretation methods for deep neural networks. Neurocomputing 2021;419:168–82. https://doi.org/10.1016/j.neucom.2020.08.011.
[43] Bassan S, Amir G, Katz G. Local vs. global interpretability: a computational complexity perspective. Proceedings of the 41st International Conference on Machine Learning, vol. 235, Vienna, Austria: JMLR.org; 2024, p. 3133–67.
[44] Speith T. A Review of Taxonomies of Explainable Artificial Intelligence (XAI) Methods. 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul Republic of Korea: ACM; 2022, p. 2239–50. https://doi.org/10.1145/3531146.3534639.

[45] Samek W, Montavon G, Vedaldi A, Hansen LK, Müller K-R, editors. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Cham: Springer International Publishing; 2019. https://doi.org/10.1007/978-3-030-28954-6.
[46] Binder A, Montavon G, Lapuschkin S, Müller K-R, Samek W. Layer-Wise Relevance Propagation for Neural Networks with Local Renormalization Layers. In: Villa AEP, Masulli P, Pons Rivero AJ, editors. Artificial Neural Networks and Machine Learning – ICANN 2016, Cham: Springer International Publishing; 2016, p. 63–71. https://doi.org/10.1007/978-3-319-44781-0_8.
[47] Samek W, Müller K-R. Towards Explainable Artificial Intelligence. In: Samek W, Montavon G, Vedaldi A, Hansen LK, Müller K-R, editors. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Cham: Springer International Publishing; 2019, p. 5–22. https://doi.org/10.1007/978-3-030-28954-6_1.
[48] Lapuschkin S, Wäldchen S, Binder A, Montavon G, Samek W, Müller K-R. Unmasking Clever Hans predictors and assessing what machines really learn. Nat Commun 2019;10:1096. https://doi.org/10.1038/s41467-019-08987-4.
[49] McDermid JA, Jia Y, Porter Z, Habli I. Artificial intelligence explainability: the technical and ethical dimensions. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 2021;379:20200363. https://doi.org/10.1098/rsta.2020.0363.
[50] Pedreschi D, Giannotti F, Guidotti R, Monreale A, Ruggieri S, Turini F. Meaningful Explanations of Black Box AI Decision Systems. Proceedings of the AAAI Conference on Artificial Intelligence 2019;33:9780–4. https://doi.org/10.1609/aaai.v33i01.33019780.
[51] Dieber J, Kirrane S. Why model why? Assessing the strengths and limitations of LIME 2020. https://doi.org/10.48550/arXiv.2012.00093.
[52] Langer M, Oster D, Speith T, Hermanns H, Kästner L, Schmidt E, et al. What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. Artificial Intelligence 2021;296:103473. https://doi.org/10.1016/j.artint.2021.103473.
[53] Sokol K, Flach P. Explainability Is in the Mind of the Beholder: Establishing the Foundations of Explainable Artificial Intelligence 2022. https://doi.org/10.48550/arXiv.2112.14466.
[54] Vilone G, Longo L. Explainable Artificial Intelligence: a Systematic Review 2020. https://doi.org/10.48550/arXiv.2006.00093.
[55] Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc.; 2017.
[56] Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics 2001;29:1189–232.
[57] Apley DW, Zhu J. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. Journal of the Royal Statistical Society Series B: Statistical Methodology 2020;82:1059–86. https://doi.org/10.1111/rssb.12377.
[58] Banzhaf W, Nordin P, Keller RE, Francone FD. Genetic programming: an introduction: on the automatic evolution of computer programs and its applications. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1998.
[59] Ferreira C. Gene Expression Programming: a New Adaptive Algorithm for Solving Problems 2001. https://doi.org/10.48550/arXiv.cs/0102027.
[60] Mangalathu S, Hwang S-H, Jeon J-S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Engineering Structures 2020;219:110927. https://doi.org/10.1016/j.engstruct.2020.110927.
[61] Zhou Y, Meng S, Lou Y, Kong Q. Physics-Informed Deep Learning-Based Real-Time Structural Response Prediction Method. Engineering 2024;35:140–57. https://doi.org/10.1016/j.eng.2023.08.011.
[62] Zhang S, Chen W, Xu J, Xie T. Use of interpretable machine learning approaches for quantificationally understanding the performance of steel fiber-reinforced recycled aggregate concrete: From the perspective of compressive strength and splitting tensile strength. Engineering

Applications of Artificial Intelligence 2024;137:109170. https://doi.org/10.1016/j.engappai.2024.109170.
[63] Chen J, Liu Y. Probabilistic physics-guided machine learning for fatigue data analysis. Expert Systems with Applications 2021;168:114316. https://doi.org/10.1016/j.eswa.2020.114316.
[64] Figueiredo E, Moldovan I, Santos A, Campos P, Costa JCWA. Finite Element–Based Machine-Learning Approach to Detect Damage in Bridges under Operational and Environmental Variations. Journal of Bridge Engineering 2019;24:04019061. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001432.
[65] Mazumder RK, Salman AM, Li Y. Reliability Assessment of Oil and Gas Pipeline Systems at Burst Limit State Under Active Corrosion. In: Matos JC, Lourenço PB, Oliveira DV, Branco J, Proske D, Silva RA, et al., editors. 18th International Probabilistic Workshop, Cham: Springer International Publishing; 2021, p. 653–60. https://doi.org/10.1007/978-3-030-73616-3_50.
[66] Liu Y, Bao Y. Review on automated condition assessment of pipelines with machine learning. Advanced Engineering Informatics 2022;53:101687. https://doi.org/10.1016/j.aei.2022.101687.
[67] Mazumder RK, Modanwal G, Li Y. Synthetic Data Generation Using Generative Adversarial Network for Burst Failure Risk Analysis of Oil and Gas Pipelines. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering 2023;9:031103. https://doi.org/10.1115/1.4062741.
[68] Kapteyn MG, Knezevic DJ, Willcox K. Toward predictive digital twins via component-based reduced-order models and interpretable machine learning. AIAA Scitech 2020 Forum, American Institute of Aeronautics and Astronautics; 2020. https://doi.org/10.2514/6.2020-0418.
[69] Vadyala SR, Betgeri SN, Matthews JC, Matthews E. A review of physics-based machine learning in civil engineering. Results in Engineering 2022;13:100316. https://doi.org/10.1016/j.rineng.2021.100316.
[70] Nguyen DH, Nguyen TH, Tran KD, Tran KP. Physics-Informed Machine Learning for Industrial Reliability and Safety Engineering: A Review and Perspective. Springer Series in Reliability Engineering 2024:5–23.
[71] Luleci F, Necati Catbas F, Avci O. CycleGAN for undamaged-to-damaged domain translation for structural health monitoring and damage detection. Mechanical Systems and Signal Processing 2023;197:110370. https://doi.org/10.1016/j.ymssp.2023.110370.
[72] Xu Y, Kohtz S, Boakye J, Gardoni P, Wang P. Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. Reliability Engineering & System Safety 2023;230:108900. https://doi.org/10.1016/j.ress.2022.108900.
[73] Zemmouchi-Ghomari L. Ontology and Machine Learning: A Two-Way Street to Improved Knowledge Representation and Algorithm Accuracy. In: Yadav A, Nanda SJ, Lim M-H, editors. Proceedings of International Conference on Paradigms of Communication, Computing and Data Analytics, Singapore: Springer Nature; 2023, p. 181–9. https://doi.org/10.1007/978-981-99-4626-6_15.
[74] Hu X, Liu K. Structural Deterioration Knowledge Ontology towards Physics-Informed Machine Learning for Enhanced Bridge Deterioration Prediction. Journal of Computing in Civil Engineering 2023;37:04022051. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001066.
[75] Wakjira TG, Alam MS, Ebead U. Plastic hinge length of rectangular RC columns using ensemble machine learning model. Engineering Structures 2021;244:112808. https://doi.org/10.1016/j.engstruct.2021.112808.
[76] Wang Z, Liu T, Long Z, Wang J, Zhang J. Predicting the drift capacity of precast concrete columns using explainable machine learning approach. Engineering Structures 2023;282:115771. https://doi.org/10.1016/j.engstruct.2023.115771.
[77] Mansouri A, Mansouri M, Mangalathu S. Interpretable machine learning model for shear strength estimation of circular concrete-filled steel tubes. The Structural Design of Tall and Special Buildings 2024;33:e2111. https://doi.org/10.1002/tal.2111.

[78] Bakouregui AS, Mohamed HM, Yahia A, Benmokrane B. Explainable extreme gradient boosting tree-based prediction of load-carrying capacity of FRP-RC columns. Engineering Structures 2021;245:112836. https://doi.org/10.1016/j.engstruct.2021.112836.
[79] Cakiroglu C, Islam K, Bekdaş G, Isikdag U, Mangalathu S. Explainable machine learning models for predicting the axial compression capacity of concrete filled steel tubular columns. Construction and Building Materials 2022;356:129227. https://doi.org/10.1016/j.conbuildmat.2022.129227.
[80] Hu T, Zhang H, Cheng C, Li H, Zhou J. Explainable machine learning: Compressive strength prediction of FRP-confined concrete column. Materials Today Communications 2024;39:108883. https://doi.org/10.1016/j.mtcomm.2024.108883.
[81] Sun Y. Estimation of compressive strength for spiral stirrup-confined circular concrete column using optimized machine learning with interpretable techniques. Mechanics of Advanced Materials and Structures 2024;31:10839–58. https://doi.org/10.1080/15376494.2023.2298232.
[82] Zarringol M, Patel VI, Liang QQ, Hassanein MF, Ahmed M. Machine-learning-based predictive models for concrete-filled double skin tubular columns. Engineering Structures 2024;304:117593. https://doi.org/10.1016/j.engstruct.2024.117593.
[83] Zhou X-G, Hou C, Peng J, Yao G-H, Fang Z. Structural mechanism-based intelligent capacity prediction methods for concrete-encased CFST columns. Journal of Constructional Steel Research 2023;202:107769. https://doi.org/10.1016/j.jcsr.2022.107769.
[84] Feng D-C, Wang W-J, Mangalathu S, Taciroglu E. Interpretable XGBoost-SHAP Machine-Learning Model for Shear Strength Prediction of Squat RC Walls. Journal of Structural Engineering 2021;147:04021173. https://doi.org/10.1061/(ASCE)ST.1943-541X.0003115.
[85] Chen Q, Yu B, Li B. Quantifying Seismic Damage in RC Walls with Image Analysis. Journal of Earthquake Engineering 2025;29:156–79. https://doi.org/10.1080/13632469.2024.2410946.
[86] Zhang H, Liu J, Wang S, Chen K, Xu L, Ma J, et al. Prediction and optimization framework of shear strength of reinforced concrete flanged shear wall based on machine learning and non-dominated sorting genetic algorithm-II. Advances in Structural Engineering 2025;28:275–95. https://doi.org/10.1177/13694332241281534.
[87] Wu Y, Zhou Y. Prediction and feature analysis of punching shear strength of two-way reinforced concrete slabs using optimized machine learning algorithm and Shapley additive explanations. Mechanics of Advanced Materials and Structures 2023;30:3086–96. https://doi.org/10.1080/15376494.2022.2068209.
[88] Mangalathu S, Shin H, Choi E, Jeon J-S. Explainable machine learning models for punching shear strength estimation of flat slabs without transverse reinforcement. Journal of Building Engineering 2021;39:102300. https://doi.org/10.1016/j.jobe.2021.102300.
[89] Wakjira TG, Al-Hamrani A, Ebead U, Alnahhal W. Shear capacity prediction of FRP-RC beams using single and ensenble ExPlainable Machine learning models. Composite Structures 2022;287:115381. https://doi.org/10.1016/j.compstruct.2022.115381.
[90] Zhang S-Y, Chen S-Z, Jiang X, Han W-S. Data-driven prediction of FRP strengthened reinforced concrete beam capacity based on interpretable ensemble learning algorithms. Structures 2022;43:860–77. https://doi.org/10.1016/j.istruc.2022.07.025.
[91] Yehia SA, Fayed S, Zakaria MH, Shahin RI. Prediction of RC T-Beams Shear Strength Based on Machine Learning. Int J Concr Struct Mater 2024;18:52. https://doi.org/10.1186/s40069-024-00690-z.
[92] Ashour AF, Alvarez LF, Toropov VV. Empirical modelling of shear strength of RC deep beams by genetic programming. Computers & Structures 2003;81:331–8. https://doi.org/10.1016/S0045-7949(02)00437-6.
[93] Ben Chaabene W, Nehdi ML. Genetic programming based symbolic regression for shear capacity prediction of SFRC beams. Construction and Building Materials 2021;280:122523. https://doi.org/10.1016/j.conbuildmat.2021.122523.

[94] Gandomi AH, Mohammadzadeh S. D, Pérez-Ordóñez JL, Alavi AH. Linear genetic programming for shear strength prediction of reinforced concrete beams without stirrups. Applied Soft Computing 2014;19:112–20. https://doi.org/10.1016/j.asoc.2014.02.007.
[95] Jia J-F, Chen X-Z, Bai Y-L, Li Y-L, Wang Z-H. An interpretable ensemble learning method to predict the compressive strength of concrete. Structures 2022;46:201–13. https://doi.org/10.1016/j.istruc.2022.10.056.
[96] Lyngdoh GA, Zaki M, Krishnan NMA, Das S. Prediction of concrete strengths enabled by missing data imputation and interpretable machine learning. Cement and Concrete Composites 2022;128:104414. https://doi.org/10.1016/j.cemconcomp.2022.104414.
[97] Liang M, Chang Z, Wan Z, Gan Y, Schlangen E, Šavija B. Interpretable Ensemble-Machine-Learning models for predicting creep behavior of concrete. Cement and Concrete Composites 2022;125:104295. https://doi.org/10.1016/j.cemconcomp.2021.104295.
[98] Ge P, Yang O, Hua X, Chen Z, He J, Liu Z, et al. Predicting bond strength of corroded reinforced concrete after high-temperature exposure: A stacking model and feature selection. Construction and Building Materials 2024;456:139290. https://doi.org/10.1016/j.conbuildmat.2024.139290.
[99] Hu T, Zhang H, Khodadadi N, Taffese WZ, Nanni A. Enhancing bond strength prediction at UHPC-NC interface: A data-driven approach with augmentation and explainability. Construction and Building Materials 2024;451:138757. https://doi.org/10.1016/j.conbuildmat.2024.138757.
[100] Ke L, Qiu M, Chen Z, Zhou J, Feng Z, Long J. An interpretable machine learning model for predicting bond strength of CFRP-steel epoxy-bonded interface. Composite Structures 2023;326:117639. https://doi.org/10.1016/j.compstruct.2023.117639.
[101] Yan F, Song K, Liu Y, Chen S, Chen J. Predictions and mechanism analyses of the fatigue strength of steel based on machine learning. J Mater Sci 2020;55:15334–49. https://doi.org/10.1007/s10853-020-05091-7.
[102] Lu G, Zhang B, Li Z, Su J, Hui Y, Wu G. An advanced ANN-based framework for seismic damage prediction and fragility analysis of reinforced concrete double-column piers in highway bridge networks. Structures 2025;78:109233. https://doi.org/10.1016/j.istruc.2025.109233.
[103] Zaker Esteghamati M, Gernay T, Banerji S. Evaluating fire resistance of timber columns using explainable machine learning models. Engineering Structures 2023;296:116910. https://doi.org/10.1016/j.engstruct.2023.116910.
[104] Mangalathu S, Karthikeyan K, Feng D-C, Jeon J-S. Machine-learning interpretability techniques for seismic performance assessment of infrastructure systems. Engineering Structures 2022;250:112883. https://doi.org/10.1016/j.engstruct.2021.112883.
[105] Yu Y, Chen Y, Liao W, Wang Z, Zhang S, Kang Y, et al. Intelligent generation and interpretability analysis of shear wall structure design by learning from multidimensional to high-dimensional features. Engineering Structures 2025;325:119472. https://doi.org/10.1016/j.engstruct.2024.119472.
[106] Gao X, Lin C. Prediction model of the failure mode of beam-column joints using machine learning methods. Engineering Failure Analysis 2021;120:105072. https://doi.org/10.1016/j.engfailanal.2020.105072.
[107] Ye M, Li L, Jin W, Tang J, Yoo D-Y, Zhou C. Interpretable ensemble machine learning models for predicting the shear capacity of UHPC joints. Engineering Structures 2024;315:118443. https://doi.org/10.1016/j.engstruct.2024.118443.
[108] Braun M, Kellner L, Schreiber S, Ehlers S. Prediction of fatigue failure in small-scale butt-welded joints with explainable machine learning. Procedia Structural Integrity 2022;38:182–91. https://doi.org/10.1016/j.prostr.2022.03.019.
[109] Karakaş S, Taşkın G, Ülker MBC. Re-evaluation of machine learning models for predicting ultimate bearing capacity of piles through SHAP and Joint Shapley methods. Neural Comput & Applic 2024;36:697–715. https://doi.org/10.1007/s00521-023-09053-3.

[110] Wang J, Ye A, Wang X, Li Y. Machine learning-driven efficient identification and analysis of seismic energy dissipation mechanisms for scoured bridge pile-group foundations in cohesionless soils. Engineering Mechanics (In Chinese) 2024.
[111] Movsessian A, Cava DG, Tcherniak D. Interpretable Machine Learning in Damage Detection Using Shapley Additive Explanations. ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg 2022;8. https://doi.org/10.1115/1.4053304.
[112] Cakiroglu C, Islam K, Bekdaş G, Nehdi ML. Data-driven ensemble learning approach for optimal design of cantilever soldier pile retaining walls. Structures 2023;51:1268–80. https://doi.org/10.1016/j.istruc.2023.03.109.
[113] Iqbal M, Zhang D, Khan MI, Zahid M, Jalal FE. Effects of Rebar Size and Volume Fraction of Glass Fibers on Tensile Strength Retention of GFRP Rebars in Alkaline Environment via RSM and SHAP Analyses. Journal of Materials in Civil Engineering 2023;35:04023318. https://doi.org/10.1061/JMCEE7.MTENG-15589.
[114] Elhishi S, Elashry AM, El-Metwally S. Unboxing machine learning models for concrete strength prediction using XAI. Sci Rep 2023;13:19892. https://doi.org/10.1038/s41598-023-47169-7.
[115] Aslam F, Farooq F, Amin MN, Khan K, Waheed A, Akbar A, et al. Applications of Gene Expression Programming for Estimating Compressive Strength of High-Strength Concrete. Advances in Civil Engineering 2020;2020:8850535. https://doi.org/10.1155/2020/8850535.
[116] Javed MF, Amin MN, Shah MI, Khan K, Iftikhar B, Farooq F, et al. Applications of Gene Expression Programming and Regression Techniques for Estimating Compressive Strength of Bagasse Ash based Concrete. Crystals 2020;10:737. https://doi.org/10.3390/cryst10090737.
[117] Chu H-H, Khan MA, Javed M, Zafar A, Ijaz Khan M, Alabduljabbar H, et al. Sustainable use of fly-ash: Use of gene-expression programming (GEP) and multi-expression programming (MEP) for forecasting the compressive strength geopolymer concrete. Ain Shams Engineering Journal 2021;12:3603–17. https://doi.org/10.1016/j.asej.2021.03.018.
[118] Inqiad WB, Javed MF, Siddique MS, Alabduljabbar H, Ahmed B, Alkhattabi L. Predicting natural vibration period of concrete frame structures having masonry infill using machine learning techniques. Journal of Building Engineering 2024;96:110417. https://doi.org/10.1016/j.jobe.2024.110417.
[119] Latif I, Banerjee A, Surana M. Explainable machine learning aided optimization of masonry infilled reinforced concrete frames. Structures 2022;44:1751–66. https://doi.org/10.1016/j.istruc.2022.08.115.
[120] Somala SN, Karthikeyan K, Mangalathu S. Time period estimation of masonry infilled RC frames using machine learning techniques. Structures 2021;34:1560–6. https://doi.org/10.1016/j.istruc.2021.08.088.
[121] Thisovithan P, Aththanayake H, Meddage DPP, Ekanayake IU, Rathnayake U. A novel explainable AI-based approach to estimate the natural period of vibration of masonry infill reinforced concrete frame structures using different machine learning techniques. Results in Engineering 2023;19:101388. https://doi.org/10.1016/j.rineng.2023.101388.
[122] Nguyen HD, LaFave JM, Lee Y-J, Shin M. Rapid seismic damage-state assessment of steel moment frames using machine learning. Engineering Structures 2022;252:113737. https://doi.org/10.1016/j.engstruct.2021.113737.
[123] Parvizi M, Nasserasadi K, Tafakori E. Development of fragility functions of low-rise steel moment frame by artificial neural networks and identifying effective parameters using SHAP theory. Structures 2023;58:105315. https://doi.org/10.1016/j.istruc.2023.105315.
[124] Latif I, Surana M, Banerjee A. Effects of material properties uncertainty on seismic fragility of reinforced-concrete frames using machine learning approach. Journal of Building Engineering 2024;86:108871. https://doi.org/10.1016/j.jobe.2024.108871.
[125] Gan Y, Chen J, Li Y, Xu Z. Prediction of progressive collapse resistance of RC frames using deep and cross network model. Structures 2023;51:800–13. https://doi.org/10.1016/j.istruc.2023.03.087.

[126] Lazaridis PC, Kavvadias IE, Demertzis K, Iliadis L, Vasiliadis LK. Interpretable Machine Learning for Assessing the Cumulative Damage of a Reinforced Concrete Frame Induced by Seismic Sequences. Sustainability 2023;15:12768. https://doi.org/10.3390/su151712768.
[127] Yan B, Ding W, Jin Z, Zhang L, Wang L, Du M, et al. Explainable machine learning-based prediction for aerodynamic interference of a low-rise building on a high-rise building. Journal of Building Engineering 2024;82:108285. https://doi.org/10.1016/j.jobe.2023.108285.
[128] Lei X, Feng R, Dong Y, Zhai C. Bayesian-optimized interpretable surrogate model for seismic demand prediction of urban highway bridges. Engineering Structures 2024;301:117307. https://doi.org/10.1016/j.engstruct.2023.117307.
[129] Gautam D, Bhattarai A, Rupakhety R. Machine learning and soft voting ensemble classification for earthquake induced damage to bridges. Engineering Structures 2024;303:117534. https://doi.org/10.1016/j.engstruct.2024.117534.
[130] Gautam D, Rupakhety R. Nonlinear tree based regression ensemble modeling for repair cost prediction in earthquake damaged RC bridges. Soil Dynamics and Earthquake Engineering 2024;187:108947. https://doi.org/10.1016/j.soildyn.2024.108947.
[131] Ben Seghier MEA, Mohamed OA, Ouaer H. Machine learning-based Shapley additive explanations approach for corroded pipeline failure mode identification. Structures 2024;65:106653. https://doi.org/10.1016/j.istruc.2024.106653.
[132] Tang Q, Cui Y, Jia J. Machine learning-based surrogate resilience modeling for preliminary seismic design. Journal of Building Engineering 2024;98:111226. https://doi.org/10.1016/j.jobe.2024.111226.
[133] Angelucci G, Quaranta G, Mollaioli F, Kunnath SK. Interpretable machine learning models for displacement demand prediction in reinforced concrete buildings under pulse-like earthquakes. Journal of Building Engineering 2024;95:110124. https://doi.org/10.1016/j.jobe.2024.110124.
[134] Demertzis K, Kostinakis K, Morfidis K, Iliadis L. An interpretable machine learning method for the prediction of R/C buildings' seismic response. Journal of Building Engineering 2023;63:105493. https://doi.org/10.1016/j.jobe.2022.105493.
[135] Fayaz J, Torres-Rodas P, Medalla M, Naeim F. Assessment of ground motion amplitude scaling using interpretable Gaussian process regression: Application to steel moment frames. Earthquake Engineering & Structural Dynamics 2023;52:2339–59. https://doi.org/10.1002/eqe.3810.
[136] Junda E, Málaga-Chuquitaype C, Chawgien K. Interpretable machine learning models for the estimation of seismic drifts in CLT buildings. Journal of Building Engineering 2023;70:106365. https://doi.org/10.1016/j.jobe.2023.106365.
[137] Ke K, Zhou X, Zhu M, Yam MCH, Zhang H. Seismic demand amplification of steel frames with SMAs induced by earthquake sequences. Journal of Constructional Steel Research 2023;207:107929. https://doi.org/10.1016/j.jcsr.2023.107929.
[138] Xing L, Gardoni P, Zhou Y, Zhang P. DNN-metamodeling and fragility estimate of high-rise buildings with outrigger systems subject to seismic loads. Reliability Engineering & System Safety 2025;253:110572. https://doi.org/10.1016/j.ress.2024.110572.
[139] Koc K, Budayan C, Ekmekcioğlu Ö, Tokdemir OB. Predicting Cost Impacts of Nonconformances in Construction Projects Using Interpretable Machine Learning. Journal of Construction Engineering and Management 2024;150:04023143. https://doi.org/10.1061/JCEMD4.COENG-13857.
[140] Aloisio A, Santis YD, Irti F, Pasca DP, Scimia L, Fragiacomo M. Machine learning predictions of code-based seismic vulnerability for reinforced concrete and masonry buildings: Insights from a 300-building database. Engineering Structures 2024;301:117295. https://doi.org/10.1016/j.engstruct.2023.117295.
[141] Chen W, Zhang L. Building vulnerability assessment in seismic areas using ensemble learning: A Nepal case study. Journal of Cleaner Production 2022;350:131418. https://doi.org/10.1016/j.jclepro.2022.131418.

[142] Zhang T, Xu W, Wang S, Du D, Tang J. Seismic response prediction of a damped structure based on data-driven machine learning methods. Engineering Structures 2024;301:117264. https://doi.org/10.1016/j.engstruct.2023.117264.
[143] Liu Z, Lin J. Combinatorial machine learning approaches for high-rise building cost prediction and their interpretability analysis. Journal of Asian Architecture and Building Engineering 2024.
[144] Wang K, Liu J, Quan Y, Ma Z, Chen J, Bai Y. Intelligent evaluation of interference effects between tall buildings based on wind tunnel experiments and explainable machine learning. Journal of Building Engineering 2024;96:110449. https://doi.org/10.1016/j.jobe.2024.110449.
[145] Liu Z, Zhuang Y. An investigation using resampling techniques and explainable machine learning to minimize fire losses in residential buildings. Journal of Building Engineering 2024;95:110080. https://doi.org/10.1016/j.jobe.2024.110080.
[146] Wang N, Xu Y, Wang S. Interpretable boosting tree ensemble method for multisource building fire loss prediction. Reliability Engineering & System Safety 2022;225:108587. https://doi.org/10.1016/j.ress.2022.108587.
[147] Zhang B, Wang K, Lu G, Guo W, Liu J, Zhang N, et al. Life-cycle seismic performance analysis of an offshore small-to-medium span bridge based on interpretable machine learning. Structures 2024;70:107511. https://doi.org/10.1016/j.istruc.2024.107511.
[148] Shabbir K, Noureldin M, Sim S-H. Data-driven model for seismic assessment, design, and retrofit of structures using explainable artificial intelligence. Computer-Aided Civil and Infrastructure Engineering 2025;40:281–300. https://doi.org/10.1111/mice.13338.
[149] Wang Q, Geng P, Wang L, He D, Shen H. Machine learning-driven feature importance appraisal of seismic parameters on tunnel damage and seismic fragility prediction. Engineering Applications of Artificial Intelligence 2024;137:109101. https://doi.org/10.1016/j.engappai.2024.109101.
[150] Wang J, Wang M, Wang X, Ye A. Quantifying post-earthquake residual vertical load-carrying capacity (VLCC) of RC bridge bents: Parametric study and development of interpretable machine learning models. Soil Dynamics and Earthquake Engineering 2025;199:109662. https://doi.org/10.1016/j.soildyn.2025.109662.
[151] Esteghamati MZ, Flint MM. Do all roads lead to Rome? A comparison of knowledge-based, data-driven, and physics-based surrogate models for performance-based early design. Engineering Structures 2023;286:116098.
[152] Hwang S-H, Mangalathu S, Shin J, Jeon J-S. Estimation of economic seismic loss of steel moment-frame buildings using a machine learning algorithm. Engineering Structures 2022;254:113877. https://doi.org/10.1016/j.engstruct.2022.113877.
[153] Zaker Esteghamati M, Baddipalli S. Efficiency and explainability of design-oriented machine learning models to estimate seismic response, fragility, and loss of a steel building inventory. Earthquake Engineering & Structural Dynamics 2024.
[154] Zaker Esteghamati M, Flint MM. Developing data-driven surrogate models for holistic performance-based assessment of mid-rise RC frame buildings at early design. Engineering Structures 2021;245:112971. https://doi.org/10.1016/j.engstruct.2021.112971.
[155] Zhang D, Chen Y, Zhang C, Xue G, Zhang J, Zhang M, et al. Prediction of seismic acceleration response of precast segmental self-centering concrete filled steel tube single-span bridges based on machine learning method. Engineering Structures 2023;279:115574. https://doi.org/10.1016/j.engstruct.2022.115574.
[156] Xu J-G, Feng D-C, Mangalathu S, Jeon J-S. Data-driven rapid damage evaluation for life-cycle seismic assessment of regional reinforced concrete bridges. Earthquake Engineering & Structural Dynamics 2022;51:2730–51. https://doi.org/10.1002/eqe.3699.
[157] Akbarnezhad M, Salehi M, DesRoches R. Application of machine learning in seismic fragility assessment of bridges with SMA-restrained rocking columns. Structures 2023;50:1320–37. https://doi.org/10.1016/j.istruc.2023.02.105.

[158] Chen M, Xin J, Tang Q, Hu T, Zhou Y, Zhou J. Explainable machine learning model for load-deformation correlation in long-span suspension bridges using XGBoost-SHAP. Developments in the Built Environment 2024;20:100569. https://doi.org/10.1016/j.dibe.2024.100569.
[159] Sun Z, Feng D-C, Mangalathu S, Wang W-J, Su D. Effectiveness Assessment of TMDs in Bridges under Strong Winds Incorporating Machine-Learning Techniques. Journal of Performance of Constructed Facilities 2022;36:04022036. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001746.
[160] Sun Z, Santos J, Caetano E, Oliveira C. Interpreting cumulative displacement in a suspension bridge with a physics-based characterisation of environment and roadway/railway loads. J Civil Struct Health Monit 2023;13:387–97. https://doi.org/10.1007/s13349-022-00647-4.
[161] Wang Y, Zheng Z, Ji D, Pan X, Tian A. Machine learning-driven probabilistic seismic demand model with multiple intensity measures and applicability in seismic fragility analysis for nuclear power plants. Soil Dynamics and Earthquake Engineering 2023;171:107966. https://doi.org/10.1016/j.soildyn.2023.107966.
[162] Abdollahi A, Li D, Deng J, Amini A. An explainable artificial-intelligence-aided safety factor prediction of road embankments. Engineering Applications of Artificial Intelligence 2024;136:108854. https://doi.org/10.1016/j.engappai.2024.108854.
[163] Murphy JD, Paal SG. A combined transfer learning physics informed interpretable machine learning approach to modelling the shear strength of concrete walls 2025.
[164] Kookalani S, Cheng B, Torres JLC. Structural performance assessment of GFRP elastic gridshells by machine learning interpretability methods. Front Struct Civ Eng 2022;16:1249–66. https://doi.org/10.1007/s11709-022-0858-5.
[165] Chen W. System vulnerability modeling and assessment in seismic-prone areas using machine learning. Nanyang Technological University, 2022. https://doi.org/10.32657/10356/163040.
[166] Johnson PM, Barbour W, Camp JV, Baroud H. Using machine learning to examine freight network spatial vulnerabilities to disasters: A new take on partial dependence plots. Transportation Research Interdisciplinary Perspectives 2022;14:100617. https://doi.org/10.1016/j.trip.2022.100617.
[167] Haggag M, Yosri A, El-Dakhakhni W, Hassini E. Interpretable data-driven model for Climate-Induced Disaster damage prediction: The first step in community resilience planning. International Journal of Disaster Risk Reduction 2022;73:102884. https://doi.org/10.1016/j.ijdrr.2022.102884.
[168] Kwak K, Lee EH. Impact of road transport system on groundwater quality inferred from explainable artificial intelligence (XAI). Science of The Total Environment 2024;917:170388. https://doi.org/10.1016/j.scitotenv.2024.170388.
[169] Movahedi A, Derrible S. Interrelationships between electricity, gas, and water consumption in large-scale buildings. Journal of Industrial Ecology 2021;25:932–47. https://doi.org/10.1111/jiec.13097.
[170] Chen W, Zhang L. An automated machine learning approach for earthquake casualty rate and economic loss prediction. Reliability Engineering & System Safety 2022;225:108645. https://doi.org/10.1016/j.ress.2022.108645.
[171] Roeslin S, Ma Q, Juárez-Garcia H, Gómez-Bernal A, Wicker J, Wotherspoon L. A machine learning damage prediction model for the 2017 Puebla-Morelos, Mexico, earthquake. Earthquake Spectra 2020;36:314–39. https://doi.org/10.1177/8755293020936714.
[172] Chen W, Zhang L. Resilience assessment of regional areas against earthquakes using multi-source information fusion. Reliability Engineering & System Safety 2021;215:107833. https://doi.org/10.1016/j.ress.2021.107833.
[173] Zhang B, Wang K, Yang C, Lu G, Li Y, He H. Enhancing seismic performance of highway bridges group with laminated rubber bearings via artificial neural networks and multi-objective genetic algorithm. Structures 2025;74:108470. https://doi.org/10.1016/j.istruc.2025.108470.
[174] Wanigarathna N, Xie Y, Henjewele C, Morga M, Jones K. Machine learning application to disaster damage repair cost modelling of residential buildings. Construction Management and Economics 2025;43:302–22. https://doi.org/10.1080/01446193.2024.2419413.
[175] Ferreira C. Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. Complex Systems 2001.

[176] Pan SJ, Yang Q. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering 2010;22:1345–59. https://doi.org/10.1109/TKDE.2009.191.
[177] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE 2021;109:43–76. https://doi.org/10.1109/JPROC.2020.3004555.
[178] Collins MP, Bentz EC, Sherwood EG. Where is Shear Reinforcement Required? Review of Research Results and Design Procedures 2008.
[179] McKenna F. OpenSees: A Framework for Earthquake Engineering Simulation. Computing in Science and Engg 2011;13:58–66. https://doi.org/10.1109/MCSE.2011.66.
[180] Pang Y, Wang X. Cloud-IDA-MSA Conversion of Fragility Curves for Efficient and High-Fidelity Resilience Assessment. Journal of Structural Engineering 2021;147:04021049. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002998.
[181] Wang X, Li Z, Shafieezadeh A. Seismic response prediction and variable importance analysis of extended pile-shaft-supported bridges against lateral spreading: Exploring optimized machine learning models. Engineering Structures 2021;236:112142. https://doi.org/10.1016/j.engstruct.2021.112142.
[182] Ding J-Y, Feng D-C. Feature selection of ground motion intensity measures for data-driven surrogate modeling of structures. Earthquake Engineering & Structural Dynamics 2024;53:1216–37. https://doi.org/10.1002/eqe.4068.
[183] Zenodo - Research. Shared. n.d. https://help.zenodo.org/docs/deposit/describe-records/reserve-doi/ (accessed December 11, 2024).
[184] Asgarkhani N, Kazemi F, Jakubczyk-Gałczyńska A, Mohebi B, Jankowski R. Seismic response and performance prediction of steel buckling-restrained braced frames using machine-learning methods. Engineering Applications of Artificial Intelligence 2024;128:107388. https://doi.org/10.1016/j.engappai.2023.107388.
[185] Zhang T, Xu W, Wang S, Du D, Tang J. Seismic response prediction of a damped structure based on data-driven machine learning methods. Engineering Structures 2024;301:117264. https://doi.org/10.1016/j.engstruct.2023.117264.
[186] Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: A corrected feature importance measure. Bioinformatics 2010;26:1340–7. https://doi.org/10.1093/bioinformatics/btq134.
[187] Mangalathu S, Hwang S-H, Jeon J-S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Engineering Structures 2020;219:110927. https://doi.org/10.1016/j.engstruct.2020.110927.
[188] Guan X, Burton H, Shokrabadi M. A database of seismic designs, nonlinear models, and seismic responses for steel moment-resisting frame buildings. Earthquake Spectra 2021;37:1199–222. https://doi.org/10.1177/8755293020971209.
[189] Lignos D, Krawinkler H. Sidesway Collapse of Deteriorating Structural Systems Under Seismic Excitations. 2012.
[190] Miranda E. Approximate Seismic Lateral Deformation Demands in Multistory Buildings. Journal of Structural Engineering 1999;125:417–25. https://doi.org/10.1061/(ASCE)0733-9445(1999)125:4(417).
[191] FEMA. Guidelines for Performance-Based Seismic Design of Buildings. 2018.
[192] Baddipalli S, Gentile R, O'Reilly G, Shahnazaryan D, Zaker Esteghamati M. Impact of Modeling Decisions on Seismic Loss and Fragility Assessment of Steel Buildings. Earthquake Spectra 2025. https://doi.org/10.1177/87552930251360751.
[193] Zaker Esteghamati M. 2 - Leveraging machine learning techniques to support a holistic performance-based seismic design of civil structures. In: Naser MZ, editor. Interpretable Machine Learning for the Analysis, Design, Assessment, and Informed Decision Making for Civil Infrastructure, Woodhead Publishing; 2024, p. 25–49. https://doi.org/10.1016/B978-0-12-824073-1.00008-3.

[194] Zaker Esteghamati M, Farzampour A. Probabilistic seismic performance and loss evaluation of a multi-story steel building equipped with butterfly-shaped fuses. Journal of Constructional Steel Research 2020;172:106187. https://doi.org/10.1016/j.jcsr.2020.106187.
[195] Zaker Esteghamati M, Kottke AR, Rodriguez-Marek A. A Data-Driven Approach to Evaluate Site Amplification of Ground-Motion Models Using Vector Proxies Derived from Horizontal-to-Vertical Spectral Ratios. Bulletin of the Seismological Society of America 2022;112:3001–15. https://doi.org/10.1785/0120220106.
[196] Zaker Esteghamati M. Toward using explainable data-driven surrogate models for treating performance-based seismic design as an inverse engineering problem. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 2025. https://doi.org/10.1098/rsta.2024.0050.
[197] Burton H. Causal inference on observational data: Opportunities and challenges in earthquake engineering. Earthquake Spectra 2023;39:54–76. https://doi.org/10.1177/87552930221125492.
[198] Burton H, Galicia Madero S, Wu C, Shams R, Nweke C. Quantifying the realized and unrealized benefits of seismic interventions using causal inference. Nat Hazards 2025. https://doi.org/10.1007/s11069-025-07531-6.
[199] Naser MZ. Causality and causal inference for engineers: Beyond correlation, regression, prediction and artificial intelligence. WIREs Data Mining and Knowledge Discovery 2024;14:e1533. https://doi.org/10.1002/widm.1533.