Machine learning-based feature importance approach for sensitivity analysis of steel frames

Hyeyoung Koh[1] and Hannah B. Blum[1,*]

[1]Department of Civil and Environmental Engineering, University of Wisconsin - Madison, Wisconsin, United States; ORCID: 0000-0002-5795-4417
[1]Department of Civil and Environmental Engineering, University of Wisconsin - Madison, Wisconsin, United States ORCID: 0000-0003-4672-0903
[*]Corresponding author: hannah.blum@wisc.edu

Abstract

This study presents a machine learning-based approach for sensitivity analysis to examine how parameters affect a given structural response while accounting for uncertainty. Reliability-based sensitivity analysis involves repeated evaluations of the performance function incorporating uncertainties to estimate the influence of a model parameter, which can lead to prohibitive computational costs. This challenge is exacerbated for large-scale engineering problems which often carry a large quantity of uncertain parameters. The proposed approach is based on feature selection algorithms that rank feature importance and remove redundant predictors during model development which improve model generality and training performance by focusing only on the significant features. The approach allows performing sensitivity analysis of structural systems by providing feature rankings with reduced computational effort.

The proposed approach is demonstrated with two designs of a two-bay, two-story planar steel frame with different failure modes: inelastic instability of a single member and progressive yielding. The feature variables in the data are uncertainties including material yield strength, Young's modulus, frame sway imperfection, and residual stress. The Monte Carlo sampling method is utilized to generate random realizations of the frames from published distributions of the feature parameters, and the response variable is the frame ultimate strength obtained from finite element analyses. Decision trees are trained to identify important features. Feature rankings are derived by four feature selection techniques including impurity-based, permutation, SHAP, and Spearman's correlation. Predictive performance of the model including the important features are discussed using the evaluation metric for imbalanced datasets, Matthews correlation coefficient. Finally, the results are compared with those from reliability-based sensitivity analysis on the same example frames to show the validity of the feature selection approach. As the proposed machine learning-based approach produces the same results as the reliability-based sensitivity analysis with improved computational efficiency and accuracy, it could be extended to other structural systems.

1 Introduction

The use of structural analysis that incorporates the effect of inelastic material and geometric properties, which is referred to as advanced analysis, inelastic analysis, and geometrically and materially nonlinear analysis with imperfections (GMNIA), has increased with significant advances in computerized structural analysis. Although there are now various sophisticated structural analysis software used by structural engineers, it is difficult to predict the actual performance of a steel frame with certainty due to ever-present uncertainties in material and geometric properties, and structural loads. For example, the shape and magnitude of geometric imperfections have a significant influence on the response of a structure, and hence need to be modeled accurately when determining the load carrying capacity of a frame building. Residual stress and elastic modulus may have a significant effect on strength by reducing the frame stiffness and influence the system reliability. In order to accurately capture structural responses, it must be understood how the uncertainties in all the factors affecting a structural system influence system performance.

Design by advanced analysis is permitted in existing steel design specifications including Section 5 of Eurocode 3 [1], Appendix 1 of AISC 360 [2], Appendix D of AS 4100 [3], and Annex O of CSA S16 [4]. The advanced analysis allows the direct modeling of system and member imperfections in a numerical model. The analysis shall consider geometric nonlinearities including geometric imperfections (P-Δ and P-δ), and stiffness reductions including partial yielding of the cross section accentuated by the presence of residual stresses. It is demonstrated that inelastic analysis can capture accurate results including the complex interactions between members of a large structural system, and can capture the beneficial system effect of load redistribution after the initial formation of plastic hinges [5–9]. Previous studies have investigated the system reliability of steel frames for system-based design by the inelastic method with a consideration of uncertainties in material and geometric imperfections [10–13]. Taras and Huemer [14] investigated the impact of different load sequences on the reliability of three resistance functions including plastic cross-sectional resistance, in- and out-of-plane buckling resistance of beam-column, and resistance of a planar portal frame structure subjected to various load combinations. The study considered uncertainties in structural loads, strength and stiffness, cross-sectional dimensions, and out-of-plumbness. Cardoso et al. [15] developed system reliability-based criteria for the design of steel rack frames, including randomness of geometric and material properties. The study derived system resistance factors for rack frames based on five series of rack frames subjected to gravity loads.

To estimate how the uncertainties influence system performance, sensitivity analysis can be conducted and several studies have proposed approaches for reliability-based sensitivity analysis. Bjerager and Krenk [16] considered the derivatives of the Hasofer-Lind index as local sensitivity of the failure probability based upon the first-order reliability method. Garnier et al. [17] estimated the sensitivities by evaluating the gradient of the probability of normal random variables obtained from Monte Carlo tools. Buonopane [18] presented reliability-based sensitivity analyses of two planar steel frames designed using first-order and second-order analysis. However, the reliability-based sensitivity analysis involves repeated evaluations of the performance function to estimate the influence of the incorporated uncertainty, which can lead to prohibitive computational resources. Many efficient reliability-based sensitivity analysis methods to reduce the computational effort have been proposed. Rubinstein and Kroese [19] introduced the application of the score function method that measures the partial derivatives of each parameter without additional simulations for the estimation of all sensitivities. The line sampling based sensitivity method was proposed by Lu et al. [20] to measure the sensitivity from the partial derivatives of the failure probability in the score function approach. Proppe [21] applied the moving particles method [22] to local sensitivity analysis, which moves samples to new locations in the design space and counts the number of moves reaching the failure region for the calculation of the failure probability.

Consideration of the large set of parameters in reliability-based sensitivity analysis is challenging due to the repeated evaluation of the performance function. Moreover, high-dimensional data increases the curse of dimensionality [23], which refers to the phenomenon arisen from too many feature variables that increase sparsity in data, storage space, and computational costs. The high-dimensional data is often resorted to feature selection in order to avoid potential problems. Feature selection techniques are broadly used in the field of data analysis and machine learning. Several papers investigated various feature selection methods and compared their performances [24–27]. Although structural design requires considering several uncertainties, the application of feature selection techniques to structural design problems is limited. Some papers developed the machine learning algorithms for developing seismic fragility curves for frame buildings [28], predicting the capacity of reinforced concrete beams [29] and cold-formed steel channels [30], compressive strength of concrete [31], failure mode of RC bridge columns [32] using the experimental or field observation data. Hwang et al. [33] investigated the effect of uncertainties on the seismic collapse risk of RC building frames using the data generated by Monte Carlo simulation. Although these papers examined the relative importance of input features, further studies may need to examine how the important features are determined depending on the failure modes and structural responses. As the application of machine learning on steel structures is fewer than other materials, it is meaningful to explore the effect of uncertainties related to the behavior of steel frames. Moreover, it is important to reduce the number of features to improve the model efficiency and accuracy in machine learning, but there is a lack of study in using feature selection techniques to structural design problems.

In this paper, the feature importance methods using machine learning technique are implemented to examine the effects of random properties on the strength of two planar steel frames: one frame fails by inelastic instability of a single member and the second frame fails by progressive yielding. Both frames have limited capability to redistribute load. The design frame strengths were designed using the inelastic method specified in AISC 360 [2]. Uncertainties of yield strength, elastic modulus, sway imperfection (out-of-plumbness), and residual stress were considered. To collect the database, finite element analyses were conducted in OpenSees and the Monte Carlo sampling method was utilized. The four feature importance methods are employed including impurity-based, permutation, SHAP [34], and Spearman's correlation coefficient [35]. The feature importance methods generate the ranking of features based on the importance score, and the features with the highest scores are considered more likely to influence structural failure. In addition, the class imbalance issue in structural design problems is examined. The model performance is evaluated by using the Accuracy and the Matthews correlation coefficient (MCC) [36], which is appropriate for the imbalanced data. Reliability-based sensitivity studies of steel frames were conducted not only to validate the proposed approach but also to provide a further understanding of how randomness in material and geometric properties affect the system strength depending on the failure mode. The aim of this paper is to (1) propose the feature importance strategy based on machine learning techniques for examining how uncertainties affect the structural failures and provide prospects in the application of feature importance techniques to sensitivity analysis of structures, (2) describe the feature selection approach that improves model accuracy and efficiency in prediction of structural failures, (3) address how to develop and evaluate the model with the class imbalanced data, and finally (4) provide undemanding solutions for using machine learning techniques to structural engineers.

2 Proposed framework for sensitivity analysis of structural systems

The traditional sensitivity analysis of structures is based on reliability analysis that computes the reliability index or probability of failure in respect of a factor under consideration to examine the effect of the factor on the structures. However, the reliability-based sensitivity analysis requires repeated evaluations of the performance function and additional simulations. If a structure to be considered has complexity such as nonlinear structural behavior or a large quantity of structural members, computational costs will increase and thereby decreasing computational efficiency. In order to avoid repeated evaluations of the performance function and to reduce the computational effort, this study proposes a framework using machine learning techniques which can be used as an alternative to the reliability-based sensitivity analysis of structures with high efficiency and accuracy.

The machine learning-based feature selection methods are employed to estimate feature importance that can be ranked based on the importance score. A feature with the maximum score is considered the most important feature. Once the rank of features is obtained, a machine learning classifier is implemented to measure the model performance when the model carries only the top-k features, where k = the number of important features. The model performance is evaluated by varying the number of features considered in the feature set. The least important features that decrease the model performance or result in the same performance can be removed from the feature set to increase the accuracy and efficiency of the model. The proposed method is demonstrated by comparing the results obtained from reliability-based sensitivity analysis.

2.1 Applied structural system

Two designs of a two-bay, two-story non-symmetric planar steel frame based on those of Ziemian [37] have been selected as examples, which are designed according to AISC 360 [2]. Fig. 1 shows the geometry, applied loads, and support conditions for the frames. The member sizes and loads of each frame are summarized in Table 1. The frames are subjected to the gravity load combination of 1.2*D* + 1.6*L*, with the live-to-dead load ratio assumed to be $L/D = 1.5$, which is a typical value [38].

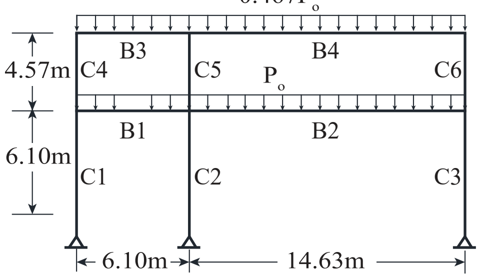

Figure 1: Frame layout

Finite element (FE) analyses of the two frames were performed with *OpenSees* [39], including material and geometric nonlinearity. An elastic-perfectly-plastic material model was used as the steel material property to improve simulation performance. Each element was subdivided into 16 displacement-based and

Table 1: Member sizes and loads for example frames

Element	Frame 1	Frame 2
C1	W6 $\times$ 20	W12 $\times$ 14
C2	W14 $\times$ 82	W14 $\times$ 99
C3	W14 $\times$ 68	W14 $\times$ 82
C4	W6 $\times$ 8.5	W10 $\times$ 12
C5	W14 $\times$ 145	W14 $\times$ 109
C6	W14 $\times$ 145	W14 $\times$ 109
B1	W30 $\times$ 132	W27 $\times$ 84
B2	W36 $\times$ 182	W36 $\times$ 135
B3	W24 $\times$ 55	W18 $\times$ 40
B4	W30 $\times$ 116	W27 $\times$ 94
Loads (P_o)	111.86 kN/m	109.45 kN/m

fiber-type elements. Connections were assumed to be fully-rigid to disregard any potential flexibility of the connections. The frames were given sway imperfections of h/500, where h = the frame height, by displacing the relevant nodes from their nominal locations. All cross-sections contain the Galambos and Ketter residual stress model [40] which has tensile stress along the entire web, tension at the web-to-flange intersection, and peak compression at flange tips equal to 0.3F_{yn} where F_{yn} = nominal material yield strength. The presence of residual stresses was considered by applying residual stress to each fiber in the cross section. A reduction factor of 0.9 was applied to F_y and E of all members in order to design members by the inelastic method as specified in Appendix 1 of AISC 360 [2]. The structures were analyzed using second-order inelastic analysis.

The load-displacement curves and the location of highly yielded zones from the inelastic analysis are presented in Fig. 2. If a section has at least 75$\%$ of cross-section area yielded, it was considered highly yielded. Analyses of the frames were performed with all properties at nominal values. The ultimate load ratio λ, the ratio of ultimate to factored design loads, is 1.08 for both frames, indicating a limited capacity of inelastic load redistribution for both frames. It is observed that Frame 2 has a clear yield point while Frame 1 exhibits a rounded curve around the yield point due to their different failure modes. There was no highly yielded zone in Frame 1 but elements of B2 and C2 located close to their connection were partially yielded with yield ratios of 63.1$\%$ and 53.5$\%$. Frame 1 collapsed due to the instability (global buckling) of the slender column, C2. The failure mode of Frame 2 is progressive yielding with six highly yielded members: B1, B2, B3, B4, C5, and C6. The FE model used in this study has the same load ratio and failure modes with the previous studies [10, 13], therefore verifying the modeling technique used herein.

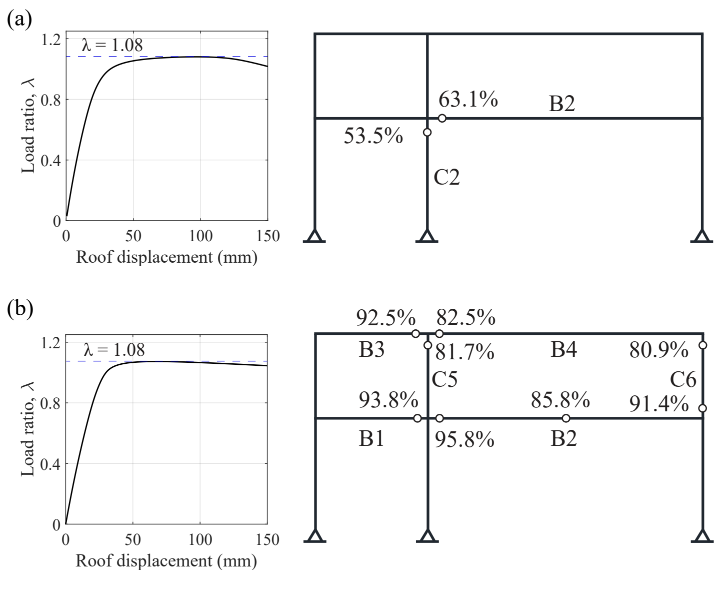

Figure 2: Load-displacement curve and location of highly yielded zones: (a) Frame 1; (b) Frame 2

2.2 Uncertainty

As it is difficult to predict the actual performance of a steel frame with certainty due to ever-present uncertainties in material and geometric properties, therefore several steel design codes [1, 2, 4] require the engineer to consider the uncertainties in the analysis even with requiring advanced structural analysis. This study considered the uncertainties in yield strength (F_y), elastic modulus (*E*), sway imperfection, and residual stress. Statistical information for these uncertainties is summarized in Table 2. Yield strength was modeled as a lognormal distribution with a mean value 1.10F_{yn} and a COV of 0.06, where F_{yn} = nominal yield strength of 248 MPa (36 ksi) [41]. Elastic modulus was modeled as a lognormal distribution with a mean value of E_n and a COV of 0.04, where E_n = nominal elastic modulus of 200 GPa (29,000 ksi) [41]. After random samples of F_y and *E* are randomly drawn from each distribution, 0.9F_y and 0.9*E* are used as material properties in the analysis in order to conduct the inelastic analysis according to Appendix 1 of AISC 360 [2]. The magnitude of sway imperfection was modeled as a lognormal distribution with a mean value of 1/770 and a COV of 0.875 [42]. After a random magnitude of imperfection has been determined, the random value is multiplied by the height of the frame to consider the imperfection of the overall frame. The first story and the second story have the same direction of sway imperfection; i.e., multistory columns have no changing direction of sway imperfection therefore the magnitude of sway imperfection accumulates continuously at each story. The direction of imperfection was distributed to have equal probability in each direction (left or right). The scale factor of residual stress (X) was modeled as a normal distribution with a mean value of 1.064 and a COV of 0.27 [43], which is based on the experimental data of Galambos and Ketter [40]. The random scale factor X is multiplied to 0.3F_{yn} to compute the peak compression stress. Once the peak compressive residual stresses are determined, the rest of the residual stresses in the cross section were set based on the residual stress pattern [40]. The frames are assumed as spatially correlated, which means that all columns or beams have the same random properties. This assumption was consistent with common construction practice [13].

Table 2: Description of feature variables

Variable	Mean	COV	Distribution	References
F_y	1.1F_{yn}	0.06	Lognormal	Bartlett et al. [41]
E	E_n	0.04	Lognormal	Bartlett et al. [41]
Sway imperfection	1/770	0.875	Lognormal	Lindner and Gietzelt [42]
X	1.064	0.27	Normal	Shayan et al. [43]

Note: F_{yn} = 248 MPa, E_n = 200 GPa

2.3 Dataset

The datasets used for training the machined learning model consist of the FE simulation results of the two example steel frames. The uncertainties are considered as input parameters. Since all columns or beams have the same properties when they are spatially correlated, two different values of each uncertainty are necessary to represent the properties of columns and beams except sway imperfection assigned to columns only. Thus, the datasets include seven independent input features: two yield strengths, two Young's moduli, two peak residual stress values, and one sway imperfection. For each feature variable, 200,000 random samples were generated by Monte Carlo simulation based on statistical information summarized in Table 2. The ultimate load ratio λ determined from the simulation is classified into two different classes; $\lambda = 1.0$, which is the probability-based limit state design criteria of AISC 360 [2], is used as the structural failure criteria. Therefore, if $\lambda < 1.0$ where design strength is less than required strength, the sample set is labeled as Class 0, which represents the structural failure, and $\lambda \geq 1.0$ is labeled as Class 1 to refer to no structural failure (structural safety). The dataset is binary because all sample sets are classified into one out of the two classes. The instances that have convergence issues in the simulation are excluded from the original dataset. Frame 1 has the convergence issues about 0.5$\%$ of the total number of simulations and Frame 2 has a few more errors than Frame 1 because of its complex failure mode, a gradual sequence of yielding. Finally, the two binary classification datasets are constructed for Frame 1 and Frame 2. The number of instances in a class according to the frames is summarized in Table. 3.

2.4 Machine learning technique

This study implemented a decision tree [44] classifier to rank features and measure the model performance, which is popular and simple to interpret. A decision tree algorithm starts from the root of the tree

Table 3: Summary of the dataset

Class label	Frame 1	Frame 2
Class 0 (failure)	763	71
Class 1 (safe)	196,586	189,252
Total	197,349	189,323

that contains the whole training set. The root node is split into leaf nodes according to a certain criterion such as impurity until no further splits can be made. When observations in a node consist of different classes, it is considered impure while a node containing instances belonging to one class only is considered pure. Gini impurity (Eq. 1) is a measurement of the likelihood of misclassification and can be a criterion for impurity in a node. If a dataset is pure, Gini impurity is 0 because there is no probability to derive incorrect classification. The model created by a decision tree predicts a target variable by learning the decision rules used for the best split that minimizes the impurity at the leaf nodes.

$$Gini\ impurity = 1 - \sum_{c} p(c|j)^2 \tag{1}$$

where $p(c|j)$ = the ratio of the observations that belongs to class c for a particular node j.

In machine learning, classification involves the two steps including (1) learning which develops a model based on training data and (2) prediction that uses the model to predict the response of the test data, which is not employed in the training. The two binary datasets of approximately 200,000 FE analysis results are used for fitting the decision tree classifier. Each dataset is divided into 50$\%$ for training and 50$\%$ for testing by random splitting in order to avoid overfitting, which refers to a modeling error that occurs when a model is too exactly fit to a training set, thus the model may not perform well with the new data (test data). The hyperparameter setting used in the decision tree classifier includes impurity for splitting criterion, 20 for tree depth, and 2 for the minimum required number of samples to split a node.

2.5 Feature importance technique

Feature selection involves reducing the number of feature variables to mitigate the curse of dimensionality. Feature selection improves the computational efficiency of the machine learning model and reduces the volume of feature space, which is a significant issue in a large dataset. Moreover, the prediction performance can be improved by removing redundant features that have a negative or no effect on prediction. There is a large number of feature selection techniques that estimate a feature importance score and provide the feature ranking based on the score of all features. This study implemented four existing feature importance methods to identify the most and least important features in steel structures: 1) impurity-based and 2) permutation, which measure feature importance based on tree-based models; 3) SHAP which is a complex machine but considers all possible pairs of features; 4) Spearman's rank correlation coefficient that is simple and does not require model training.

2.5.1 Impurity-based importance

The impurity-based feature importance method requires model training to derive the importance score. This method compares the impurities before and after a node j then averages the impurity decrease caused by the splits based on a feature, x_i, which is the Mean Decrease in Impurity (MDI). This method considers a feature that results in a large MDI as important for class prediction. As MDI uses statistics obtained from the training set, this method may be limited to examine the features that are important in the test set, which is unknown to the model. The impurity-based feature importance can be computed by Eq. 2 [45]:

$$J_{impurity}(x_i) = \frac{\sum_{i \in N^{(i)}} (\text{Impurity before node } k - \text{Impurity after node } k)}{|N^{(i)}|} \tag{2}$$

where $N^{(i)}$ = the number of nodes in a classifier in which a split is carried out based on x_i.

2.5.2 Permutation importance

The permutation approach depends on the prediction from fitted models similar to the impurity-based approach but performs repeated permutations to estimate feature importance. This approach starts with training a model to obtain a baseline model performance. After fitting the model, the samples in a single feature column are randomly shuffled, then the model performance is evaluated. The difference between the model accuracy obtained from the models using the original dataset, which is the baseline performance, and the dataset with permuted values of x_i is considered as the importance of x_i (Eq. 3). An important feature for prediction leads to a large decrease in accuracy because the important feature information needed for prediction is absent when the values of the feature are permuted. This study repeated shuffling ten times for a feature x_i and used the average of $J_{permutation}(x_i)$ as the importance of each feature. The permutation method can overestimate the importance of correlated features because features are assumed as independent when fitting the baseline model.

$$\begin{aligned} J_{permutation}(x_i) = &\ \text{accuracy for dataset without permutation} \\ &- \text{accuracy for permuted dataset of } x_i \end{aligned} \tag{3}$$

2.5.3 SHAP

SHAP (SHapley Additive exPlanations) proposed by Lundberg and Lee [34] is carried out to determine the important features. The SHAP algorithm indicates how much each feature contributes to the response variable based on the predictions for linear models trained on all feature subsets. The SHAP value of a feature x_i can be computed by Eq. 4:

$$J_{SHAP}(x_i) = \sum_{S \subseteq F \setminus i} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)] \tag{4}$$

where F = the set of all features, S = all feature subsets without x_i

First, predictions from a model $f_{S\cup\{i\}}$ trained on a feature subset S including a feature x_i and another model f_S excluding the feature are compared. The effect of a feature x_i on the prediction can be interpreted as the difference of the predictions, $f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)$. The SHAP importance score is interpreted as a weighted average of all possible differences. When x_i is dependent with other features in a subset, keeping x_i can affect the predictive performance. Thus, all possible subsets are used to achieve the SHAP value. The running time of the SHAP algorithm exponentially increases with the number of features because the models are retrained on all possible feature subsets.

2.5.4 Spearman's rank correlation

As Spearman's rank correlation coefficient [35] can quantify the relationship between each feature and the response variable, it is used to measure feature importance. This approach identifies feature importance by ranking correlation coefficients of features. The coefficient has a range from -1 as the perfect negative correlation to 1 as the perfect positive correlation. When a feature has a coefficient close to 1 or -1, the feature is considered as important. A coefficient of 0 indicates that the two variables are uncorrelated. Unlike the previous three feature importance approaches that achieve feature rankings after training a model, feature ranking by Spearman's correlation can be obtained without a model fitting, thereby increasing computational efficiency.

2.6 Evaluation metric

Accuracy and the Matthews correlation coefficient (MCC) are used to evaluate the predictive performance. Accuracy is the most popular metric in binary classification datasets but it provides optimistically biased results on the majority class especially when the dataset is imbalanced [46]. Class imbalance refers to a problem in machine learning classification where each class accounts for an unequal portion of the data, which may lead to poor predictive performance. The MCC is demonstrated as a reliable metric that overcomes the class imbalance problem by Chicco and Jurman [47]. Since the datasets used in this study are highly imbalanced, this study employs both the Accuracy and MCC as evaluation measures, which are computed based on the confusion matrix M that describes the performance of a classifier on a test dataset:

$$M = \begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix} \tag{5}$$

where TP (True Positive) = the number of actual positives that are correctly predicted positives, TN (True Negative) = the number of actual negatives that are correctly predicted negatives, FN (False Negative) = the number of actual positives that are incorrectly predicted negatives, and FP (False Positive) = the number of actual negatives that are incorrectly predicted positives. In this study, for example, TN represents a failure in a test set that is predicted as safe.

Accuracy is calculated by the ratio of the number of correctly predicted instances to all instances (Eq. 6). If a classifier is applied on an imbalanced dataset, the classifier is biased towards the majority class [48]. Moreover, a trivial classifier learns the majority class only and attributes the label to all instances, thereby no instance for predicted negatives, i.e., FN = TN = 0 [47]. The MCC considers the relationship between actual and predicted values and is computed by Eq. 7. The values of MCC ranges from -1 (perfect misclassification) to 1 (perfect classification), while 0 means random guessing. The MCC produces a high score only if all of the four confusion matrix categories are predicted well.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \tag{7}$$

3 Machine learning-based feature importance results

A decision tree algorithm is implemented to measure feature importance and model accuracy. Each feature importance method trained on 50$\%$ of the data to rank features and used the remaining 50$\%$ to evaluate the model performance. The predictive performance of the classifiers is evaluated by computing the Accuracy and MCC. The nomenclature is used to reference the feature names. The letter ‘c’ following the feature name refers to columns and ‘b’ represents beams in a frame. Residual stress and sway imperfection are abbreviated to ‘rs’ and ‘sway’, respectively. As all columns have the same magnitude of sway imperfection when spatially correlated, the feature for sway imperfection assigns a name without a letter. For example, $\mathrm{F_y}$-b is the yield strength of beams and rs-c indicates the residual stress of columns.

3.1 Imbalanced classification

In common structural engineering practice, structures are designed to have the target reliability index β on the order of 2 to 6 [38], corresponding to the range of probability of failure P_f between 9.9×10^{-10} and 2.3×10^{-2} calculated from $P_f = \Phi(-\beta)$, where Φ = standard normal cumulative density function. As the failure probability can be estimated by the number of failure (Class 0) out of the total number of simulations, datasets for structural design problems become extremely imbalanced.

The degree of class imbalance can be measured by the no-information rate, which is the ratio between the number of the majority class and the total instances in the confusion matrix. The highly imbalanced dataset has a rate close to 1. The rate of the balanced binary dataset is 0.5 because it has the same size as the two classes. In this study, the no-information rates are 0.996 and 1.0 for Frame 1 and Frame 2, respectively, thereby severely imbalanced data. It is more challenging to model with a severe imbalance of the classes because the minority class can be ignored entirely.

Several sampling techniques are developed to address class imbalance problems such as undersampling and oversampling. Undersampling deletes examples from the majority class in the training set, but it can pass over important information during removing examples. Oversampling simply duplicates examples from the minority class and certain examples can be crowded at a specific location, thereby leading to overfitting. SMOTE (Synthetic Minority Over-sampling Technique) [49], one of the improved oversampling techniques, creates new synthetic data from the minority class rather than simply duplicates the data. This study employed SMOTE as the class imbalance compensation method.

Table 4: Feature importance scores using the imbalanced dataset of Frame 2

	Impurity-based		Permutation		SHAP		Spearman	
Rank	Feature	Score	Feature	Score	Feature	Score	Feature	Score
1	sway	0.50	sway	5.5×10^{-4}	sway	1.2×10^{-3}	sway	0.03
2	F_y-c	0.38	F_y-c	3.9×10^{-4}	F_y-c	6.8×10^{-4}	F_y-c	0.02
3	F_y-b	0.10	F_y-b	9.0×10^{-5}	F_y-b	1.8×10^{-4}	F_y-b	0.01
4	rs-b	0.02	rs-b	1.0×10^{-5}	rs-b	3.0×10^{-5}	E-b	0
5	E-c	0	E-c	0	E-c	0	rs-c	0
6	E-b	0	E-b	0	E-b	0	rs-b	0
7	rs-c	0	rs-c	0	rs-c	0	E-c	0

As Frame 2 has the higher no-information rate than Frame 1 due to fewer failures, it is selected to determine the appropriate data for measuring feature importance between balanced and imbalanced class distributions. The original dataset is imbalanced and the dataset generated by SMOTE is balanced, which has an equal number of the two classes. The correlation matrices of Frame 2 are given in Fig. 3 which shows the correlation coefficients between each feature and the binary outputs. Since the original dataset is severely imbalanced, all coefficients between the features and response variable are approximately equal to 0 which represents no correlation between the class labels and features. However, the correlation matrix of the balanced data displays a broad range of coefficients from -0.04 to 0.82, presenting both strong and weak correlations. As summarized in Table 4, the importance scores of Frame 2 using the imbalanced dataset have negligible values, except the scores from the impurity-based method which is less affected by the class imbalance. It seems that the imbalanced dataset can be used for measuring feature importance because of the identical orders for the top-ranked features regardless of the method. However, the magnitudes of importance score are approximately equal to zero even for the first feature, thereby lacking in information for the assessment of significant features. This study employed the balanced dataset for training that measures the feature importance, which provides more information about the relationships between features and the response variable.

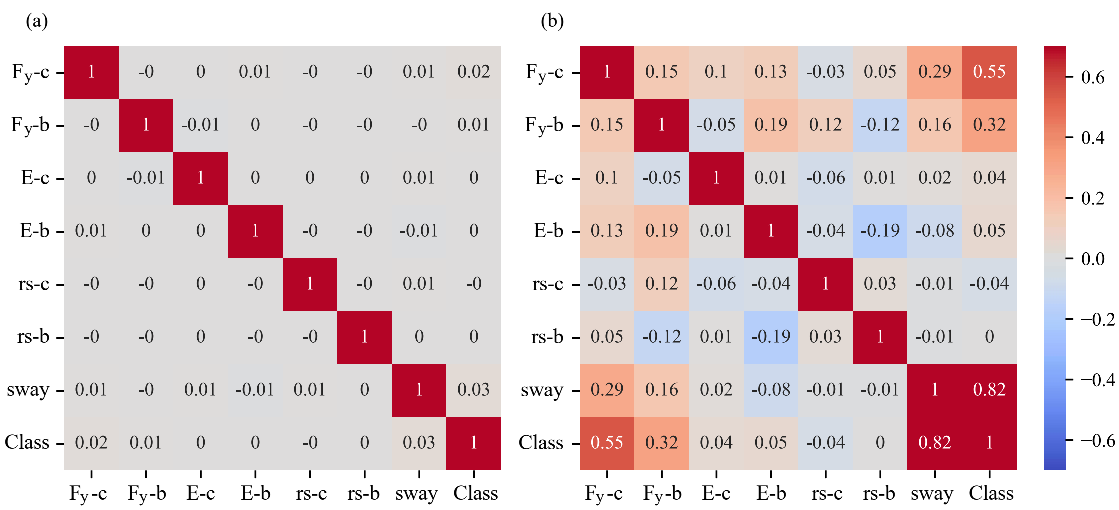

Figure 3: (a) Correlation matrix of Frame 2 using (a) imbalanced data (b) balanced data

3.2 Feature importance

The rank and magnitude of feature importances for Frame 1 derived from the impurity-based, permutation, SHAP, and Spearman's correlation methods are shown in Fig. 4. The horizontal bars indicate the feature importance scores computed based on the equations discussed previously in Section 2.5. The features are ranked in descending order. With regard to the ranking by Spearman's correlation, the features are ordered based on the absolute value of the correlation coefficients because the negative correlation also indicates the relationship between two variables. Although various methods were carried out to measure feature importance, Frame 1 has the same feature rankings when the method that requires model fitting is performed such as impurity-based, permutation, and SHAP. Spearman's rank resulted in the same order as the other methods except for the three least important features. Frame 1, which fails by the instability of a single column, rates sway imperfection and yield strength of columns ($\mathrm{F_y}$-c) as the top two features. Elastic modulus and residual stress of columns, which correspond to E-c and rs-c, are ranked next but they have negligible scores in comparison to the top features. The features that describe properties of beams such as E-b, $\mathrm{F_y}$-b, and rs-b are bottom-ranked because columns have an impact on the Frame 1 behavior rather than beams. Since the correlated scenario applies the same properties on all columns or beams, it is reasonable that the properties related to columns are high-ranked. The impurity-based importance method measured $\mathrm{F_y}$-b, E-b, and rs-b as zero, as shown in Fig. 4a, which indicates that removing the features from model training could improve the predictive performance and computational efficiency.

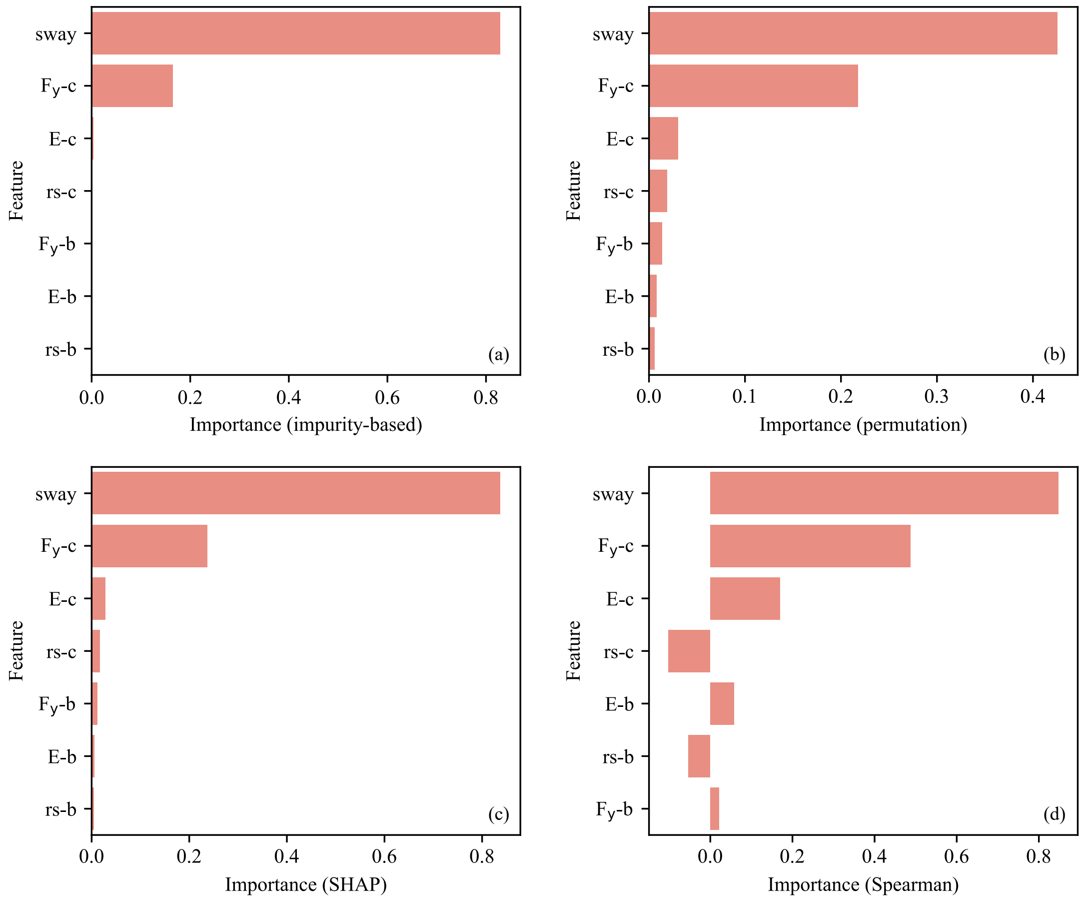

Figure 4: Feature importance ranking of Frame 1 derived from (a) impurity-based (b) permutation (c) SHAP (d) Spearman correlation

Fig. 5 shows the rank of features for the Frame 2 dataset. Sway imperfection is ranked as the most important feature by all feature importance methods. Yield strength of columns and beams are ranked after sway imperfection except for the permutation method which ranks the yield strength of beams at fourth. SHAP and Spearman's rank showed the same order of features (Fig. 5c and d). Contrary to Frame 1 where most methods derived the same ranking, Frame 2 obtained different feature orders between the methods. Comparing the importance scores between Frame 1 and Frame 2, the magnitude of importance smoothly decreases as the ranking decreases in Frame 2 while only the high-ranked two features are significant in Frame 1. In other words, the beam yield strength, ranked at the third or fourth place for Frame 2, is also important in Frame 2 where as for Frame 1 the bottom-ranked features have negligible scores, indicating that the yield strength of all structural members has an important effect on Frame 2. The failure mode of Frame 2, gradual sequence of yielding, implies that the yielding of several members has a significant effect on the frame behavior. It is found that the number of structural members related to structural failure influences the magnitude of importance. The more features affecting the failure, the more features are considered important but with a smaller magnitude. Moreover, if a frame has a complex failure mode, it is difficult to obtain consistent feature ranking from various methods even with the balanced data.

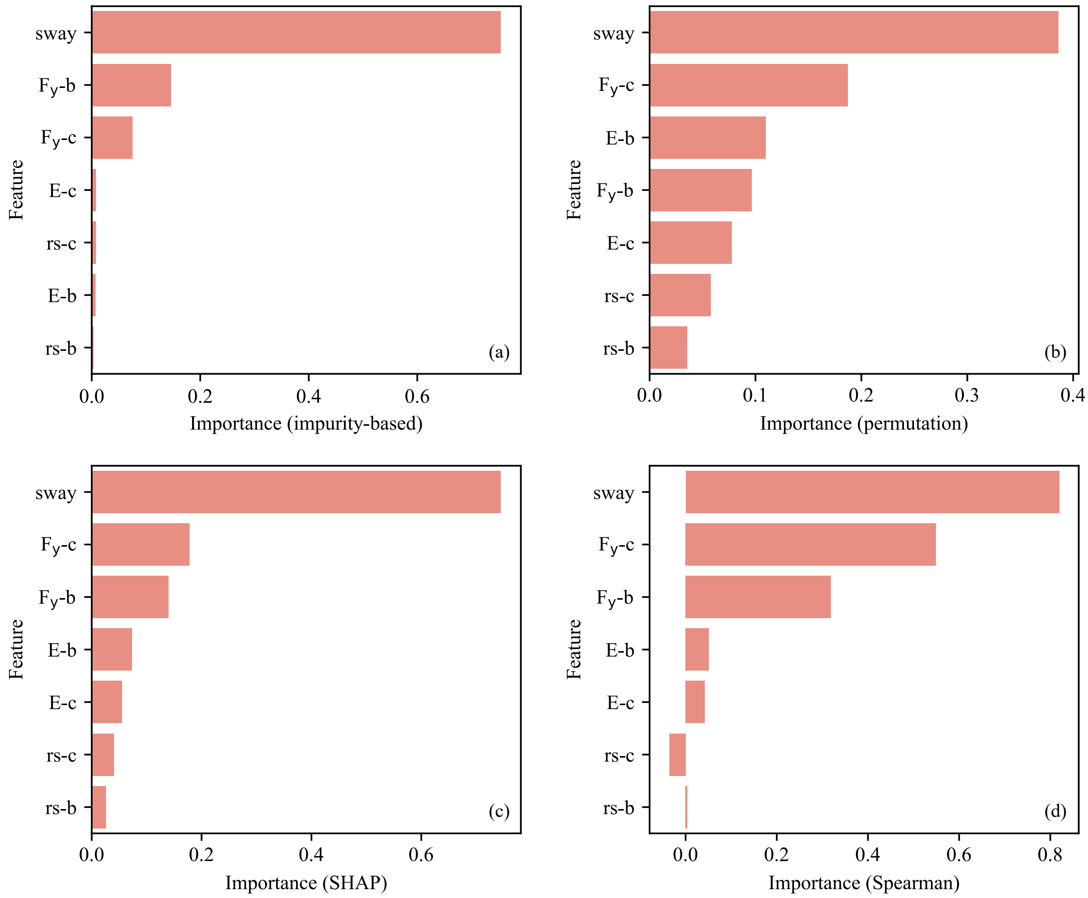

Figure 5: Feature importance ranking of Frame 2 derived from (a) impurity-based (b) permutation (c) SHAP (d) Spearman correlation

3.3 Performance evaluation

The model evaluation must be carried out based on the test dataset which is not used in the training to obtain a reliable performance of the fitted models. That is, unknown data must be used for the evaluation. 50% of the data were oversampled to compensate for the class imbalance and used for the model training. The remaining 50% of the data are assigned to the test set, which is still imbalanced. Although the balanced dataset for both training and testing showed the best performance for predicting structural responses [28], the test set should keep original data points because the real world data cannot be expected to be balanced. Both Accuracy and the Matthews correlation coefficient are used for the evaluation metric of the predictive performance due to the imbalanced test dataset.

The Accuracy and MCC curves for a decision tree model trained using the top-k features of Frame 1 as ranked by impurity-based, permutation, SHAP, and Spearman are shown in Fig. 6. The methods that measure feature importance using predictions such as impurity-based, permutation, and SHAP showed an identical performance because they had the same feature order of Frame 1. Although Spearman's correlation had a different rank in terms of the bottom three features, the performance is similar with the other methods because the model with three features already reached the nearly perfect Accuracy. The model with one feature shows the Accuracy value of 0.943 then it increases abruptly to 0.997 when two features are included in a feature set. As Frame 1 has two features that are considered as the most influential, which are sway and F_y-c, the classifier showed good accuracy with only the two features. Accuracy maintains 0.998 when three or more features are added to the feature set. The estimate of the performance by Accuracy is overoptimistic.

The Matthews correlation coefficient (Fig. 6b) provides a lower score than Accuracy because it addresses the optimistically biased estimate caused by class imbalance. MCC provides a high score only when all four categories in the confusion matrix obtained good results. It is observed that the MCC score rapidly increases when the number of features increases from one to two. When the model trained using the top three features, the MCC curve has the highest score, 0.81. After the MCC score reaches the peak, the score falls up to 0.78 until all features are added. This indicates that the least important features are redundant and can be removed to improve the computational efficiency in addition to the model performance. For example, the impurity-based feature importance provided zero scores for the four bottom-ranked features and therefore they might be considered as not necessary for training.

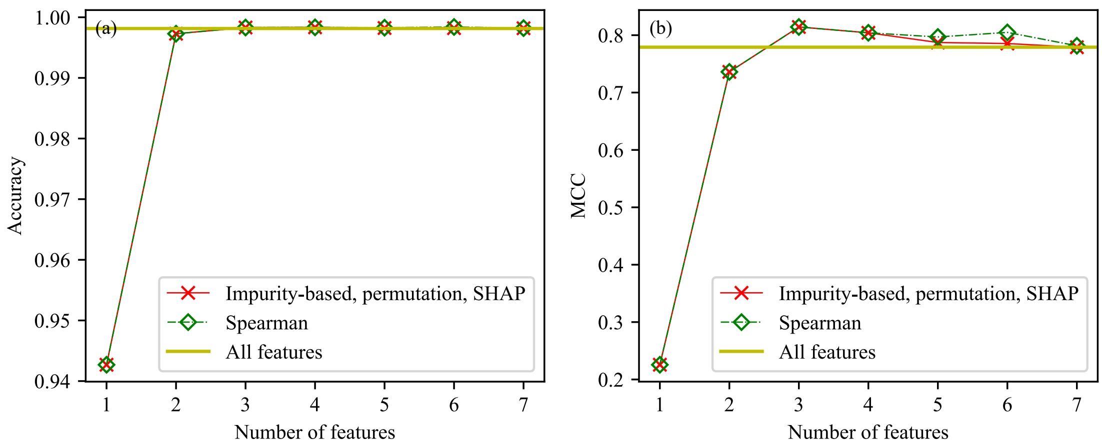

Figure 6: Model performance of Frame 1 (a) Accuracy (b) MCC

Fig. 7a shows the outcome Accuracy of Frame 2 generated from a decision tree classifier trained using a different number of features. Comparing across different feature importance methods, the impurity-based model that ranked the beam yield strength as the second most important has slightly better performance than the other methods. SHAP and Spearman's correlation show the same curve because they generated the same feature rank. In contrast to Frame 1 that outperforms when the number of selected feature is two, Frame 2 reaches the highest Accuracy, 0.999, when trained using three or more features. Permutation importance ranked the yield strength of beams as the fourth most important, thus the model was able to attain the full Accuracy after including that feature. It is observed that all important features should be included in training to obtain the best accuracy. Accuracy again provided inflated results that are misleading values.

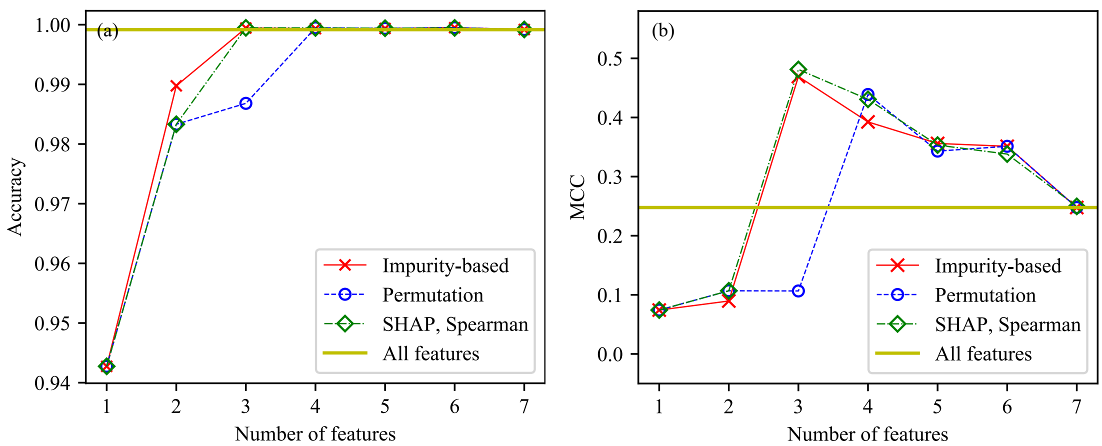

Figure 7: Model performance of Frame 2 (a) Accuracy (b) MCC

The outcome MCC scores across the number of features are shown in Fig. 7b. As Frame 2 has a larger class imbalance rate than Frame 1, the overall MCC score of Frame 2 is low compared to that of Frame 1 because the test dataset of Frame 2 has fewer minority class points due to the higher imbalanced ratio. The low score of MCC indicates that it correctly evaluates the predictive performance by addressing the class imbalance problem. When the dataset includes three or four features, the model leads to the best classification performance than the models with more than four features. The MCC score decreases down to 0.25 from the peak score 0.48 when all features are considered. Redundant features can increase irrelevant information and complexity of a model, thus decreasing the model performance.

4 Reliability-based sensitivity study

Reliability-based sensitivity studies on the two example frames were conducted to demonstrate the application of the proposed feature importance method to sensitivity analysis. Each analysis contained random realizations of the property under consideration while the other properties remained as nominal values. The effect of random properties on the strength and behavior of the frames was investigated. 40,000 structural analyses were performed for each uncertainty under consideration. The total number of simulations required by reliability-based sensitivity analysis increases in proportion to the number of uncertainties because each analysis has different random values of the property under consideration while all other properties remain as nominal values. However, the simulations used for the machine learning datasets consider all uncertainties at once, i.e., all uncertainties are random in the simulation, thus additional simulations are not necessary to be performed. For each random property, the probability of failure denoted P_f was estimated by Eq. 8 [50]:

$$P_f = P(\lambda < 1.0) = n/N \tag{8}$$

where λ = ultimate load ratio, n = number of simulations in which the frame failed, and N = total number of simulations. If sensitivity analysis resulted in no cases of $\lambda < 1.0$, a normal probability plot [51] was used as an alternative to estimate P_f.

Fig. 8 and Fig. 9 show the strength distributions derived by sensitivity analysis. The vertical axes of all strength histograms were normalized to the probability density function. Each plot includes the values of mean, COV, P_f, and β computed based on the distribution. The strength distributions for yield strength and Young's modulus were modeled as a normal distribution. The spatial correlation of sway imperfection has an obvious effect on the strength distributions. The frames are non-symmetric therefore gravity loads will always result in frame sway in one direction. When the sway imperfections of all columns are correlated and are in the same direction as the deflections resulting from gravity loads, the strength can be modeled as a smallest extreme Type I distribution while the case of the column sway in opposite direction as that resulting from gravity loads can be modeled as a largest extreme Type I distribution. Thus, the correlation of sway imperfection leads to a bi-modal strength distribution. For example, in Fig. 8c, the red line indicates the smallest extreme Type I distribution and the blue line indicates the largest extreme Type I distribution. The same direction between the frame sway and deflections results in a lower mean strength, leading to a lower boundary of the strength distribution, thereby resulting in more structural failures. The Frame 1 strength for random residual stress was modeled as a smallest extreme Type I distribution while the strength of Frame 2 was modeled as a normal distribution with a small COV. Likely the failure mode affects the frame strength distribution.

As shown in Fig. 8a, the randomness of yield strength produced λ below 1.0 for Frame 1, indicating that the frame is sensitive to yield strength. Since Frame 1 fails by the instability of C2, the frame is sensitive to Young's modulus, which could affect second-order bending moments of the frame. Thus, the results for Frame 1 has a larger COV than that of Frame 2 and experienced failures while Frame 2 showed no failures (Fig. 8b and Fig. 9b). Random sway imperfection provides the lowest β for Frame 1, which implies that the frame is sensitive to the factor related to column behavior. P_f of Frame 1 with random residual stress was approximately equal to 0, thus negligible impact on the failure of Frame 1. Frame 1 results have a larger COV than Frame 2 results because random sway imperfection and residual stress can increase second-order bending moments of the frame, leading to the instability of columns. To sum up, the sensitivity analysis results showed that sway imperfection provided the lowest β followed by yield strength, Young's modulus, and residual stress. It is identical with the feature rankings of Frame 1 derived by the feature importance approach, as previously shown in Fig. 4. The top-4 features were sway imperfection, $\mathrm{F_y}$-c, E-c, and rs-c in descending order of the importance scores. Furthermore, the feature importance results provided further information that columns are more influential than beams in resulting in frame failures by listing the beam properties at the bottom of the feature rankings.

For Frame 2, as shown in Fig. 9a and c, the uncertainties in yield strength and sway imperfection generated failures where λ is less than 1.0. Sway imperfection led to the largest P_f among the uncertainties and the

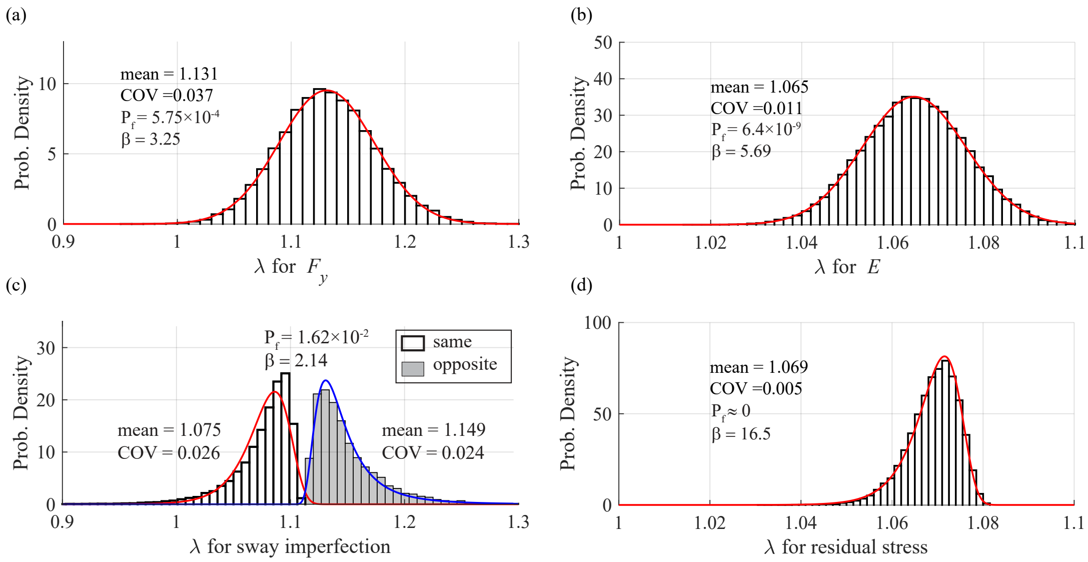

Figure 8: Strength distribution of Frame 1 with random (a) yield strength (b) elastic modulus (c) sway imperfection (d) residual stress

distribution of yield strength has a clearly larger variability than the other random properties, showing that the frame is sensitive to sway imperfection and yield strength. In contrast to Frame 1 where random variations in the elastic modulus resulted in failures, Frame 2 has no failures by random variations in the elastic modulus because the frame is less affected by the factors related to the second-order bending moments due to its failure mode of progressive yielding. Random variations in residual stress resulted in the smallest COV (Fig. 9d), indicating that it has the least significant effect on the frame behavior. Frame 2 strength distributions showed that sway imperfection provided the largest P_f followed by yield strength, Young's modulus, and residual stress. As previously shown in Fig. 5, the four highest-ranked features from the feature importance approach include sway, $\mathrm{F_y}$-c, $\mathrm{F_y}$-b, and E-b (or E-c by the impurity-based method). Residual stress is bottom-ranked by all feature importance methods. The sensitivity analysis showed that only sway imperfection and yield strength led to failures and residual stress had negligible impacts; therefore it is confirmed that the feature importance method can identify the significant factors. Although Frame 2 had slightly different feature rankings from the various importance methods due to its complex failure mode, the most and least important features matched the reliability-based sensitivity analysis results.

Regarding the feature importance results of Frame 1, SHAP, impurity-based, and permutation methods generated results which best matched with the sensitivity analysis. Features related to column properties are top-ranked while beam properties are bottom-ranked, reflecting the failure mode of instability of a single column. The beam yield strength was ranked first among the beam properties, which reflected the sensitivity analysis results that yield strength is more influential than elastic modulus and residual stress. The feature order of Frame 2 by SHAP and Spearman ranked the yield strengths of columns and beams after the sway imperfection. Moreover, the beam and column residual stresses are ranked at the bottom, which mirrors the same results as the sensitivity analysis. In summary, among the four feature importance techniques, SHAP showed the best identical results with the sensitivity analysis of both frames.

When an uncertain property leads to a larger P_f and a lower β in a reliability-based sensitivity study, the property has a significant effect on the system failure. If the feature has a large importance score from the feature importance method, the feature has a more significant effect on frame behavior than the other properties. By comparing the identical top-ranked features between the reliability-based sensitivity method and all feature importance methods employed in this study, it is demonstrated that the proposed machine learning feature importance framework can be used for sensitivity analysis of structures. Using the feature importance method is not only straightforward to examine the significance of features by comparing importance scores, but it can also determine which structural members (e.g. columns and beams) have large impacts on structural failure without additional evaluation of performance.

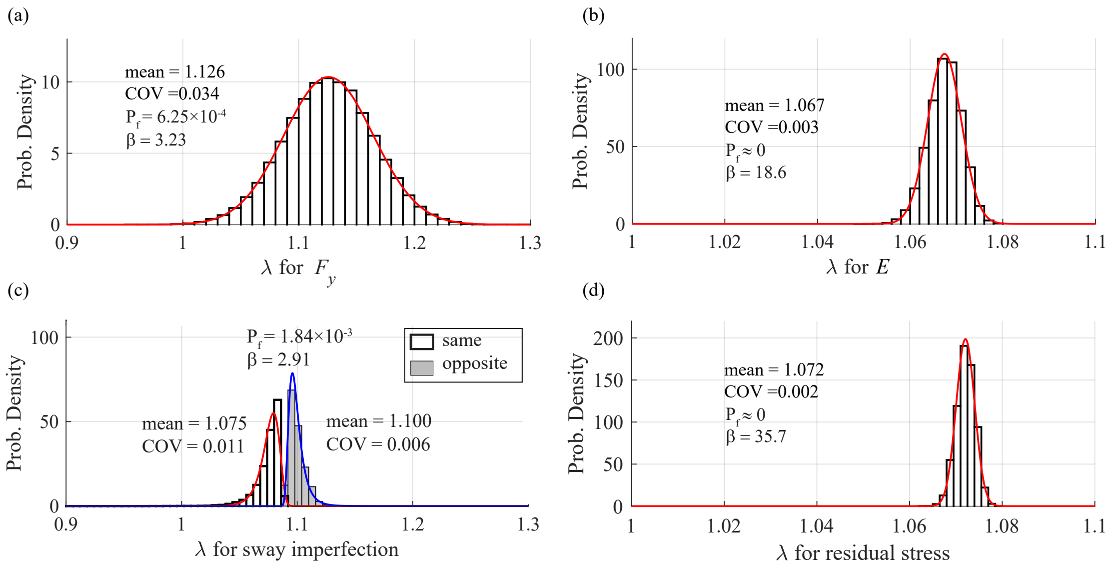

Figure 9: Strength distribution of Frame 2 with random (a) yield strength (b) elastic modulus (c) sway imperfection (d) residual stress

5 Conclusion

A machine learning-based approach for sensitivity analysis of steel structures is proposed, based upon feature importance measured by existing feature selection methods including impurity-based, permutation, SHAP, and Spearman's correlation. The datasets used in this study consist of the finite element analysis results of two designs of a nonsymmetric planar steel frame which have different failure modes. The finite element models include the uncertainties in material yield strength, Young's modulus, sway imperfection, and residual stress. This study evaluates the sensitivity of uncertainties affecting structural failures by comparing the feature importance score derived from the feature importance methods. It was identified that the feature importance results reflect the failure mode well because the most important features were matched with the factors resulting in the frame collapse. When a frame fails by inelastic instability of a single column, the features related to column properties were highly ranked while the beam properties were bottom-ranked. When a frame fails by yielding of multiple members, the yield strength of all structural members was top-ranked in the feature rankings. A complex failure mode such as progressive yielding led to the inconsistent orders of the least important features between implemented feature selection methods. Accuracy and the Matthews correlation coefficient score, used as a measure of the predictive performance of the model, verified the efficiency and accuracy of the proposed approach. Structural engineering problems involve the class imbalance due to the design criteria that provide significantly fewer failures than safe structures. The model evaluation by the MCC score was appropriate for the dataset for structural design, which has an imbalanced classification problem. The suggested feature selection method showed high accuracy with the MCC score even with the severely imbalanced dataset.

Reliability-based sensitivity studies were conducted to provide a further understanding of how uncertainties in material and geometric properties affect the strength and performance of the frame in addition to demonstrate the proposed feature importance method. The probability of failure and reliability index derived from the traditional method are compared with the feature rankings. The proposed feature importance approach yields the same results as the reliability-based sensitivity analysis, i.e., the features ordered by the reliability index are identical with the ranking generated by the feature importance score. The use of the feature selection method could be more efficient than traditional methods in identifying the significant factors affecting the structural behavior because the traditional method requires repeated simulations to examine the sensitivity of a factor. Moreover, it is not challenging to assess how large the impact of a feature is on structural failure by comparing the importance scores.

The potential of using machine learning technique for sensitivity analysis of structures is investigated. Although the proposed framework is examined based on the structural responses of non-symmetric planar steel frames, it is demonstrated that the application of machine learning-based feature selection methods has the efficiency and accuracy to evaluate how the factors affect the structural failures. The general procedure of the proposed approach can be used for sensitivity analysis on diverse structural systems. However, further research should be performed to support the application of the proposed framework on various structural systems because factors that influence structural behavior and failure, such as materials, load sequence, and complex system behavior, may complicate the analysis.

References

[1] EN 1993-1-1, *Eurocode 3: Design of steel structures - Part 1-1: General rules and rules for buildings*. European Committee for Standardisation, 2005.

[2] ANSI/AISC, Chicago, IL, *Specification for Structural Steel Buildings*, 2016.

[3] AS 4100, *Steel Structures*. Standards Australia, 1998.

[4] CSA S16:19, *Design of Steel Structures*. Canadian Standards Association, 2019.

[5] R. D. Ziemian, W. McGuire, and G. Deierlein, “Inelastic limit states design. I: Planar frame studies,” *J Struct Eng*, vol. 118, no. 9, pp. 2532–2549, 1992.

[6] W.-F. Chen and S.-E. Kim, *LRFD steel design using advanced analysis*, vol. 13. CRC press, 1997.

[7] D. White and J. Hajjar, “Stability of steel frames: The cases for simple elastic and rigorous inelastic analysis/design procedures,” *Eng Struct*, vol. 22, no. 2, pp. 155–167, 2000.

[8] N. S. Trahair and S. L. Chen, “Out-of-plane advanced analysis of steel structures,” *Eng Struct*, vol. 25, no. 13, pp. 1627–1637, 2003.

[9] C. Ngo-Huu, S. E. Kim, and J. R. Oh, “Nonlinear analysis of space steel frames using fiber plastic hinge concept,” *Eng Struct*, vol. 29, no. 4, pp. 649–657, 2007.

[10] S. G. Buonopane and B. W. Schafer, “Reliability of steel frames designed with advanced analysis,” *J Struct Eng*, vol. 132, no. 2, pp. 267–276, 2006.

[11] H. Zhang, R. Ellingwood, B., and K. J. R. Rasmussen, “System reliabilities in steel structural frame design by inelastic analysis,” *Eng Struct*, vol. 81, pp. 341–348, 2014.

[12] W. Liu, K. J. R. Rasmussen, and H. Zhang, “Systems reliability for 3d steel frames subject to gravity loads,” *Struct*, vol. 8, pp. 170–182, 2016.

[13] H. Zhang, H. Liu, B. R. Ellingwood, and K. J. Rasmussen, “System reliabilities of planar gravity steel frames designed by the inelastic method in aisc 360-10,” *J Struct Eng*, vol. 144, no. 3, p. 04018011, 2018.

[14] A. Taras and S. Huemer, “On the influence of the load sequence on the structural reliability of steel members and frames,” *Struct*, vol. 4, pp. 91–104, 2015.

[15] F. S. Cardoso, H. Zhang, and K. J. R. Rasmussen, “System reliability-based criteria for the design of steel storage rack frames by advanced analysis: Part II – reliability analysis and design applications,” *Thin-Walled Struct*, vol. 141, pp. 725–739, 2019.

[16] P. Bjerager and S. Krenk, “Parametric sensitivity in first order reliability theory,” *J Eng Mech*, vol. 115, no. 7, pp. 1577–1582, 1989.

[17] J. Garnier, A. Omrane, and Y. Rouchdy, “Asymptotic formulas for the derivatives of probability functions and their monte carlo estimations,” *Eur J Oper Res*, vol. 198, no. 3, pp. 848–858, 2009.

[18] S. G. Buonopane, “Strength and reliability of steel frames with random properties,” *J Struct Eng*, vol. 134, no. 2, pp. 337–344, 2008.

[19] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*, vol. 10. John Wiley & Sons, 2016.

[20] Z. Lu, S. Song, Z. Yue, and J. Wang, “Reliability sensitivity method by line sampling,” *Struct Saf*, vol. 30, no. 6, pp. 517–532, 2008.

[21] C. Proppe, “Local reliability based sensitivity analysis with the moving particles method,” *Reliab Eng Syst Saf*, vol. 207, p. 107269, 2021.

[22] C. Walter, "Moving particles: A parallel optimal multilevel splitting method with application in quantiles estimation and meta-model based algorithms," *Struct Saf*, vol. 55, pp. 10–25, 2015.

[23] R. E. Bellman, *Adaptive control processes: a guided tour*. Princeton university press, 1961.

[24] I. Inza, P. Larranaga, R. Blanco, and A. J. Cerrolaza, "Filter versus wrapper gene selection approaches in dna microarray domains," *Artif Intell Med*, vol. 31, no. 2, pp. 91–103, 2004.

[25] Z. Zhu, Y.-S. Ong, and M. Dash, "Wrapper–filter feature selection algorithm using a memetic framework," *IEEE Trans Syst Man Cybern*, vol. 37, no. 1, pp. 70–76, 2007.

[26] Y. Aphinyanaphongs, L. D. Fu, Z. Li, E. R. Peskin, E. Efstathiadis, C. F. Aliferis, and A. Statnikov, "A comprehensive empirical comparison of modern supervised classification and feature selection methods for text categorization," *J Assoc Inf Sci Technol*, vol. 65, no. 10, pp. 1964–1987, 2014.

[27] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Comput Stat Data Anal*, vol. 143, p. 106839, 2020.

[28] J. Kiani, C. Camp, and S. Pezeshk, "On the application of machine learning techniques to derive seismic fragility curves," *Comput Struct*, vol. 218, pp. 108–122, 2019.

[29] J. Rahman, K. S. Ahmed, N. I. Khan, K. Islam, and S. Mangalathu, "Data-driven shear strength prediction of steel fiber reinforced concrete beams using machine learning approach," *Eng Struct*, vol. 233, p. 111743, 2021.

[30] V. V. Degtyarev, "Neural networks for predicting shear strength of cfs channels with slotted webs," *J Constr Steel Res*, vol. 177, p. 106443, 2021.

[31] P. Chopra, R. K. Sharma, M. Kumar, and T. Chopra, "Comparison of machine learning techniques for the prediction of compressive strength of concrete," *Adv Civ Eng*, vol. 2018, 2018.

[32] S. Mangalathu and J.-S. Jeon, "Machine learning–based failure mode recognition of circular reinforced concrete bridge columns: Comparative study," *J Struct Eng*, vol. 145, no. 10, p. 04019104, 2019.

[33] S.-H. Hwang, S. Mangalathu, J. Shin, and J.-S. Jeon, "Machine learning-based approaches for seismic demand and collapse of ductile reinforced concrete building frames," *J Build Eng*, vol. 34, p. 101905, 2021.

[34] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.

[35] C. Spearman, "The proof and measurement of association between two things," *Am J Psychol*, vol. 15, no. 1, pp. 72–101, 1904.

[36] B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," *Biochim Biophys Acta*, *Protein Struct*, vol. 405, no. 2, pp. 442–451, 1975.

[37] R. D. Ziemian, *Advanced methods of inelastic analysis in the limit states design of steel structures*. PhD thesis, Cornell University, Ithaca, NY, 1990.

[38] B. R. Ellingwood, "Lrfd: implementing structural reliability in professional practice," *Eng Struct*, vol. 22, no. 2, pp. 106–115, 2000.

[39] S. Mazzoni, F. McKenna, M. H. Scott, G. L. Fenves, *et al.*, "Opensees command language manual," 2006.

[40] T. V. Galambos and R. L. Ketter, "Columns under combined bending and thrust," *J Eng Mech Div*, vol. 85, no. 2, pp. 1–30, 1959.

[41] F. M. Bartlett, R. J. Dexter, M. D. Graeser, J. J. Jelinek, B. J. Schmidt, and T. V. Galambos, "Updating standard shape material properties database for design and reliability," *Eng J AISC*, vol. 40, no. 1, pp. 2–14, 2003.

[42] J. Lindner and R. Gietzelt, "Imperfektionsannahmen für stützenschiefstellungen," *Der Stahlbau*, vol. 52, no. 4, pp. 97–101, 1984.

[43] S. Shayan, K. J. R. Rasmussen, and H. Zhang, "Probabilistic modelling of residual stress in advanced analysis of steel structures," *J Constr Steel Res*, vol. 101, pp. 407–414, 2014.

[44] J. R. Quinlan, "Induction of decision trees," *Mach Learn*, vol. 1, no. 1, pp. 81–106, 1986.

[45] A. J. Izenman, *Modern multivariate statistical techniques: Regression, classification and manifold learning*. New York, USA: Springer, 2013.

[46] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation," in *Australasian joint conference on artificial intelligence*, pp. 1015–1021, Springer, 2006.

[47] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020.

[48] R. Blagus and L. Lusa, "Class prediction for high-dimensional class-imbalanced data," *BMC Bioinformatics*, vol. 11, no. 1, pp. 1–17, 2010.

[49] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *J Artif Intell Res*, vol. 16, pp. 321–357, 2002.

[50] R. Melchers and A. T. Beck, *Structural reliability analysis and prediction*. Hoboken, NJ: John Wiley & Sons Ltd, 3rd ed., 2018.

[51] J. M. Chambers, *Graphical methods for data analysis*. CRC Press, 1983.