**Safety, Reliability, Risk, Resilience and Sustainability of Structures and Infrastructure**
12th Int. Conf. on Structural Safety and Reliability, Vienna, Austria, 6–10 August 2017
Christian Bucher, Bruce R. Ellingwood, Dan M. Frangopol (Editors)


# Application of the Random Forests Machine Learning in Assessing the Post-Earthquake Structural Safety of Damaged Buildings

Yu Zhang[a], Henry V. Burton[b], Han Sun[c] and Mehrdad Shokrabadi[d]

[a,c,d]Ph.D. Candidate, Department of Civil and Environmental Engineering, University of California, Los Angeles
[b]Assistant Professor, Department of Civil and Environmental Engineering, University of California, Los Angeles

**Abstract**: This paper utilizes a machine learning algorithm, Random Forests, to establish a mapping from response and damage patterns of damaged buildings to their post-earthquake structural safety. Residual collapse capacity is used to quantify structural safety and classify the safe and unsafe states, of which distinct response and damage patterns are characterized. The assessment performance is also discussed.

## 1 Introduction

Assessing the structural safety of buildings damaged by earthquake shaking is essential to preventing further casualties, estimating seismically-induced financial losses, devising repair strategies and evaluating urban resilience [1]–[4]. ATC-20 [5], [6] provides guidelines for post-earthquake visual inspection to rapidly evaluate building structural safety, which has been widely used after U.S. earthquakes [6], [7] as well as in many other countries. Based on the work of Porter et al. [8] and Mitrani-Reiser [9], FEMA P58 [10], [11] assigns the unsafe placard if a pre-defined triggering fraction of components in a particular damage state at a certain floor (or story) is exceeded, but these judgement-based triggering fractions are not explicitly linked to the reduction in building safety. Recently, the reduction in the collapse capacity of mainshock damaged buildings has been used as a metric for assessing the post-earthquake structural safety and occupiability of damaged buildings. Incremental dynamic analysis (IDA) [12]–[14] are performed using sequential ground motions to quantify the reduction in collapse capacity. Maffei et al. [15], [16] developed four sets of post-earthquake occupiability criteria, which differ based on the metric used to quantify the reduction in collapse safety of the damaged building. Yeo and Cornell [17] used the time-varying aftershock hazard at a given site to compute an equivalent constant collapse rate, which decreases with time after the occurrence of the mainshock. However, no direction link was made between component-level damage and the safety state of the building. Raghunandan et al. [18] quantified the increase in vulnerability to collapse of mainshock-damaged buildings, where individual system- and component-level damage indicator, such as beam and column plastic rotation and transient and residual story drift ratio (SDR), is linked to the residual collapse capacity using linear regression. The interaction between damage indicators is not considered.

This paper presents a machine-learning-based methodology to assess post-earthquake structural safety using the “response patterns” and “damage patterns”. The response patterns are characterized through various engineering demand parameters (EDPs) and the damage patterns are

obtained using the simulated visual damage states in key structural components. Non-linear response history analyses (NRHAs) are performed in sequential ground motions. The "damage generating" (DG) ground motions are used to generate response and damage patterns within the building. IDAs are then carried out using the "collapse capacity testing" (CCT) ground motions to evaluate the reduction in structural safety. Random Forests algorithm is employed to map the response and damage patterns of a damaged building to its associated safety state classified by an acceptable threshold of residual collapse capacity. The proposed methodology is demonstrated on a 4-story reinforced concrete (RC) special moment-frame (SMF) building. Results show distinct yet partially overlapped response and damage patterns for the safe and unsafe states. Although no clear boundary could be found for any individual predictor, the Random Forests algorithm is able to recognize the various response and damage patterns in the high-dimension predictor space. More importantly, the limit state defined by the building being unsafe to occupy following an earthquake is predicted with satisfactory performance.

# 2 Overview of Methodology

A schematic overview of the methodology used to assess post-earthquake structural safety is shown in Figure 1. Starting with an intact structure, five distinct yet fully integrated steps are used to illustrate the assessment framework. The outcome of this assessment is the predicted structural safety state conditioned on the structural response demands (from instrumentation) and/or available observed physical damage (through field inspections).

The first step describes the process of using a set of DG ground motions to create samples of the damaged structure from which response and damage patterns will be extracted. The response patterns or distribution of EDPs is obtained directly from NRHA. Subjecting the intact structure to a single ground motion scaled to a specific spectral intensity will produce a single distinct response pattern. Multiple response patterns with different levels and distributions of response demands are obtained by using a suite of DG ground motions scaled to incrementally increasing spectral intensities. Damage patterns are simulated using structural component damage fragility functions which relate local EDPs to the probability of exceeding a given damage state. Monte Carlo Simulation is used to generate multiple damage patterns for a single ground motion and spectral intensity. A single damage pattern is described by each structural component assigned a single discrete damage state.

The collapse capacity of the damaged structure is assessed through the application of IDAs using sequential ground motions in the second step (Step 2). Each DG record, which is used as the first ground motion in the sequence, is followed by an IDA using a set CCT ground motions. The median first-mode spectral acceleration corresponding to the collapse point ($\hat{S}a_{col,DMG}$) is used as the measure of residual collapse safety of the damaged building. In the third step, the collapse capacity of the intact structure is assessed by conducting single-record IDAs using the CCT ground motions. The median collapse capacity ($\hat{S}a_{col,INT}$) is also used as the measure of collapse safety for the intact structure. The ratio ($\kappa$) of $\hat{S}a_{col,DMG}$ to $\hat{S}a_{col,INT}$ is used as a quantitative measure of the increased collapse vulnerability or the reduction in the collapse capacity of the damage structure. The damaged building is classified as safe or unsafe to occupy by comparing to a pre-established threshold ($\kappa_{\min}$), which represents the minimum acceptable reduced collapse safety.

The fourth step uses (1) the response and damage patterns generated in step 1 and (2) their associated post-earthquake structural safety states determined based on the reduction in collapse safety (comparing $\kappa$ to $\kappa_{\min}$) obtained from steps 2 and 3 to establish a classification problem. The machine learning algorithm is then used to construct the predictive models. Subsequently, any arbitrary response and/or damage pattern can be used to probabilistically predict the safety state of an earthquake-damaged structure (Step 5).

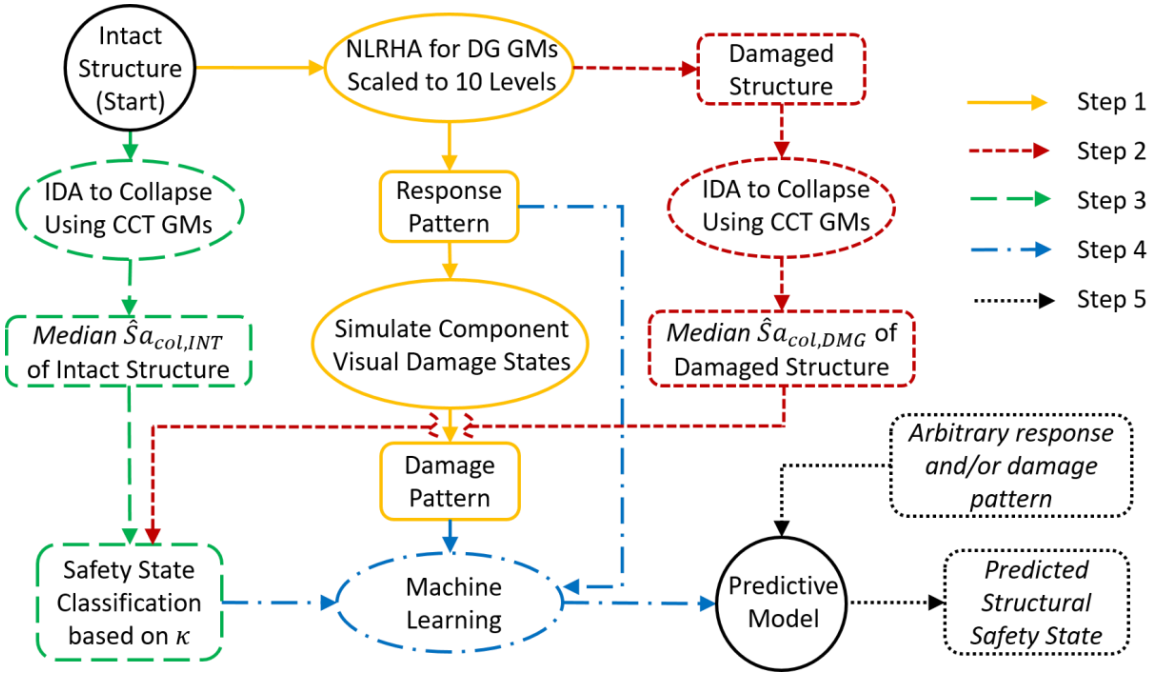



*Figure 1: Schematic of post-earthquake structural safety assessment methodology*

## 3 Establishing Post-Earthquake Safety State Criteria

A 4-story perimeter frame building developed by Haselton et al. [19] (identified with building design ID 1013) is used throughout this paper to demonstrate the proposed framework. The building has an 8-in flat slab floor system with a square plan of 120 $ft$ by 120 $ft$. A height of 15 $ft$ is used for the first story and 13 $ft$ for all the upper stories.

The ground motions used for NRHA are selected based on the conditional mean spectra [20], [21], which are computed for the building site (118.162º W, 33.996º N) using probabilistic seismic hazard deaggregation. DG and CCT ground motions are selected with a mean ε of 1.47 for large-magnitude short distance events. A total of 98 ground motions is used for the DG set by randomly sampling one of the horizontal components of each record. This is done to avoid having ground motions with highly correlated spectral shapes that would induce similar (and redundant) response patterns. For each of the DG ground motions, ten incrementally-increasing maximum story drift ratio $(SDR_{DG})$ levels are targeted, ranging from 0.5% to 5% at increments of 0.5%. These $SDR_{DG}$ values serve as proxies for the possible states of building damage under a DG ground motion. For each DG ground motion, log-normal collapse fragility curves are fitted for the ten target damage levels, while accounting for record-to-record variability and structural modeling uncertainty [10], [11]. 32 ground motion pairs are selected for the CCT set. The collapse performance is evaluated by performing IDAs using the two components in each

pair of the CCT ground motions and taking the lower of the two collapse intensities. 980 samples of damage patterns are recorded with their associated $\kappa's$, of which 55 were excluded because collapse occurred during the DG analysis.

Establishing an acceptable value of $\kappa_{\min}$ is partly a policy decision and is not fully addressed in this study. However, the observed trend between $\kappa$ and $SDR_{DG}$ can inform the choice of $\kappa_{\min}$. Figure 2a shows a plot of $\kappa$ and $SDR_{DG}$ generated by DG ground motions. Figure 2b shows a statistical summary including the median $\kappa's$ and their 90% confidence intervals. Here it can be observed that the absolute value of the slope of the median line remains roughly constant up to about $SDR_{DG} = 2\%$, after which there is a sharp increase. Hence $\kappa_{\min}$ is taken as the median value of $\kappa$ (0.95) corresponding to $SDR_{DG} = 2\%$ in this study.

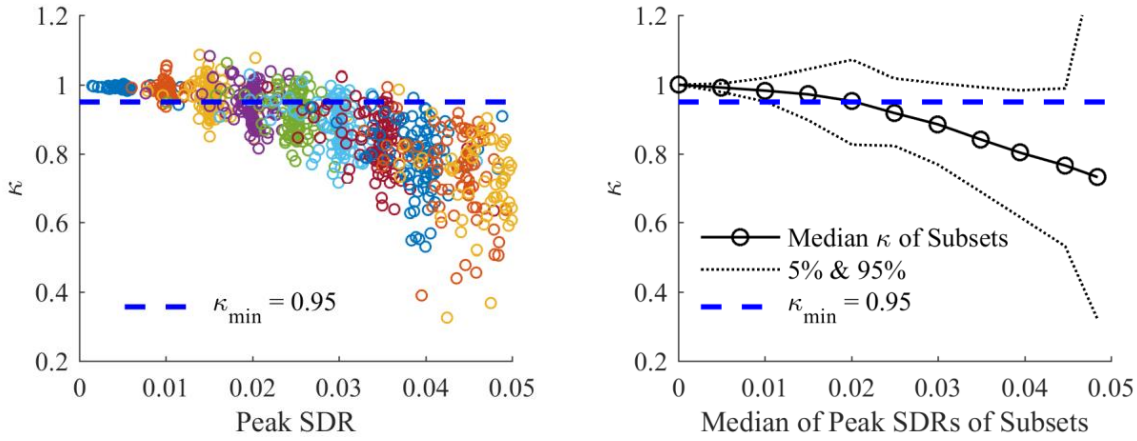



*Figure 2: Median collapse capacity for (a) individual samples and (b) sample subsets by target damage levels*

# 4 Classified Structural Response and Damage Patterns

The probability distributions of the peak SDRs for all the 935 samples classified as safe and unsafe by the proposed $\kappa_{\min}$ is shown in Figure 3. In the first story, the 90% confidence interval for peak drift demands is from 0.4% to 2.1% for the safe state and 1.5% to 4.6% for the unsafe state, and that of the 2nd story is 0.5% to 2.9% for the safe state and 1.6% to 5.0% for the unsafe state, which indicates that the distribution for the two safety states is quite different. In contrast, the 3rd and 4th stories have significant overlap in the distribution of peak story drift ratios for the safe and unsafe samples, for example, the 4th story have 90% confidence intervals of 0.2% to 1.5% and 0.3% to 1.6% respectively. The separation of the global EDP distribution in the first two stories suggests that they are more likely to be used as split boundaries that distinguish between safe and unsafe states.

For each of the 935 samples, 100 damage patterns are generated using Monte Carlo simulation. A single damage pattern is described by each component being assigned a damage state. Each damage measure (DM) (or mode of component-level damage) consists of a pre-defined number of unique damage states [10]. The damage states for beam (DM1) and column (DM2) flexural damage are described in Table 1, together with the median EDP and dispersion that defines the lognormal fragility functions. These fragility parameters are computed according to FEMA P58

[10],[11]. The median rotation demand for the flexural damage states is defined relative to the capping rotation at peak strength, $\theta_c$.

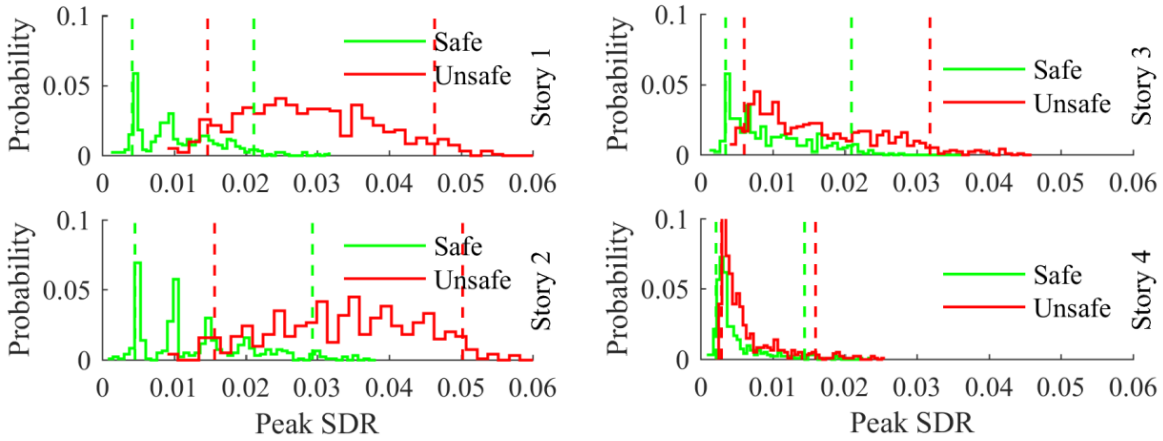



*Figure 3: Distribution of peak SDRs for the damaged building classified as safe and unsafe*

*Table I. Structural component damage measures and median* EDP *and dispersion associated with each damage state*

| Damage Measure ID | Damage Measure | Damage State ID | Damage State Description | Median EDP | Dispersion |
|---|---|---|---|---|---|
| DM1 | Beam Flexural Damage | DM1-1 | Residual crack widths > .06 in | $0.3\theta_c$ | 0.4 |
| | | DM1-2 | Concrete spalling and exposed rebar | $0.7\theta_c$ | |
| | | DM1-3 | Concrete crushing and buckling or fracture of rebar | $1.0\theta_c$ | |
| DM2 | Column Flexural Damage | DM2-1 | Residual crack widths > .06 in | $0.25\theta_c$ | 0.4 |
| | | DM2-2 | Concrete spalling and exposed rebar | $0.55\theta_c$ | |
| | | DM2-3 | Concrete crushing and buckling or fracture of rebar | $0.8\theta_c$ | |

In Figure 4, the distribution of different damage states in exterior beam and column hinges for the safe and unsafe states is compared using the 93,500 samples with distinct damage patterns.

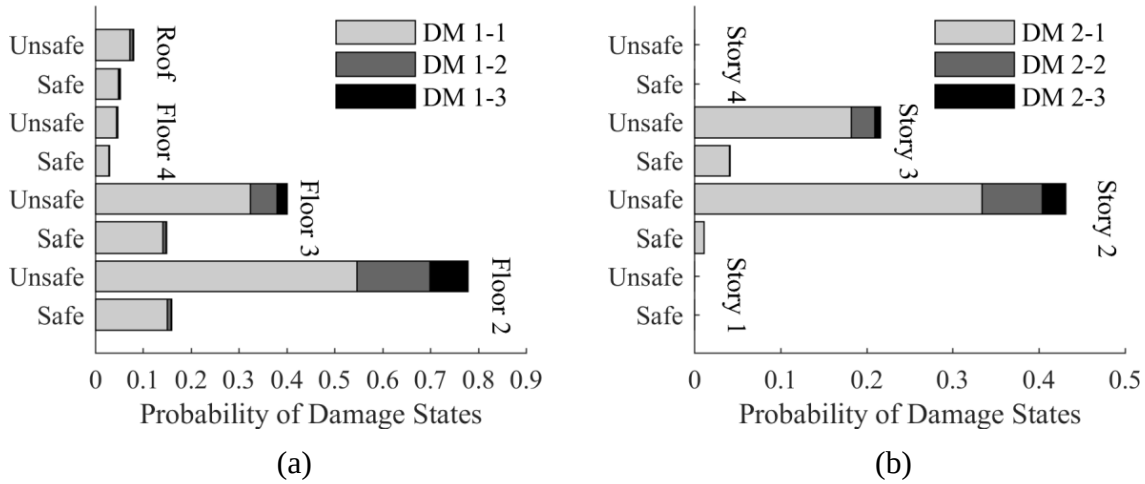



*Figure 4: Distribution of damage states for safe and unsafe samples at (a) exterior beam end and (b) top column end*

Beam flexural damage is highest in the two lower floors. For the safe state, 14.0% and 16.0% of the samples corresponding to the 2nd and 3rd floors respectively are in the lowest damage state (DM1-1). For those at the 3rd and 4th floors, only 2.9% and 5.2% respectively are in DM1-1. Almost no higher damage states occur among the safe samples. For the unsafe state,

54.7%, 15.2% and 7.8% of beams at the 2nd floor are in damage states DM1-1, DM1-2 and DM1-3, respectively. The proportions of those damage states in the 3rd floor are 32.4%, 5.6% and 2.1%, respectively. For both the safe and unsafe states, observable damage to the top column hinges only occurs in the 2nd and 3rd story. For the safe state, DM2-1 occurs in 1.1% and 4.0% of the samples corresponding to the 2nd and 3rd stories respectively. Much more damage occurs in the samples associated with the unsafe state. 33.4%, 7.0% and 2.8% of top column hinges in the 2nd story and 18.2%, 2.7% and 0.7% in the 3rd story, are in DM2-1, DM2-2 and DM2-3, respectively. In general, the distinct damage patterns between the safe and unsafe states provides the basis for building the advanced prediction models using machine learning algorithms.

# 5 Safety State Prediction using Random Forests Machine Learning

Predictive models are built using two different datasets: 1) 935 classified response patterns that comprise only global EDPs and 2) 93500 damage patterns sampled from classified local EDPs. Each dataset is partitioned into two subsets: 75% of the samples are used for training and the remaining 25% are held-out for testing the performance of proposed models. 10-fold cross-validations are performed using training subsets to find the optimal predictive models with the least amount of overfitting given specific tuning parameters.

Since the unsafe state is unfavorable and critical, it is defined in this paper as the "positive class" while the safe state as the "negative class". The sensitivity or true positive rate, is defined as the proportion of the unsafe samples in the testing dataset that are also predicted to be unsafe. The specificity of a prediction model, which is also referred to as the true negative rate, is expressed as the proportion of known safe samples that is correctly predicted. The accuracy, which combines the sensitivity and specificity, is computed as the ratio between the sum of all the true positive and true negative samples and the number of the total samples. The false positive rate represents the proportion of the safe samples which are recognized as unsafe by the predictive model, and is equal to one minus the specificity.

Random Forests [22] are based on a considerable large set of individual decision trees [23]. By recursively partition the data space, each tree explores and learns the patterns of the data set and is able to predict outcomes. To grow a tree, all the predictors are lined up and their corresponding split points are tested using a Gini Index [23]. A greedy approach, which always choose the feature and its split with the lowest cost, is used to divide the space recursively. A stop criterion, such as the depth of the tree, is needed to balance the need to capture enough characteristics and avoid overfitting of the training data. A simplified decision tree from the Random Forests model constructed using the training subset of the damage pattern dataset is presented in Figure 5. Start with the root node, the tree first checks the damage state at Hinge 1 of Beam 1 at Floor 2. It predicts unsafe state with a probability of 88% if damage state is larger or equal to 1, otherwise it provides safe assessment with a probability of 72%. In order to improve the prediction performance, the tree continue to check Hinge 2 of Beam 3 at Floor 2, and predicts safe state with improved probability of 83% if it is no damaged; and more importantly, it also recognizes those with damage states larger or equal to 1 as unsafe, which is an update from the initial prediction. The tree keeps growing to take more predictors into consideration and split at the optimal points, and finally end up in five leaf nodes. It is able to give predictions with partial or all of the 4 predictors in the decision nodes.

A decision tree could be very sensitive to the specific dataset and the predictions generated for different subsets of the same training dataset could be quite different. This problem can be addressed by assembling bootstrap-sampled subsets of the training data. Bootstrap is a statistical technique that involves randomly sampling from a dataset to create a series of sub-datasets [24], [25]. Each random sample is placed back into the original dataset such that multiple (or no) instances of a particular sample can be included in the sub-dataset. Based on these resampled datasets, bootstrap aggregation, also known as bagging [26], is an ensemble method, which seeks to achieve better predictions by combining multiple less accurate models. While bagging addresses some of the limitations of a basic tree model, the method still has an inherent defect that prevents it from producing the optimal predictive model. Bagging tree can result in highly correlated tree structures due to the application of the greedy algorithm, where all of the original predictors are considered at every split of every tree [27]. In order to overcome this problem, the Random Forests [22] algorithm was developed with a critical modification to Bagging to reduce such tree correlation. At each split point during the growth of the tree, Random Forests only applies the greedy algorithm to a randomly selected proportion of the original predictors.

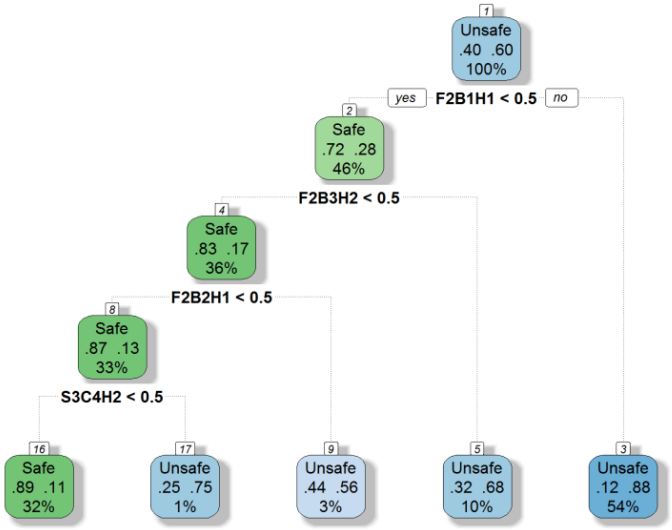
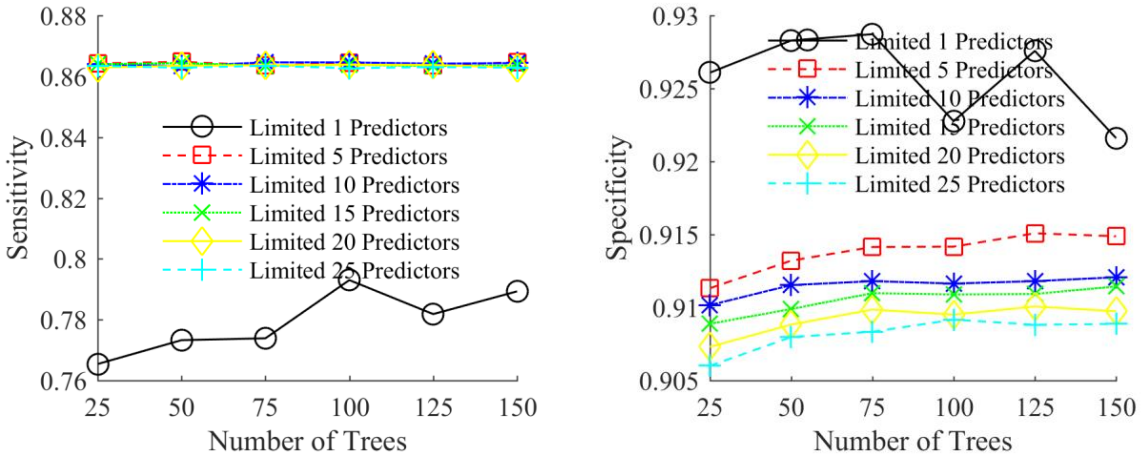



*Figure 5: Example tree structure, where FiBjHk denotes the damage states at Hinge k of Beam j at Floor i , and SiCjHk denotes the damage states at Hinge k of Column j in Story i .*

Figure 6 illustrates the parameter tuning process for the Random Forests models trained using the damage patterns. It can be seen that the model with only 1 predictor at each split has the lowest performance for sensitivity but highest performance for specificity, while all the other models are comparable. Based on these observations, the optimal model allows 5 predictors to be considered at split points and uses 150 trees. Table II summarizes the performance measures including sensitivity, specificity and accuracy for the final Random Forests models using 25% held-out testing datasets. Predictions obtained are compared with their corresponding reference (real) states. 240 testing samples with response patterns are employed in the final CART model, where 76 true negative and 143 true positive predictions are observed. Only 17 cases are recognized as unsafe (negative) when they are actually safe (positive), while 4 unsafe ones are misclassified as safe. An overall accuracy of 91.3% is achieved, and more importantly, the false negative rate is only 1.7%, giving a satisfied low rate of risky predictions. 22900 testing damage

patterns are used in the final model, of which 7287 and 12913 correct predictions are found for the safe and unsafe states respectively. The false positive rate is 14.3% and the more critical false negative rate is 10.3%.




(a) (b)

*Figure 6: Parameter tuning for Random Forests models trained using damage patterns*

*Table II. Performance measures of final models based on testing datasets*

| | Model using response patterns | | Model using damage patterns | |
|---|---|---|---|---|
| | Reference negative | Reference positive | Reference negative | Reference positive |
| Predicted negative | 76 | 4 | 7287 | 1487 |
| Predicted positive | 17 | 143 | 1213 | 12913 |
| Sensitivity | 97.3% | | 89.7% | |
| Specificity | 81.7% | | 85.7% | |
| Accuracy | 91.3% | | 88.2% | |
| 2.5% | 86.9% | | 87.8% | |
| 92.5% | 94.5% | | 88.6% | |

The predictive models based on the response patterns have an overall slightly better performance the damage pattern models. A possible explanation is that additional randomness is introduced by using component fragility curves and Monte Carlo simulation to generate the damage pattern from the EDPs. Moreover, the precision of the predictor information is somewhat reduced when continuous EDPs are converted to discrete damage states. With that being said, it should be noted that exact or real EDP values are assumed in the response pattern models even though the field instruments used to record these EDPs may introduce measurement errors, which could potentially lower the performance. This issue is not addressed in the current study. More importantly, unlike the response patterns which require pre-installed instruments, the damage patterns can be more easily obtained after an earthquake through visual inspection.

# 6 Conclusion

Random Forests Machine learning algorithm is implemented to map the response and damage patterns to their classified structural safety states, where the predictor space is recursively partitioned to capture the underlying relationship. Previously developed post-earthquake structural

evaluation methods have based the safety state criteria on the exceedance of pre-defined response demand levels or damage state ratios within individual structural component groups. In contrast to them, the outcome of the proposed framework consists of classification trees, of which each comprises a high-dimension space of response and damage states partitioned into multiple subspaces. Given an arbitrarily observed response and/or damage pattern of a damaged structure, each tree is intelligently searched to find the matching subspace, which will serve as the basis for classifying the safety state of the structure. The final result is a discrete probability distribution of the structural safety state, which serves as an indicator of the confidence of the prediction.

The proposed methodology is applied to a 4-story RC SMF building located in a region of high seismicity. The results show that although different response and damage patterns exist for the safe and unsafe structural states, they cannot be separated by assigning a certain criterion for any individual engineering demand parameter or damage measure. Such a criterion would result in considerable overlap between the two safety states. The predictive models presented in this study, which are trained using machine learning algorithms, are able to provide predictions with high performance in terms of accuracy, sensitivity and specificity. The prediction accuracy is 91.3% when response patterns are used in the Random Forests model, and 88.2% accuracy is achieved for damage patterns. Relatively high sensitivity is also observed in the prediction models, which is critical for reducing the rate of high-risk erroneous predictions i.e. model predicts safe when building is unsafe.

The proposed framework could be used for rapid probabilistic assessment of whether a damaged building is safe to reoccupy following an earthquake. Additionally, the trees generated by the machine learning algorithms could be used to prioritize field inspections following an earthquake. Moreover, the probabilistic safety state predictions could be used in community resilience evaluations and individual building life-cycle performance assessment and optimization.

## Acknowledgement

The research presented in this paper is supported by the National Science Foundation CMMI Research Grant No. 1538866.

## References

[1] G. P. Cimellaro, A. M. Reinhorn, and M. Bruneau, “Framework for analytical quantification of disaster resilience,” *Eng. Struct.*, vol. 32, no. 11, pp. 3639–3649, 2010.

[2] G. P. Cimellaro, A. M. Reinhorn, and M. Bruneau, “Performance-based metamodel for healthcare facilities,” *Earthq. Eng. Struct. Dyn.*, vol. 40, no. 11, pp. 1197–1217, Sep. 2011.

[3] I. Iervolino and M. Giorgio, “Stochastic Modeling of Recovery from Seismic Shocks,” pp. 1–9, 2015.

[4] H. V. Burton, G. Deierlein, D. Lallemant, and T. Lin, “Framework for Incorporating Probabilistic Building Performance in the Assessment of Community Seismic Resilience,” *J. Struct. Eng.*, vol. 142, no. 8, p. C4015007, Aug. 2016.

[5] ATC, *ATC-20 Procedures for Post-earthquake building safety evaluation procedures*. Redwood, CA: Applied Technology Council, 1995.

[6] ATC, *ATC-20-2, Addendum to the ATC-20 Postearthquake Building Safety Evaluation Procedures*. Redwood, CA, 1995.

[7] I. Robertson, T. Kindred, and E. Lau, “Compilation of Observations of the October 15 , 2006,” 2006.
[8] K. A. Porter, A. S. Kiremidjian, and J. S. LeGrue, “Assembly-based vulnerability of buildings and its use in performance evaluation,” *Earthquake Spectra*, vol. 17, no. 2. pp. 291–312, 2001.
[9] J. Mitrani-Resier, S. Wu, and J. L. Beck, “Virtual Inspector and its application to immediate pre-event and post-event earthquake loss and safety assessment of buildings,” *Nat. Hazards*, vol. 81, no. 3, pp. 1861–1878, Apr. 2016.
[10] FEMA, *Seismic Performance Assessment of Buildings Volume 1 – Methodology*. Washington, DC: Federal Emergency Management Agency, 2012.
[11] FEMA, *Seismic Performance Assessment of Buildings Volume 2 – Implementation Guide*. Washington, DC: Federal Emergency Management Agency, 2012.
[12] D. Vamvatsikos and C. Allin Cornell, “Incremental dynamic analysis,” *Earthq. Eng. Struct. Dyn.*, vol. 31, no. 3, pp. 491–514, 2002.
[13] D. Vamvatsikos and C. A. Cornell, “The Incremental Dynamic Analysis and Its Application To Performance-Based Earthquake Engineering,” *Eur. Conf. Earthq. Eng.*, p. 10, 2002.
[14] M. Dolsek, “Incremental dynamic analysis with consideration of modeling uncertainties,” *Earthq. Eng. Struct. Dyn.*, vol. 38, no. 6, pp. 805–825, 2009.
[15] M. Joe, T. Karl, D. Mohr, and H. William, “Test Applications of Advanced Seismic Assessment Guidelines,” *Pacific Earthq. Eng. Res. Cent.*, no. October, 2006.
[16] J. Maffei, K. Telleen, and Y. Nakayama, “Probability-based seismic assessment of buildings, considering post-earthquake safety,” *Earthq. Spectra*, vol. 24, no. 3, pp. 667–699, 2008.
[17] G. L. Yeo and C. A. Cornell, “Building tagging criteria based on aftershock PSHA,” *Proceeding 13WCEE, Vancouver, Canada. Pap.*, vol. 3283, no. 3283, 2004.
[18] M. Raghunandan, A. B. Liel, and N. Luco, “Aftershock collapse vulnerability assessment of reinforced concrete frame structures,” *Earthq. Eng. Struct. Dyn.*, vol. 44, no. 3, pp. 419–439, Mar. 2015.
[19] C. B. Haselton and G. G. Deierlein, *Assessing seismic collapse safety of modern reinforced concrete moment-frame buildings*. PEER Report 2007/08, Pacific Earthquake Engineering Research Center, Berkeley, CA, 2008.
[20] J. W. Baker and C. A. Cornell, “Spectral shape, epsilon and record selection,” *Earthq. Eng. Struct. Dyn.*, vol. 35, no. 9, pp. 1077–1095, 2006.
[21] J. W. Baker, “Conditional Mean Spectrum: Tool for Ground-Motion Selection,” *J. Struct. Eng.*, vol. 137, no. 3, pp. 322–331, 2010.
[22] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[23] R. A. Breiman, L., Friedman, J., Stone, C. J., & Olshen, *Classification and regression trees*. CRC press, 1984.
[24] B. Efron and R. Tibshirani, “Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy,” *Stat. Sci.*, pp. 54–75, 1986.
[25] R. Kohavi and others, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 1995, vol. 14, no. 2, pp. 1137–1145.
[26] L. Breiman, “Bagging predictors,” *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
[27] M. Kuhn and K. Johnson, *Applied predictive modeling*. Springer, 2013.