World Conference on
Timber Engineering
Oslo 2023

DATA-DRIVEN APPROACH FOR ASSESSMENT OF SEISMIC DAMAGE IN WOOD BUILDINGS

Sifat Muin[1], Asif Iqbal[2], John W. van de Lindt[3]

ABSTRACT: Advances in modern computing techniques and capabilities have been utilized in development of new structural health monitoring techniques that utilize a significant volume of data from instrumentation. These data-driven methods open new possibilities for the damage assessment of wood (or any) buildings. A method known as feature engineering is a key step in developing such tools. This paper presents a feature engineering procedure for wood buildings based on data recorded during seismic events and utilizes a full-scale wood shake table test data set and damage inspection results to illustrate the procedure.

KEYWORDS: Data-driven Damage Assessment, Seismic Performance, Resiliency, Wood Buildings, Feature Engineering, Cumulative absolute velocity

1 INTRODUCTION

Over the past decade, significant progress has been made in structural health monitoring through the application of data-driven damage assessment. Data-driven seismic damage assessment evaluates the potential damage to structures or infrastructure caused by earthquakes using data and analytical techniques. Advances in computing technologies, remote sensing, and data science paved the way for new and improved damage assessment tools, many of which are machine learning algorithm-based [1-3]. These ML-SHM techniques, which utilize structural vibration response as the primary data type, have four stages: *health definition*, *data acquisition*, *feature engineering*, and *machine-learned damage assessment*.
Arguments for the damage prediction power of ML have been made in several studies [4-7]. These studies have focused on addressing feature engineering through extracting and selecting features that result in more effective ML model training. This feature engineering step is particularly crucial for ML-based seismic damage assessment as data from the damage class is sparse. Hence, a dataset that provides vibration data from damaged structures is extremely valuable for the development of data-driven damage assessment tools. In this study, feature engineering concepts are applied to a dataset acquired from a series of shake table tests of a four-story full-scale soft-story wood-frame building model. The study's primary goal is to identify the most suitable features for damage assessment of such buildings so that ML-based algorithms would provide the most accurate predictions.

2 FEATURE ENGINEERING

In feature engineering (FE), domain knowledge and data mining techniques are used to identify the most useful features from the data to enhance the performance of ML models. For seismic damage assessment, classifiers or regression models may be used with these features to predict the extent of damage caused to the structure. The data required for the FE are acceleration response histories (RH) and corresponding damage states (DS), or RH-DS pairs. Through *feature extraction*, *exploratory data analysis* ((EDA), and *feature selection*, a vector of damage features is selected that is most suitable for damage classification and localization.

Feature extraction is the process of transforming the measured data into a helpful alternative form. In this step, the RH-DS dataset is transformed into a feature-DS dataset. This transformation reduces the dimension of the dataset keeping the most informative and salient features for damage assessment from the raw data. The extracted damage features include vibration characteristics that have been studied as structural damage indicators by researchers in the past [3]. Past work has shown that these features, referred to as “intensity measures” or “engineering demand parameters” in earthquake engineering, are related to the structural response and are thus likely correlated to damage from a physics-based perspective. Examples of these damage features include peak ground acceleration, peak ground velocity, peak ground displacement, peak responses, cumulative absolute velocity (CAV) [8,9], standardized CAV (CAV_{STD}) [10], Arias Intensity (AI), etc.

[1] Sifat Muin, University of Southern California, USA, muin@usc.edu
[2] Asif Iqbal, University of Northern British Columbia, Canada, asif.iqbal@unbc.ca
[3] John W. van de Lindt, Colorado State University, USA, jwv@colostate.edu

https://doi.org/10.52202/069179-0284

Exploratory data analysis (EDA) is the process of studying the *feature-DS* dataset in detail. EDA aims to gain insights into the underlying structure, patterns, and relationships in the data and to identify any potential problems or outliers. EDA often involves visualizing the data using graphs, plots, histograms, and running summary statistics. By exploring the data this way, researchers can make informed decisions about which features and models to use for further analysis. Figure 1 shows an example of EDA where CAV shows a clear pattern with ductility. This set of data was collected using a finite element model of a building and the displacement ductility was used as a measure of damage. Consequently, it was successfully used as a damage indicator for concrete and steel buildings.

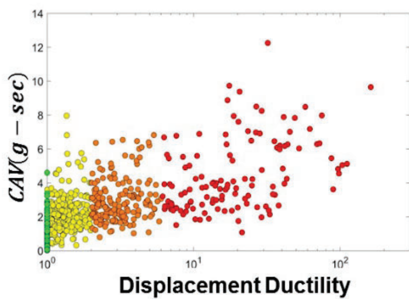

Figure 1. *EDA of CAV showing distinct pattern with increasing ductility (damage) for a reinforced concrete structure. [8]*

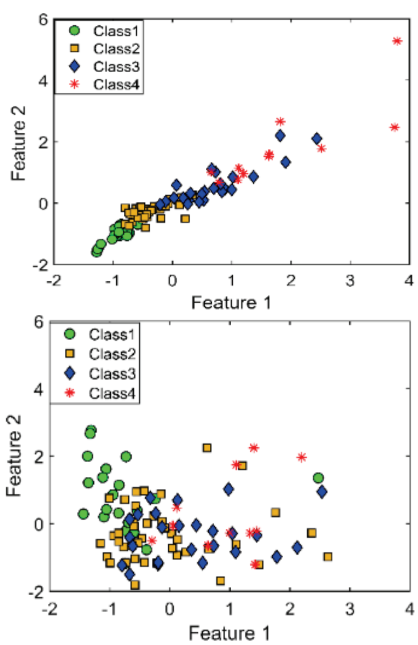

Figure 2. *Discrimination ability of features with high (top) and low (bottom) Fisher scores.*

Feature selection is the process of identifying a subset of the original feature set, which increases the learning efficiency without compromising the classification performance. A filter-based feature selection method is chosen herein called the Fisher Score (FS) because it is computationally inexpensive.). FS determines the relative ability of features to discriminate between categorical classes in a classification model. A larger Fisher score indicates that the feature is more discriminative (Figure 2). Although the Fisher score is known to have limitations as it considers each feature separately and therefore cannot reveal mutual information between features, it is widely used as a heuristic algorithm for feature selection [11].

In this study, the classification model aims to determine which class of DS is indicated by the feature vector. The DSs are ordered in this problem, i.e., if DS0 refers to undamaged and DS1 refers to minor damage, then DS1 is ranked higher than DS0. Due to the ordering of the classes in this model, feature selection that considers *ordinality* is preferred. Thus, an ordinal variant of the Fisher score, the *Ordinal Fisher Score*, is used for feature selection [12]. The ordinal Fisher score (F_{OR}) for the i^{th} feature, x_i is computed following a specific procedure. First, the class discrimination term, $F_O(x_i)$, is defined as follows,

$$F_O(x_i) = \sum_{k=1}^{K}\sum_{j=1}^{K}|k-j| \cdot d^i(C_k, C_j)/\{(K-1)\sum_{k=1}^{K}\left(\sigma_k^i\right)^2 \qquad (1)$$

where the class C_k indicates data from one of the K classes, $k, j \in \{1, 2, \ldots, K\}, \sigma_k^i$ is the variance of the i^{th} feature in class k, and $d^i\ (C_k\ , C_j\ )$ is the distance between classes $C_k\ and\ C_j$. It is computed using the Hausdorff distance which allows for use of nonlinear, multimodal, or non-normal data, and shown in [12] to perform best for ordinal classification. Second, an ordinality term, $O_R\ (x_i)$, incorporating the relative distance between classes, is defined as follows,

$$O_R(x_i) = \sum_{k=1}^{K-2}\sum_{j=k+1}^{K-1}\sum_{h=j+1}^{K} I(d^i(C_k, C_h) - d^i(C_k, C_j) > 0) / \sum_{j=2}^{K-1}(K-j) \qquad (2)$$

where $I(\cdot)$ is the indicator function (1 if the inner condition is true & = 0 otherwise). This $O_R(x_i)$ score measures the number of ordinal requirements fulfilled for the i^{th} feature. Finally, the two terms $F_O(x_i)\ and\ O_R\ (x_i)$ are combined in a weighted sum as follows,

$$F_{OR}(x_i) = \alpha \cdot F_O(x^i) + (1-\alpha) \cdot O_R(x_i), \alpha \in (0,1) \qquad (3)$$

The candidate damage features are evaluated using Equation (3) with $\alpha = 0.1$ to identify which features best discriminate between the different classes of DS. Selecting features that can identify ordinality is crucial for accurate damage assessment. Hence, the ordinality term, $O_R(x_i)$, is given a higher weight $(1 - \alpha = 0.9)$ compared to the class discrimination term, $F_O(x_i)$. The features with the highest Ordinal Fisher Scores constitute the final damage feature set.

3 FULL WOOD BUILDING APPLICATION

The methodology described above is applied to data acquired from a shake table test program on a four-storied full-scale wood building (Figure 3) tested at the University of California San Diego in 2013 under the project titled "Seismic Risk Reduction for Soft-Story Woodframe buildings (NEES-Soft)" [13, 14]. The model represents many thousands of buildings throughout California, and the United States, generally built before 1970 and many as early as the 1920s, following conventional construction practices of the time which are no longer satisfactory by today's seismic codes and standards. The NEES-Soft project included a major testing program to validate performance-based seismic retrofit methodology developed for these buildings. The penultimate experiment of this program was a collapse test sequence. Data from the collapse test is utilized in this investigation as the structure experienced increasing levels of damage.

Figure 3. *NEES-Soft model building*

3.1 MODEL DESCRIPTION AND TESTING

The building in this study was a 370 m^2 (4,000 sqft) four-story light wood-frame building with a "soft-story" archetype most common in San Francisco Bay Area. It had a plan dimension of 7.3 m x 11.6 m and a total height of 10.9 m. The first story (soft story) mainly serves as a parking area with big wall openings for parking doors. The other three upper stories facilitate two typical two-bedroom apartment units [15]. In order to monitor the behavior of the building during testing, the building was instrumented with two string potentiometers and four accelerometers on the first floor, and three accelerometers on the roof. Moreover, six video cameras monitored the movement of the building. Figure 4 presents the location and details of the instrumentation used in the series of tests leading to the collapse of the test building.

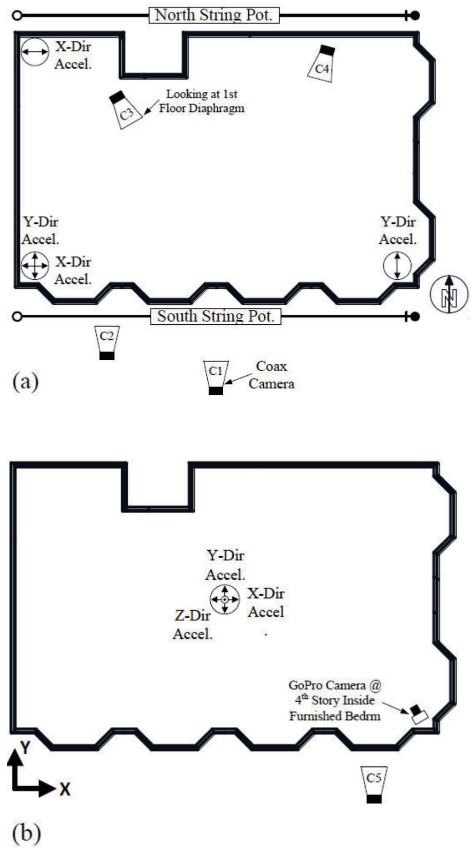

Figure 4. *Instrumentation plan a) first story b) fourth story*

To study the collapse mechanism and behavior of this type of at-risk building, the building was subjected to a range of ground motions with different scaling. Three different ground motions with different intensities were selected. The selections were such that they would provide a range of earthquake records based on differences in ground displacement, even if the seismic intensities as determined through spectral acceleration were similar. The ground motions were then scaled to spectral accelerations ranging from Sa = 0.4g (33% of the design-based earthquake level) to Sa =1.8g (maximum credible earthquake (MCE) level). Table 1 presents the ground motions and test sequences with the corresponding peak ground acceleration (PGA) and peak ground displacement (PGD) for each test.

3.2 DEVELOPMENT OF RH-DS DATABASE

The RH-DS database is created from the acceleration response histories of the eight tests and their corresponding observed damage conditions. The RH data is acquired from the six accelerometers that measured the horizontal responses of the first story and roof. The ground motion RHs were computed from the feedback of the shaketable.

The DS data is based on damages reported during the testing. Due to safety regulations, no damage inspection

https://doi.org/10.52202/069179-0284

and repair was conducted between each consecutive test; therefore, the structural and non-structural damage accumulated during the entire collapse test program. However, building period calculation and residual displacement measurements indicated the level of damage after each test. It was observed that the period of the building increased significantly after Test-4 due to permanent structural damage [16]. But no residual displacements were observed even after Test 5. Test 6 led to extensive permanent damage to the building, bringing it to the verge of collapse with a residual displacement of 350 mm. Then, it was subjected to the same ground motion scaled to Sa=0.9g to evaluate the aftershock performance of the building (Test 7); however, the building did not collapse even with about 406 mm (16 in.) residual drift and 2.3 degrees of residual rotation. The building was then subjected to Test 8, which led to the collapse of the building. Based on these observations, the damage states are classified into four categories, shown in Table 1. These are:

i. DS0 when no damages are reported.
ii. DS1 when the period increase is reported.
iii. DS2 when residual displacement is observed.
iv. DS3 when the building collapsed.

Table 1: *Testing details and damage states*

Seismic Test ID	Eq record	Sa (g)	PGA (g)	PGD (mm)	Damage State
1	Cape Medocino - Rio	0.40	0.21	13.1	DS0
2	Cape Medocino - Rio	0.90	0.44	29.4	DS0
3	Cape Medocino - Rio	1.20	0.56	39.2	DS0
4	Cape Medocino - Rio	1.80	0.90	58.8	DS1
5	Loma Prieta - Gilroy	1.80	0.98	72.1	DS1
6	Superstition Hills	1.80	0.86	277	DS2
7	Superstition Hills	0.90	0.42	138	DS2
8	Superstition Hills	1.80	0.86	277	DS3

4 RESULTS AND DISCUSSION

4.1 FEATURE EXTRACTION:

In the feature extraction phase, the RH-DS database is transformed into a feature-DS database. Six individual features are computed from the RHs. Table 2 shows the feature name and their mathematical definition. For each test, these six features are calculated for the ground response, first story response, and the roof response making the total number of features 18. These 18 features are studied in the EDA phase and utilized in the feature selection phase with corresponding DS shown in Table 1 as their respective class.

Table 2: *Description of candidate damage features*

Feature	Definition
Peak acceleration (PA)	$max(\|\ddot{u}(t)\|)$
Peak velocity (PV)	$max(\|\dot{u}(t)\|)$
Peak displacement (PD)	$max(\|u(t)\|)$
Arias intensity (AI)	$\frac{\pi}{2g}\int_0^T [\ddot{u}(t)]^2 dt$
Cumulative absolute velocity (CAV)	$\int_0^T \|\ddot{u}(t)\| dt$
Standardized CAV (CAV_{std})	$\sum_{i=1}^{N} H(PA_i - 0.025)\int_{i-1}^{i} \|\ddot{u}(t)\| dt$ $H(x) = Heaveside\ function = \begin{cases} 0\ , x < 0 \\ 1, otherwise \end{cases}$ $N: non - overlapped$ 1 sec $intervals\ with\ PA_i$

4.2 EXPLORATORY DATA ANALYSIS

In this phase, the extracted features are studied to understand their relationship to damage and to point out any outliers. Two different tasks were performed as EDA. Firstly, the relationship between CAV profiles and damage states is investigated. Then all 18 features and their progression with damage are studied.

Figures 5 and 6 show each test's CAV time series and Normalized CAV time series. The normalized CAV is computed by dividing the CAV values by the final CAV, making the ultimate value 1. The responses are calculated from acceleration data of X-direction from the first floor. Figure 5 shows that the damage-inducing tests (4,6,8) have high CAV values (>0.8 g-sec). The normalized CAV plots (Figure 6) show a baseline behavior for tests 1, 2, 3, 4 and a departure from this baseline pattern for tests 5,6,7,8. In previous studies, such deviations coincided with damage [8]. Similar behavior is observed for roof response as well. The figures show that CAV-based

https://doi.org/10.52202/069179-0284

features should be considered as candidate damage features.

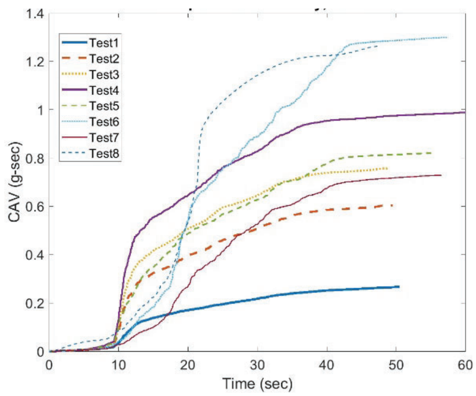

Figure 5. *CAV profile at first floor for X-dir response*

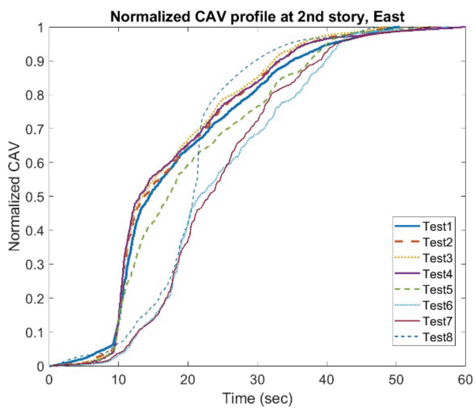

Figure 6. *Normalized CAV profile at first floor for X-dir response*

Figures 7, 8, and 9 show the relationship of the features with the damage states for ground, first floor, and roof response respectively through box plots. The ends of the box plots indicate the quartiles and the line in the middle is the median. The cross inside the box represent the mean. The whiskers (if any) indicate outliers. All ground motion features (Figure 7) depict some trend with increasing damage states. However, for the first floor (Figure 8) and roof responses (Figure 9) the trend is unclear due to the excessively high feature values of DS3. Moreover, the box plots in Figure 8 shows outliers for DS3. Such data, if used in feature selection will produce undesirable outcome. Therefore, DS3 data was not utilized in the feature selection phase following the findings of EDA.

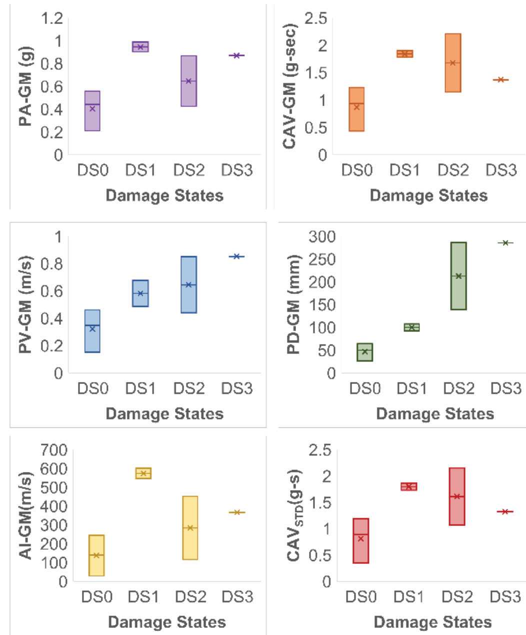

Figure 7. *Features vs damage states computed from the shake table response (ground motion)*

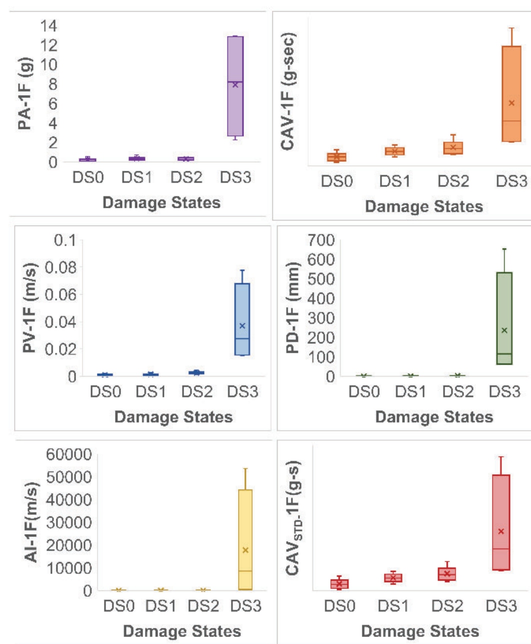

Figure 8. *Features vs damage states computed from the first floor response*

https://doi.org/10.52202/069179-0284

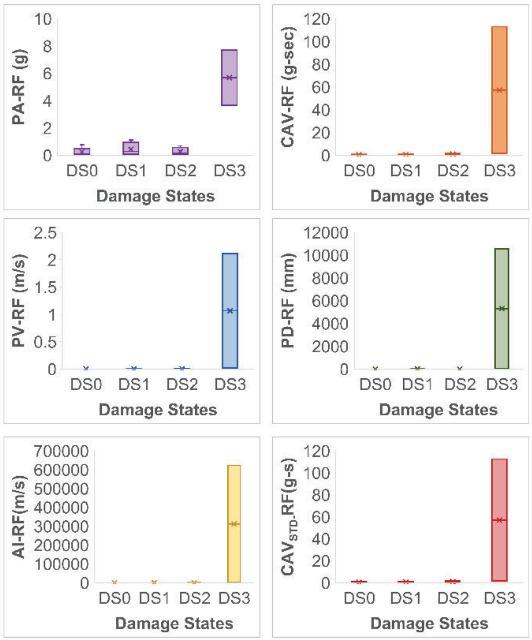

Figure 9. *Features vs damage states computed from the roof response*

4.3 FEATURE SELECTION

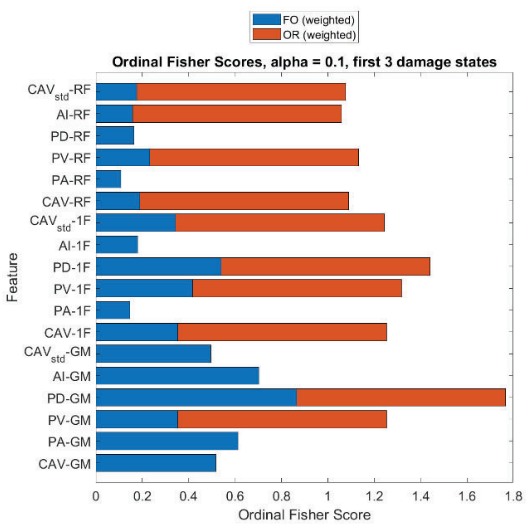

Figure 10. *Ordinal Fisher Scores of the candidate damage features*

In feature selection phase, Fisher score is calculated for all 18 features using Equation 1, 2, and 3. Three damage classes were considered (DS0, DS1, and DS2). Figure 10 shows the ordinal fisher score (F_{OR}). The blue portion is weighted the class discriminatory part and red is the weighted ordinality term. Higher weight is given to the ordinality term so that the features that can identify order of the DS scores higher.

Table 3: *Ordinal Fisher score of the candidate features*

Features	Fisher Score
PD-GM	1.77
PD-1F	1.44
PV-1F	1.32
PV-GM	1.25
CAV-1F	1.25
CAVstd-1F	1.24
PV-RF	1.13
CAV-RF	1.09
CAVstd-RF	1.08
AI-RF	1.06
AI-GM	0.70
PA-GM	0.61
CAV-GM	0.52
CAVstd-GM	0.50
AI-1F	0.18
PD-RF	0.16
PA-1F	0.15
PA-RF	0.11

Table 3 lists the F_{OR} from highest to lowest. Peak displacement of the ground scores the highest among all the features. The high correlation of ground displacement and damage is a well-established observation in the literature. The first floor's peak displacement and peak velocity are the second and third-highest-scoring features. CAV and CAVstd of the first floor, along with the peak velocity of the ground motion, are ranked next with almost the same values. This higher ranking of the CAV features agrees well with Figure 5 and Figure 6 observations. Overall the first-floor features ranked higher than the roof features. This behavior may be due to the "soft-story" mechanism of the structure. One interesting observation is that even though some features (PA-GM, AI-GM) showed a specific trend to damage (Figure 7), their F_{OR} was low. Figure 10 shows that these features have high discriminatory terms but very low ordinality terms bringing down the overall score. Therefore, these features would not be suitable for damage classification.

The results imply that for damage assessment of "soft-story" wood-frame buildings, peak displacement and peak velocity of the ground motion are crucial indicators. Moreover, peak displacement, peak velocity, CAV, and CAVstd of the first-floor response will improve the overall performance of an ML model developed toward damage classification. It should be noted that computing displacement values from acceleration by double integrating can sometimes produce erroneous peak values, particularly for low signal-to-noise ratio measurements. CAV values, CAVstd specifically by definition, avoid such noise levels. Therefore, having these features with the peak values will provide a more reliable outcome.

5 CONCLUSIONS

In this study, feature engineering is applied to a dataset acquired from a series of shake table tests of a soft-story wood-frame building. The building was subjected to increasing levels of ground motion until it collapsed, providing a valuable RH-DS dataset. All three phases of feature engineering have been applied to the data set. During the feature extraction phase, peak values of acceleration, velocity, and displacement, and energy representing features such as CAV and AI are extracted. In the exploratory data analysis phase, CAV-based features of the first floor have shown an obvious pattern with increasing damage states. In this phase, all the features are studied for the ground, first floor, and roof in order to identify any outliers. It was discovered that the DS3 values were outliers compared to other DSs. Therefore, in the feature selection phase, DS3 data was left out.

In feature selection, the ordinal Fisher score is used as identifying the order of the DSs is essential to accurate damage assessment. The analysis showed that the peak displacement of the ground and first floor are the most important features. Peak velocities of the ground and the first floor were also critical. Moreover, CAV and CAVstd of the first floor were also identified as important. Therefore, when developing an ML model for damage assessment of a wood-frame soft-story building, these six features should be used.

ACKNOWLEDGEMENT

The work done for the NEES-Soft project was supported by the National Science Foundation under Grant No. CMMI-1314957 (NEES Research) and NEES Operations. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the investigators and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Sohn, Hoon, et al. "A review of structural health monitoring literature: 1996–2001." Los Alamos National Laboratory, USA 1 (2003): 16.

[2] Fan, Wei, and Pizhong Qiao. "Vibration-based damage identification methods: a review and comparative study." Structural health monitoring 10.1 (2011): 83-111.

[3] Farrar, C. R., & Worden, K. (2012). Structural health monitoring: a machine learning perspective. John Wiley & Sons.

[4] Lam, Heung-Fai, Ka-Veng Yuen, and James L. Beck. "Structural health monitoring via measured Ritz vectors utilizing artificial neural networks." Computer-Aided Civil and Infrastructure Engineering 21.4 (2006): 232-241.

[5] Figueiredo, Eloi, et al. "Machine learning algorithms for damage detection under operational and environmental variability." Structural Health Monitoring 10.6 (2011): 559-572.

[6] Abdul-Aziz, Ali, et al. "Rotor health monitoring combining spin tests and data-driven anomaly detection methods." Structural Health Monitoring 11.1 (2012): 3-12.

[7] Babajanian Bisheh, Hossein, et al. "Damage detection of a cable-stayed bridge using feature extraction and selection methods." Structure and Infrastructure Engineering 15.9 (2019): 1165-1177.

[8] Muin, Sifat, and Khalid M. Mosalam. "Cumulative absolute velocity as a local damage indicator of instrumented structures." Earthquake Spectra 33.2 (2017): 641-664.

[9] Muin S., Mosalam K.: Human-machine collaboration framework for structural health monitoring and resiliency. Engineering Structures, 235: 112084, 2021.

[10] Campbell, Kenneth W., and Yousef Bozorgnia. "A comparison of ground motion prediction equations for Arias intensity and cumulative absolute velocity developed using a consistent database and functional form." Earthquake Spectra 28.3 (2012): 931-941.

[11] Gu, Quanquan, Zhenhui Li, and Jiawei Han. "Generalized fisher score for feature selection." arXiv preprint arXiv:1202.3725 (2012).

[12] Pérez-Ortiz, María, et al. "Fisher score-based feature selection for ordinal classification: a social survey on subjective well-being." Hybrid Artificial Intelligent Systems: 11th International Conference, HAIS 2016, Seville, Spain, April 18-20, 2016, Proceedings 11. Springer International Publishing, 2016.

[13] Bahmani, Pouria, et al. "Experimental seismic behavior of a full-scale four-story soft-story wood-frame building with retrofits. I: Building design, retrofit methodology, and numerical validation." Journal of Structural Engineering 142.4 (2016): E4014003.

[14] van de Lindt, John W., et al. "Experimental seismic behavior of a full-scale four-story soft-story wood-frame building with retrofits. II: Shake table test results." Journal of Structural Engineering 142.4 (2016): E4014004.

[15] Bahmani P., van de Lindt J., Iqbal A., Rammer D.: Mass Timber Rocking Panel Retrofit of a Four-Story Soft-Story Building with Full-Scale Shake Table Validation. *Buildiings.* 7(2): 48, 2017

[16] Bahmani, Pouria; van de Lindt, John; Mochizuki, Gary; Pryor, Steve; Gershfeld, Mikhail; tian, jingjing; Symans, Michael; Rammer, Douglas, (2013), "4-Story Full-Scale Soft-Story Woodframe Building Test at UCSD", DesignSafe-CI [publisher], doi: 10.4231/D3F76678B https://doi.org/10.4231/D3F76678B

 https://doi.org/10.52202/069179-0284