Journal of Building Engineering 97 (2024) 110772

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Incorporation of machine learning into multiple stripe seismic fragility analysis of reinforced concrete wall structures

Hoang D. Nguyen [a], Chanyoung Kim [b], Young-Joo Lee [b], Myoungsu Shin [b, *]

[a] Department of Civil and Environmental Engineering, Faculty of Engineering, University of Alberta, Edmonton, Alberta, T6G 1H9, Canada
[b] Department of Civil, Urban, Earth and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), 50 UNIST-gil, Ulsan, 44919, Republic of Korea

ARTICLE INFO

Keywords:
Fragility analysis
R/C wall structures
Multiple stripe analysis
Machine learning
Random forest
Ensemble machine learning models

ABSTRACT

This study proposes a novel procedure that incorporates machine learning (ML) into the multiple stripe analysis (MSA) approach to efficiently produce seismic fragility curves for reinforced concrete (R/C) shear walls in building frame systems. The proposed procedure aims to mitigate computational challenges associated with the original MSA approach. In this context, ML models were developed for predicting the failure probability of R/C walls subjected to ground motions based on a specified threshold of maximum interstory drift ratio (MIDR). Specifically, the result of each numerical analysis, taken as the output variable of the ML models, was classified as either "B" (Below) or "E" (Exceeding) to indicate whether the MIDR of R/C walls was below or exceeding the specified threshold. This binary categorization was then used to calculate the failure probability points, which are necessary to derive fragility curves per the MSA approach. Data for training and testing the ML models were generated from nonlinear time history analyses of 46 distinct R/C walls subjected to 1000 ground motions. The R/C walls varied in height from four to 40 stories, and the ground motions included far-field, near-field pulse, and near-field no-pulse types. Four well-established ML methods, including random forest (RF), extreme gradient boosting, light gradient boosting machine, and categorical boosting, were considered. The performances of the ML models were compared using a confusion matrix. Based on this comparison, the RF model was selected and incorporated into the proposed procedure. Subsequently, the proposed approach was demonstrated to create the seismic fragility function of a new R/C wall structure. This study highlights the potential of ML applications in optimization problems within the earthquake engineering domain.

1. Introduction

Fragility analysis plays an important role in assessing and mitigating risks and vulnerabilities associated with various hazards of building structures, including earthquakes, hurricanes, or floods [1–3]. Within the realm of earthquake engineering, fragility functions provide the failure probability of a structure subjected to earthquakes of a given value of ground motion intensity measure (IM), such as peak ground acceleration (PGA). Importantly, each fragility curve is constructed with relation to a considered limit state represented by an engineering demand parameter (EDP), such as the maximum interstory drift ratio (MIDR). Furthermore, fragility analysis is integrated into performance-based design to inform design decisions that achieve the desired performance objectives of a

* Corresponding author.
E-mail address: msshin@unist.ac.kr (M. Shin).

https://doi.org/10.1016/j.jobe.2024.110772
Received 1 April 2024; Received in revised form 13 August 2024; Accepted 15 September 2024
Available online 21 September 2024

structure [4–6]. Serving as an effective method for evaluating their seismic performance, fragility analyses have been applied to a wide range of structures, such as steel moment frames [7], reinforced concrete building structures [8], and highway bridges [9].

Conventional fragility analysis methods have involved techniques such as the probabilistic seismic demand model (PSDM), incremental dynamic analysis (IDA), and multiple stripe analysis (MSA). In all these methods, the process begins with constructing a numerical model of a structure and selecting a set of ground motions. Following this, the parameters representing the EDP and ground motion IM are chosen. The PSDM [7,10] involves conducting nonlinear time history analyses of structures subjected to unscaled ground motions. A formulation is then established to present the relationship between the EDPs and IMs. The fragility curve is typically assumed to follow a lognormal cumulative distribution function. In the context of IDA [11–13], the process involves scaling ground motions with multiple scale factors to obtain various levels of IM. The IDA curves are generated by establishing the relationship between the IM values of scaled ground motions and the EDP. The ground motions are scaled until the EDP reaches the considered value, and a separate IDA curve is generated for each ground motion. The IM values are determined from the IDA curves associated with the designated EDP value, and then the fragility curve is derived assuming that the IM values follow a lognormal cumulative distribution function. As for the MSA approach [14,15], the ground motions are scaled at various IM values. At each IM level, the probability of failure is computed by dividing the number of cases where the EDP exceeds a considered value by the total number of ground motions. The fragility curve is then constructed by connecting all the failure probability points using the maximum likelihood estimation (MLE).

The summary above indicates that the conventional fragility analysis methods are straightforward and accessible for implementation. However, they require a large number of nonlinear time history analyses that usually involve significant computational complexity and cost. Recently, machine learning (ML) methods have risen as a promising solution for mitigating the computational cost associated with nonlinear time history analysis [16–21]. Some notable studies can be listed as follows. Nguyen et al. [22] proposed an extreme gradient boosting (XGBoost) model for predicting the seismic drift responses of steel moment frames. Zhang et al. [23] developed ML models (i.e., classification and regression tree, random forest (RF)) to classify the damage states of a 4-story reinforced concrete (R/C) special moment frame building. Nguyen et al. [24] suggested a RF model among eight considered models for predicting the seismic damage states of steel moment frames. Hwang et al. [25] explored ML models for identifying the collapse limit states of 4- and 8-story R/C moment frames. Nguyen et al. [26] proposed XGBoost and gradient boosting regression tree models for predicting the fragility functions of steel moment frames. More ML applications in earthquake engineering can be found in the literature [27,28].

Some studies have demonstrated the potential of ML models in generating seismic fragility functions [29–31]. However, most of these models were trained and developed relying on limited datasets from a single structure, and thus their applicability was constrained to that specific structure, undermining their effectiveness across varied configurations. This study aims to address this limitation by developing a comprehensive predictive model that can generalize across diverse structural configurations, while tackling computational cost challenges associated with traditional fragility analysis methods. The main objective of the current study is to introduce a novel procedure that integrates ML models into the MSA approach for producing seismic fragility functions for R/C wall structures in building frame systems. The proposed procedure uses binary categorization derived from the ML model to estimate failure probabilities at different IM levels of ground motion. Thus, this approach avoids the need for extensive nonlinear time history analyses, reducing computational demands substantially. Furthermore, we augment our study by providing a code package along with detailed guidelines of the proposed procedure to facilitate users in generating the seismic fragility curves of R/C wall structures.

To incorporate the most effective ML model for the proposed procedure, we developed and compared four ML models involving RF, XGBoost, light gradient boosting machine (LightGBM), and categorical boosting (CatBoost). The dataset for developing the four ML models was created through nonlinear time history analyses of 46 R/C wall structures under 1000 ground motions, yielding a total of 46,000 distinct analyses. These structures varied in height from 4 to 40 stories, and the ground motion set comprised 50 pairs of records, scaled to PGA values from 0.1g to 1.0g at 0.1g intervals. Each derived fragility curve corresponds to a specific EDP value (e.g., MIDR of 1 %). and in this study, various threshold MIDR values ranging from 0.5 % to 10 % were considered in the fragility analysis. This extensive data collection and analysis effort was designed to encompass a broad spectrum of structural behaviors, thereby enhancing the generalizability of the developed ML models. Finally, the proposed procedure was applied to generate seismic fragility functions for a new R/C wall structure, serving as an independent validation of the approach.

2. Research methodologies

2.1. Multiple stripe analysis approach

In this study, the MSA approach [14] was utilized to produce the fragility curves of building frame systems in which all the seismic forces are resisted by R/C shear walls. The MIDR was designated as the EDP, and the peak ground acceleration (PGA) was taken as the ground motion IM, as these parameters have been widely employed in previous studies [32–34]. The seismic fragility analysis procedure using the MSA approach can be summarized as follows.

Step 1: Choose a range of PGA values, and subsequently scale ground motions to match the chosen PGA levels.
Step 2: Perform nonlinear time history analyses of R/C wall structures subjected to the scaled ground motions and obtain their MIDRs corresponding to each PGA level.
Step 3: Set the threshold value of MIDR (e.g., 2 % as shown in Fig. 1a) for fragility analysis, as a fragility curve is specifically defined for a particular threshold value.

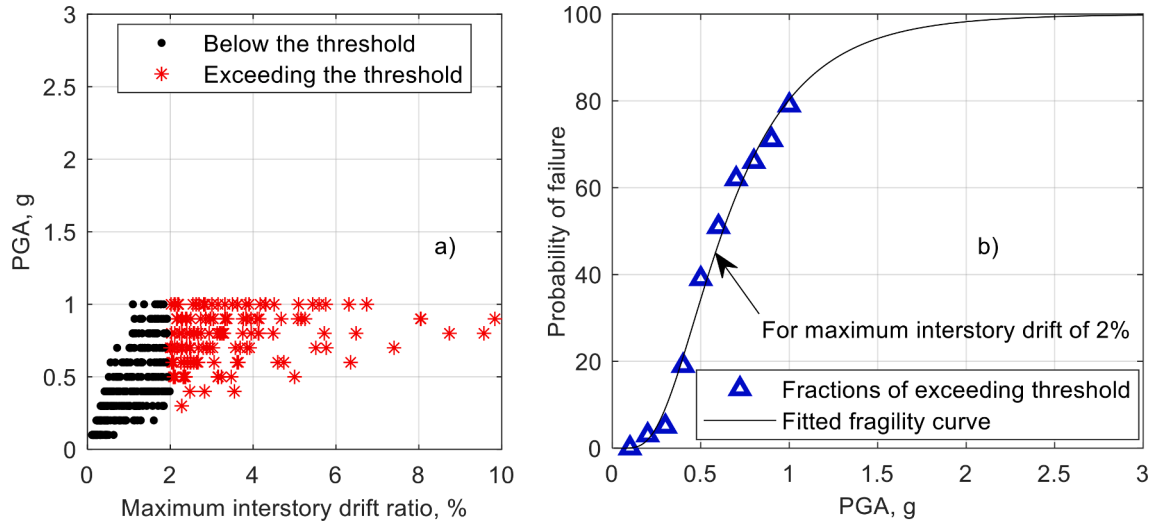


Fig. 1. – a) Example MSA analysis results and b) fitted fragility curve obtained from failure probabilities of analyses exceeding the threshold.

Step 4: Determine the failure probability of the structures at each PGA level taken equal to the ratio of the number of cases that exceed the threshold (i.e., cases in which the MIDR is higher than 2 %, indicated by red asterisks in Fig. 1a) to the total number of analyses (i.e., ground motions).

Step 5: Plot the failure probability of the structures against each PGA level as illustrated by a blue triangle in Fig. 1b. Then, derive the fragility curve that best fits the failure probabilities of the structures at all PGA levels using the maximum likelihood method [35–37], depicted as the solid line in Fig. 1b.

2.2. Development of a predictive model based on ML techniques

Fig. 2 illustrates the process of developing a predictive model using ML techniques [26]. Firstly, a dataset including input and output variables in a given problem is required for training and testing ML models. This dataset can be obtained through experimental tests or generated from numerical simulations. Subsequently, the entire dataset is then partitioned into two sets for training and testing purposes. The training dataset is used to train the ML models, which leads to the establishment of predictive models. Conversely, the testing dataset is used to evaluate the performance of the predictive models. Performance comparisons are then executed among the considered ML models based on the output results predicted from the testing dataset. If the predictive models demonstrate creditable performance, the model with the highest accuracy is proposed as the predictive model. Otherwise, alternative ML techniques may be considered, or the reevaluation of input variables in the dataset may be necessary.

Following the above process, there are four critical tasks to address when implementing ML techniques in any case studies [38,39]. Initially, acquiring a comprehensive dataset is essential to ensure the effectiveness of predictive models. For this purpose, we conducted extensive nonlinear time history analyses on 46 R/C wall structures subjected to 1000 ground motions. Detailed information regarding the considered R/C wall configurations and ground motions is explained in Sections 3.1 and 3.2, respectively. Secondly, the careful selection of input variables is also an important step in developing the ML models. In this investigation, two distinct groups of input variables were analyzed to determine the importance of ground motion type in the development of the ML models and the fragility analysis. The methodology of selecting the input and output variables is detailed in Section 3.3. The third pivotal task is the choice of ML techniques. We implemented four widely recognized ML techniques, described in Section 4.1, to ensure a thorough comparison. Lastly, it is vital to properly define evaluation criteria for the performance of the predictive models. In the case of a clas-

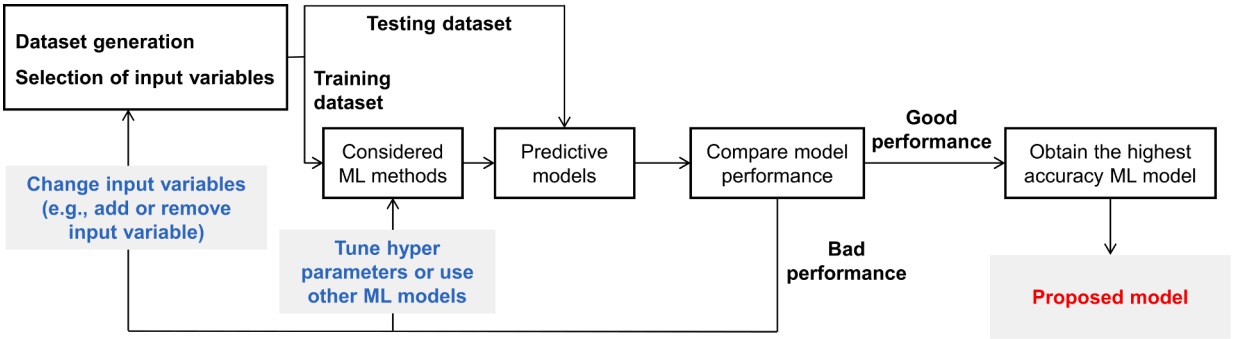


Fig. 2. – Procedure for developing a predictive model based on ML techniques.

sification problem, the confusion matrix [21,25] is typically used for this assessment, and its fundamental concept is discussed in Section 4.2.

3. Dataset constitution

3.1. Archetypes of R/C wall structures and their numerical models

This study employed 46 R/C wall structures designed by Mafari et al. [40,41]. These R/C wall structures have typical configurations, such as the number of stories ranging from 4 to 40, a consistent bay width of 9.15 m in all bays, and a consistent story height of 3.05 m across all levels. Fig. 3 illustrates (a) the typical floor plan view and (b) the basement plan view. The 4-story and 8-story configurations have two and three basement levels, respectively, while all the taller configurations include four basement levels. It is noted that this study examined R/C wall structures in building frame systems only, where all seismic forces are resisted by the shear walls [42]. All the R/C wall structures were designed in collaboration with the Earthquake Engineering Committee of the Structural Engineering Association of Washington [41].

According to Mafari et al. [40,41], the R/C wall structures were designed and detailed in accordance with ACI 318-14 [43], considering standard loads and load combinations in ASCE 7–16 [42]. Dead load, comprising both self-weight and superimposed loads, was applied as 6.2 kPa for typical floors. Meanwhile, the live load was specified as 2.4 kPa. Regarding the earthquake-resistant design, the equivalent lateral force procedure outlined in ASCE 7–16 [40] was followed. The structures were assumed to be located in Seattle, United States, with a shear wave velocity of 500 m/s (soil class C). The short-period spectral acceleration and 1s spectral acceleration were assumed to be 1.12g and 0.49g, respectively. For the materials, the compressive strength of 55.2 MPa was used for concrete, and the nominal yield stress of 414 MPa was utilized for reinforcement.

In order to establish a comprehensive design framework for R/C wall structures, Marafi et al. [40] explored various design criteria to create 46 distinct R/C wall structures. This process involved setting the interstory drift limit at three different levels—1.25 % for an enhanced code performance, 1.5 % as a median case for generating a broader range of design scenarios, and 2 % for the minimum code performance per ASCE 7–16 [42]. Moreover, the seismic load was taken equal to the minimum load required by the design code [42], or increased by 25 % and 50 % of the minimum load. These variations led to the formation of seven groups of R/C wall structures, each with distinct design specifications. Detailed descriptions for the seven groups of R/C wall structures can be found in the literature [40,41].

To generate the dataset used for developing ML models, we constructed and analyzed 2-dimensional (2D) nonlinear models of the 46 R/C wall structures using OpenSees software [44], following the modeling approach suggested by Marafi et al. [41]. This approach was initially introduced by Pugh et al. [45] and was based on data from 30 experimental tests. Marafi et al. [41] refined Pugh et al.'s approach [45] by adding displacement-based beam-column elements (DBCEs) that simulate the combined axial and bending behavior of RC walls. Although we recognize other efficient methods [46,47], comparing different modeling approaches (e.g., dis-

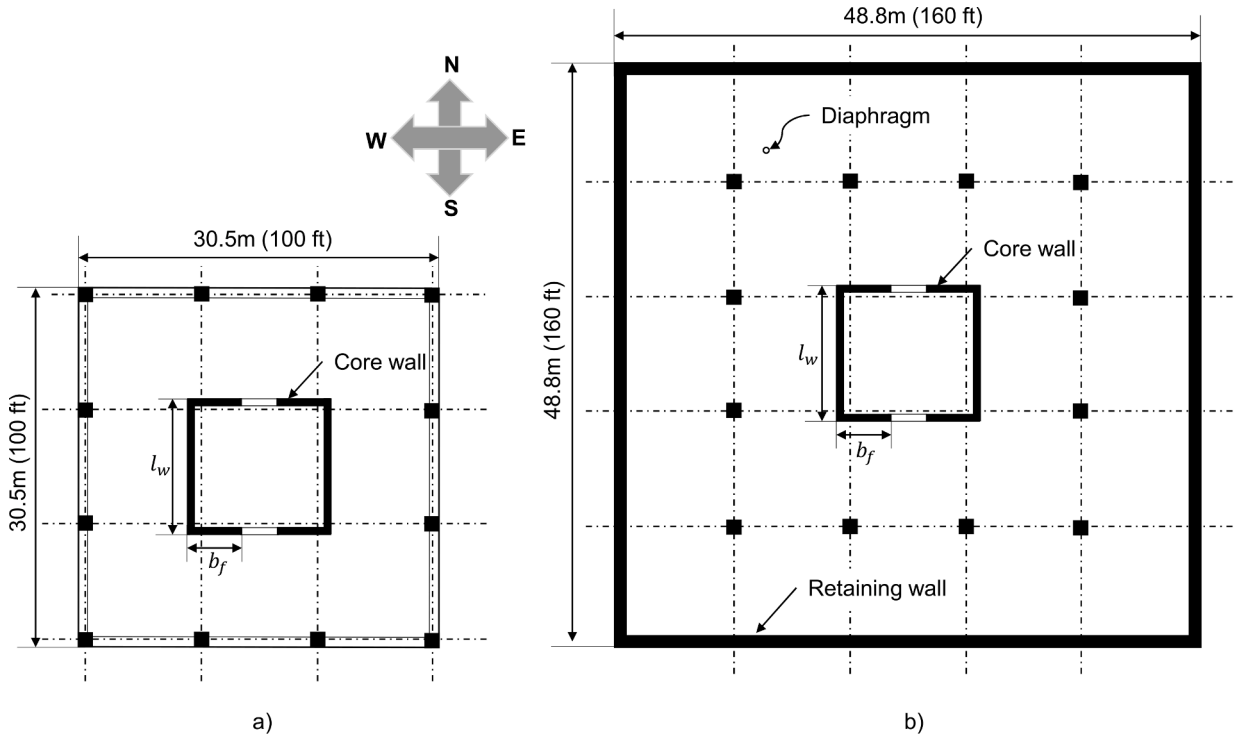


Fig. 3. –Plan views of a) typical floors and b) basement floors [40,41].

placement-based vs. force-based) is out of the scope of this study. In the 2D models, the effects of torsion and bidirectional loading were neglected, and only the response of walls in the N-S direction in Fig. 3 was analyzed in this study. This was modeled using two separate C-shaped walls, assumed to function together as an I-shaped wall.

Fig. 4 illustrates the construction of the nonlinear numerical model. Each story of the walls was represented by six DBCEs, as shown in Fig. 4a. Each of the DBCE elements has five integration points (Fig. 4b). To simulate shear deformation along the height of the walls, a linear elastic shear spring (using a zero-length element in OpenSees) was incorporated into each DBCE element (Fig. 4b). The stiffness of the shear spring was determined by multiplying the shear modulus of concrete with the effective shear area and then dividing by the length of the wall element. Additionally, elastic springs were used to model the behaviors of retaining walls and diaphragms at and below the ground floor (Fig. 4a). Axial springs were used to represent the stiffness of the diaphragms, while shear springs represented the stiffness of the retaining walls. The stiffnesses of the axial and shear springs were calculated according to the recommendations in Tables 4–3 of the Tall Buildings Initiative [48]. The nonlinear behavior of the R/C cross section was modeled using the fiber section approach, illustrated in Fig. 4c. The fiber section was applied at every integration point to simulate the axial and flexural behaviors of the cross-section. In all the models, each fiber represented a width of 0.1524 m (6 inches). As an example, for the 32-story model in group 1, 54 fibers were utilized spanning a length of 8.2296 m (324 inches). To consider the P-delta effect due to gravity loads on the entire structure, a leaning column was included at each story, and connected to the wall by a rigid truss element at each floor level. Nonlinear stress-strain behaviors of concrete and reinforcing steel were simulated using *Concrete02* and *Steel02* material models in OpenSees. It should be noted that cyclic strength deterioration was not considered in the *Steel02* model. Further information on the construction of the numerical model can be found in the literature [41].

To validate the numerical model used in this study, a comparative analysis was conducted between the numerical and experimental results on the cyclic responses of R/C walls. The experimental test conducted by Shegay et al. [49] was used for this purpose. The tested wall had a height of 3.5 m and a cross-section measuring 2.25 m in width and 0.2 m in thickness, as illustrated in Fig. 5a. Detailed information for the design specifications and experimental setup is provided in the literature [49] (see Fig. 5a).

Fig. 5b compares the cyclic responses of the R/C wall [49] derived from the experiment and the numerical model (denoted as OpenSees in the figure). It can be seen that the cyclic response of the numerical model closely accords to that observed in the experiment. Specifically, the difference in the maximum base moment is minimal, 1.57 % at most. These minor discrepancies are considered negligible, underscoring the numerical model's precision. This confirms the validity of the numerical modeling approach and its reliability for the dataset generation required to develop the ML models in this study.

3.2. Selected ground motions

In this investigation, 50 pairs of ground motions introduced in ATC-63 [50] were employed; each pair consists of two individual components in orthogonal directions. Of them, 22 pairs are categorized as far-field (FF), 14 pairs as near-field with-pulse (NF-Pulse), and 14 pairs as near-field without-pulse (NF-No Pulse) ground motions. All the ground motions possessed a magnitude exceeding 6.5, PGA greater than 0.2g, and peak ground velocity (PGV) exceeding 15 cm/s. Furthermore, the distance from source to site for FF ground motions is or greater than 10 km, while it is less than 10 km for NF-Pulse and NF-No Pulse ground motions. All the ground

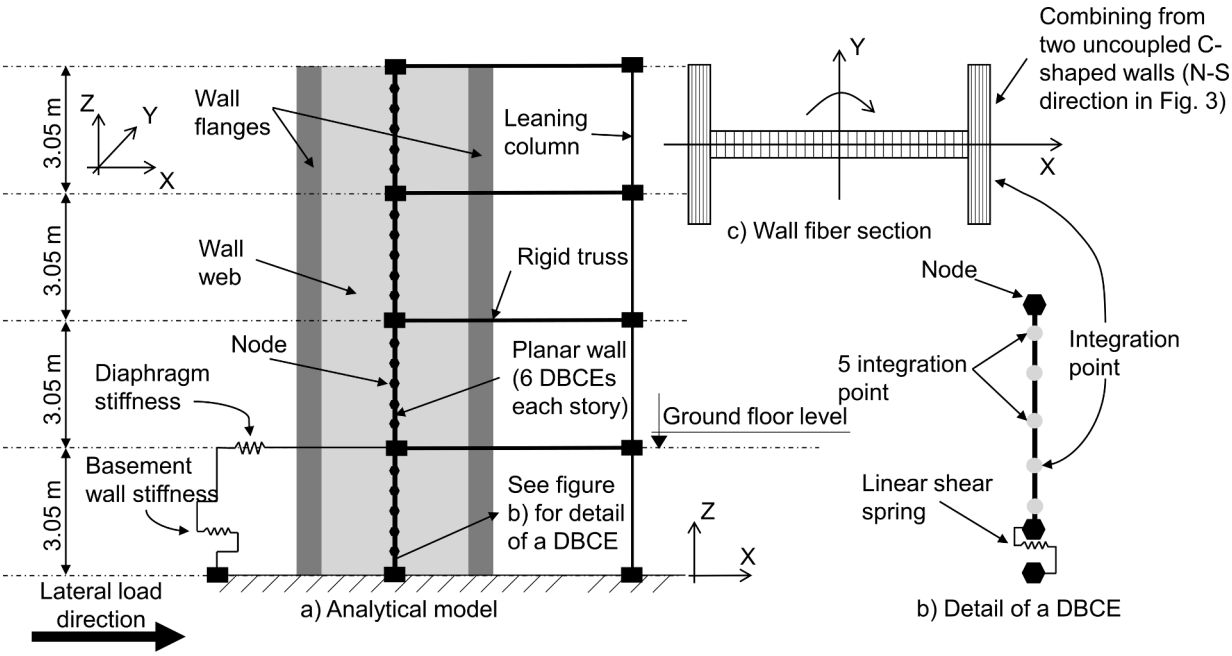


Fig. 4. – Schematic illustration of the OpenSees numerical model for R/C wall structures [41].

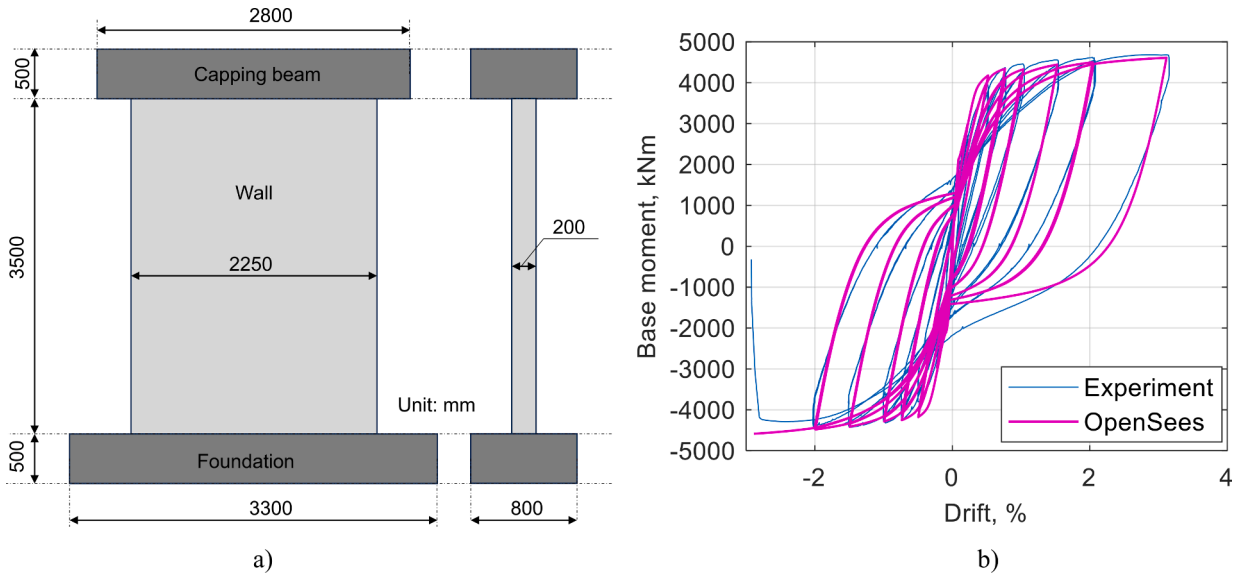


Fig. 5. – a) Elevation and side views of the experimental specimens and b) cyclic responses of numerical modes against experiment conducted by Shegay et al. [49].

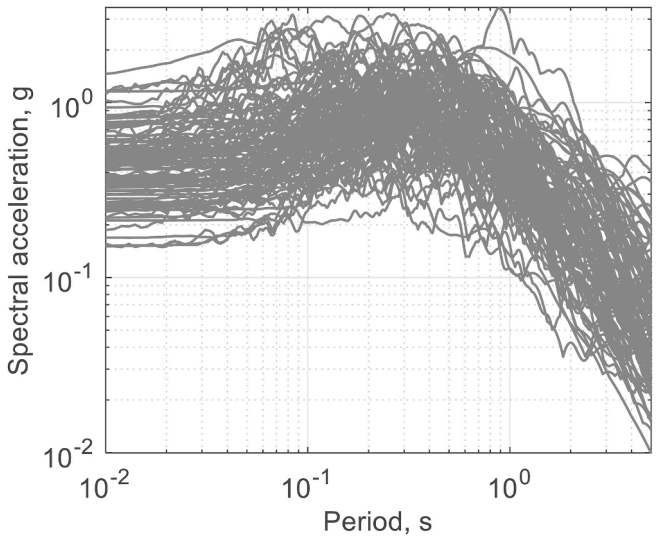
motions were recorded on rock or stiff soil conditions with a shear wave velocity greater than 180 m/s Fig. 6 presents 5%-damped response spectra for all the 100 individual ground motions.

In this study, the results predicted by ML models were integrated into the MSA approach. Therefore, the PGA values of ground motions are expected to fall within a stripe as well as should have a wide range. To this regard, we scaled the collection of 100 ground motions, each to have ten PGA values ranging from 0.1 g to 1.0 g in increments of 0.1 g. This scaling procedure led to a total of 1000 ground motions. Further details on the selected ground motions can be found in Appendix A in FEMA P695 [51].

3.3. Input and output variables of ML models

In this study, the dataset for training and testing ML models was generated using nonlinear time history analysis. The combination of the 46 R/C wall structures (Section 3.1) with the 1000 ground motions (Section 3.2) yielded a total of 46,000 analyses. Two different sets of input variables were examined to identify the significance of ground motion type in the development of the ML models as well as the fragility analysis. Specifically, the second dataset incorporates the ground motion type (i.e., FF, NF-Pulse, or NF-No Pulse) in addition to the first dataset. Moreover, considering these two input sets can enhance the applicability of the developed ML models.

The rationale behind selecting input variables for the development of ML models in this study was to emulate the modal response spectrum analysis [52,53]. Therefore, the chosen input variables are associated with the structural periods of the walls and the response spectra of the considered extensive set of seismic ground motions. In the first dataset, we selected the first five periods of R/C



Figs. 6. 5%-damped response spectra of 100 unscaled ground motions.

wall structures to capture the impact of higher modes on the structural response. In addition, we included the PGA and spectral accelerations at 1 s–6 s, at 1-s intervals, to fully represent the response spectra of ground motions. The PGA, representing the spectral acceleration at nearly zero period, and the spectral acceleration at 6 s, corresponding to the longest period observed in the 46 R/C wall structures (approximately 5.327 s), were included to cover the entire range of fundamental periods pertinent to the structures under consideration.

In the development of a fragility curve in this study, the MIDR was considered as a performance parameter to define the limit state of R/C wall structures. The minimum and maximum MIDRs in the dataset were 0.036 % and 14.88 %, respectively, noting that we set the threshold of 15 % to stop the analysis to minimize the convergent issues of large-displacement nonlinear analyses in OpenSees. Meanwhile, the average and standard deviation of the MIDR results were 2.16 % and 2.29 %, respectively. In the fragility analysis, each fragility curve corresponds to a defined MIDR. In this study, the threshold for MIDR varied between 0.5 % and 10 % in increments of 0.1 % for each structural configuration. This extensive range was designed to cover a broad spectrum of structural behaviors, thereby increasing the applicability of the developed ML models. In other words, by considering the MIDR value as an input with a wide range, the developed model can generate fragility curves for various threshold values. In summary, the first dataset comprised a total of 13 input variables: five periods of structures, seven spectral accelerations of ground motions, and the MIDR threshold. On the other hand, the second dataset comprised 14 input variables, adding the ground motion type (FF, NF-Pulse, or NF-No Pulse) to the first dataset. It is important to note that the two input sets considered are suitable for both the preliminary analysis and post-earthquake fragility assessment of building structures. This is because the response spectra can be derived from actual ground motion records (for post-earthquake assessment) or from design standards and codes (for preliminary analysis). This versatility highlights the advantage of the input sets examined in this study.

The output variable of the ML models was set to classify the states of R/C wall structures subjected to ground motions based on a predefined MIDR threshold (varied between 0.5 % and 10 % in increments of 0.1 %). Specifically, the state of an R/C wall structure against a ground motion is classified as either “B” (Below) or “E” (Exceeding) to indicate whether the MIDR is below or exceeds the specified threshold. This binary categorization was then used to estimate the failure probability of the structures for a given PGA, necessary for deriving the fragility curve through the MSA approach.

Combining 46,000 nonlinear analyses with 96 MIDR threshold values resulted in a total of 4,416,000 data points for the development of the ML models. Table 1 summarizes the boundaries of the input variables. It is worth of noting that the ML models will generally achieve better predictions when interpolating within the trained data ranges [16]. The procedure incorporating predicted results from a ML model into the MSA approach is presented in Section 6.

4. Machine learning methods

4.1. Theoretical overview of selected ML methods

In this study, we employed four ensemble ML models: RF, XGBoost, LightGBM, and CatBoost, to classify the state (i.e., “B” or “E”) of R/C wall structures. All considered ML models are based on the ensemble learning method, which combines the outputs of multiple decision trees (i.e., also known as weak learners). The ensemble model development process is depicted in Fig. 7. The primary difference between these ensemble models lies in the construction of their decision trees, achieved through either bagging or boosting techniques. The bagging method constructs decision trees in parallel, allowing them to learn independently from each other, while the boosting method builds trees sequentially, with each tree learning from the errors of the previous ones.

RF model [54] is an advanced form of the bagging method, which combines bagging with a random selection of features to improve model diversity and accuracy. Specifically, in each individual tree, the RF model aims to identify the optimal set of input variables based on the given dataset. XGBoost, LightGBM, and CatBoost, on the other hand, employ the boosting method to construct decision tree. The XGBoost model was developed by Chen and Guestrin in 2016 [55]. XGBoost constructs individual trees using the

Table 1
– Statistical analysis of the input variables.

Input variable no.	Description	Min	Max	Mean
1	PGA (g)	0.100	1.000	0.550
2	Spectral acceleration at 1 *s* (g) - Sa (1s)	0.027	3.526	0.665
3	Spectral acceleration at 2 *s* (g) - Sa (2s)	0.009	2.652	0.333
4	Spectral acceleration at 3 *s* (g) -Sa (3s)	0.006	2.084	0.212
5	Spectral acceleration at 4 *s* (g) - Sa (4s)	0.004	1.188	0.143
6	Spectral acceleration at 5 *s* (g) - Sa (5s)	0.003	1.116	0.108
7	Spectral acceleration at 6 s (g) - Sa (6s)	0.002	0.797	0.079
8	First natural period (s) - *T*1	0.552	5.327	2.283
9	Second natural period (s) - *T*2	0.100	0.933	0.399
10	Third natural period (s) - *T*3	0.049	0.346	0.154
11	Fourth natural period (s) - *T*4	0.044	0.192	0.093
12	Fifth natural period (s) - *T*5	0.028	0.124	0.071
13	Ground motion type (only for the second dataset) - GM Type	FF: “1” NF-Pulse: “2” NF-No Pulse: “3”		
14	Threshold value for MIDR (%) – performance parameter	0.5	10	5.25

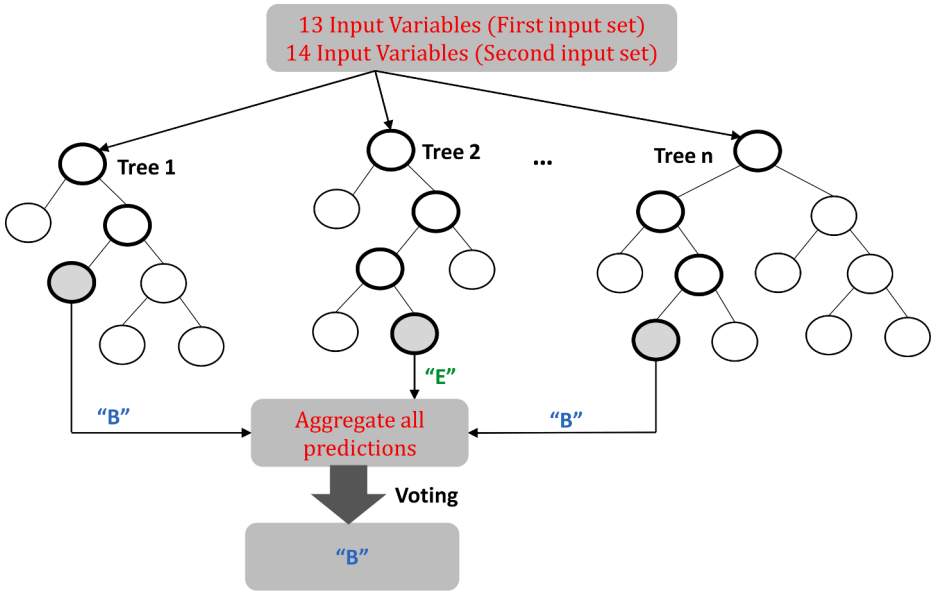


Fig. 7. – Example of an ensemble model.

level-wise (horizontal) growth approach. It employs an iterative approach where each new tree is adjusted based on the residuals—the differences between the observed and predicted values—from the previous tree. In addition, the regularization term is added into the objective function to reduce the overfitting issue.

LightGBM [56], while based on a concept similar to the XGBoost model, distinguishes itself by employing a leaf-wise (vertical) tree growth approach. This key distinction allows LightGBM to often achieve a more significant reduction in loss at each iteration, which can result in a more accurate predictive model. Moreover, LightGBM is known for its faster processing times when compared to XGBoost and is capable of maintaining high levels of accuracy.

CatBoost [57], another gradient boosting-based method, differentiates itself by constructing symmetric (balanced) trees. This means that the conditions for splitting leaves in subsequent trees are consistent with those in previous trees. Additionally, within the CatBoost framework, a uniform feature-split pair technique is applied to all nodes at the same level. This approach is designed to minimize loss within the objective function. The balanced growth of trees not only contributes to faster prediction times but also acts as a regularization mechanism, effectively reducing the risk of overfitting. Detailed information about the four ensemble ML models examined in this study can be found in the literature [20,24].

4.2. Criteria to evaluate performance of the ML models

In ML applications, particularly for classification problems, the evaluation of predictive models is often performed using a confusion matrix [21,24,58]. Fig. 8 illustrates the concept of the confusion matrix and includes the formulas required to compute various accuracy metrics. In this research, we utilized four key performance indicators — classification accuracy, precision, recall, and the F1-score — to assess and compare the efficacy of the predictive models.

Classification accuracy is a vital metric for assessing the global performance of a model in classification problems. It measures the proportion of instances where the model correctly predicts the target label. This is calculated by dividing the number of correct predictions by the total number of predictions made. Meanwhile, precision and recall are metrics used to evaluate the performance of predictive models on specific class predictions. Precision measures the proportion of true positive predictions in all positive predic-

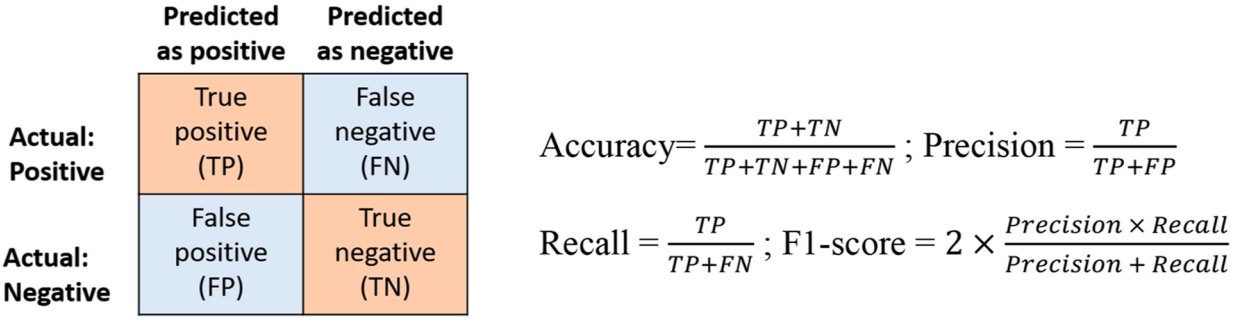


Fig. 8. – Concept of confusion matrix.

tions made by the model. This metric is especially critical in situations where the consequences of false positives (FP) are more severe than false negatives (FN).

Recall, on the other hand, is a measure of actual observations that are predicted correctly, i.e., how many observations of positive class are actually predicted as positive. Additionally, recall proves to be an advantageous metric in cases where the cost of FN is of higher concern than that of FP. The F1-score, conversely, represents a harmonic mean of both precision and recall metrics. In simpler terms, the F1 score attains a harmonious balance between the precision and recall values within the classified model. In other words, F1-score is a better metric for an imbalanced dataset. More information on the confusion matrix can be found in the literature [24].

5. Development of predictive models for classifying states of R/C wall structures

In this section, we developed four ML models to predict the states (i.e., B or E) of R/C wall structures utilizing the two input sets discussed in Section 3.3. The ML models were developed using the training dataset, which constitutes 60 % of the entire dataset, while the testing dataset, representing the remaining 40 %, was employed to assess model performance. Figs. 9 and 10 illustrate the confusion matrices for the testing dataset, along with accuracy, precision, recall, and F-1 score parameters that define model performance, for the first and second input sets, respectively. It should be noted that the input parameters (e.g., number of trees, maximum depth) defining the ML models were set to their default values [54–57]. Hyperparameter tuning was not conducted as the performances of the developed models were outstanding, as demonstrated below.

Fig. 9 reveals that all four ML models demonstrated satisfactory performance in predicting the states of R/C wall structures using the first input set. Notably, the RF model emerged as the top performer, achieving an impressive accuracy of 0.99 and an F1-score averaging 0.98. Close behind, the CatBoost model showed strong results with an accuracy of 0.98 and an average F1-score of 0.96. The LightGBM model, while still performing well, achieved the lowest metrics of the group, with an accuracy of 0.96 and an average F1-score of 0.93. Although the performance varied among the models, the disparities in global performance among all the considered ML models were relatively small.

Regarding individual predictions, the RF models consistently exhibited the highest values for both precision and recall parameters. Specifically, precision values of 0.99 and 0.97 were achieved for "B" and "E" states, while recall values reached 0.99 and 0.97, respectively. In contrast, the LightGBM model scored the lowest on these metrics. It is worth noting that the ML models performed better in predicting "B" state compared to "E" state. This difference in performance can be explained by the fact that the dataset contains a relatively smaller proportion of "E" state compared to "B" state. This observation was consistent across all developed ML models. However, the RF model showed very small differences in precision and recall values between "B" and "E" states (i.e., 0.99 and 0.97), underscoring its outstanding performance in this comparison once again. In other words, the RF model demonstrated great performance, even when dealing with imbalanced datasets.

Consistent with the findings in the first input set, the RF model maintained superior performance in the second input set, as demonstrated in Fig. 10. It outperformed other models in overall accuracy as well as in individual prediction metrics. The CatBoost model followed closely, showing strong performance, whereas the LightGBM model displayed the least impressive results among the group.

Given its superior performance in both global and individual predictions across the two input sets, the RF model is recommended as the most accurate for predicting the states of R/C wall structures in this study. It is notable that each ML model has its own advan-

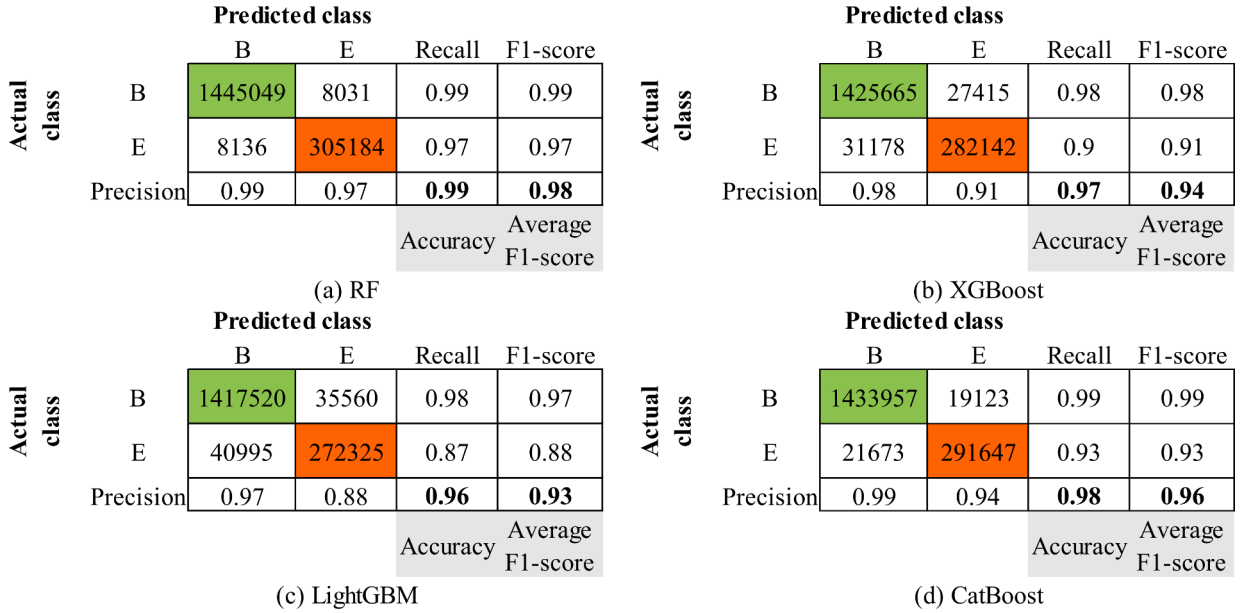


Fig. 9. Confusion matrices (testing dataset) of four ML models in the first input set.

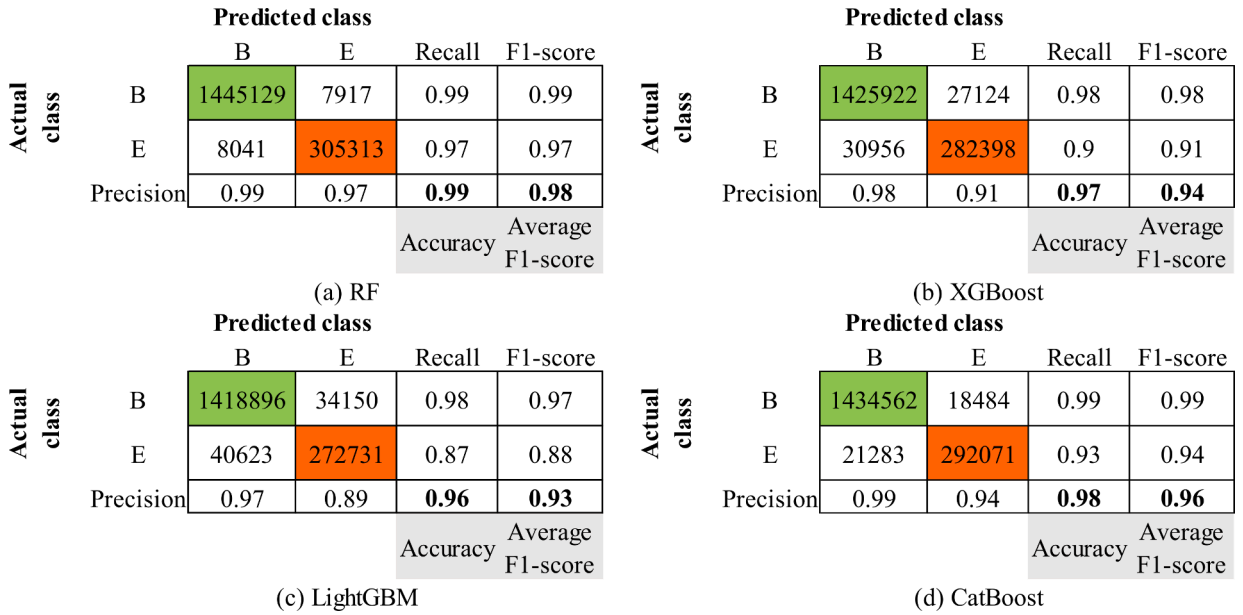


Fig. 10. Confusion matrices (testing dataset) of four ML models in the second input set.

tages, and there is no single ML model that guarantees the highest prediction accuracy across all case studies. For instance, Nguyen et al. [21] demonstrated that the LightGBM model was more effective than the RF model in predicting damage states in underground box tunnels due to earthquakes. Kourehpaz et al. [59] found that the gradient-boosting model outperformed the RF model. Meanwhile, the RF model outperforms the other models in the current study. Given these variations, it is essential to explore and compare multiple ML models to determine the most accurate for a specific application.

Fig. 11 shows the relative importance of input variables extracted from two RF models corresponding to the two input sets. It can be seen that the MIDR has the most significant influence in both input sets. This underscores the reliability of the RF models, as the MIDR threshold was crucial for determining the output variables ("B" or "E") in the dataset. In the second input set, the ground motion type (denoted as GM Type in Fig. 11) has a minimal effect on the predictions. In the comparison of model performance between the two considered input sets, the RF model consistently yielded identical results across all accuracy parameters. Consequently, either the first input set, which excludes ground motion characteristics, or the second, which includes them—can be effectively em-

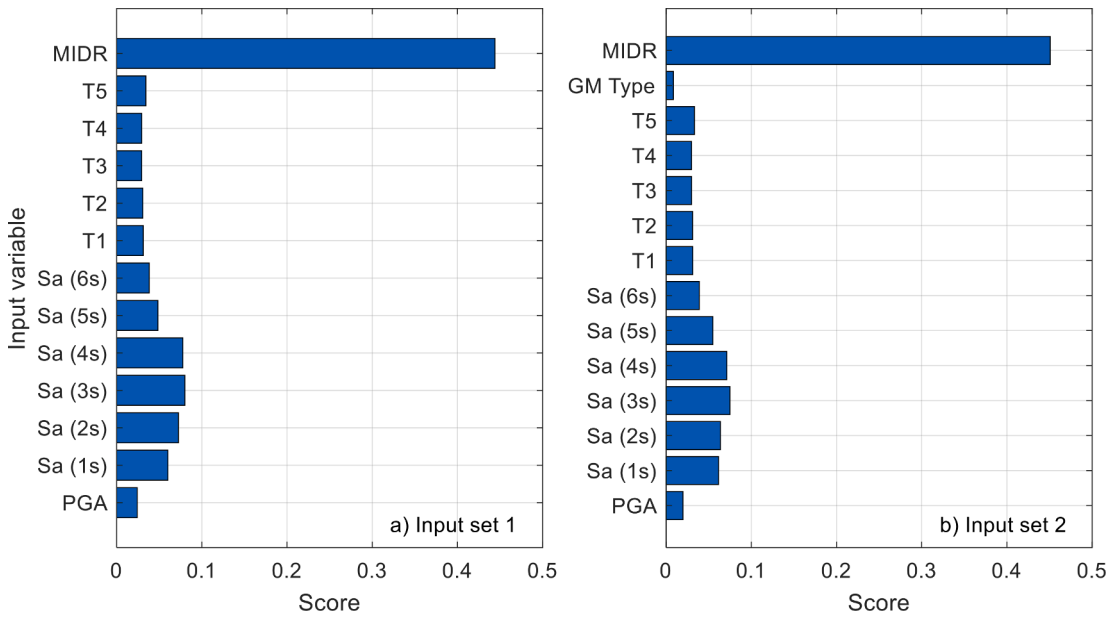


Fig. 11. Relative importance of input variables from RF model: a) Input set 1 and b) Input set 2.

ployed to predict the states of R/C wall structures. The implementation of these two developed RF models associated with two input sets will be discussed in the subsequent section.

6. Incorporating proposed ML models into MSA approach

This section introduces a novel procedure that utilizes the outcomes of the proposed ML model (see Section 5) into the MSA approach. Fig. 12 displays a visual illustration of the proposed procedure and demonstrates the way how we integrate the ML model results into the MSA approach to derive the fragility curves of R/C wall structures in building frame systems. The process begins with preparing an input dataset containing either 13 input variables or 14 variables, as presented in Section 3.3. During this step, the ground motions are scaled with PGA values ranging from 0.1 g to 1.0 g, and the spectral accelerations at different periods are calculated based on the scaled ground motions. Subsequently, this prepared input dataset is input into the predictive model (i.e., RF model) to produce the corresponding output variables. Note that separate RF models are suggested for the first and second input sets. The outcomes predicted by the RF model are categorized for each PGA level, as explained in Step 3. Finally, the failure probability is calculated at each PGA level and the fitted fragility curve is then obtained (for details on the MSA approach, see Section 2.1).

To illustrate the proposed procedure, we randomly selected a 12-story R/C wall structure from a pool of 46 configurations. Fragility curves were generated from two RF models with two input sets. For the RF model associated with the second input set, we specifically considered the FF case (labeled as "1″ in the input set) for demonstration purposes. In this specific case, an MIDR threshold of 1 % was assumed to develop the fragility curves. Fig. 13a and b illustrate the relationship between PGAs and spectral accelerations at 1 s for the first and second input sets, respectively. The figures also display the outputs from the respective RF models, categorized as “B” and “E”. Finally, fragility curves for a 12-story R/C wall structure are generated using both the proposed procedure and the original MSA approach, which incorporates results from nonlinear time history analyses using OpenSees (i.e., denoted as “OS” in Fig. 14). Figs. 14a and b present the fragility curves for the first and second input sets, respectively.

The example shown in Fig. 14 clearly demonstrates that the fragility curves obtained through the proposed procedure closely align with those derived from the original MSA approach, indicating a good agreement between the two methods. To further provide a comprehensive comparison of the results from the RF model, we assessed the performance of proposed procedure against the original MSA approach through PGA values associated with different levels of probability of failure, including 5 %, 50 %, and 95 %. This evaluation was conducted for the two input sets. For the first input set, this involved a total of 4416 data points (46 structures × 96 threshold values), while for the second input set, there were 13,248 data points (46 structures × 96 threshold values × 3 ground motion types). It is noted that the structures that did not reach the failure as defined by the specified thresholds of MIDR were excluded from this assessment.

Figs. 15 and 16 illustrate the relationship between PGAs derived from the proposed procedure and those obtained using the original MSA approach, for failure probabilities of 5 %, 50 %, and 95 % applied to both the first and second input sets. The figures also display the coefficients of determination (R^2), showcasing that the proposed procedure consistently yields outstanding results, as evi-

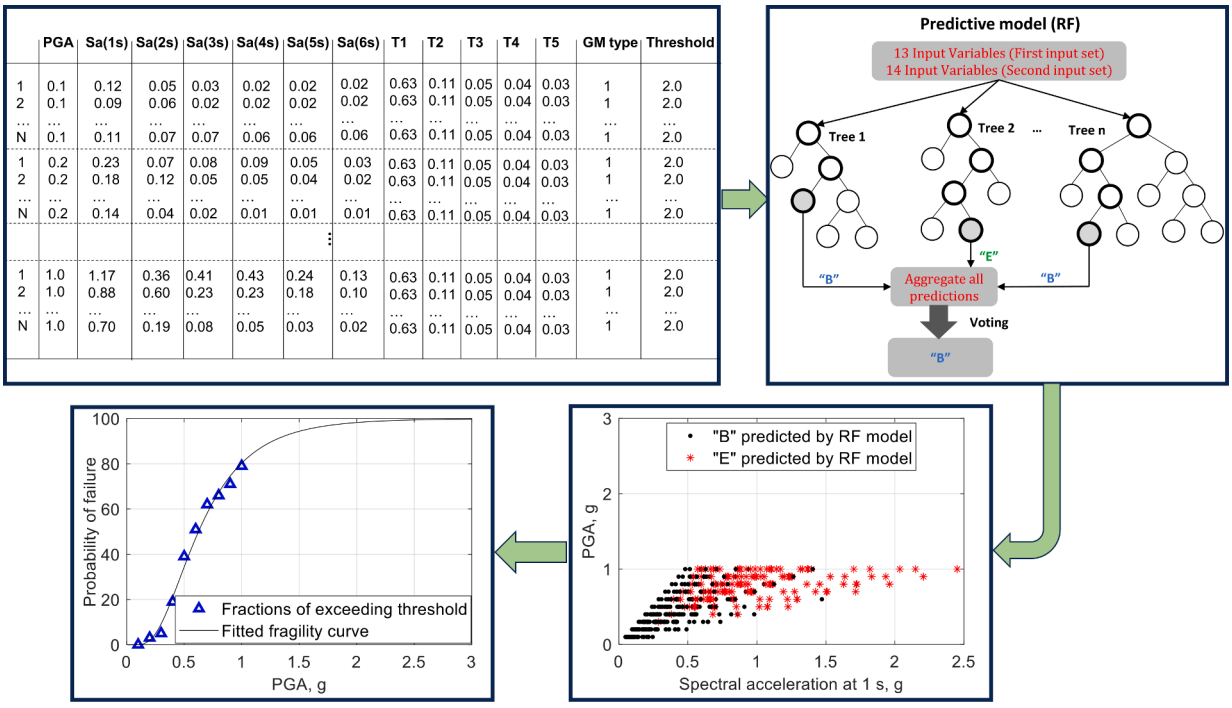


Fig. 12. – Procedure of incorporating the suggested ML model to the MSA approach.

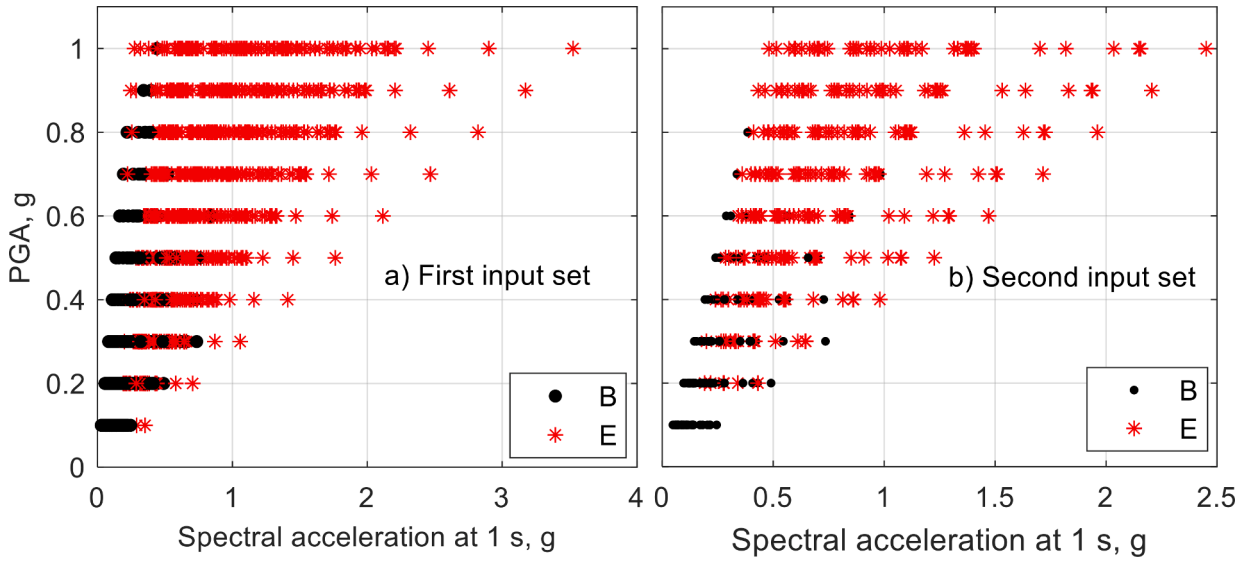


Fig. 13. – Relationship of PGA and spectral acceleration at 1 s: a) first and b) second input sets.

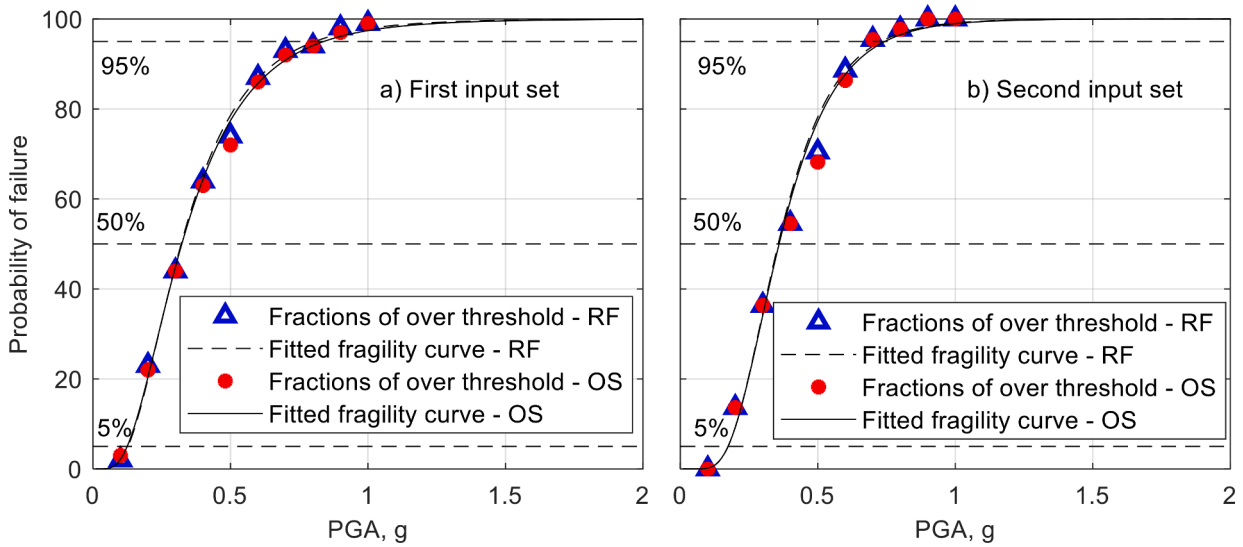


Fig. 14. – Fragility curves in the cases of a) first and b) second input sets.

denced by the very high R^2 values, which indicate a strong positive correlation between the outcomes of the two methods. It is important to emphasize that the comparisons in this section (Figs. 15 and 16) are intended to assess the performance of RF models integrated into the proposed procedure within the dataset used to develop the RF models. The performance of the proposed procedure when applied to a new structure, which was not used to generate the dataset, will be explored in the following section.

7. Application of proposed procedure for fragility analysis of an arbitrary R/C wall structure

In this study, we demonstrate the efficacy of the proposed procedure using a 24-story R/C wall structure, which was detailed in the prior work of Mafari et al. [41]. It is important to note that this specific R/C wall structure was not included in the dataset used to develop RF models. This building was designed and detailed to withstand all seismic loads in compliance with ACI 318-14 [43] and met the minimum requirements of the Equivalent Lateral Force Procedure outlined in ASCE 7–10 [60].

As an initial step, determining the first five natural periods of the R/C wall structures is required, as detailed in Table 2. Subsequently, PGA and spectral accelerations at intervals of 1 s–6 s are computed from the considered ground motions. For this application, 1000 ground motions introduced in Section 3.2 were utilized. To obtain a solid comparison, the fragility curves were developed with three thresholds of MIDR. These thresholds, which are summarized in Table 3, were based on the recommendations outlined in FEMA P-58 [61]. In this investigation, the MIDR thresholds of 0.5 %, 1.5 %, and 2.7 % were employed for negligible damage, moderate damage, and collapse cases, respectively.

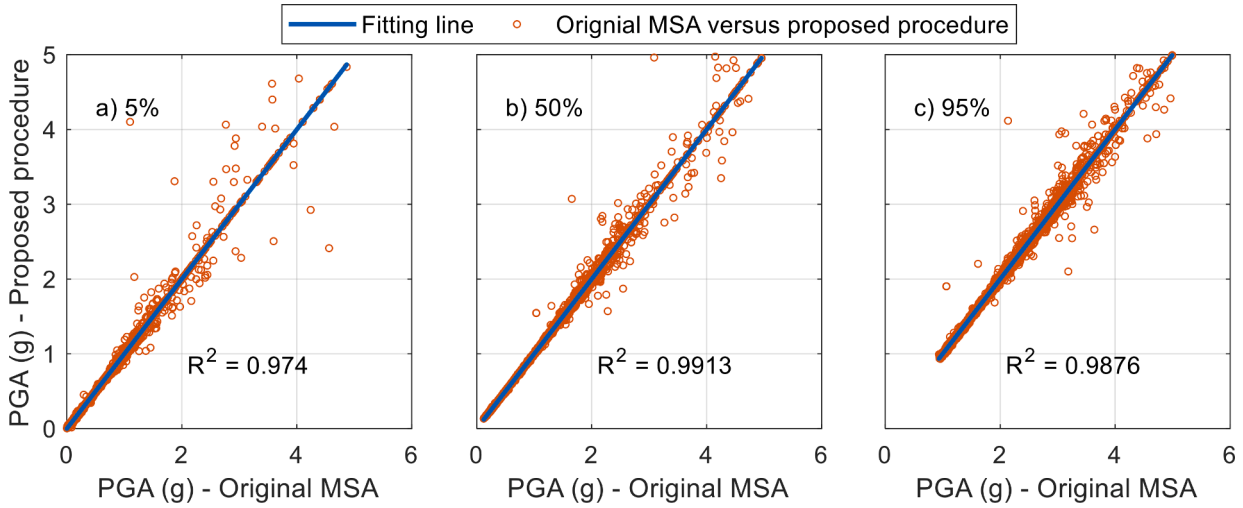


Fig. 15. – PGA values at a) 5 %, b) 50 %, and c) 95 % failure probabilities of MSA approach versus those of the proposed procedure in the first input set.

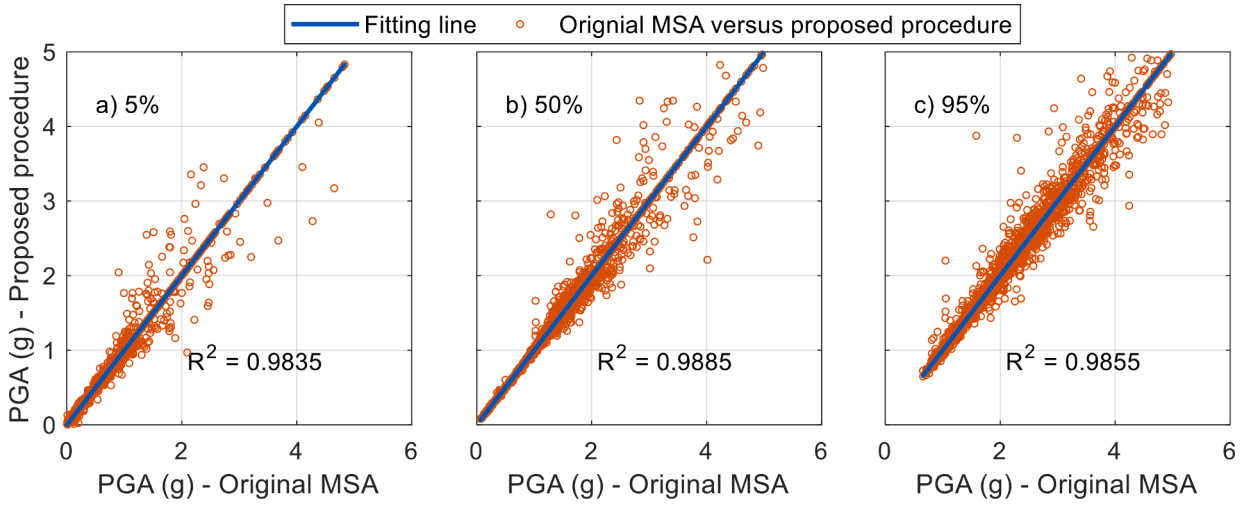


Fig. 16. – PGA values at a) 5 %, b) 50 %, and c) 95 % failure probabilities of MSA approach versus those of the proposed procedure in the second input set.

Table 2
– First five periods of 24-story R/C wall structure.

Input variable no.	Description	Value
1	First natural period (s)	1.156
2	Second natural period (s)	0.211
3	Third natural period (s)	0.090
4	Fourth natural period (s)	0.064
5	Fifth natural period (s)	0.055

Table 3
– Definition of damage state of R/C wall structures.

MIDR	Damage state (DS)
MIDR <1 %	DS1: Negligible damage
1 % $\leq$ MIDR $\leq$2.6 %	DS2: Minor to moderate damage
MIDR >2.6 %	DS3: Severe damage to collapse

Table 4
– Comparisons of proposed procedure and MSA approach for PGA values at 50 % probability of failure.

Input case	Damage state	PGA value at 50 % probability of failure (g)	
		Proposed procedure	MSA approach
First input set	DS1	0.178	0.170
	DS2	0.450	0.508
	DS3	0.864	0.997
Second input set - FF	DS1	0.196	0.190
	DS2	0.508	0.591
	DS3	1.113	1.289
Second input set – NF-Pulse	DS1	0.128	0.121
	DS2	0.300	0.330
	DS3	0.541	0.598
Second input set – NF-No Pulse	DS1	0.213	0.200
	DS2	0.583	0.603
	DS3	1.222	1.323

Fig. 17 presents the fragility curves for three damage states generated by both the proposed procedure and the original MSA approach, applied to the first and second input sets. The results indicate that the proposed procedure generally presented the favorable outcomes of the fragility curves across both input sets. It is noteworthy that the fragility curves can be rapidly generated using the proposed procedure. In contrast, the execution of 1000 analyses required for the original MSA approach took 73 h to run using 24 parallel cores on a Dell 13th Gen Intel® Core™ i9-13900KF.

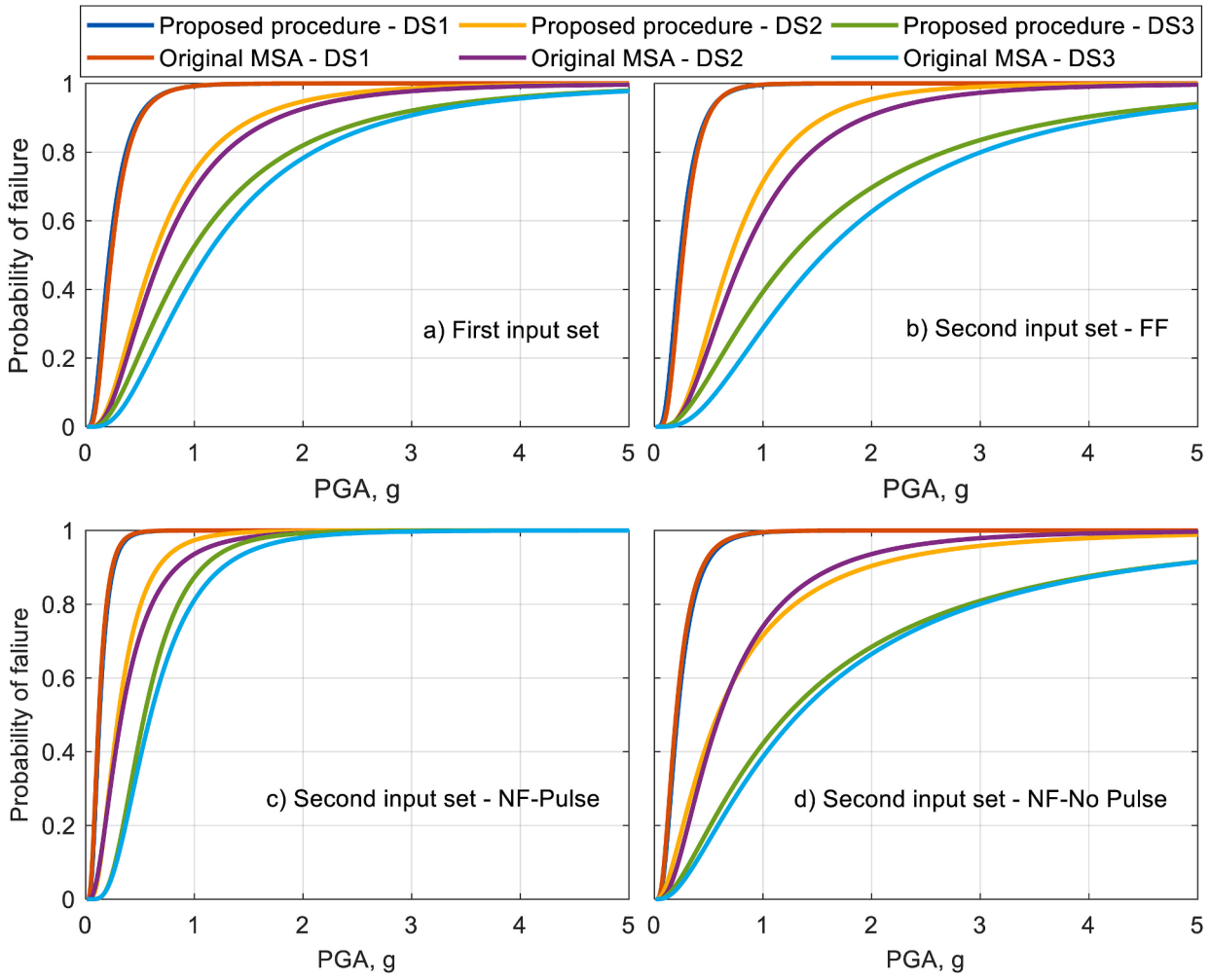


Fig. 17. – Fragility curves of 24-story R/C wall structures.

To obtain a solid comparison between the proposed procedure and the original MSA approach, the PGA values corresponding to a 50 % probability of failure were compared across all cases, as presented in Table 4. It can be seen that the discrepancies in PGA values at 50 % probability of failure between these two methods in all investigated cases are relatively small (an average of approximately 9.32 %). However, when considering higher threshold values of MIDR, the divergence in fragility curves becomes more pronounced. This trend is likely due to the complex behaviors of R/C walls in the nonlinear range and the limited data available for high MIDR values (with an average of 2.16 %) in the dataset, which can result in less accurate predictions from the developed RF model. In addition, the fragility curves generated by the developed model with and without considering ground motion types exhibit notable differences. These differences are apparent in both the proposed procedure and the original MSA approach. Therefore, it is important to incorporate a diverse set of ground motions encompassing various types when developing the fragility function. This approach not only enhances the model's robustness but also underscores the applicability of the proposed procedure with the two input sets considered in this study.

To facilitate accessibility and implementation of the proposed procedure to generate the fragility curve for R/C wall structures, a code package, along with accompanying guidelines, is provided via the following link (IncorporatingMLToMSAAppoach.git).

8. Conclusions

In this study, we proposed a novel procedure that integrates ML into the MSA approach for constructing seismic fragility curves for R/C wall structures in building frame systems. This procedure could mitigate computational challenges associated with the original MSA approach. The ML models were developed to classify the damage state of R/C wall structures subjected to ground motions. The binary categorization was then used to calculate failure probability points needed to derive the fragility curve per the MSA approach. Four well-established ML methods, including RF, XGBoost, LightGBM, and CatBoost, were considered. Two sets of input variables were examined to identify the significance of ground motion type (FF, NF-Pulse, or NF-No Pulse) in the development of the ML models as well as the fragility analysis. The performance evaluation of the four developed ML models and the proposed procedure incorporating the suggested ML models for fragility analysis yielded the following conclusions.

- All four ML models showcased remarkable predictive capabilities. Among them, the RF model presented the greatest performance, achieving an accuracy metric of 0.99 and an average F1-score of 0.98 on the testing dataset for both input sets. Thus, the RF model was selected for integration into the proposed procedure.
- In addition, the RF models corresponding to the two input sets showed consistent results across all accuracy metrics. This implies that either of the two developed RF models can be used for the proposed procedure, regardless of whether the ground motion type is considered or not. This also indicates that the ground motion type is not significant in developing the ML models in this case.
- To assess the effectiveness of the proposed procedure incorporating the developed RF models, we compared with the results of the original MSA approach using PGA values corresponding to various levels of the probability of failure: 5 %, 50 %, and 95 %. The results consistently demonstrate that the proposed procedure delivers outstanding performance.
- As an external validation, the seismic fragility analysis of a new 24-story R/C wall structure was conducted using the proposed procedure. The results clearly demonstrate that the proposed procedure incorporating the developed RF models allowed for the rapid construction of fragility curves and consistently produced favorable outcomes. On the other hand, the discrepancy between the fragility analysis results from the proposed procedure and the original MSA approach became pronounced when employing a higher threshold value of MIDR.
- A code package along with comprehensive guidelines of the proposed procedure is provided to make it easier for users to access for generating the seismic fragility curves of R/C wall structures.

It is notable that the proposed procedure is limited to the fragility analysis with just one pair of EDP (i.e., MIDR) and IM (i.e., PGA). A similar procedure may be used for other pairs of EDP (e.g., maximum top drift or peak floor acceleration) and IM (e.g., spectral acceleration at first natural period). In addition, covering a wider range of ground motion intensities and generating more wall configurations from 24 to 40 stories could enrich the dataset with higher MIDR values, potentially enhancing the performance of the ML model. Furthermore, the analysis of R/C wall structures in this study assumed a complete coupled action between wall segments in perpendicular directions, although R/C core walls may exhibit a considerable shear lag under lateral forces. The numerical models also should be validated at higher MIDR values and account for the cyclic strength deterioration of reinforcing steel. The authors intend to explore these limitations in subsequent research.

CRediT authorship contribution statement

Hoang D. Nguyen: Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Chanyoung Kim:** Writing – original draft, Methodology, Formal analysis. **Young-Joo Lee:** Writing – review & editing, Validation, Supervision. **Myoungsu Shin:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

I am the corresponding author of the manuscript entitled "**Incorporation of Machine Learning Into Multiple Stripe Seismic Fragility Analysis of Reinforced Concrete Wall Structures**." I am writing this letter to claim this manuscript has not been published or presented elsewhere in part or in entirety and is not under consideration by another journal. We have read and understood

your journal's policies, and we believe that neither the manuscript nor the study violates any of these. There are no conflicts of interest to declare.

Acknowledgment

This research was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korean government (MOTIE) (Grant No. 20224B10300010), and the BK21plus Program (Grant No. 4299990613923) funded by the Ministry of Education and National Research Foundation of Korea.

Data availability

Data will be made available on request.

References

[1] J.W. van de Lindt, M. Taggart, Fragility analysis methodology for performance-based analysis of wood-frame buildings for flood, Nat. Hazards Rev. (2009), https://doi.org/10.1061/(asce)1527-6988(2009)10:3(113).
[2] D. Lallemant, A. Kiremidjian, H. Burton, Statistical procedures for developing earthquake damage fragility curves, Earthq Eng Struct Dyn (2015), https://doi.org/10.1002/eqe.2522.
[3] O.M. Nofal, K. Amini, J.E. Padgett, J.W. van de Lindt, N. Rosenheim, Y.M. Darestani, et al., Multi-hazard socio-physical resilience assessment of hurricane-induced hazards on coastal communities, Resilient Cities Struct (2023), https://doi.org/10.1016/j.rcns.2023.07.003.
[4] C.B. Haselton, G.G. Deierlein, Assessing Seismic Collapse Safety of Modern Reinforced Concrete Moment-Frame Buildings, PEER Report 2007-08, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, 2007.
[5] A. Liel, Assessing the collapse risk of California's existing reinforced concrete frame structures: metrics for seismic safety decisions, PhD thesis. Stanford University, 2008.
[6] L.F. Ibarra, H. Krawinkler, Global collapse of frame structures under seismic excitations. Rep. No. TB 152, the John A. BlumeEarthquake Engineering Center, Stanford Univ., Stanford, CA, 2005.
[7] H.D. Nguyen, M. Shin, J.M. LaFave, Optimal intensity measures for probabilistic seismic demand models of steel moment frames, J. Build. Eng. 65 (2023) 105629, https://doi.org/10.1016/j.jobe.2022.105629.
[8] K. Kostinakis, A. Athanatopoulou, Incremental dynamic analysis applied to assessment of structure-specific earthquake IMs in 3D R/C buildings, Eng. Struct. (2016), https://doi.org/10.1016/j.engstruct.2016.07.007.
[9] J.E. Padgett, B.G. Nielson, R. DesRoches, Selection of optimal intensity measures in probabilistic seismic demand models of highway bridge portfolios, Earthq Eng Struct Dyn 37 (2008) 711–725, https://doi.org/10.1002/eqe.782.
[10] J.W. Baker, Efficient analytical fragility function fitting using dynamic structural analysis, Earthq. Spectra 31 (2015) 579–599, https://doi.org/10.1193/021113EQS025M.
[11] D. Vamvatsikos, C. Allin Cornell, Incremental dynamic analysis, Earthq Eng Struct Dyn 31 (2002) 491–514, https://doi.org/10.1002/eqe.141.
[12] D. Vamvatsikos, M. Fragiadakis, Incremental dynamic analysis for estimating seismic performance sensitivity and uncertainty, Earthq Eng Struct Dyn 39 (2010) 141–163, https://doi.org/10.1002/eqe.935.
[13] B. Ghanbari, M. Fathi, A.H. Akhaveissy, Fragility curves for reinforced concrete (RC)/steel vertical hybrid frame structure under mainshock–aftershock sequences, J Struct Integr Maint (2023), https://doi.org/10.1080/24705314.2023.2220946.
[14] F. Jalayer, Direct Probabilistic Seismic Analysis: Implementing Non-linear Dynamic Assessments, Stanford University, California, 2003.
[15] B.A. Bradley, A generalized conditional intensity measure approach and holistic ground-motion selection, Earthq Eng Struct Dyn (2010), https://doi.org/10.1002/eqe.995.
[16] H.D. Nguyen, N.D. Dao, M. Shin, Machine learning-based prediction for maximum displacement of seismic isolation systems, J. Build. Eng. 51 (2022) 104251, https://doi.org/10.1016/j.jobe.2022.104251.
[17] S. Mangalathu, J.S. Jeon, Stripe-based fragility analysis of multispan concrete bridge classes using machine learning techniques, Earthq Eng Struct Dyn 48 (2019) 1238–1255, https://doi.org/10.1002/eqe.3183.
[18] H.D. Nguyen, J. Kim, M. Shin, Development of ensemble machine learning models for evaluating seismic demands of steel moment frames, Steel Compos. Struct. 44 (2022) 49–63, https://doi.org/10.12989/scs.2022.44.1.049.
[19] X. Guan, H. Burton, M. Shokrabadi, Z. Yi, Seismic drift demand estimation for steel moment frame buildings: from mechanics-based to data-driven models, J. Struct. Eng. 147 (2021) 04021058, https://doi.org/10.1061/(asce)st.1943-541x.0003004.
[20] N.V. Nguyen, H.D. Nguyen, N.D. Dao, Machine learning models for predicting maximum displacement of triple pendulum isolation systems, Structures 36 (2022) 404–415, https://doi.org/10.1016/j.istruc.2021.12.024.
[21] V.-Q. Nguyen, H.D. Nguyen, F. Petrone, D. Park, Rapid damage state classification for underground box tunnels using machine learning, Struct Infrastruct Eng (2023) 1–14, https://doi.org/10.1080/15732479.2023.2266709.
[22] H.D. Nguyen, N.D. Dao, M. Shin, Prediction of seismic drift responses of planar steel moment frames using artificial neural network and extreme gradient boosting, Eng. Struct. 242 (2021) 112518, https://doi.org/10.1016/j.engstruct.2021.112518.
[23] Y. Zhang, H.V. Burton, H. Sun, M. Shokrabadi, A machine learning framework for assessing post-earthquake structural safety, Struct. Saf. 72 (2018) 1–16, https://doi.org/10.1016/j.strusafe.2017.12.001.
[24] H.D. Nguyen, J.M. LaFave, Y.-J. Lee, M. Shin, Rapid seismic damage-state assessment of steel moment frames using machine learning, Eng. Struct. 252 (2022) 113737, https://doi.org/10.1016/j.engstruct.2021.113737.
[25] S.H. Hwang, S. Mangalathu, J. Shin, J.S. Jeon, Machine learning-based approaches for seismic demand and collapse of ductile reinforced concrete building frames, J. Build. Eng. 34 (2021), https://doi.org/10.1016/j.jobe.2020.101905.
[26] H.D. Nguyen, Y.-J. Lee, J.M. LaFave, M. Shin, Seismic fragility analysis of steel moment frames using machine learning models, Eng. Appl. Artif. Intell. 126 (2023) 106976, https://doi.org/10.1016/j.engappai.2023.106976.
[27] Y. Xie, M. Ebad Sichani, J.E. Padgett, R. DesRoches, The promise of implementing machine learning in earthquake engineering: a state-of-the-art review, Earthq. Spectra 36 (2020) 1769–1801, https://doi.org/10.1177/8755293020919419.
[28] H. Sun, H.V. Burton, H. Huang, Machine learning applications for building structural design and performance assessment: state-of-the-art review, J. Build. Eng. 33 (2021), https://doi.org/10.1016/j.jobe.2020.101816.
[29] J. Kiani, C. Camp, S. Pezeshk, On the application of machine learning techniques to derive seismic fragility curves, Comput. Struct. 218 (2019) 108–122, https://doi.org/10.1016/j.compstruc.2019.03.004.
[30] H.D. Nguyen, C. Kim, K. Lee, M. Shin, Development of data-driven models to predict seismic drift response of RC wall structures: An application of deep neural networks, Soil Dyn. Earthq. Eng. 186 (2024) 108952, https://doi.org/10.1016/j.soildyn.2024.108952.
[31] D.W. Jia, Z.Y. Wu, Seismic fragility analysis of RC frame-shear wall structure under multidimensional performance limit state based on ensemble neural network, Eng. Struct. (2021), https://doi.org/10.1016/j.engstruct.2021.112975.
[32] H.D. Nguyen, M. Shin, M. Torbol, Reliability assessment of a planar steel frame subjected to earthquakes in case of an implicit limit-state function, J. Build. Eng. 32 (2020), https://doi.org/10.1016/j.jobe.2020.101782.

[33] M. Serdar Kirçil, Z. Polat, Fragility analysis of mid-rise R/C frame buildings, Eng. Struct. (2006), https://doi.org/10.1016/j.engstruct.2006.01.004.
[34] D.-S. Moon, Y.-J. Lee, S. Lee, Fragility analysis of space reinforced concrete frame structures with structural irregularity in plan, J. Struct. Eng. 144 (2018) 04018096, https://doi.org/10.1061/(asce)st.1943-541x.0002092.
[35] M. Shinozuka, M.Q. Feng, J. Lee, T. Naganuma, Statistical analysis of fragility curves, J. Eng. Mech. (2000), https://doi.org/10.1061/(asce)0733-9399(2000)126:12(1224).
[36] J. Baker, C.A. Cornell, Vector-Valued Ground Motion Intensity Measures for Probabilistic Seismic Demand Analysis. PEER Report 2006-08, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, 2006.
[37] M.W. Khan, M. Usman, S.H. Farooq, M. Zain, S. Saleem, Effect of masonry infill on analytical fragility response of RC frame school buildings in high seismic zone, J Struct Integr Maint (2021), https://doi.org/10.1080/24705314.2020.1865624.
[38] V.L. Tran, D.K. Thai, D.D. Nguyen, Practical artificial neural network tool for predicting the axial compression capacity of circular concrete-filled steel tube columns with ultra-high-strength concrete, Thin-Walled Struct. 151 (2020), https://doi.org/10.1016/j.tws.2020.106720.
[39] V.H. Ho, C.T. Nguyen, H.D. Nguyen, H.S. Oh, M. Shin, S.Y. Kim, Hydrogenated graphene with tunable Poisson's ratio using machine learning: implication for wearable devices and strain sensors, ACS Appl. Nano Mater. (2022), https://doi.org/10.1021/acsanm.2c01950.
[40] N.A. Marafi, A.J. Makdisi, J.W. Berman, M.O. Eberhard, Design strategies to achieve target collapse risks for reinforced concrete wall buildings in sedimentary basins, Earthq. Spectra (2020), https://doi.org/10.1177/8755293019899965.
[41] N.A. Marafi, A.J. Makdisi, M.O. Eberhard, J.W. Berman, Impacts of an M9 cascadia subduction zone earthquake and Seattle basin on performance of RC core wall buildings, J. Struct. Eng. (2020), https://doi.org/10.1061/(asce)st.1943-541x.0002490.
[42] ASCE, Minimum design loads for buildings and other structuresASCE/SEI 7-16, ACI, Farmington Hills, MI, 2016.
[43] ACI Committee 318, in: Building Code Requirements for Structural Concrete, 2014.
[44] S. Mazzoni, F. McKenna, M.H. Scott, G.L. Fenves, Open system for earthquake engineering simulation (OpenSees), Pacific Earthq Eng Res Cent 465 (2006).
[45] J.S. Pugh, L.N. Lowes, D.E. Lehman, Nonlinear line-element modeling of flexural reinforced concrete walls, Eng. Struct. (2015), https://doi.org/10.1016/j.engstruct.2015.08.037.
[46] A.A. Correia, J.P. Almeida, R. Pinho, Force-based versus displacement-based formulations in the cyclic nonlinear analysis of RC frames. 14th World Conf. Earthq. Eng., 2008.
[47] E. Spacone, F.C. Filippou, F.F. Taucer, Fibre beam-column model for non-linear analysis of R/C frames: Part I. Formulation, Earthq Eng Struct Dyn (1996) 2–9, https://doi.org/10.1002/(SICI)1096-9845(199607)25:7<711::AID-EQE576>3.0.CO.
[48] R. Hamburger, J. Moehle, J. Baker, J. Bray, C.B. Crouse, G. Deierlein, et al., Guidelines for Performance-Based Seismic Design of Tall Buildings, University of California, Berkeley, CA, 2017.
[49] Alex Shegay, Christopher Motter, Rick Henry, K. Elwood, University of auckland high axial load wall tests, https://doi.org/10.17603/DS2609N, 2018.
[50] G.G. Deierlein, A.B. Liel, C.B. Haselton, C.A. Kircher, ATC 63 methodology for evaluating seismic collapse safety of archetype buildings, in: Struct. Congr. 2008, Reston, VA, American Society of Civil Engineers, 2008, pp. 1–10, https://doi.org/10.1061/41016(314)48.
[51] ATC, Quantification of building seismic performance factors, Fema P695 (2009) 421.
[52] European Committee for Standardization, Eurocode 8: Design of structures for earthquake resistance - Part 1: General rules, seismic actions and rules for buildings (EN 1998-1: 2004), European Committee for Standardization, Brussels, 2004.
[53] American Society of Civil Engineers, Minimum design loads and associated criteria for buildings and other structures. Reston, VA, American Society of Civil Engineers (2017), https://doi.org/10.1061/9780784414248.
[54] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[55] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 13–17– Augu (2016) 785–794, https://doi.org/10.1145/2939672.2939785.
[56] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, et al., LightGBM: a highly efficient gradient boosting decision tree, Adv. Neural Inf. Process. Syst. 30 (2017) 3149–3157.
[57] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, Catboost: unbiased boosting with categorical features, Adv. Neural Inf. Process. Syst. 31 (2018) 6639–6649.
[58] S. Mangalathu, H. Sun, C.C. Nweke, Z. Yi, H.V. Burton, Classifying earthquake damage to buildings using machine learning, Earthq. Spectra 36 (2020) 183–208, https://doi.org/10.1177/8755293019878137.
[59] P. Kourehpaz, C. Molina Hutt, Machine learning for enhanced regional seismic risk assessments, J. Struct. Eng. (2022), https://doi.org/10.1061/(asce)st.1943-541x.0003421.
[60] ASCE, Minimum Design Loads for Buildings and Other Structure, ASCE/SEI 7–10, Am Soc Civ Eng Reston, VA, 2010.
[61] R.O. Hamburger, FEMA P58 seismic performance assessment of buildings, NCEE 2014 - 10th U.S. Natl. Conf. Earthq. Eng. Front. Earthq. Eng. (2014), https://doi.org/10.4231/D3ZW18S8N.