Postprint of: Kazemi F., Asgarkhani N., Jankowski R., Machine learning-based seismic fragility and seismic vulnerability assessment of reinforced concrete structures, SOIL DYNAMICS AND EARTHQUAKE ENGINEERING Vol. 166 (2023), 107761, DOI: 10.1016/j.soildyn.2023.107761

Machine learning-based seismic fragility and seismic vulnerability assessment of reinforced concrete structures

F. Kazemi[a,*], N. Asgarkhani[a], R. Jankowski[a]

[a] *Faculty of Civil and Environmental Engineering, Gdańsk University of Technology, ul. Narutowicza 11/12, 80-233 Gdansk; Poland.*

Abstract

Many studies have been performed to put quantifying uncertainties into the seismic risk assessment of reinforced concrete (RC) buildings. This paper provides a risk-assessment support tool for purpose of retrofitting and potential design strategies of RC buildings. Machine Learning (ML) algorithms were developed in Python software by innovative methods of hyperparameter optimization, such as halving search, grid search, random search, fine-tuning method, and the *k*-fold cross-validation, to derive the seismic fragility curve for accelerating seismic risk assessment. Proposed ML methods significantly reduced the computational effort compared to conventional procedure of seismic fragility assessment. The prediction results can be combined with considered hazard curves for the purpose of seismic risk assessment of RC buildings. To prepare the training dataset, Incremental Dynamic Analyses (IDAs) were performed on 165 RC frames to achieve 1121184 data points. Performance indicators showed that the algorithms of Artificial Neural Networks (ANNs), Extra-Trees Regressor (ETR), Extremely Randomized Tree Regressor (ERTR), Bagging Regressor (BR), Extreme Gradient Boosting (XGBoost), and Histogram-based Gradient Boosting Regression (HGBR) had higher performance, which achieved acceptable accuracy and fitted to actual curves. In addition, Graphical User Interface (GUI) was introduced as a practical tool yet reliable for seismic risk assessment of RC buildings.

Keywords: Seismic risk assessment; Machine learning method; Seismic fragility curve prediction; Seismic limit-state capacity; Seismic vulnerability assessment; Reinforced concrete.

* Corresponding author: Farzin Kazemi, Faculty of Civil and Environmental Engineering, Gdańsk University of Technology, ul. Narutowicza 11/12, 80-233 Gdansk, Poland.
E-mail address: farzin.kazemi@pg.edu.pl (F. Kazemi).

1. Introduction

Previous earthquakes proved that lateral load predictions of such capacity-based seismic codes could not fulfill seismic design expectations, and there is a need to shift the performance-based design methods. Therefore, there is possible to target a specific performance level and its probable risks and financial damages. Economic and social losses after seismic excitations are considered as an important factor that many researchers attempted to provide a general procedure for evaluating. The FEMA P-58 [1] developed by Applied Technology Council (ATC) is an effort toward the performance-based seismic guideline. This criterion developed the seismic collapse fragility and hazard curves, and quantitative performance assessment of structures. Seismic vulnerability and risk assessment of buildings have become more important research areas due to their importance for restoring the functionality of buildings after seismic events. Regarding this fact, the probabilistic approaches have been used for determining fragility functions that are useful for seismic vulnerability assessment of buildings [2]. Feng et al. [3, 4] proposed a PDEM-based method to determine the fragility function without pre-defined distribution of the demands assuming different failure modes and limit states. Cao et al. [5] compared four probabilistic fragility analyses and concluded that the least squares regression method is more efficient than other methods for its trends and accuracy assuming different limit states. Multiple‐stripe analysis, which performs analysis with similar Intensity Measures (IMs) [6, 7], and Incremental Dynamic Analysis (IDA), which performs analysis until the seismic collapse of structures assuming different IMs [8-10], are two popular methods to account record-to-record variability for evaluating seismic fragility curves. Seismic performance level assessment of buildings is also known as a preliminary vulnerability assessment of buildings that helps engineers to estimate the performance of constructed buildings for retrofitting purposes. Cao et al. [11] used fragility curves to investigate the retrofitting influence of a novel kind of buckling-restrained brace implemented on the RC frames. Moreover, Asgarkhani et al. [12] proposed optimal retrofitting strategy using viscous dampers based on the seismic performance level of steel and Reinforced Concrete (RC) structures. Kazemi et al. [13-15] performed a comprehensive investigation on the seismic limit-state and collapse capacity prediction of steel and Reinforced Concrete (RC) frames assuming collision as external loads on the floor levels. They proposed the modification factors that can be used to predict the influence of a collision on the seismic limit-state and collapse capacity. In addition, they investigated the effects of infill masonry wall on the seismic performance of steel structures [16].

The seismic fragility curve can be considered as failure probability of structure regarding seismic performance level corresponding to the IMs. This curve can be widely used for any type of structure to show its failure probability. Yazdanpanah et al. [17] proposed a new procedure based on the Morlet and complex Morlet wavelets to estimate the seismic fragility curve of pounding structures. Machine Learning (ML) methods have been employed for civil engineering problems, and researchers have suggested some prediction models to estimate the failure of beam-column joints [18], the failure mode of RC frames with infill walls [19], classification of reinforced masonry shear walls [20], the failure mode of RC shear walls [21], shear strength of RC beams [22], and derivation of seismic fragility curves [23]. Some studies have used ML methods to estimate the seismic fragility curves of structures using regression or classification techniques [24-27]. Mitropoulou and Papadrakakis [28] used Artificial Neural Networks (ANNs) for the fragility assessment of RC structures using 20 patterns for training the algorithms. Giovanis et al. [29] used IDAs with a Monte Carlo method to train ANNs for creating the IDA curves for

purpose of seismic fragility curve assessment. ML approaches were used for damage identification and fragility analysis had the advantage of reducing the computational cost while proposing results with higher accuracy of estimation. Most studies used predefined seismic limit states or only a limited number of structures and input parameters [see 30-32]. Hwang et al. [33] proposed a prediction model for estimating seismic response and seismic collapse of RC frames using both regression and classification methods implemented in boosting algorithms. Sainct et al. [34] proposed a seismic fragility curve estimator based on the Support Vector Machine (SVM) classifier with an active learning algorithm. Tang et al. [35] used 500 hypothetical steel frames to create a training model for seismic risk assessment based on the four ML algorithms of ANNs, SVM, Classification and regression tree, and Random Forest (RF).

This study is focused on the comprehensive investigation using numerical simulations to propose a novel process for seismic risk assessment of RC buildings based on ML algorithms. In this study, there is no limitation for selecting the intensity level of IM and seismic limit-state threshold, while to show the ability of the proposed method, the results of only four prescribed seismic limit states have been presented. Compared with existing ML models, the random selection of seismic threshold is adopted, and twenty improved ML algorithms with innovative hyperparameter optimization methods were considered to predict seismic risk by exploring different ML advantages. The proposed ML-based models improved the surrogate model of seismic risk assessment. Although a limited number of data points can reduce the reliability of ML-based models, this study uses numerous data points (more than 1121184 data points) achieved by performing IDAs on 165 RC buildings assuming different types of input features to provide a powerful tool for seismic risk estimation of RC buildings.

2. ML methods

2.1. Artificial neural network

ANNs can be trained for different types of engineering problems as well as unpredictable seismic responses. ANNs have three main parts of the input layer, which receives the parameters of data points, hidden layers, which are the part for functions and can be more than one layer, and the output layer, which is the outcome of the prediction. Fig. 1 illustrates the ANNs and its main parts. In the hidden layer, algorithms can have different layers and activation functions to achieve the higher result with lower bias [36].

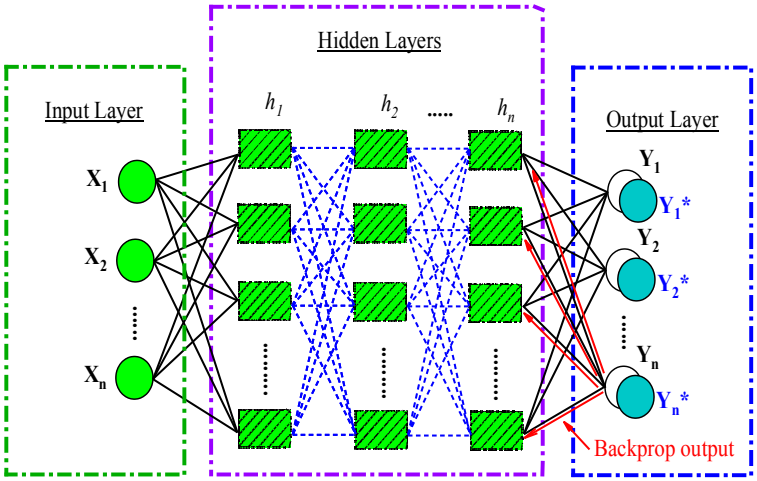


Fig. 1. Artificial Neural Networks.

To improve the algorithm, the hidden layers and type of functions play a crucial role, which is no direct formula to know about the number of hidden layers. In this study, three hidden layers were chosen based on trial and error, and the activation function of the Rectified Linear Unit (ReLU) was selected for the first and second layers, while the Sigmoid function was selected for the third layer [37]. In addition, the ANNs were developed by the back-propagation and feed-forward methods. The back-propagation ANNs receive the values of input (e.g. X_i) and target data (e.g. Y_i). For each neuron in the hidden layer, the function (e.g. ReLU) calculates the output, which is the input for the next layer (feed-forward), and this calculation affect the prediction error (e.g. Y_i-Y_i^*). Then, the back-propagation calculates the error (i.e. loss) and updates the weight of the neurons. This process will improve the prediction accuracy. To have a better comparison, a Multi-layer Perceptron Regressor (MLPR) was also considered with a linear activation function [38]. It can be noted that due to huge number of data points, forward and backward data point selection approach were developed to improve the speed of calculations, in which, the prediction model was fitted with small data points and the number of data points were increased or decreased (i.e. forward and backward) until no changing happened to model metrics.

2.2. Random forest

The RF algorithm has the ability of bagging ensemble to use multiple models with different training datasets selected from main data to achieve higher prediction accuracy. Fig. 2 illustrates the bagging principle applied in the RF algorithm. The ability of random selection of data can be used to solve overfitting issues in parallel processing of the structure-related uncertainty problems. Recently, different types of decision tree algorithms developed that can be effectively used in earthquake engineering. Extra-Trees Regressor (ETR) uses the random selection of data with different models of selection that improves the accuracy of prediction result. In addition, Extremely Randomized Tree Regressor (ERTR) has the ability of random selection while the speed of calculations was improved compared to ETR [39]. In addition, Bagging Regressor (BR) can be applied to decision tree methods to avoid overfitting problems [40]. BR can work with strong models to reduce the variance by aggregating the individual estimations and forming them as a final estimation.

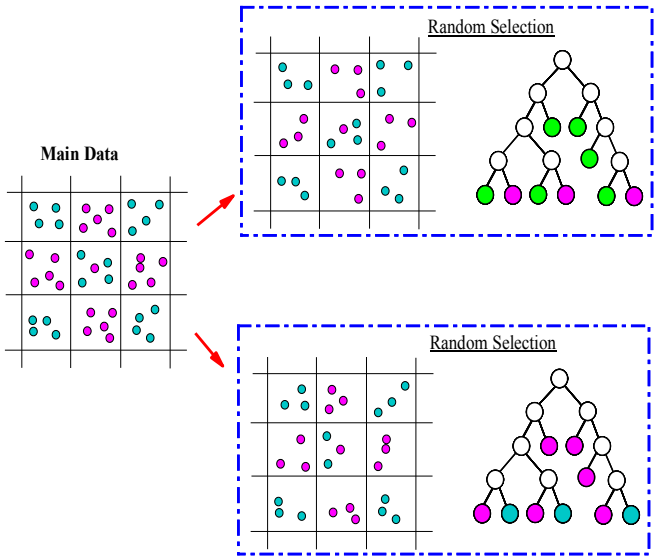


Fig. 2. Bagging principle applied in RF method.

2.3. Boosting algorithms

Boosting algorithms, mostly known as Gradient Boosting Machine (GBM), employ a methodology to improve the performance of algorithms using weak learners in sequential order to minimize the loss function. Assuming the first probability equal to 0.5, Equation (1) can be used for GBM to compute the value of V that can be used for comparing the results of estimation until finding the optimal model:

$$V = \frac{\sum_{i=1}^{n} (\text{Observed}_i - \text{Predicted}_i)}{\sum_{i=1}^{n} (\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i))} \quad (1)$$

The improved type of GBM, which has a regularization parameter, λ, to control the small leaves in the algorithm, named as Extreme Gradient Boosting (XGBoost), which assumes the following equation to compute the value of V [36, 38]:

$$V = \frac{\sum_{i=1}^{n} (\text{Observed}_i - \text{Predicted}_i)}{\sum_{i=1}^{n} (\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)) + \lambda} \quad (2)$$

In this study, the GBM and XGBoost were improved with a fine-tuning method of hyperparameter selection with the ability to change the parameters of trees to achieve higher values of prediction. In addition, the Histogram-based Gradient Boosting Regression (HGBR) that uses the quantization method to split the data features was improved with a fine-tuning method (i.e. HGBR(FN)), which had a higher computing time compared to GBM [41]. Moreover, Adaptive Boosting (AdaBoost), which uses the methodology of combining strong learners (e.g.

RF) to achieve data weights for estimation of the target, was implemented in Python software [42].

2.4. Support vector machine

The Support Vector Machine (SVM) method has the ability to use boundary conditions based on the distances from an assumed hyperplane in both 2D and 3D spaces. To control the vectors, the parameter of v can be assumed, and the algorithm is named the Nu-Support Vector Regression (NuSVR). Moreover, in this study, the Linear Support Vector Regression (LSVR) was implemented by assuming the loss and penalties function [43].

2.5. Other prediction algorithms

To investigate all possible ML algorithms for investigating the best prediction model, important regressor models are used, such as Voting Regressor (VR), which uses the average of predictions made by base regressors, Linear Regression (LR), which uses a linear function to explore the relation of input and output data, and Gamma Regressor (GR), which uses an inverse function to combine data points. Stacking Regressor (SR) is a method to use a regressor for computing the output of individual estimations (see [44]). K-Nearest neighbors (KNR) is a simple algorithm that finds the numbers of nearest neighbors and determines their distances. Partial Least Squares Regression (PLSR) is a methodology to reduce the variables to have a small set of predictors for performing a regression between inputs and outputs [45].

2.6. Data resampling methods

In this research, all proposed algorithms were improved with resampling procedure, such as the k-fold cross-validation presented in Fig. 3, which has the ability to ensure that different portions of data were used in the training and testing dataset. Previous studies showed that k=10 is the most applied value in the ML [46], which has the best results as well as the speed of calculation among other values. In addition, the capability of algorithms for data selecting were improved with methods of halving search, grid search, and random search, which are presented in Fig. 4. Due to having a large number of data points (i.e. 1121184 data points), the grid search and random search can help to enhance the time of execution, since these approaches have the ability to use a selected part of data points (see Fig. 4). It is noteworthy that, for each of the algorithms, the data resampling methods introduced in this section were implemented and the best method was selected for the ML algorithm. It is worth mentioning that, in order to improve the capability of aforementioned algorithms, innovative methods of hyperparameter optimization, such as pipe-line and fine-tuning methods were implemented in Python software [36-38].

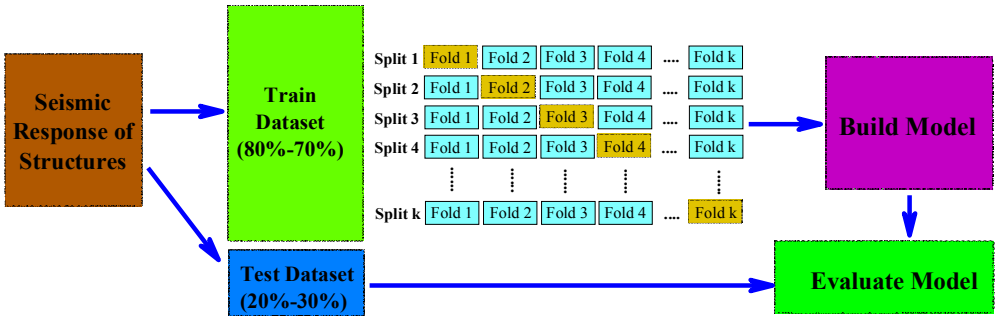


Fig. 3. The *k*-fold cross-validation method.

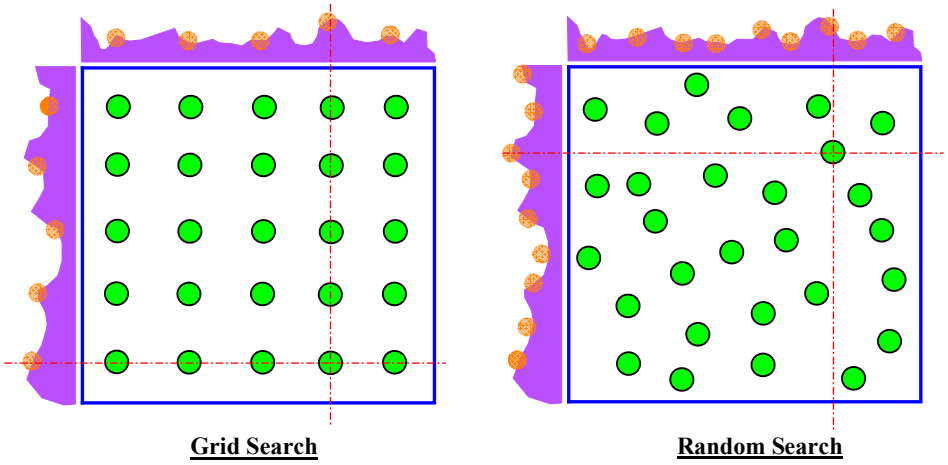


Fig. 4. Grid search and random search methods.

3. Structural modeling

The purpose of this research is to propose an ML-based seismic risk assessment methodology for the evaluation of RC buildings. To prepare datasets for ML algorithms, different RC buildings, having two to twelve-story elevations (i.e. 2-Story, 3-Story, 4-Story, 5-Story, 6-Story, 7-Story, 8-Story, 9-Story, 10-Story, 11-Story, and 12-Story buildings) assuming five types of bays (e.g. one to five-bay frames) with the length of 5 m, 6.1 m, and 7.6 m, were modeled. ETABS 2016 software was used for structural modeling of buildings assuming the site with high-risk category and soil type D, and seismic parameters of SD_1=0.6g and SDs=1.0g (i.e. SD_S and SD_1 are spectral response acceleration parameters at 0.2 and 1.0 seconds). Considering the special RC moment frames as a lateral resisting system, the design parameters of R=8, Cd=5.5, and Ω=3 were chosen following ASCE 7-16 [47]. For modelling, the structural plan presented in Fig. 5 was used with dead and live loads of 8.4 kN/m^2 and 2.4 kN/m^2, respectively, considering the concrete compressive strength of 345 MPa. In addition, the elements of beams and columns were defined using elastic beam-column element with zero-length in both ends of element (see Fig. 6) having the trilinear backbone curve behavior presented in Fig.5. To accurately model the effects of beam-to-column connection, the panel zone was modeled using the procedure proposed by Haselton and Deierlein [48]. In addition, the accuracy of the models were verified according to the models used by Haselton and Deierlein [48]. For brevity, the structural design of beams and columns of RC frames having a bay length of 6.1 m is illustrated in Figs. A-1 to A-3 in *Appendix*. To perform collapse analysis, the procedure considered by Haselton and Deierlein [48] and Kazemi et al. [13-15] was used to model two-dimensional frames in Opensees [49]. According to this procedure, each gravity columns was assumed as a leaning column to consider the P-delta effects [50-53] (see Fig. 6).

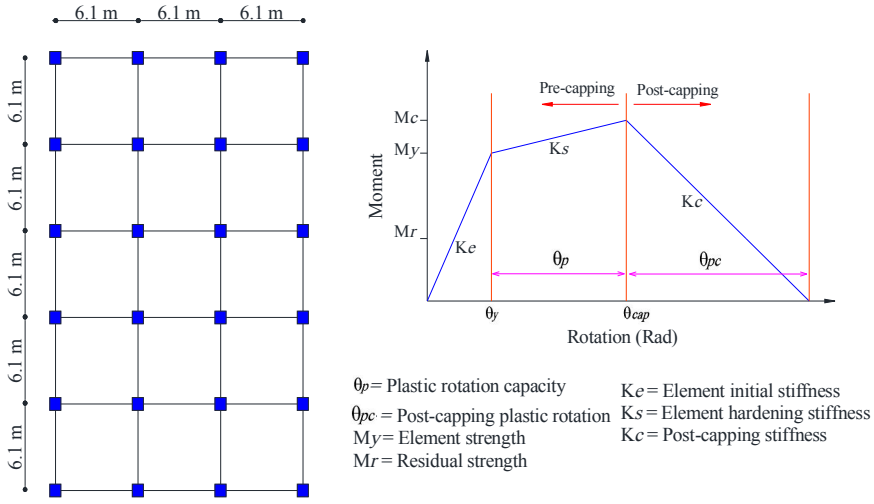


Fig. 5. Plan of RC frames and concentrated plasticity model used for the modeling process.

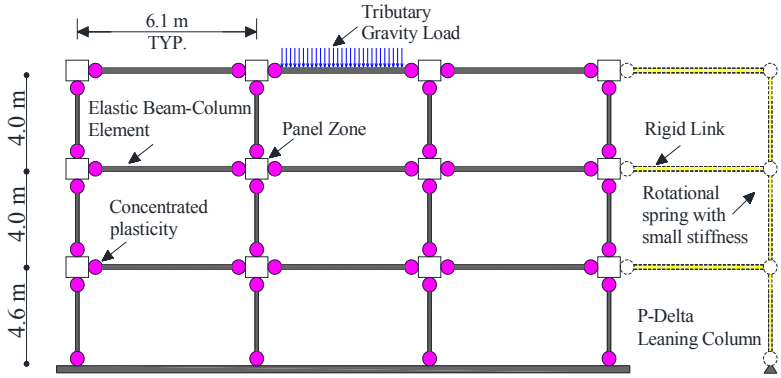


Fig. 6. Schematic view of the modeling process in Opensees [49].

It is noted that in severe earthquakes, the deterioration at large displacements, which leads to sideway collapse, is important. To accurately model the structures with possibility of the stiffness and strength deterioration and sideway collapse, concentrated plastic hinges were used (see Fig 5). To model structures, the concentrated plasticity model, that considers the hysteretic deterioration of the structural elements, was assumed in both ends of the elastic beam-column element. To assume the plasticity of the elements, the trilinear backbone curve model developed by Ibarra et al. [54] was used that can be defined based on the parameters shown in Fig. 5. This model was implemented in Opensees [49] by Altoontash [55] and it was used by some researchers (e.g. see [5, 10, 43]).

ML methods were trained with 1121184 data points achieved by performing IDAs on the 165 RC frames. It is noteworthy that the structural features can affect the seismic response of the structures. Therefore, all structural features of fundamental period, bay length, number of bays, total and story elevation, number of stories, spectral acceleration in the fundamental period, $Sa(T_1)$, story weight, dead and live loads, the RSN number, and record direction were assumed as input parameters in the training dataset. To perform IDAs, $Sa(T_1)$ and the Maximum Interstory

Drift Ratio (Max. IDR) were considered as Intensity Measure (IM) and structural demanding parameter, respectively, assuming three subsets of far-fault (including 44 records), near-fault pulse-like (including 28 records), and no-pulse (including 28 records) ground motions introduced by FEMA-P695 [56] (see Tables A-4B and A-6B in [56] for the details of the ground motions). It should be noted that, in order to reduce the time of analysis, the hunt & fill algorithm with three sub-steps was implemented in Opensees [49], in which, performing analysis was started with the small amplitude and increased by sub-steps to achieve the collapse state. Therefore, for each of the records, the number of analysis may vary due to different collapse states. For managing the analysis and processing the datasets, MATLAB [57] software was used. The IDA curves of the 3-Story, 5-Story, 7-Story, and 9-Story RC frames assuming five bays with a length of 7.6 m including pulse-like records are presented in Fig. 7. The IDA curves can be consider to obtain the fragility curves assuming the lognormal distribution of seismic capacities of RC frames in the prescribed Max. IDR.

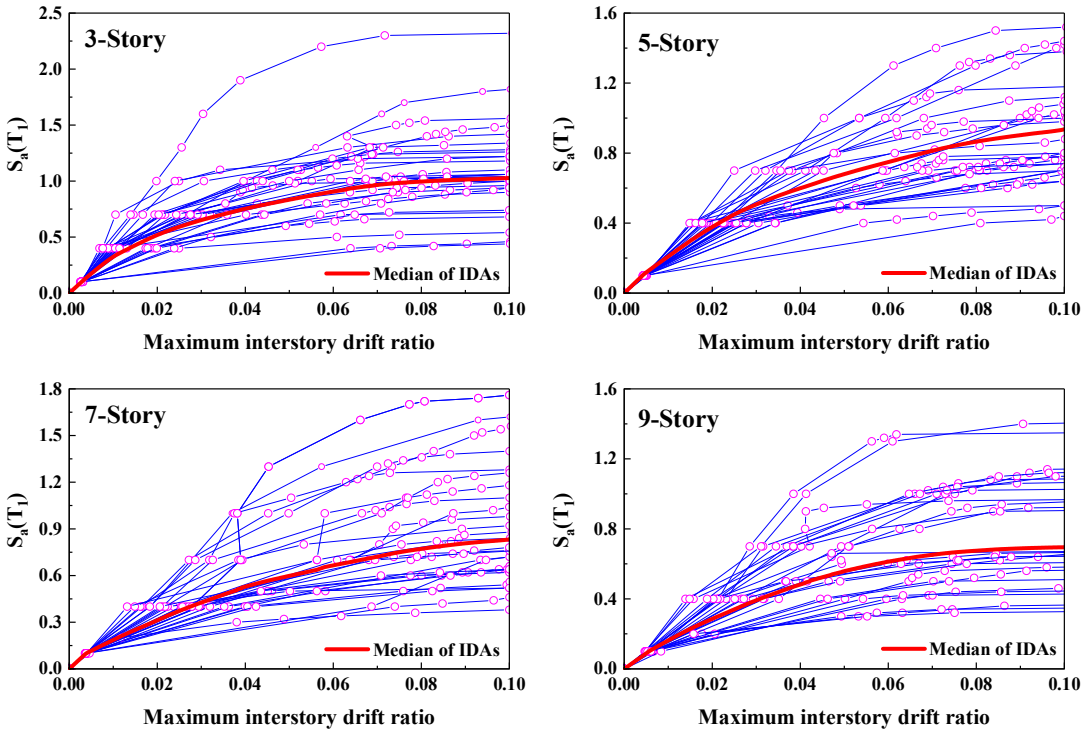


Fig. 7. IDA curves achieved for the RC frames assuming five bays with a length of 7.6 m including pulse-like records.

4. Seismic vulnerability assessment

A probability distribution can be widely used for describing probabilistic events. Two main curves that assign a possibility to each possible outcome can be known as Probability Density Function (PDF) and Cumulative Distribution Function (CDF), which is also known as the fragility curve. Fig. 8 presents the PDF and CDF curves.

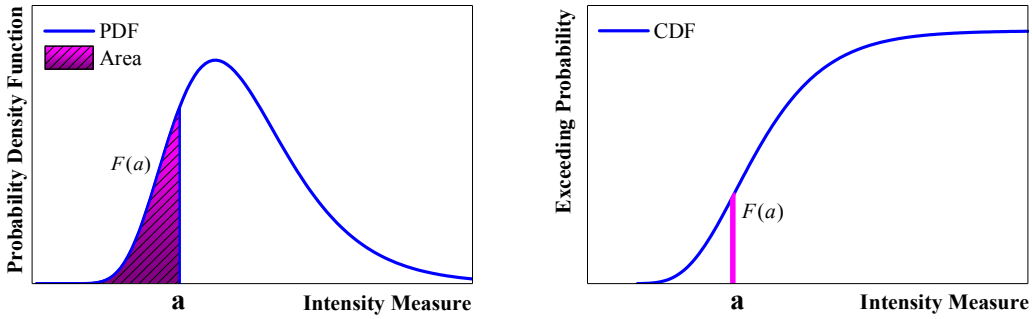


Fig. 8. The PDF and CDF curves.

According to Fig. 8, each value of the CDF curve is equal to the area under the PDF curve to the left of the value (highlighted part in PDF curve). Both of the curves can be used for probability assessments in earthquake engineering. Therefore, both curves were defined as targets of prediction. The probability of exceeding a specific Performance Level (PL) in the assumed IM can be determined as follows:

$$P(PL \mid IM) = P\{PL \geq (IO, LS, CP, C) \mid (IM = Sa(T_1)\} = \Phi\left(\frac{\ln \mathrm{Sa} - \ln \overline{\mathrm{Sa}}_{PL}}{\sigma_{\ln Sa_{PL}}}\right) \tag{3}$$

where $\Phi()$ is known as CDF of the standard normal distribution, and the standard deviation and logarithmic mean are shown by $\sigma_{\ln Sa_{PL}}$ and $\ln \overline{\mathrm{Sa}}_{PL}$, respectively. To obtain these values, the following formula can be used [58]:

$$\ln \overline{\mathrm{Sa}}_{PL} = \frac{1}{n}\sum_{i=1}^{n}(\ln Sa_{PL_i}) \tag{4}$$

$$\sigma_{\ln Sa_{PL}} = \left[\frac{1}{n-1}\sum_{i=1}^{n}(\ln \mathrm{Sa}_{PL_i} - \ln \overline{\mathrm{Sa}}_{PL})^2\right]^{0.5} \tag{5}$$

where n is the number of earthquake events and $\ln \mathrm{Sa}_{PL_i}$ is the natural logarithmic value of Sa in the selected PL assuming the ith earthquake (for more detail, see [58]). The seismic risk in each selected PL can be calculated by the mean annual frequency, λ_{PL}, according to the following formula:

$$\lambda_{PL} = \int_0^{\infty} P(PL \mid Sa) \times \left| d\lambda_{Sa}(Sa) \right| \tag{6}$$

$$\lambda_{PL} = \int_0^{\infty} P(PL \mid Sa) \times \left| \frac{d\lambda_{Sa}(Sa)}{d(Sa)} \right| \times d(Sa) \tag{7}$$

Equation (7) was rewritten from Equation (6) by dividing and multiplying *d(Sa)* on the right side. The term of $\left|\frac{d\lambda_{Sa}(Sa)}{d(Sa)}\right|$ presents the slope of the assumed hazard curve; thus, it can be determined numerically by changing the equation to the following formula:

$$\lambda_{PL} = \sum_{i=1}^{n} P(PL \mid Sa_i) \times \left| \frac{d\,\lambda_{Sa}(Sa_i)}{d(Sa)} \right| \times \Delta(Sa) \tag{8}$$

Fig. 9 presents the graphical description of achieving the mean annual frequency, λ_{PL}, in the selected PL using the fragility curve and the hazard curve, which shows the mean annual frequency of exceeding earthquake intensities at the specific site (e.g. the site of structures, California). In each selected PL, the fragility curve describes the probability of PL (e.g. life safety) conditioned on the intensity of the earthquake.

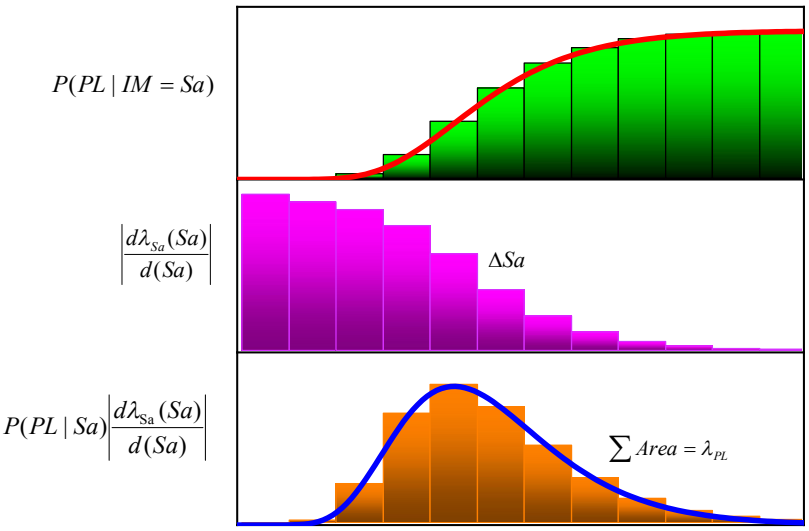


Fig. 9. Fragility curve of assumed PL, hazard curve in numerical derivative, and λ_{PL} deaggregation curve.

λ_{PL} presents the mean yearly rate of the PL, which is used to describe the occurrence probability (e.g. collapse occurrence) over the *t* years, *R(t)*, considering a Poisson process according to:

$$R(t) = 1 - e^{-\lambda_{PL} t} \tag{9}$$

It should be noted that the *R(t)* was widely determined for collapse state to calculate the safe remaining lifetime for a building. For instance, assuming the collapse probability conditioned for a structure in the particular site is equal to 10% in the maximum considered earthquake (MCR). It is important to estimate the safe remaining lifetime of a building for allowable performance levels. Therefore, by rewriting Equation (9) and assuming 1% probability of collapse in 50 years, in accordance with FEMA P155 [59], and the *R(t)* equal to 0.01, the safe remaining lifetime of a structure, t_R, considering the collapse state as PL can be calculated as follows:

$$t_R = 0.01 \lambda_{collapse}^{-1} \tag{10}$$

The seismic risk assessment and estimating the safe remained lifetime, t_R, require complex modeling processes and analytical difficulties. The main difficulty is to determine the fragility curve, while the hazard curve can be achieved according to the site of structures. Therefore, this study aims to propose a prediction model based on the ML methodology to estimate the PDF and

CDF curves of the RC frames in four considered PL, which can be used for determining λ_{PL} and t_R.

5. Analytical procedure

This study explores a surrogate prediction model to reduce the analytical efforts and computational time for seismic risk assessment of RC buildings. After modeling the RC buildings in ETABS 2016 software and validating models in Opensees [49], IDAs were performed to obtain IDA curves of RC frames based on the assumed record subsets. Then, the training and testing datasets were prepared based on the aforementioned structural features of RC frames, and the PDF and CDF curves according to three prescribed allowable Max. IDR values of 1.0%, 2.0%, and 4.0% that shows the PL of Immediate Occupancy (IO), Life Safety (LS), and Collapse Prevention (CP) [60]. In addition, the flat part of the IDA curve that shows the seismic sideway collapse of structures was assumed as a total collapse (C) limitation. It should be noted that these four prescribed allowable Max. IDR were used for comparing the reliability of prediction models, while there is no limitation in the prediction models to select the intensity level and seismic limit-state threshold. Fig. 10 presents the analytical procedures used for ML-based seismic risk assessment of RC buildings. First box is related to the modeling process of the RC structures and preparing the data points for ML models, and second box presents the process used in this research to estimate the λ_{PL} and $R(t)$ using predicted PDF and CDF curves.

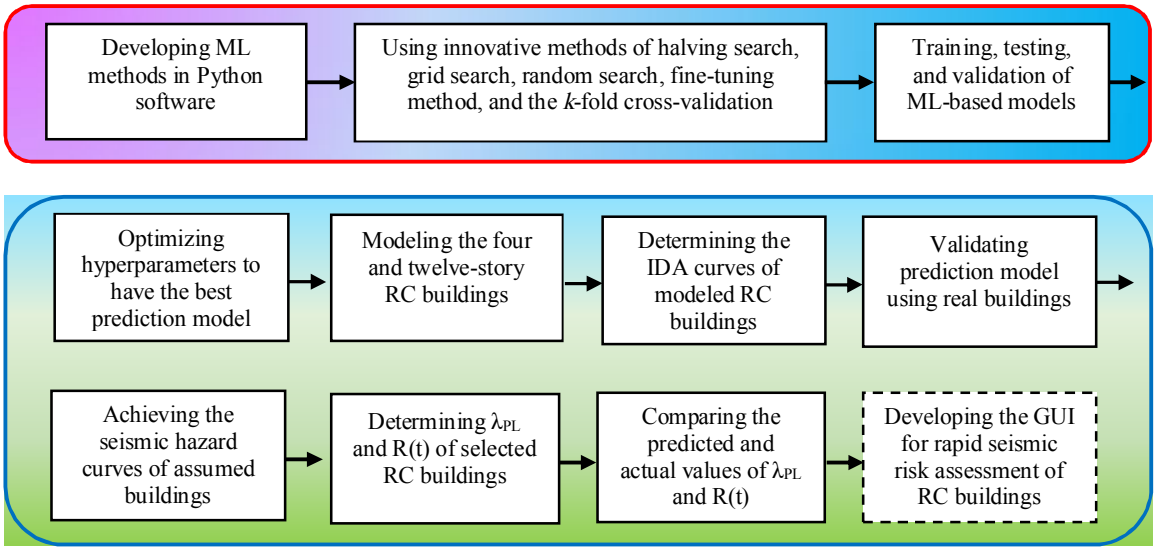


Fig. 10. Analytical procedures used for ML-based seismic risk assessment of RC buildings.

Although many structural features such as pulse-like effects of ground motions may affect the result of the prediction, forward and backward feature selection method were used to distinguish as low as possible features, in which the model metrics remained unchanged. This method can be used to verify the selected features and reduce the number of features that can be facilitated the use of prediction models for users. In general view, feature importance can identify the score of each input features based on their influence at predicting a target. The trial and error method was used to explore the important structural features that may affect the results of the prediction. Using

different innovative methods of hyperparameter optimization, such as halving search, grid search, random search, fine-tuning method, and the *k*-fold cross-validation for model evaluation showed that the main features of $Sa(T_1)$, fundamental period, number of stories, number of bays, story weight, bay length, and total elevation can play a key role in prediction.

Fig. 11 illustrates the relative importance of input features determined by trial and error for PDF (left) and CDF (right) curves. For the PDF curve, $Sa(T_1)$, fundamental period, and the number of stories achieved scores of 70.7%, 17.18%, and 5.14%, respectively, while for the CDF curve, the scores of 57.46%, 20.38%, and 11.73%, respectively, were determined. It can be observed that three features of $Sa(T_1)$, fundamental period, and the number of stories had higher scores compared to other features in both curves.

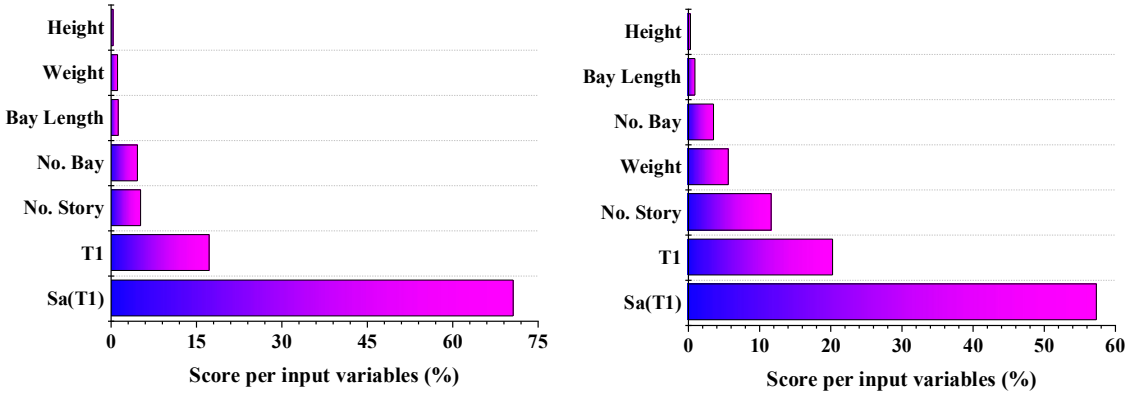


Fig. 11. Relative importance of input features determined by trial and error for PDF (left) and CDF (right) curves.

To compare the results of prediction, statistical metrics of Mean Average Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2), were determined for twenty ML algorithms. The following formula can be used for calculations:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\text{Actual}_i\text{-Predicted}_i)^2 \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{Actual}_i - \text{Predicted}_i)^2} \tag{12}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\text{Actual}_i - \text{Predicted}_i\right| \tag{13}$$

Table 1 illustrates the statistical metrics determined for predicting the CDF curve of the 8-Story RC frame assuming five bays with a length of 6.1 m including no-pulse records.

Table. 1. Statistical metrics for predicting the CDF curve of the 8-Story RC frame assuming five bays with a length of 6.1 m including no-pulse records.

Model	Training					Testing				
	R^2	MSE (10^{-2})	RMSE (10^{-2})	MAE (10^{-2})	Score	R^2	MSE (10^{-2})	RMSE (10^{-2})	MAE (10^{-2})	Score
XGBoost	0.999	0.016	0.264	0.845	18	0.998	0.030	1.719	1.330	24

RF	0.980	0.252	5.024	3.194	44	0.993	0.096	3.092	2.310	40
BR	0.999	0.001	0.304	0.155	8	0.999	0.007	0.852	0.652	9
ETR	0.999	0.003	0.003	0.002	8	0.999	0.012	0.082	0.683	12
GBM	0.997	0.035	1.881	1.254	36	0.997	0.037	1.931	1.511	28
HGBR	0.999	0.011	1.051	0.685	23	0.999	0.006	0.783	0.581	7
HGBR(FN)	0.999	0.016	1.249	0.798	28	0.995	0.059	2.428	1.703	36
AdaBoost	0.976	0.306	5.531	4.604	48	0.988	0.151	3.891	3.510	44
KNR	0.999	0.005	0.694	0.419	19	0.947	0.695	8.335	5.721	55
PLSR	0.570	5.514	23.482	19.308	77	0.536	6.064	24.624	22.935	77
SR	0.856	1.854	13.616	7.811	63	0.914	1.124	10.604	7.027	59
VR	0.905	1.213	11.013	9.059	61	0.901	1.300	11.400	10.082	64
LR	0.623	4.838	21.996	18.090	70	0.610	5.100	22.583	20.038	73
GR	0.699	8.086	28.435	24.800	77	0.674	9.868	31.413	28.982	77
MLPR	0.993	0.084	2.902	2.027	40	0.997	0.038	1.946	1.589	32
SVM	0.957	0.553	7.437	6.147	56	0.950	0.651	8.067	7.094	54
NuSVR	0.967	0.425	6.522	5.479	52	0.970	0.393	6.271	5.281	48
LSVR	0.600	5.127	22.642	17.515	72	0.630	4.845	22.010	17.401	69
ERTR	0.999	0.002	0.484	0.219	16	0.999	0.009	0.967	0.632	15
ANNs	0.999	0.008	0.884	0.662	24	0.999	0.008	0.886	0.769	17

The score marker was used to rate each algorithm compare to others with a number from 1 to 20. The ranking list provided by sorting the R^2 in the highest to lowest values and for the MAE, RMSE, and MSE in the lowest to highest values, which shows the better the performance, the smaller the indicator. Finally, scores for four metrics are collected in training and testing datasets to show the reliability of the methods. It is noteworthy that the best method should achieve the lowest score which means its performance is better than others. According to scores calculated for the training dataset, algorithms of the BR, ETR, ERTR, XGBoost, KNR, HGBR, ANNs, HGBR(FN), and GBM achieved scores of 8, 8, 16, 18, 19, 23, 24, 28, and 36, respectively. For the testing dataset, the HGBR, BR, ETR, ERTR, ANNs, XGBoost, GBM, MLPR, and HGBR(FN) algorithms achieved scores of 7, 9, 12, 15, 17, 24, 28, 32, and 36, respectively. To conclude the comparison, these algorithms are introduced as the best prediction model for training and testing datasets. In addition, it is observed that the algorithms of VR, SR, LR, LSVR, GR, and PLSR had lower scores and cannot be able to be a prediction model.

To better illustrate the results, all predicted CDF curves are plotted with the actual curve. Fig. 12(a)-(c) shows the predicted CDF curves of the 8-Story RC frame assuming five bays with a length of 6.1 m including no-pulse records using twenty aforementioned ML models. According to Fig. 12(a), seven algorithms of the XGBoost, RF, BR, ETR, GBM, HGBR, and HGBR(FN), have the ability to estimate the CDF curves, while the XGBoost, BR, ETR, and HGBR algorithms were fitted on the actual CDF curve. In Fig. 12(b), there is no algorithm to exactly fit the actual CDF curve. In Fig. 12(c), three algorithms of the MLPR, ERTR, and ANNs have the ability to estimate the actual CDF curve of the 8-story RC frame. According to the sensitivity analysis that was done by plotting the predicted CDF curves, the XGBoost, BR, ETR, MLPR, ERTR, ANNs, and HGBR algorithms have higher accuracy of curve fitting, and these seven improved algorithms were used as ML-based prediction model to estimate the PDF and CDF curves.

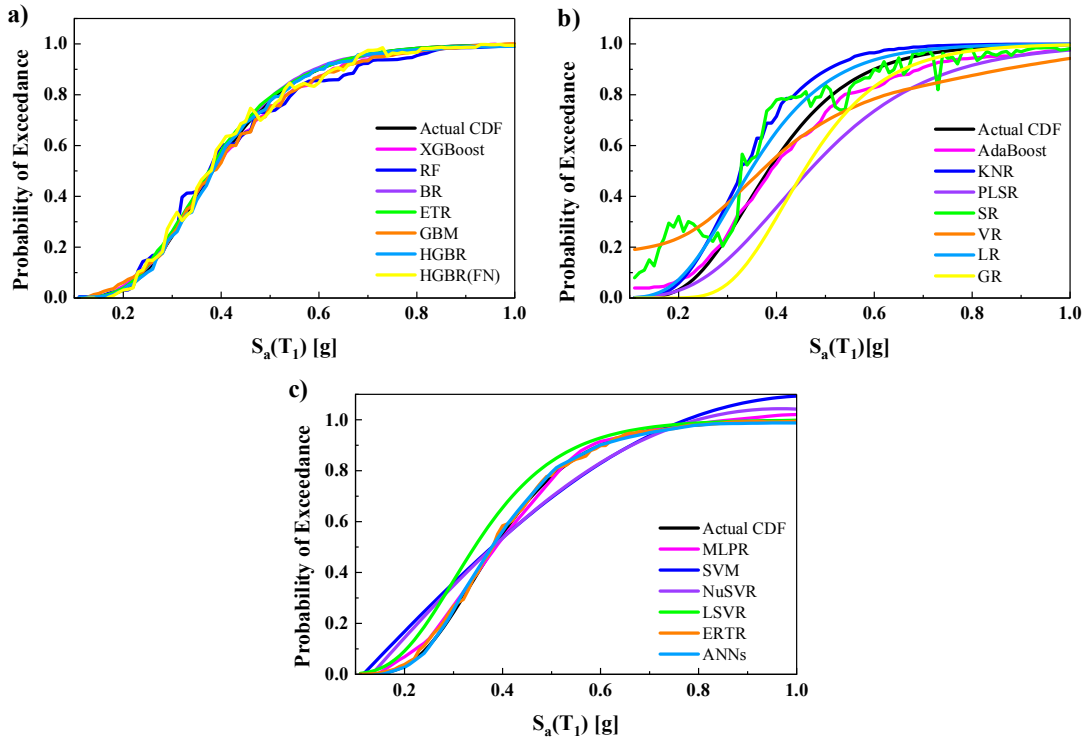


Fig. 12. Predicted CDF curves of the 8-Story RC frame assuming five bays with a length of 6.1 m including no-pulse records.

6. Performance evaluation of models

According to *Section 4*, eight algorithms had acceptable results in both train and test datasets and can be assumed as a prediction model. In this section, to better evaluate the analysis results, three performance indicators of mean absolute relative error (MARE), mean squared relative error (MSRE), and root mean squared relative error (RMSRE) were also determined to evaluate the performance of ML algorithms for a new set of the testing dataset to predict CDF and PDF curves. The MARE, MSRE, and RMSRE can be calculated as follows [36, 61]:

$$MARE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\text{Actual}_i - \text{Observed}_i}{\text{Observed}_i}\right| \tag{14}$$

$$MSRE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\text{Actual}_i - \text{Observed}_i}{\text{Observed}_i}\right|^2 \tag{15}$$

$$RMSRE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\text{Actual}_i - \text{Observed}_i}{\text{Observed}_i}\right|^2} \tag{16}$$

Table 2 illustrates the results of loss functions for eight selected algorithms assuming the 6-Story RC frame assuming four bays with the length of 7.5 m in the performance level of LS including no-pulse records. For the predicted CDF curve, the algorithms of ETR, ERTR, BR,

ANNs, HGBR, and XGBoost had total scores of 17, 20, 26, 29, 31, and 36, respectively, while two algorithms of HGBR(FN) and GBM achieved higher scores of 44 and 49, respectively. For the predicted PDF curve, the algorithms of ERTR, ETR, ANNs, BR, HGBR, and XGBoost had total scores of 16, 17, 27, 31, 32, and 37, respectively, while two algorithms of GBM and HGBR(FN) achieved higher scores of 45 and 47, respectively. Therefore, for predicting both curves of CDF and PDF, the performance of six algorithms is higher than for other assumed ML algorithms, and they are introduced as the best prediction models.

Table 2. Performance indicators of predicted CDF and PDF curves of the 6-Story RC frame assuming four bays with the length of 7.5 m in the performance level of LS including no-pulse records.

Model	R^2	MSE (10^{-2})	RMSE (10^{-2})	MAE (10^{-2})	MARE	MSRE	RMSRE	Score
	CDF as a testing dataset							
XGBoost	0.988	0.149	3.863	2.770	0.387	2.729	1.652	36
BR	0.992	0.101	3.182	2.121	0.287	0.862	0.928	26
ETR	0.998	0.029	1.714	1.316	0.148	0.315	0.561	17
GBM	0.963	0.474	6.887	4.708	5.497	1247.35	35.31	49
HGBR	0.992	0.106	3.254	2.151	0.293	0.996	0.998	31
HGBR(FN)	0.983	0.222	4.711	3.259	0.741	19.453	4.411	44
ERTR	0.997	0.036	1.894	1.289	0.175	0.572	0.756	20
ANNs	0.999	0.006	0.786	0.613	0.446	12.267	3.502	29
	PDF as a testing dataset							
XGBoost	0.978	1.124	10.604	8.147	0.401	0.320	0.566	37
BR	0.977	1.180	10.862	7.589	0.218	0.227	0.476	31
ETR	0.997	0.177	4.203	3.394	0.166	0.098	0.314	17
GBM	0.943	2.942	17.152	11.620	0.377	0.438	0.662	45
HGBR	0.964	1.855	13.620	9.738	0.295	0.187	0.433	32
HGBR(FN)	0.944	2.901	17.031	11.326	0.472	0.664	0.815	47
ERTR	0.993	0.370	6.082	3.946	0.129	0.041	0.202	16
ANNs	0.989	0.589	7.673	5.227	0.243	0.225	0.474	27

Since different values of hyperparameters can considerably influence the results and performance of algorithms, the hyperparameters of algorithms were optimized in order to achieve higher percentage of prediction. For reproducibility of the algorithms, Table 3 illustrates the hyperparameters used for the definition of the six selected algorithms.

Table. 3. Hyperparameters of the best prediction ML algorithms.

Model	Parameter	Value	Model	Parameter	Value	Model	Parameter	Value
XGBoost	No. of estimators	2500	HistGBR(FN)	No. of estimators	3000	ERTR	No. of estimators	2000
	Learning rate	0.01		Learning rate	0.001		Min sample leaf	10
	Min sample leaf	10		Min sample leaf	8		Min sample split	2
	Min sample split	10		Min sample split	8		Random state	3000
	Max depth	4		Max depth	4	ANNs	Loss	MSE
GBM and HistGBR	No. of estimators	3000	BR	No. of estimators	300		Hidden layer 1	10 (Relu)
	Learning rate	0.01		Random state	0		Hidden layer 2	10 (Relu)
	Min sample leaf	10	ETR	No. of estimators	300		Hidden layer 3	5 (Sigmoid)
	Min sample split	10		Random state	0		Learning rate	Invscaling
	Max depth	5		Min sample leaf	5		Optimizer	Adam
				Min sample split	10		Max iterations	3000

Figs. 13 and 14 present the predicted CDF and PDF curves of the 6-Story RC frames assuming

five types of bays with the length of 7.5 m in the performance level of LS including no-pulse records, respectively. It can be observed that the predicted curve by ANNs precisely fitted the actual curve.

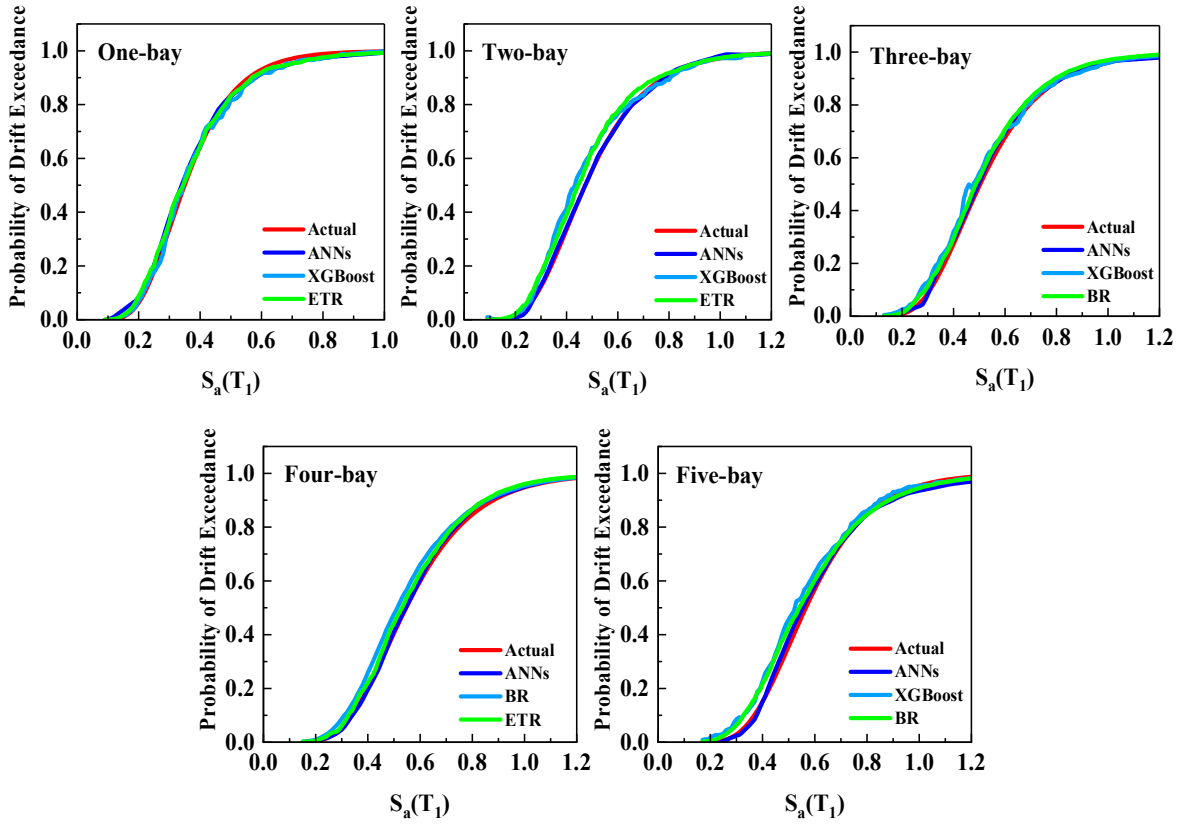
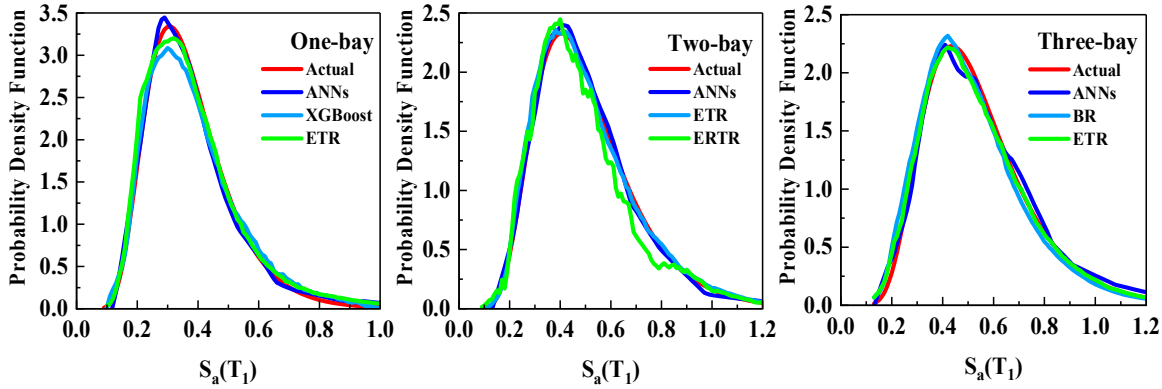


Fig. 13. Predicted CDF curves of the 6-Story RC frames with the length of 7.5 m in the performance level of LS including no-pulse records.



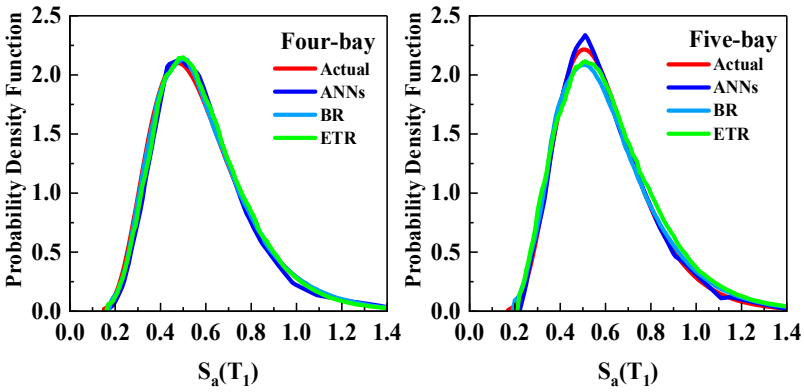


Fig. 14. Predicted PDF curves of the 6-Story RC frames with the length of 7.5 m in the performance level of LS including no-pulse records.

To present the normally distributed collected data points from analyses, the Quantile-Quantile (Q-Q) plot, which shows the standard normality of data points, was determined [62]. Fig. 15 presents the Q-Q plot related to the CDF and PDF data points. As shown, the CDF and PDF follow the normal distribution due to approximately well-spread points in the bisector.

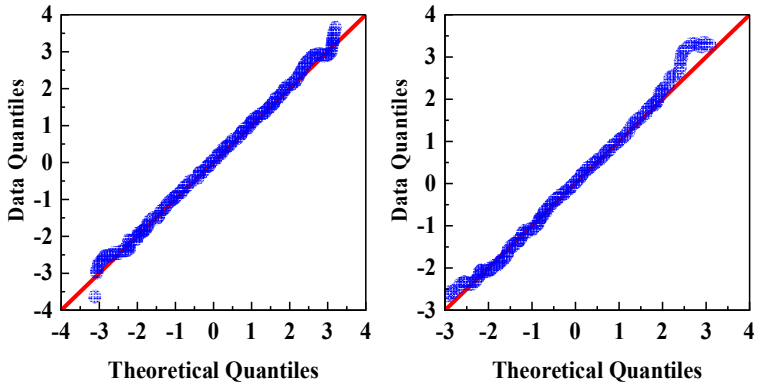


Fig. 15. Q-Q plot of normality test corresponding to CDF (left) and PDF (right) data points.

7. General ability of ML-based model

In Section 3, the models were defined to provide a wide range of the RC structures for training and testing the ML methods. According to previous sections, the prediction models have shown their acceptable accuracy of estimation based on the aforementioned RC structures. In this section, the validation of prediction models is discussed to use ML-based models in a general way for seismic risk assessment of RC buildings. Fig. 16 presents the structural plan of the four-story RC frame used by Haghollahi and Behnamfar [63], and the twelve-story RC frame used by Allahvirdizadeh et al. [64], which are used for estimating seismic risk assessment.

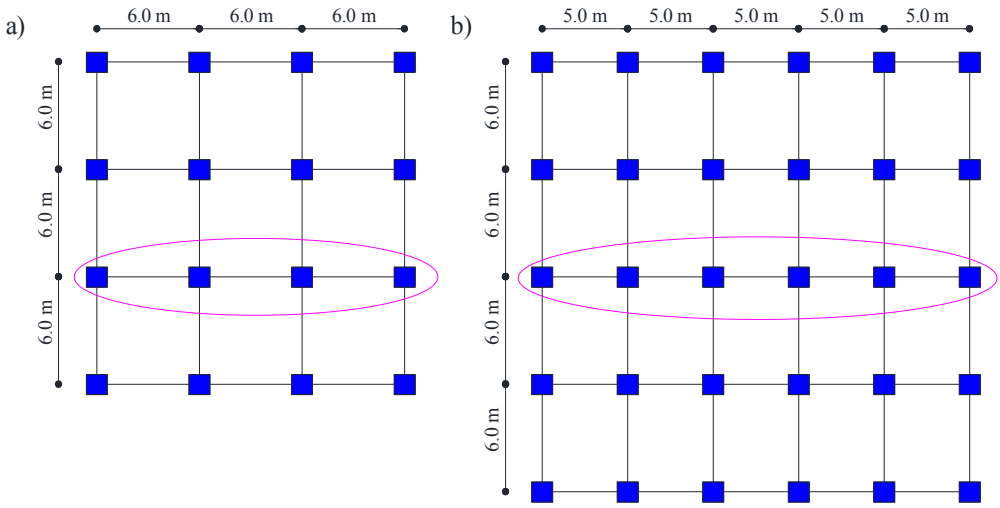


Fig. 16. Structural plan of a) the four-story RC frame used by Haghollahi and Behnamfar [63], and b) the twelve-story RC frame used by Allahvirdizadeh et al. [64].

The selected structures are used to check the applicability of the ML-based models for predicting the CDF and PDF curves. For designing the four-story frame, the length of bays and each story elevation were assumed 6 m and 4 m, respectively. The frame was assumed as a residential building with a special lateral-resisting frame founded on soil type D, having dead, and live loads of 7 kN/m^2 and 2.5 kN/m^2, respectively.

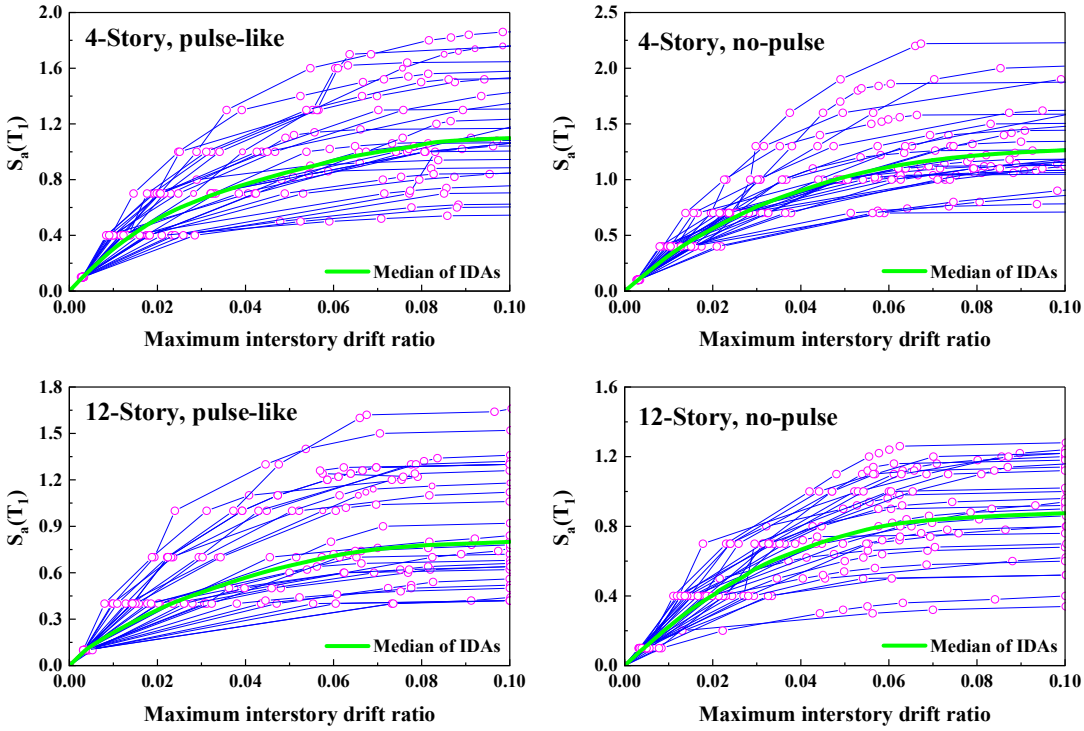


Fig. 17. IDA curves of the four and twelve-story RC frames.

The location of the building is assumed in California with seismic parameters of S_s and S_1 equal to 1.5g and 0.6g, respectively (for more detail see [63]). The twelve-story frame is assumed for residential use with story height and bay length equal to 3.2 m and 5 m, respectively. The floors were assumed one-way slab, and seismic design parameters of S_s and S_1 equal to 1.96g and 0.69g, respectively, were selected regarding a site in California with soil type D (for more detail see [64]). The fundamental period of four and twelve-story RC frames are equal to 1.33 sec and 1.72 sec, respectively. To determine the PDF and CDF of both frames, the IDAs were performed considering three record subsets using Opensees [49] software. Fig. 17 presents the IDA curves of the four and twelve-story RC frames in two record subsets.

The accuracy of proposed prediction models is the main topic of this section. Thus, the PDF and CDF curves of the selected RC frames are converted into the target of the testing dataset. Then, the proposed ML methods are used for estimating the CDF and PDF curves. It is worth noting that the training dataset was assumed from the results of 165 RC frames subjected to three record subsets. Figs. 18 and 19 present the ML-based CDF (left) and PDF (right) curves compared to those of the four and twelve-story RC frames assumed in this section (i.e. shown by actual legends) subjected to no-pulse records, respectively. It can be observed that the ML algorithms have the ability of curve fitting and can be used as a powerful tool for predicting the CDF and PDF curves that can be used for seismic risk assessment of the general structures. It should be said that there are no unique algorithms to estimate all CDF or PDF curves; therefore, the aforementioned best models can be considered, and the best-fitted curve can be used for seismic risk assessment. Similar results were observed assuming other records.

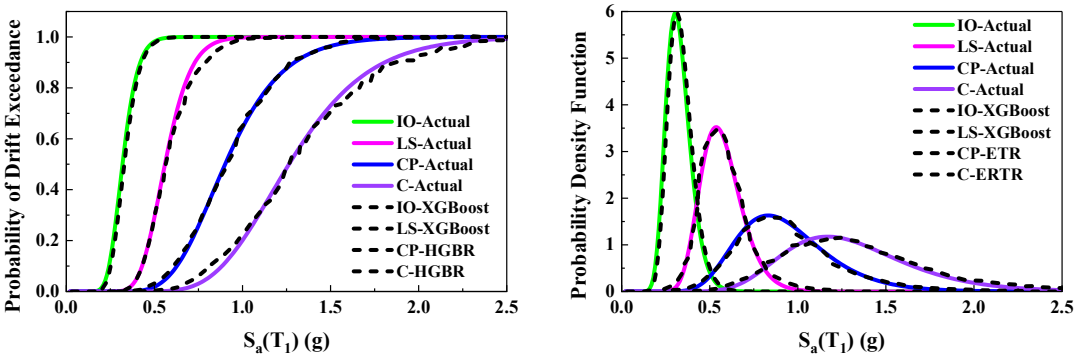


Fig. 18. ML-based CDF (left) and PDF (right) curves of the four-story RC frame under no-pulse records.

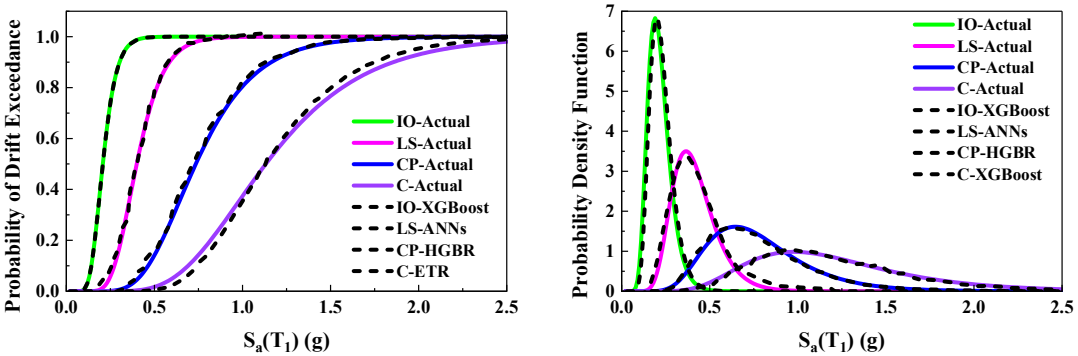


Fig. 19. ML-based CDF (left) and PDF (right) curves of the twelve-story RC frame under no-pulse records.

To determine the seismic risk of the assumed RC frames, the hazard curves were obtained from the USGS hazard tool application [65], and the procedure proposed by Eads [66] was considered to extract the hazard curves of $Sa(T_1$=1.33 sec) and $Sa(T_1$=1.72 sec) corresponding to the four and twelve-story RC frames. Fig. 20 presents the seismic hazard curves of the four and twelve-story RC frames assuming the selected site.

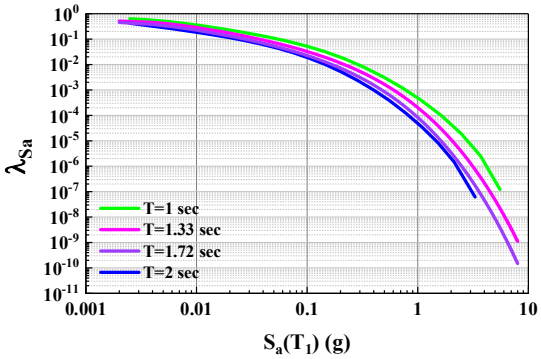


Fig. 20. Seismic hazard curves of the four and twelve-story RC frames assuming the selected site.

To achieve λ_{PL} deaggregation curves, the procedure introduced in Fig. 9 was used. Figs. 21 and 22 compare the predicted and actual λ_{PL} deaggregation curves of the four and twelve-story RC frames determined by ML algorithms.

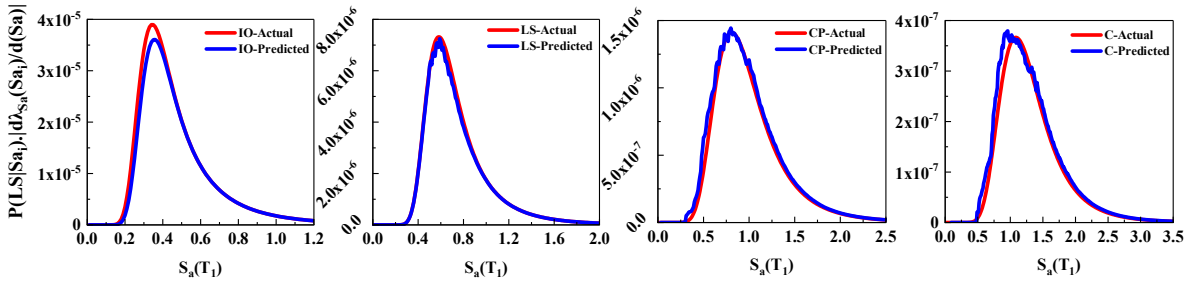


Fig. 21. Seismic risk assessment of the four-story RC frame, λ_{PL} deaggregation curves assuming four seismic limit states.

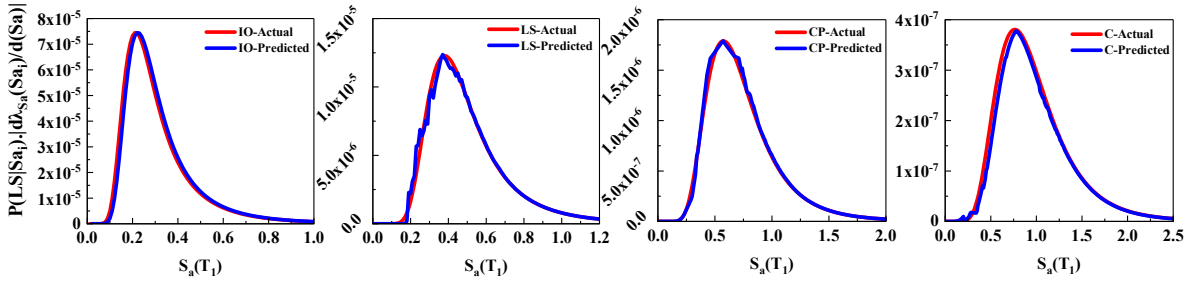


Fig. 22. Seismic risk assessment of the twelve-story RC frame, λ_{PL} deaggregation curves assuming four seismic limit states.

According to the figures, it is obvious that the predicted λ_{PL} deaggregation curves have acceptable accuracy and fit the actual curves in four selected limit-states. Therefore, in the next step, these curves were used for seismic risk assessment. The mean yearly rate of the PL, λ_{PL}, which can be used to describe a probability of occurrence (e.g. four seismic limit-states of IO, LS, CP, and C) over the *t* years, *R(t)*, was calculated based on the Equation of (8) and (9), respectively. Figs. 23 and 24 compare the ML-based predictions of λ_{PL} (left) and *R(t)* (right) for the four and twelve-story RC frames.

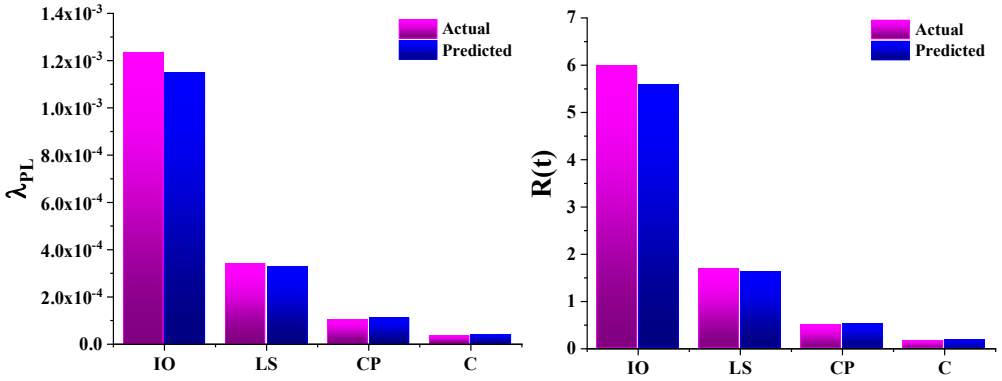


Fig. 23. Comparison of ML-based predictions of λ_{PL} (left) and *R(t)* (right) for the four-story RC frame.

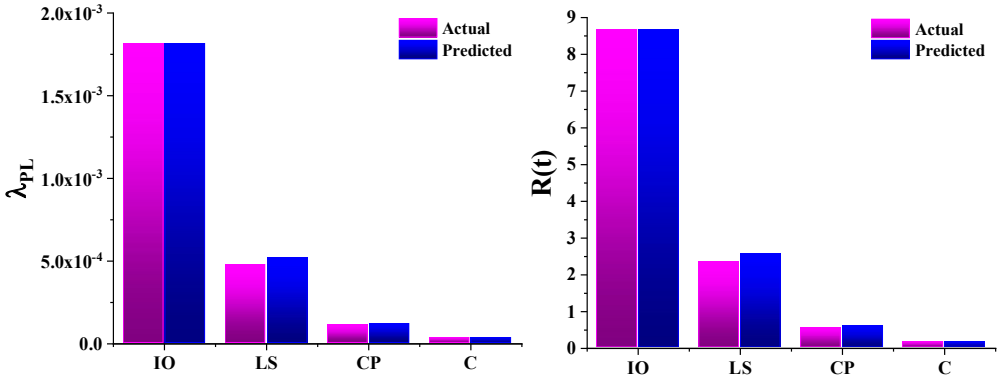


Fig. 24. Comparison of ML-based predictions of λ_{PL} (left) and *R(t)* (right) for the twelve-story RC frame.

For brevity, only the results of no-pulse records were presented, while similar trends and figures were obtained for other records. Therefore, ML models have acceptable capability and reliability for prediction. Table 4 illustrates the results of the $Sa(T_1)$ corresponding to the maximum value of the λ_{PL} deaggregation curves, λ_{PL}, and *R(t)* achieved by actual curves and ML-based predicted curves assuming four seismic performance levels including no-pulse records. By comparing the values of $Sa(T_1)$ corresponding to the maximum value of the λ_{PL} deaggregation curves in both four and twelve-story RC frames assuming four seismic limit-states of IO, LS, CP, and C, it can be seen that the predicted values and actual values have a good agreement, which proves the accuracy of the ML-based models. Similarly, there are small differences in the predicted and actual values of λ_{PL} and *R(t)*. In general conclusion, it can be said that the proposed ML-based prediction models have the ability for predicting the λ_{PL} deaggregation curves, which can be used for determining the values of λ_{PL} and *R(t)*. By validating the ML models with the case

study RC structures, proposed ML-based models can be used as a preliminary tool to evaluate the seismic risk of RC structures assuming different types of structural parameters.

Table 4. Results of the $Sa(T_1)$ corresponding to the maximum value of the λ_{PL} deaggregation curves, λ_{PL}, and $R(t)$ assuming four seismic performance levels including no-pulse records.

$Sa(T_1)$	Four-story				Twelve-story			
	IO	**LS**	**CP**	**C**	**IO**	**LS**	**CP**	**C**
Actual value	0.34	0.58	0.81	1.08	0.21	0.38	0.57	0.76
Predicted value	0.35	0.59	0.80	0.96	0.22	0.37	0.57	0.74
λ_{PL}	Four-story				Twelve -story			
Actual value	1.23E-3	3.39E-4	1.01E-4	3.32E-5	5.985	1.681	0.504	0.166
Predicted value	1.15E-3	3.28E-4	1.07E-4	3.62E-5	5.600	1.627	0.527	0.169
$R(t)$	Four-story				Twelve -story			
Actual value	1.82E-3	4.72E-4	1.07E-4	3.05E-5	8.675	2.334	0.537	0.152
Predicted value	1.81E-3	5.18E-4	1.15E-4	3.04E-5	8.663	2.557	0.574	0.149

8. Graphical user interface

To prepare the accessibility of results for use as a preliminary evaluation of the seismic risk of RC structures, Graphical User Interface (GUI) was developed based on the selected important structural features (see Fig. 25) as input parameters, and the plotting ability of the CDF and PDF curves on the selected limit-state threshold. The GUI has the ability to predict the curves without limitation on the random selection of seismic threshold; thus, in any seismic performance limitation (e.g. life safety), it is possible to have the CDF and PDF curves. According to Fig. 25, four predefined seismic performances were used. In order to provide a guideline for user, the 'Default' option can fill in the input parameters. Then, by introducing the seismic hazard curve of the selected RC structure using the 'Browse' option, the predicted CDF curve can be used for determining the λ_{PL} deaggregation curve, and the value of λ_{PL} according to Equation (8). By assuming the $R(t)$ as input parameter (e.g. $R(t)$ equal to 0.01 assuming 1% probability of collapse in 50 years), it is possible to estimate the safe remained lifetime, t_R, for collapse state according to Equation (10). The users should consider the reliability of predictions made by ML models discussed in this research. The GUI was introduced to mitigate computing time and complicated analysis for seismic risk assessment of RC structures (new updates including the database will be added).

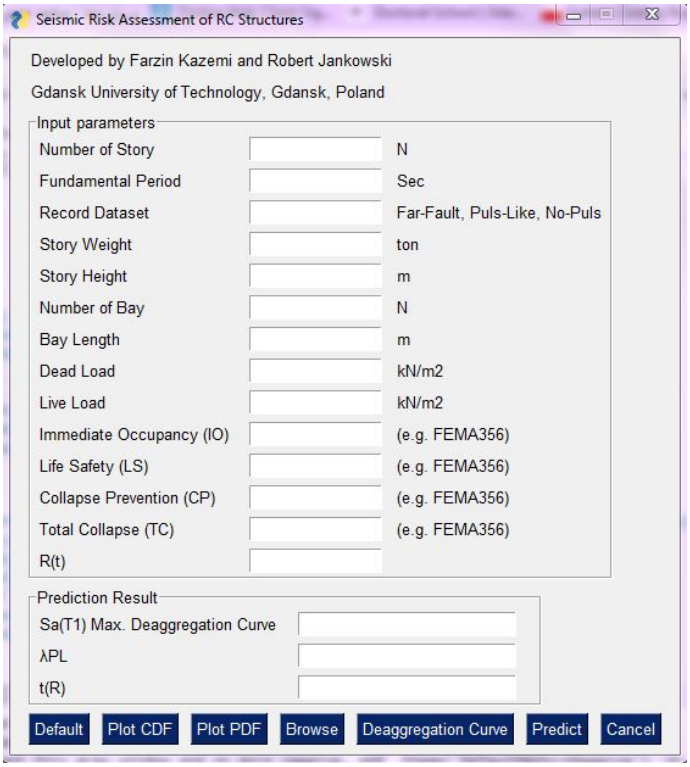


Fig. 25. Graphical user interface for the preliminary seismic risk assessment of RC structures available at https://github.com/FarzinKazemi

9. Conclusions

This research proposes an ML-based prediction model for seismic risk and vulnerability assessment of RC buildings. Seismic risk assessments require the CDF curve of the structure that can be determined by performing IDAs assuming severe earthquakes. IDAs need high-speed computers and, in most cases, complicate the modeling process to assess collapse state. To accelerate the seismic risk assessment, ML algorithms were developed by innovative methods of hyperparameter optimization, such as halving search, grid search, random search, fine-tuning method, and the *k*-fold cross-validation, to derive the PDF and CDF curves of RC frames. In other words, the CDF and PDF curves derived from ML-based prediction models are used for seismic risk assessment to speed the procedure and reduce the time and effort. For this purpose, twenty ML algorithms were developed and trained based on the 1121184 data points achieved by performing IDAs on the 165 RC frames. Following points illustrate the main results:

- The relative importance of input features showed that three structural features of $Sa(T_1)$, fundamental period, and the number of stories had higher effects, as compared to other features. For the PDF curve, $Sa(T_1)$, fundamental period, and the number of stories achieved scores of 70.7%, 17.18%, and 5.14%, respectively, while for the CDF curve, the scores of 57.46%, 20.38%, and 11.73%, respectively, were determined.
- According to sensitivity analysis performed on twenty ML algorithms using four statistical metrics of R^2, MSE, RMSE, and MAE, the algorithms of BR, ETR, ERTR, XGBoost, KNR, HGBR, ANNs, HGBR(FN), and GBM with scores of 8, 8, 16, 18, 19, 23, 24, 28, and 36,

respectively, were the best models for the train dataset. For the test dataset, the HGBR, BR, ETR, ERTR, ANNs, XGBoost, GBM, MLPR, and HGBR(FN) algorithms achieved scores of 7, 9, 12, 15, 17, 24, 28, 32, and 36, respectively. In addition, seven improved algorithms of the XGBoost, BR, ETR, MLPR, ERTR, ANNs, and HGBR had higher accuracy of curve fitting in both the PDF and CDF curves.

- According to seven statistical metrics of R^2, MSE, RMSE, MAE, MARE, MSRE, and RMSRE, the algorithms of ETR, ERTR, BR, ANNs, HGBR, and XGBoost had total scores of 17, 20, 26, 29, 31, and 36, respectively, while two algorithms of HGBR(FN) and GBM achieved higher scores of 44 and 49, respectively, for predicting CDF curve. In addition, the algorithms of ERTR, ETR, ANNs, BR, HGBR, and XGBoost had total scores of 16, 17, 27, 31, 32, and 37, respectively, while two algorithms of GBM and HGBR(FN) achieved higher scores of 45 and 47, respectively, for predicting PDF curve. Therefore, the performance of six algorithms is higher than for other assumed ML algorithms, and they are introduced as the best prediction models.
- To generalize the achievements, two case studies focused on RC buildings were conducted. According to the results, the predicted λ_{PL} deaggregation curves had acceptable accuracy and they fitted to the actual curves in four selected limit-states. The results illustrated that the $Sa(T_1)$ corresponding to the maximum value of the λ_{PL} deaggregation curves, λ_{PL}, and *R(t)* achieved by actual curves and ML-based predicted curves had a good agreement, which proves the accuracy of the ML-based models.
- By validating the ML models with the case studies dealing with RC structures, the proposed ML-based models can be used as a preliminary tool for estimating the seismic risk of RC structures. To conveniently access the results of this study, GUI was introduced to prepare the predicted CDF and PDF curves, the $Sa(T_1)$ corresponding to the maximum value of the λ_{PL} deaggregation curves, λ_{PL}, and *R(t)* values, which facilitates the seismic risk assessment of existing or newly constructed RC structures.

Acknowledgment

Numerical calculations were carried out at the Tri-City Academic Supercomputer and Network (*CI TASK*) in Gdańsk, Poland.

Appendix

The structural design parameters of the RC frames having a bay length of 6.1 m are summarized in Figs. A-1 to A-3.

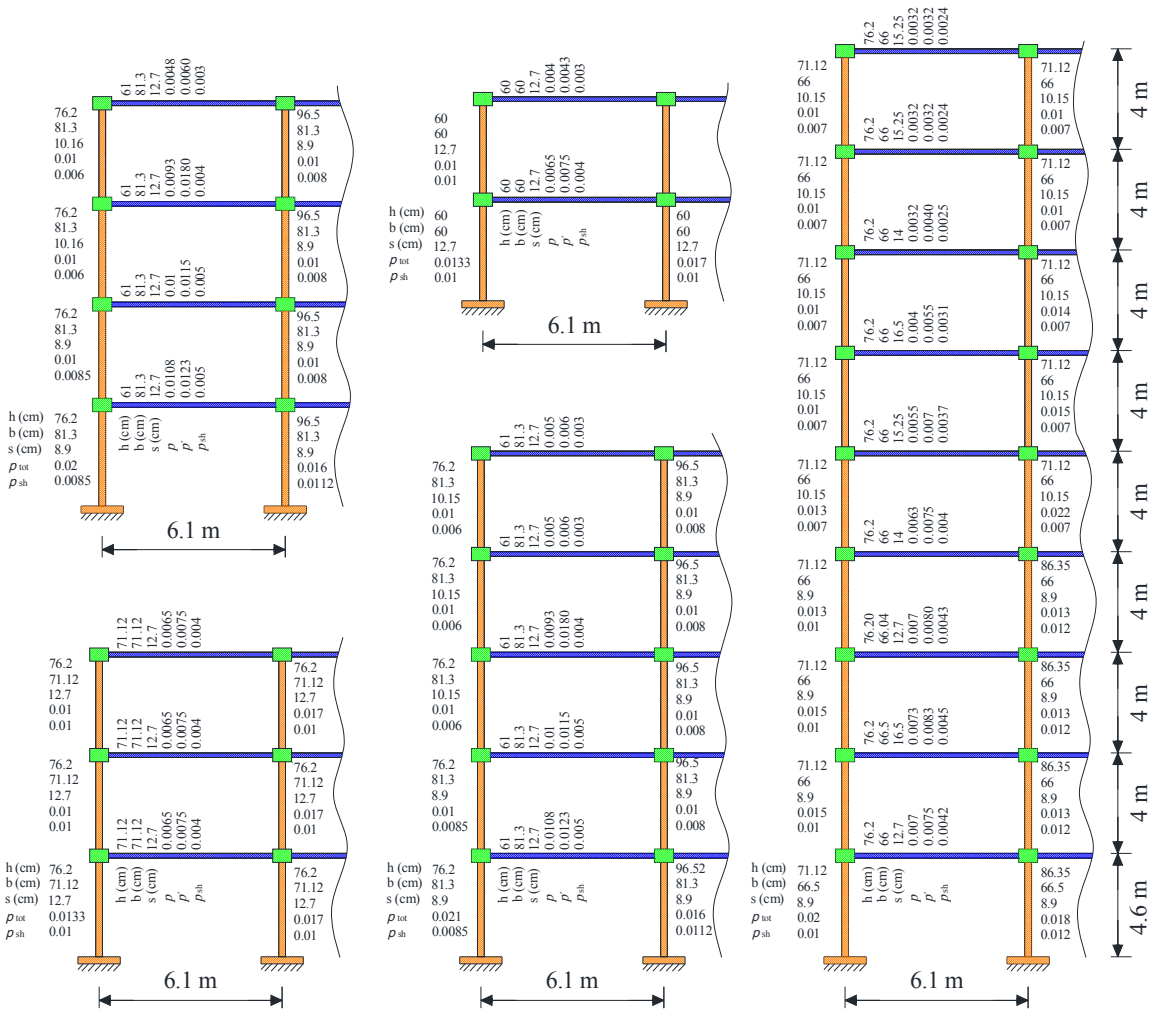


Fig. A-1. Design parameters of the 2-Story, 3-Story, 4-Story, 5-Story, and 9-Story RC frames with 6.1 m bay length.

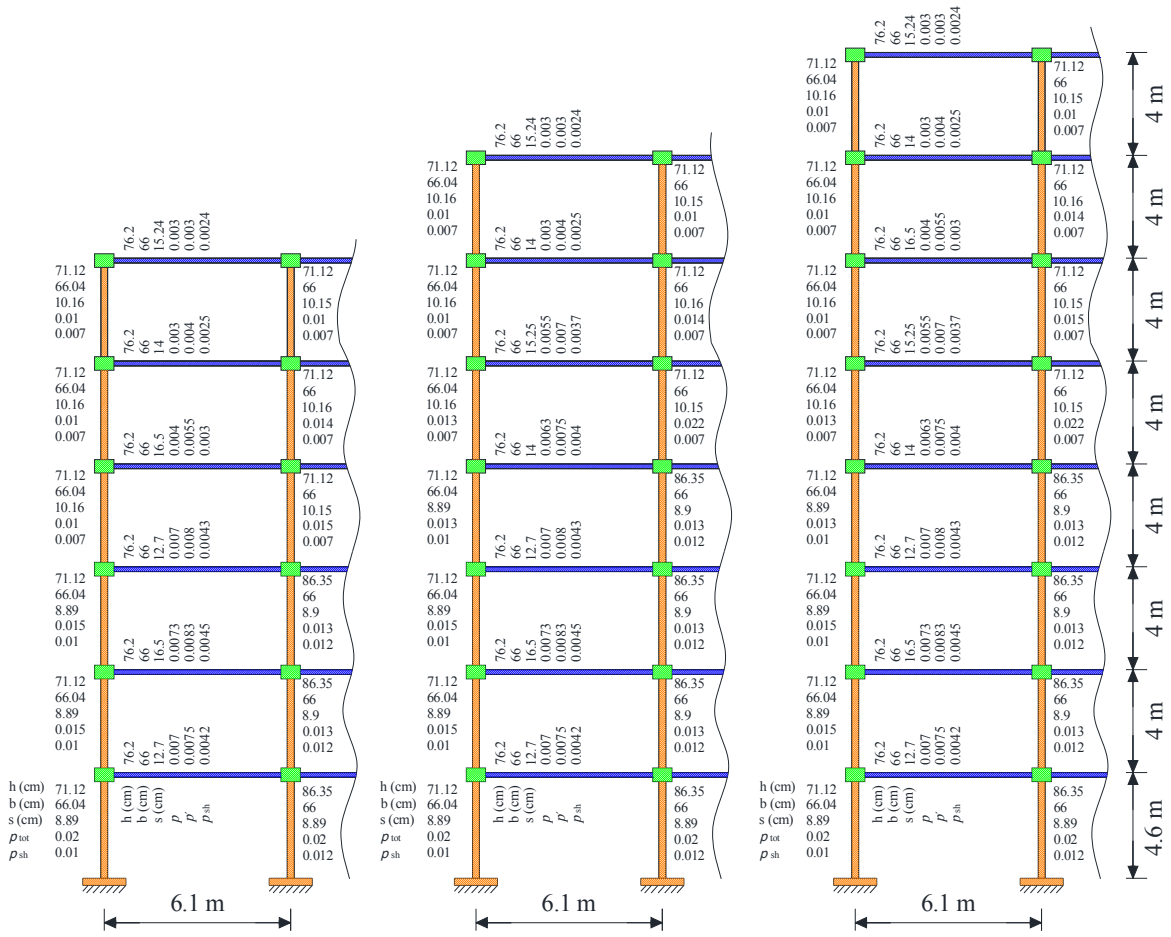


Fig. A-2. Design parameters of the 6-Story, 7-Story, and 8-Story RC frames with 6.1 m bay length.

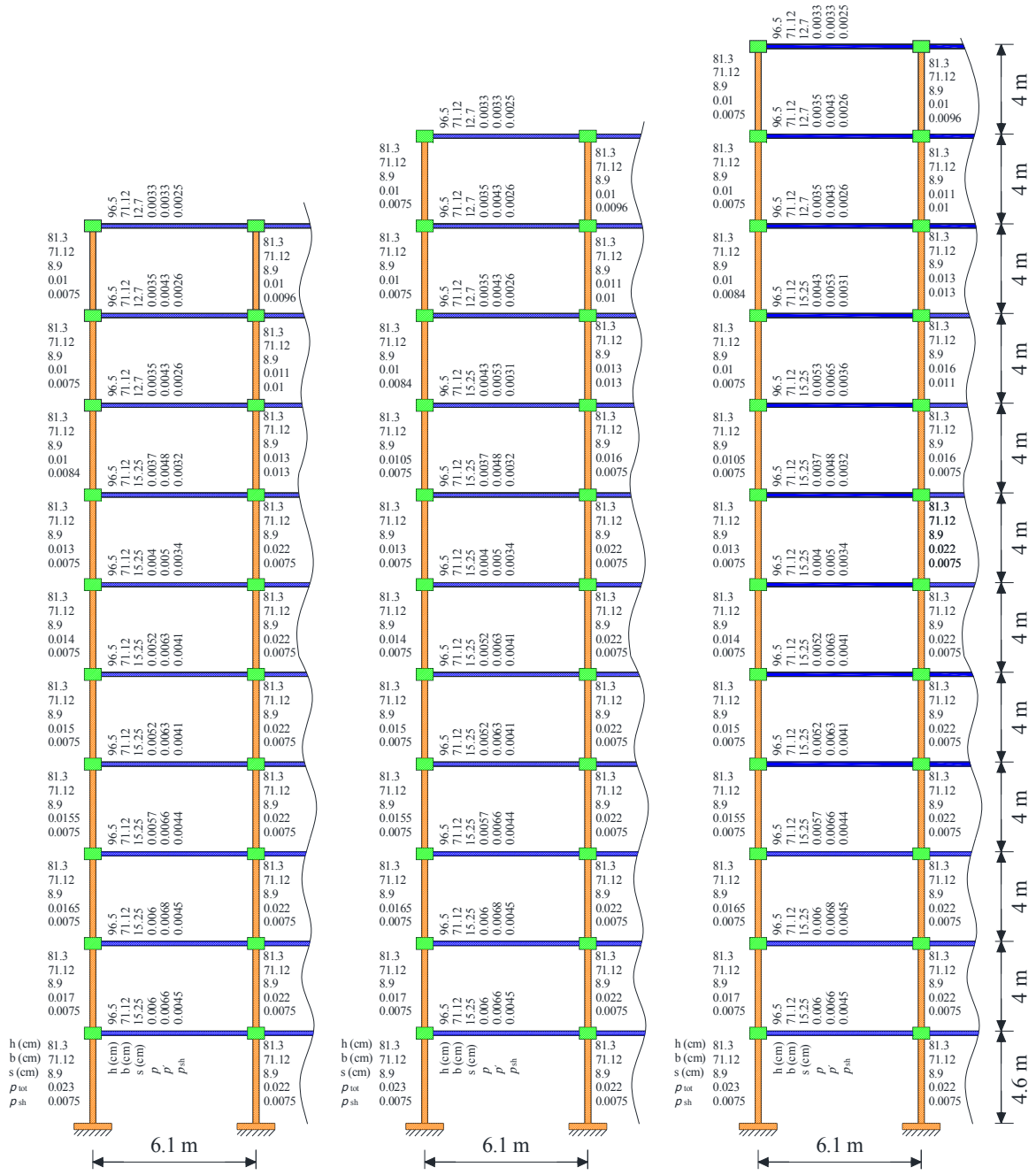


Fig. A-3. Design parameters of the 10-Story, 11-Story, and 12-Story RC frames with 6.1 m bay length.

References

[1] FEMA, P58-1. Seismic performance assessment of buildings Volume 1-Methodology. Applied Technology Council. California: Redwood City 2012.

[2] Kassem, M. M., Nazri, F. M., Farsangi, E. N., Ozturk, B. Improved vulnerability index methodology to quantify seismic risk and loss assessment in reinforced concrete buildings. Journal of Earthquake Engineering, 2021, 1-36.
[3] Feng, D. C., Cao, X. Y., Wang, D., Wu, G. A PDEM-based non-parametric seismic fragility assessment method for RC structures under non-stationary ground motions. Journal of Building Engineering, 2022, 105465.
[4] Feng, D. C., Cao, X. Y., Beer, M. An enhanced PDEM-based framework for reliability analysis of structures considering multiple failure modes and limit states. Probabilistic Engineering Mechanics, 70, 2022, 103367.
[5] Cao, X. Y., Feng, D. C., Li, Y. Assessment of various seismic fragility analysis approaches for structures excited by non-stationary stochastic ground motions. Mechanical Systems and Signal Processing, 186, 2023, 109838.
[6] Scozzese, F., Tubaldi, E., Dall'Asta, A. Assessment of the effectiveness of Multiple-Stripe Analysis by using a stochastic earthquake input model. Bulletin of Earthquake Engineering, 18(7), 2020, 3167-3203.
[7] Asgarkhani, N., Yakhchalian, M., Mohebi, B. Evaluation of approximate methods for estimating residual drift demands in BRBFs. Engineering Structures, 224, 2020, 110849.
[8] Mohebi, B., Yazdanpanah, O., Kazemi, F., Formisano, A. Seismic damage diagnosis in adjacent steel and RC MRFs considering pounding effects through improved wavelet-based damage-sensitive feature. Journal of Building Engineering, 33, 2021, 101847.
[9] Kazemi, F., Mohebi, B., Yakhchalian, M. Evaluation the P-delta effect on collapse capacity of adjacent structures subjected to far-field ground motions. Civil Engineering Journal, 4 (5), 2018, 1066. doi:10.28991/cej-0309156.
[10] Mohebi B., Kazemi F., Asgarkhani N., Ghasemnezhadsani P., and Mohebi A. Performance of Vector-valued Intensity Measures for Estimating Residual Drift of Steel MRFs with Viscous Dampers," International Journal of Structural and Civil Engineering Research, Vol. 11, No. 4, pp. 79-83, 2022. doi: 10.18178/ijscer.11.4.79-83.
[11] Cao, X. Y., Feng, D. C., Wang, Z., Wu, G. Parametric investigation of the assembled bolt-connected buckling-restrained brace and performance evaluation of its application into structural retrofit. Journal of Building Engineering, 48, 2022, 103988.
[12] Asgarkhani, N., Kazemi, F., Jankowski, R. Optimal retrofit strategy using viscous dampers between adjacent RC and SMRFs prone to earthquake-induced pounding. Archives of Civil and Mechanical Engineering, 23(1), 1-26, 2023.
[13] Kazemi, F., Mohebi, B., Yakhchalian, M. Predicting the seismic collapse capacity of adjacent structures prone to pounding. Canadian Journal of Civil Engineering, 47(6), 2020, 663-677.
[14] Kazemi, F., Mohebi, B., Jankowski, R. Predicting the seismic collapse capacity of adjacent SMRFs retrofitted with fluid viscous dampers in pounding condition. Mechanical Systems and Signal Processing, 161, 2021, 107939.
[15] Kazemi, F., Jankowski, R. Enhancing seismic performance of rigid and semi-rigid connections equipped with SMA bolts incorporating nonlinear soil-structure interaction, Engineering Structures, 274, 2023, 114896.
[16] Kazemi, F., Asgarkhani, N., Jankowski, R. Probabilistic assessment of SMRFs with infill masonry walls incorporating nonlinear soil-structure interaction. Bulletin of Earthquake Engineering, 1-32, 2022.
[17] Yazdanpanah, O., Mohebi, B., Kazemi, F., Mansouri, I., Jankowski, R. Development of fragility curves in adjacent steel moment-resisting frames considering pounding effects through improved wavelet-based refined damage-sensitive feature. Mechanical Systems and Signal Processing, 173, 2022, 109038.
[18] Mangalathu, S., Jeon, J. S. Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques. Engineering Structures, 160, 2018, 85-94.
[19] Huang, H., Burton, H. V. Classification of in-plane failure modes for reinforced concrete frames with infills using machine learning. Journal of Building Engineering, 25, 2019, 100767.
[20] Siam, A., Ezzeldin, M., El-Dakhakhni, W. Machine learning algorithms for structural performance classifications and predictions: Application to reinforced masonry shear walls. In Structures (Vol. 22, pp. 252-265). Elsevier, 2019.
[21] Mangalathu, S., Jang, H., Hwang, S. H., Jeon, J. S. Data-driven machine-learning-based seismic failure mode identification of reinforced concrete shear walls. Engineering Structures, 208, 2020, 110331.
[22] Feng, D. C., Wang, W. J., Mangalathu, S., Hu, G., Wu, T. Implementing ensemble learning methods to predict the shear strength of RC deep beams with/without web reinforcements. Engineering Structures, 235, 2021, 111979.

[23] Kiani, J., Camp, C., Pezeshk, S. On the application of machine learning techniques to derive seismic fragility curves. Computers & Structures, 218, 2019, 108-122.
[24] Lagaros, N. D., Tsompanakis, Y., Psarropoulos, P. N., Georgopoulos, E. C. Computationally efficient seismic fragility analysis of geostructures. Computers & Structures, 87(19-20), 2009, 1195-1203.
[25] Zhou, Y., Zhang, Y., Pang, R., Xu, B. Seismic fragility analysis of high concrete faced rockfill dams based on plastic failure with support vector machine. *Soil Dynamics and Earthquake Engineering*, *144*, 2021, 106587.
[26] Liu, Z., Li, S., Guo, A., Li, H. Comprehensive functional resilience assessment methodology for bridge networks using data-driven fragility models. *Soil Dynamics and Earthquake Engineering*, *159*, 2022, 107326.
[27] Yan, Y., Xia, Y., Yang, J., Sun, L. Optimal selection of scalar and vector-valued seismic intensity measures based on Gaussian Process Regression. Soil Dynamics and Earthquake Engineering, *152*, 2022, 106961.
[28] Mitropoulou, C. C., Papadrakakis, M. Developing fragility curves based on neural network IDA predictions. Engineering Structures, 33(12), 2011, 3409-3421.
[29] Giovanis, D. G., Fragiadakis, M., Papadopoulos, V. Epistemic uncertainty assessment using incremental dynamic analysis and neural networks. Bulletin of Earthquake Engineering, 14(2), 2016, 529-547.
[30] Ferrario, E., Pedroni, N., Zio, E., Lopez-Caballero, F. Bootstrapped Artificial Neural Networks for the seismic analysis of structural systems. Structural Safety, 67, 2017, 70-84.
[31] Liu, Z., Zhang, Z. Artificial neural network based method for seismic fragility analysis of steel frames. KSCE Journal of Civil Engineering, 22(2), 2018, 708-717.
[32] Morfidis, K., Kostinakis, K. Comparative evaluation of MFP and RBF neural networks' ability for instant estimation of r/c buildings' seismic damage level. Engineering Structures, 197, 2019, 109436.
[33] Hwang, S. H., Mangalathu, S., Shin, J., Jeon, J. S. Machine learning-based approaches for seismic demand and collapse of ductile reinforced concrete building frames. Journal of Building Engineering, 34, 2021, 101905.
[34] Sainct, R., Feau, C., Martinez, J. M., Garnier, J. Efficient methodology for seismic fragility curves estimation by active learning on Support Vector Machines. Structural Safety, 86, 2020, 101972.
[35] Tang, Q., Dang, J., Cui, Y., Wang, X., Jia, J. Machine learning-based fast seismic risk assessment of building structures. Journal of Earthquake Engineering, 1-22, 2021.
[36] Kazemi, F., Jankowski, R. Machine learning-based prediction of seismic limit-state capacity of steel moment-resisting frames considering soil-structure interaction, Computers & Structures, 274, 2023,106886.
[37] Dehestani, A., Kazemi, F., Abdi, R., Nitka M. Prediction of fracture toughness in fibre-reinforced concrete, mortar, and rocks using various machine learning techniques. Engineering Fracture Mechanics, 2022, 108914.
[38] Kazemi, F., Asgarkhani, N., Jankowski, R. Predicting seismic response of SMRFs founded on different soil types using machine learning techniques, Engineering Structures, 2023, 114953.
[39] Geurts, P., Ernst, D., Wehenkel, L. Extremely randomized trees. Machine learning, 63(1), 3-42.
[40] Breiman, L. Bagging predictors. Machine learning, 24(2), 123-140, 2006.
[41] Todorov, B., Billah, A. M. Machine learning driven seismic performance limit state identification for performance-based seismic design of bridge piers. Engineering Structures, 255, 2022, 113919.
[42] Feng, D. C., Liu, Z. T., Wang, X. D., Chen, Y., Chang, J. Q., Wei, D. F., Jiang, Z. M. Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. Construction and Building Materials, 230, 2020, 117000.
[43] Drucker, H., Burges, C. J., Kaufman, L., Smola, A., Vapnik, V. Support vector regression machines. Advances in neural information processing systems, 9, 1996.
[44] Wolpert, D. H. Stacked generalization. Neural networks, 5(2), 1992, 241-259.
[45] Höskuldsson, A. PLS regression methods. Journal of chemometrics, 2(3), 1988, 211-228.
[46] Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai, Vol. 14, No. 2, pp. 1995, 1137-1145.
[47] ASCE 7-16. Minimum design loads and associated criteria for buildings and other structures. Reston, VA: American Society of Civil Engineers, 2017.
[48] Haselton, C. B., Deierlein, G. G. Assessing seismic collapse safety of modern reinforced concrete frame buildings. PEER report, 8, 2007.
[49] McKenna, F., Fenves, G. L., Filippou, F. C., Scott, M. H. Open system for earthquake engineering simulation (OpenSees). Berkeley, Pacific Earthquake Engineering Research Center, University of California, 2016.

[50] Kazemi, F., Mohebi, B., Yakhchalian, M. Enhancing the seismic performance of adjacent pounding structures using viscous dampers. In The 16th European Conference on Earthquake Engineering (16ECEE), 18-21, June, Thessaloniki, Greece, 2018.
[51] Yakhchalian, M., Asgarkhani, N., Yakhchalian, M. Evaluation of deflection amplification factor for steel buckling restrained braced frames. Journal of Building Engineering, 30, 2020, 101228.
[52] Yakhchalian, M., Yakhchalian, M., Asgarkhani, N. An advanced intensity measure for residual drift assessment of steel BRB frames. Bulletin of Earthquake Engineering, 19(4), 2021, 1931-1955.
[53] Kazemi, F., Asgarkhani, N., Manguri, A., Jankowski, R. Investigating an Optimal Computational Strategy to Retrofit Buildings with Implementing Viscous Dampers. In International Conference on Computational Science (pp. 184-191). Springer, Cham, 2022. https://doi.org/10.1007/978-3-031-08754-7_25.
[54] Ibarra, L. F., Medina, R. A., Krawinkler, H. Hysteretic models that incorporate strength and stiffness deterioration. Earthquake engineering & structural dynamics, 34(12), 2005, 1489-1511.
[55] Altoontash, A. Simulation and Damage Models for Performance Assessment of Reinforced Concrete Beam-Column Joints, Dissertation, Department of Civil and Environmental Engineering, Stanford University, 2004.
[56] Applied Technology Council, United States. Federal Emergency Management Agency (FEMA-P695). Quantification of building seismic performance factors. US Department of Homeland Security, 2009.
[57] MATLAB/Simulink as a Technical Computing Language. Engineering computations and modeling in MATLAB, 2018.
[58] Kitayama, S., Constantinou, M. C. Probabilistic collapse resistance and residual drift assessment of buildings with fluidic self-centering systems. Earthquake Engineering & Structural Dynamics, 45(12), 2016, 1935-1953.
[59] Federal Emergency Management Agency (FEMA P155). Rapid visual screening of buildings for potential seismic hazards: Supporting documentation. Government Printing Office, 2015.
[60] FEMA-356. Prestandard and commentary for the seismic rehabilitation of buildings. Washington, DC: Federal Emergency Management Agency, 2000.
[61] Göçken, M., Özçalıcı, M., Boru, A., Dosdoğru, A. T. Integrating metaheuristics and artificial neural networks for improved stock price prediction. Expert Systems with Applications, 44, 2016, 320-331.
[62] H.C. Thode, Testing for Normality, Marcel Dekker, INC, New York, NY, 2002. ISBN:0-8247-9613-6.
[63] Haghollahi, S., Behnamfar, F. Performance evaluation of special RC moment frames against collapse considering soil–structure interaction. International Journal of Geomechanics, 20(2), 2020, 04019176.
[64] Allahvirdizadeh, R., Khanmohammadi, M., Marefat, M. S. Probabilistic comparative investigation on introduced performance-based seismic design and assessment criteria. Engineering Structures, 151, 2017, 206-220.
[65] United States Geological Survey, https://www.usgs.gov/programs/earthquake-hazards/hazards.
[66] Eads, L. Seismic collapse risk assessment of buildings: effects of intensity measure selection and computational approach. Stanford University, 2013.