



# Machine learning-based seismic response and performance assessment of reinforced concrete buildings

F. Kazemi<sup>1</sup> · N. Asgarkhani<sup>1</sup> · R. Jankowski<sup>1</sup>

Received: 19 November 2022 / Revised: 2 February 2023 / Accepted: 26 February 2023  
© The Author(s) 2023

## Abstract

Complexity and unpredictability nature of earthquakes makes them unique external loads that there is no unique formula used for the prediction of seismic responses. Hence, this research aims to implement the most well-known Machine Learning (ML) methods in Python software to propose a prediction model for seismic response and performance assessment of Reinforced Concrete Moment-Resisting Frames (RC MRFs). To prepare 92,400 data points of training dataset for developing data-driven techniques, Incremental Dynamic Analyses (IDAs) were performed considering 165 RC MRFs with two-, to twelve-Story elevations having the bay lengths of 5.0 m, 6.1 m, and 7.6 m assuming near-fault seismic excitations. Then, important structural features were considered in datasets to train and test the ML-based prediction models, which were improved with innovative techniques. The results show that improved algorithms have higher  $R^2$  values for estimating the Maximum Interstory Drift Ratio ( $IDR_{max}$ ), and two improved algorithms of artificial neural networks and extreme gradient boosting can estimate the Median of IDA curves (M-IDAs) of RC MRFs, which can be used to estimate the seismic limit-state capacity and performance assessment of existing or newly constructed RC buildings. To validate the generality and accuracy of the proposed ML-based prediction model, a five-Story RC building with different input features was used, and the results are promising. Therefore, graphical user interface is introduced as user-friendly tool to help researchers in estimating the seismic limit-state capacity of RC buildings, while reducing the computational cost and analytical efforts.

**Keywords** Machine learning method · Maximum interstory drift ratio · Seismic limit-state capacity · Predicting seismic performance · Seismic probabilistic assessment

## 1 Introduction

The vulnerability of a building can be evaluated either by in-situ technique of data analysis with non-destructive methods, known as structural health monitoring, or numerical analysis of structural models. The main idea of using such methods is to evaluate the building performance in the operating condition. Although in-situ technique can provide a wide range of data, some practical limitations such

as implementing the sensors or actuators and mechanical problems during the time can prevent the performance assessment of structures [1–3]. Therefore, this method can be improved by response prediction methods for buildings subjected to seismic excitations.

Nowadays, the seismic probabilistic assessment of a building needs to perform complicate analysis using precise finite element model, which may need a time-consuming process for evaluating different limit states (e.g., see [4, 5]). Due to the unpredictable nature of ground motions, it is essential to predict the nonlinear structural response during seismic loads to take precautions for reducing the probability of collapse risk. There are some approaches that can be employed to perform nonlinear analysis. The nonlinear static analysis, known also as pushover analysis, can provide information about the base shear versus top floor displacement. While the nonlinear time history analysis uses the pre-recorded earthquakes and performs analysis considering the scale factors defined based on the

✉ F. Kazemi  
farzin.kazemi@pg.edu.pl

N. Asgarkhani  
neda.asgarkhani@pg.edu.pl

R. Jankowski  
jankowr@pg.edu.pl

<sup>1</sup> Faculty of Civil and Environmental Engineering, Gdańsk University of Technology, Ul. Narutowicza 11/12, 80-233 Gdansk, Poland

acceleration spectrum prescribed by design code. Hence, the most accurate approaches of estimating seismic response are conducted by the nonlinear time history analysis and Incremental Dynamic Analysis (IDA) using prior seismic events and finite element methods [6–8]. The prediction of seismic response using these approaches need to prepare complex models and perform time-consuming analysis, while using simplified models (e.g., single-degree of freedom model) are computationally efficient with low performance and behavior compared to the real structures. Therefore, there is a need to introduce a novel Machine Learning (ML)-based method to efficiently and accurately predict the seismic response of RC frames.

Finding the seismic capacity of buildings can help engineers to find a preliminary prediction for the performance levels of the designed building. Kazemi et al. [9, 10] proposed factors for modifying and estimating the collapse capacity of colliding steel Moment-Resisting Frames (MRFs) and colliding Reinforced Concrete (RC) and steel frames [11]. It should be noted that the proposed factors were achieved from complex modeling and analysis; therefore, there is a need to propose a prediction model to avoid such prohibitively complex analysis. Recently, ML algorithms are applied in many civil engineering areas such as failure mode of steel base-plate connection [12], damage identification of bridge [13], damage state of steel frames [14], and RC beams [15]. ML methods are divided into two main parts of supervised and unsupervised algorithms, which the seismic response prediction can be considered as supervised learning using training and testing datasets with the possibility of assuming  $n$ -features for  $n$ -samples. Therefore, in this method, it is possible to take the important features into account [16–18]. Huang et al. [19] proposed a backpropagation neural network to predict the seismic response of structures. Yinfeng et al. [20] used the Support Vector Machine (SVM) algorithm for predicting the nonlinear time history response of structures. Then, Lagaros and Papadrakakis [21] improved neural networks for predicting the nonlinear time history response of a three-dimensional building using six seismic excitations. De Lautour and Omenzetter [22] developed a methodology for estimating the structural responses using pattern recognition of damages. ML algorithms are in the interest of some researchers to use for nonlinear modal analysis [23], predicting seismic responses for achieving fragility curves [24], and predicting maximum displacements of isolated pendulum system [25]. Oh et al. [26] developed a neural network model for predicting the seismic response of buildings based on the correlation of records using 2700 artificial records. Luo and Paal [27] proposed a novel artificial methodology for seismic response prediction of RC structures using 272 RC columns datasets.

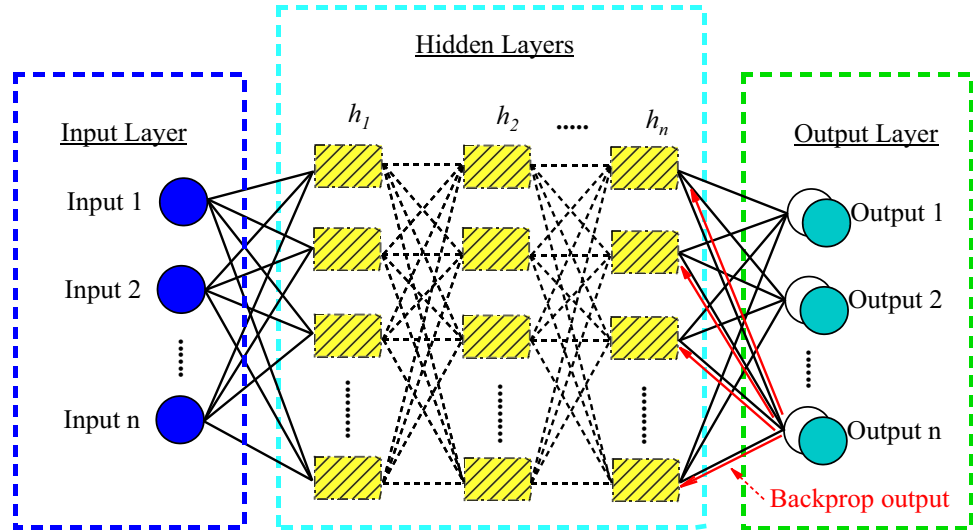
It is confirmed that there is no unique formula for the prediction of Maximum Interstory Drift Ratio ( $IDR_{max}$ ) and Median of IDA curves (M-IDAs) for any type of RC buildings. The purpose of this research is to develop a powerful ML-based tool with employing the innovative data sampling and hyperparameter optimization methods such as fine-tune method, halving search strategy, grid search method, and  $k$ -fold cross-validation. For this purpose, a wide range of data points containing 165 RC MRFs with different length and number of bays were numerically determined to prepare training dataset. Then, the ML-based prediction model can be used for estimating the seismic response and seismic limit-state capacities of RC buildings that can be further applied for a preliminary estimation of  $IDR_{max}$  and M-IDAs of existing and newly constructed buildings. The seismic response prediction results would help designers to find out the behavior of the designed building, and regarding the behavior, it is possible to control the performance of structural elements for postponing the seismic damages. In other words, estimating the  $IDR_{max}$  can be used for predicting the maximum deformation of buildings, and predicting the  $Sa(T_1)$  of M-IDAs can be applied for seismic performance levels assessment. Finally, the results of research were used for introducing an estimation tool based on the developed ML algorithms.

## 2 Structural response prediction model

### 2.1 Artificial neural network

Due to the high ability of Artificial Neural Networks (ANNs) for prediction, they can be trained for different problems, such as positioning site facilities [28], the seismic limit-state performance of bridge piers [29], estimating the fracture toughness of rocks [30], optimizing the consumption of energy [31], estimating the compressive strength of steel fiber-reinforced concrete [32], seismic vulnerability assessment of RC frames [33], and seismic response prediction of structures [34]. ANNs contain three main parts of the input layer, hidden layers, and output layer, which are connected by some nonlinear function with the adjusted weight. The weight of each neuron can increase or decrease the strength of connection for purpose of minimizing the loss function or error (i.e., the difference between the predicted and actual values). Backward and forward propagation methods can be used for recalculating the weights of each neuron in the previous iteration to minimize the error; then, the process can be repeated with new adjusted weights to achieve a reliable model. The backward propagation method is presented in Fig. 1. In this study,  $IDR_{max}$  and  $Sa(T_1)$  were defined as targets for backward and forward propagation ANNs. Moreover, Multi-layer Perceptron Regressor (MLPReg)

**Fig. 1** Functioning of the back-propagation method in ANNs

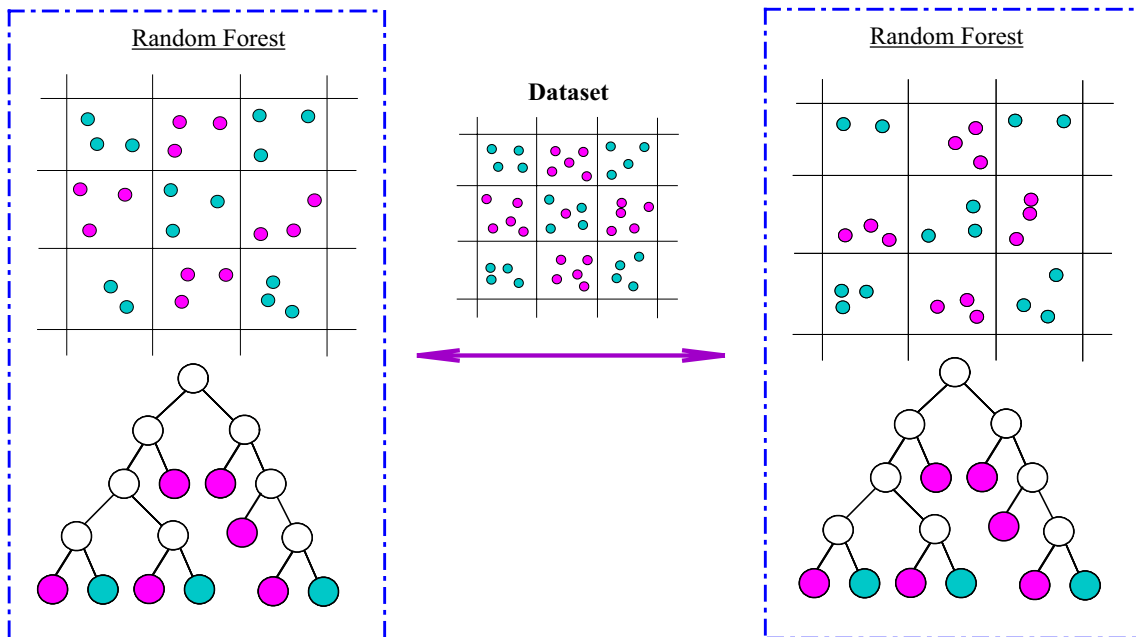


considers the linear function to predict seismic responses of RC structures.

**2.2 Random decision forest**

Random decision Forest (RF) can be employed for both regression and classification problems. RF algorithm uses an ensemble multiple bagging models parallel to a different train subset from train data, and achieves the final result based on the majority voting. Figure 2 presents the RF algorithm with the bagging principle.

Although the RF algorithm can be classified as a decision tree, the RF method considers subsets of data to solve the overfitting problem while selecting random observations instead of a set of formulas [35]. It should be noted that different parameters were selected by trial and error to find the lower bias and higher variance values to overcome the overfitting problem and achieve an optimized prediction model. Moreover, different types of RF algorithms known as An Extra-Trees Regressor (ETReg), which randomly selects decision tree to fit input data, and An Extremely Randomized Tree Regressor (ERTReg), which uses the random tree selection to improve the calculations



**Fig. 2** RF algorithm with the bagging principle

speed [36], and Bagging Regressor (BReg), which aggregate individual predictions were used to find the best prediction model [37].

### 2.3 Boosting algorithms

Boosting principle is another way of using RF methods. In this principle, weak learners combine in sequential order to create a strong model with higher accuracy of prediction. Adaptive Boosting (AdaBoost) algorithm combines strong base learners such as decision trees with a single split to weight the data points for improving the accuracy of estimation [38]. Gradient Boosting Machine (GBM) comes from the idea of improving the weak learners to enhance their final results by minimizing the loss function. Moreover, Histogram-based Gradient Boosting Regression (HistGBR) considers the quantization method for splitting the features for prediction with a higher speed compared to GBM. To control the accuracy of the results, the following formula can be used considering the initial probability equal to 0.5, and in each step, the value can be compared with the previous step to find an optimized model.

$$\text{Value} = \frac{\sum_{i=1}^n \text{Residual}_i}{\sum_{i=1}^n (\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i))}, \quad (1)$$

$$\text{Residual}_i = (\text{Actual} (\text{IDR}_{\max i}) - \text{Predicted} (\text{IDR}_{\max i})), \quad (2)$$

$$\text{Value} = \frac{\sum_{i=1}^n \text{Residual}_i}{\sum_{i=1}^n (\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)) + \lambda}. \quad (3)$$

Extreme Gradient Boosting (XGBoost) is an improved algorithm of GBM with a regularize factor,  $\lambda$ , to reduce the effectiveness of small leaves [39, 40]. In this study, a fine-tune XGBoost model was used to change the trees

number and parameters to find the best target based on the following formula:

### 2.4 Support vector machine

Support Vector Machine (SVM) is selected as a decision boundary method with the capability of using hyperplane based on the marginal distances for two-dimensional and three-dimensional spaces [41]. In addition, Nu-Support Vector Regression (NuSVR), which considers the  $\nu$  parameter as the controlling number of vectors [42], and Linear Support Vector Regression (LSVR), which considers functions for loss and penalties [43], were assumed to find a suitable model for estimating  $\text{IDR}_{\max}$  and  $\text{Sa}(T_1)$ . To enhance the performance of ML methods during the training, and reduce the risk of losing the important datasets, the  $k$ -fold cross-validation was employed. Figure 3 presents the  $k$ -fold cross-validation methodology, in which, training and testing datasets are 70–80% and 30–20% of total data points, respectively [44]. It is worth mentioning that the  $k$ -fold cross-validation with different  $k$  was employed for assumed ML algorithms to find the suitable  $k$  with higher performance.

### 2.5 Regressors models

Some important regression algorithms can be used for  $\text{IDR}_{\max}$  prediction, which is a supervised regression model while not included in the abovementioned category. For

example, these models are not using the hidden layers ability (i.e., ANNs) or boosting methods (i.e., XGBoost); therefore, this subsection is defined to include the ML algorithms used in this research with different ability of predictions.

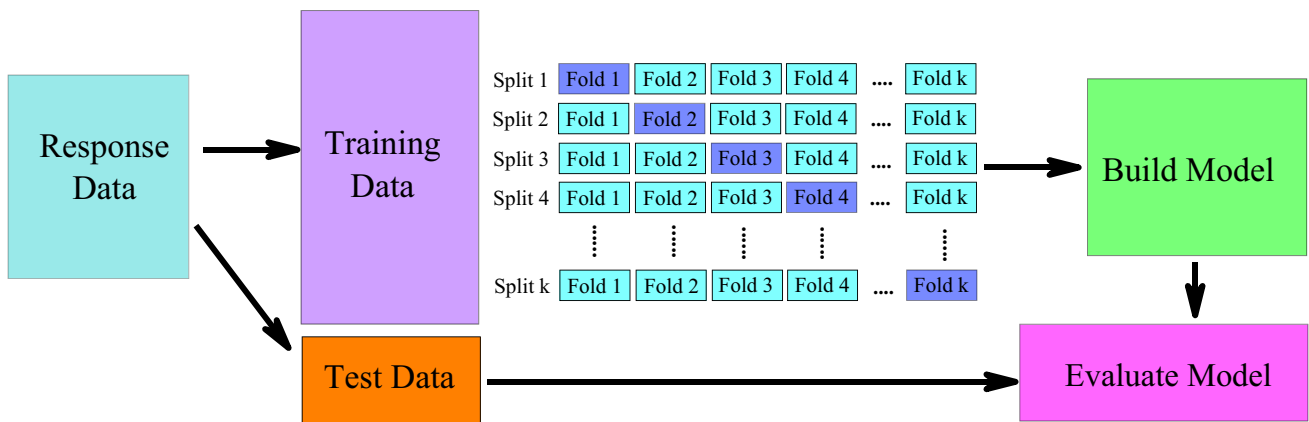


Fig. 3 Architecture of  $k$ -fold cross-validation

Response Prediction in Voting Regressor (VReg) is based on the average of the individual results, while K-Nearest neighbor Regression (KNN) assume linear estimation on the mean of data points. On the other hand, Gaussian Process Regression (GPR) renormalizes the targets to find a zero mean for the maximum log marginal of data points. Linear Regression (LReg) considers a linear estimation model to minimize a target of residual sum defined as squares of predicted and actual values. In addition, Gamma Regressor (GReg) uses the strategy of combining data points with an inverse function and their logarithmic unit deviance [45]. The algorithm that uses the strength of estimators for finding the final estimator to solve the prediction model is known as Stacking Regressor (SReg) (see more detail [46]). Partial Least Squares Regression (PLSReg) is another regression model that has the ability to assume maximum multidimensional direction for data points to achieve fundamental relations between inputs and outputs [47]. Since Python libraries provide a great possibility for developing the ML algorithms as well as the free access of this software, the Python software as a general-purpose programming language is selected for implementing ML methods. Therefore, all assumed ML algorithms were developed in Python software and different resampling strategies, such as fine-tune method, halving search strategy, grid search method, and  $k$ -fold cross-validation were used to improve them as a prediction model.

### 3 Modeling process

To train the ML algorithms, eleven types of RC buildings including two to twelve-floor elevations (i.e., 2-, to 12-Story buildings) having three bay length types (i.e., 5 m, 6.1 m, and 7.6 m) with the plan presented in Fig. 4 were assumed. All buildings modeled in ETABS software based on the assumption of soil type D, acceleration parameters of  $SD_1 = 0.6$  g and  $SD_s = 1.0$  g for the construction site of high seismic, and design parameters of  $R = 8$ ,  $C_d = 5.5$ , and

$\Omega = 3$  in accordance with ASCE 7-16 [48]. It is noteworthy that the acceleration parameters of the construction site were achieved based on the USGS website [49]. In addition, a floor dead load of  $8.4$  kN/m<sup>2</sup> and a floor live load of  $2.4$  kN/m<sup>2</sup> were applied to all floor levels of buildings. To design structural elements, the concrete compressive strength of  $34.5$  MPa (i.e., 5 ksi, see Table 6–2 in reference [50]) was used [51]. Details of structural elements of RC frames assuming the bay length of 6.1 m were presented in Figs. 5, 6 and 7. To perform collapse analysis, all buildings were modeled as two-dimensional RC frames in Opensees [52] assuming the leaning column for those gravity columns not included in models to consider the P-delta effects [53–56]. In addition, the two-dimensional frames were modeled and verified with their corresponding buildings considering modeling procedures used by Haselton and Deierlein [50] and Kazemi et al. [9–11, 57, 58]. According to these procedures, plastic hinge models for simulating seismic collapse presented in Fig. 4 were developed by Ibarra et al. [59] and Altoontash [60]. It should be noted that for considering the real condition of RC buildings, all panel zones were modeled, and concentrate plastic hinge models were used in the ends of structural elements with possibility of achieving seismic collapse (for more detail on modeling see [50]).

To train the ML algorithms, 165 RC MRFs were assumed to have one-, two-, three-, four-, and five-bays, and 2-, to 12-Story elevations having the bay lengths of 5 m, 6.1 m, and 7.6 m. To assess  $IDR_{max}$  in different intensity measures and seismic limit-state curves of all 165 RC MRFs, IDAs were performed based on the spectral acceleration in the period of the structure,  $Sa(T_1)$ , as intensity measure, and  $IDR_{max}$  as engineering demand parameter, considering near-fault Pulse-like (PL), and No-Pulse (NP) records introduced by FEMA-P695 [61]. To perform IDAs, an algorithm was developed to implement the hunt and fill methodology using both Opensees [52] and MATLAB [62] software to reduce the time of analysis. It is worth mentioning that the programming code was developed in MATLAB [62] to control the

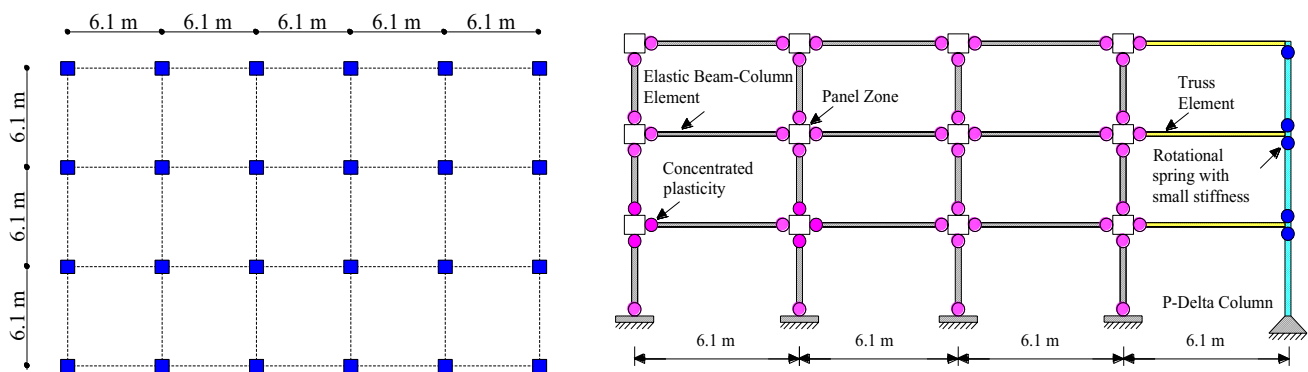


Fig. 4 RC MRFs plan and concentrated plasticity approach employed to model buildings

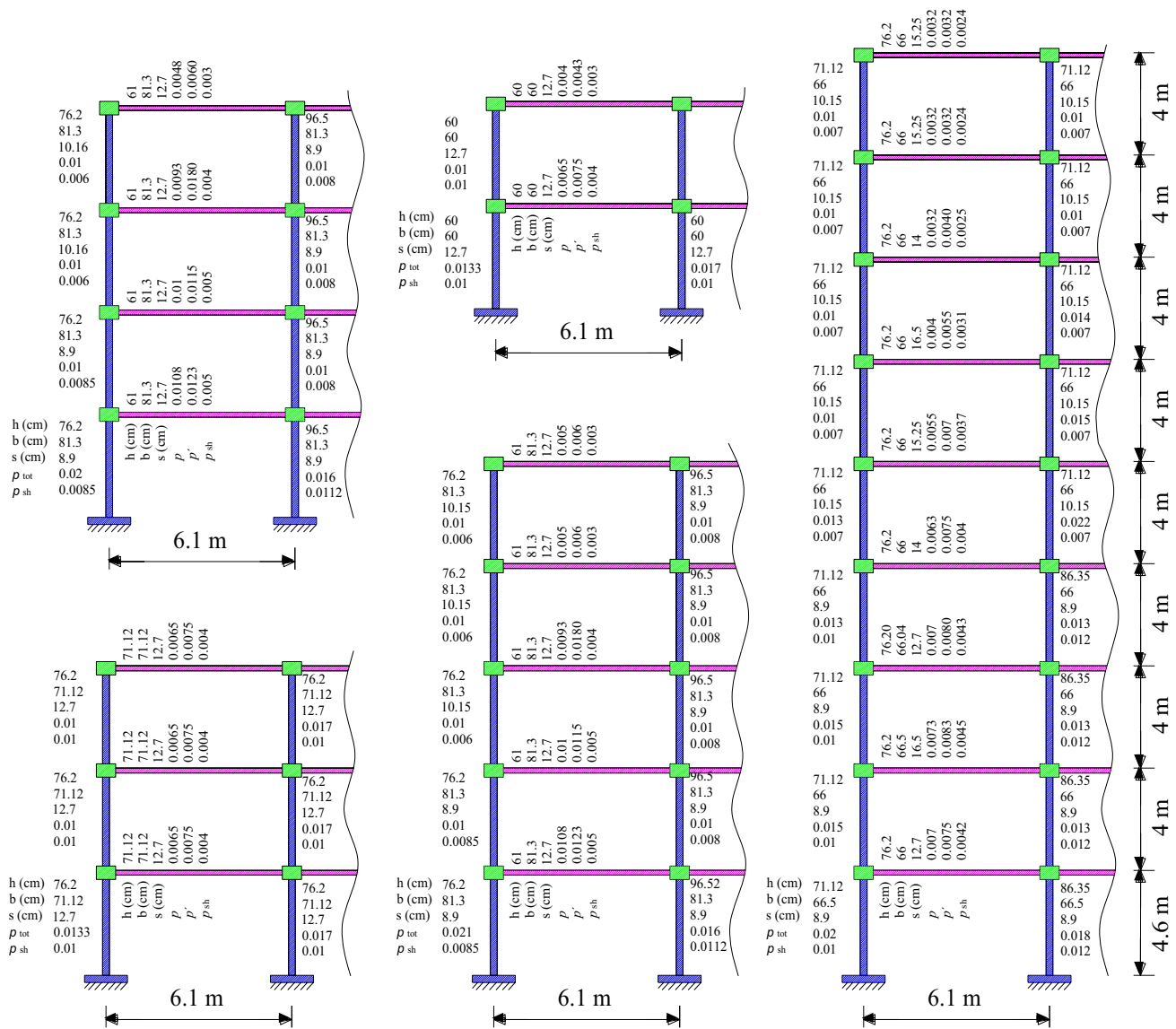


Fig. 5 Structural documentation of the 2-Story, 3-Story, 4-Story, 5-Story, and 9-Story RC MRFs

entire analysis procedure; and in addition, to post-process the results of the analysis. Figure 8 presents the IDA curves of the 2-Story, 4-Story, 8-Story, and 12-Story RC frames having three bays with 6.1 m length including NP records. It should be noted that there is no restriction on the increasing steps of the intensity measure selection in this study; therefore, the results are distributed with different ranges of the  $Sa(T_1)$ .

The training datasets were prepared with important features of weight, aspect ratio, reinforcement ratio for beams and columns, story number, bay length and the total height of RC frames,  $Sa(T_1)$ , the direction and RSN number of record, fundamental period ( $T_1$ ), and  $IDR_{max}$  in each step of the analysis, which achieved based on the trial and error. In addition, for seismic response prediction models, the

$IDR_{max}$  of selected RC frames was considered as a target in the test dataset, and for seismic limit-state capacity prediction models, the  $Sa(T_1)$  of M-IDAs of selected RC frames were considered as a target of prediction in the testing dataset. Therefore, two main training datasets were considered to train and test the prediction models. In addition, 92,400 data points were considered in the training dataset that were achieved by performing IDAs.

### 4 Analytical procedure

The main purpose of this study is to train ML algorithms for accurate prediction of the  $IDR_{max}$  and the seismic limit-state capacity of RC frames using M-IDAs (e.g., presented

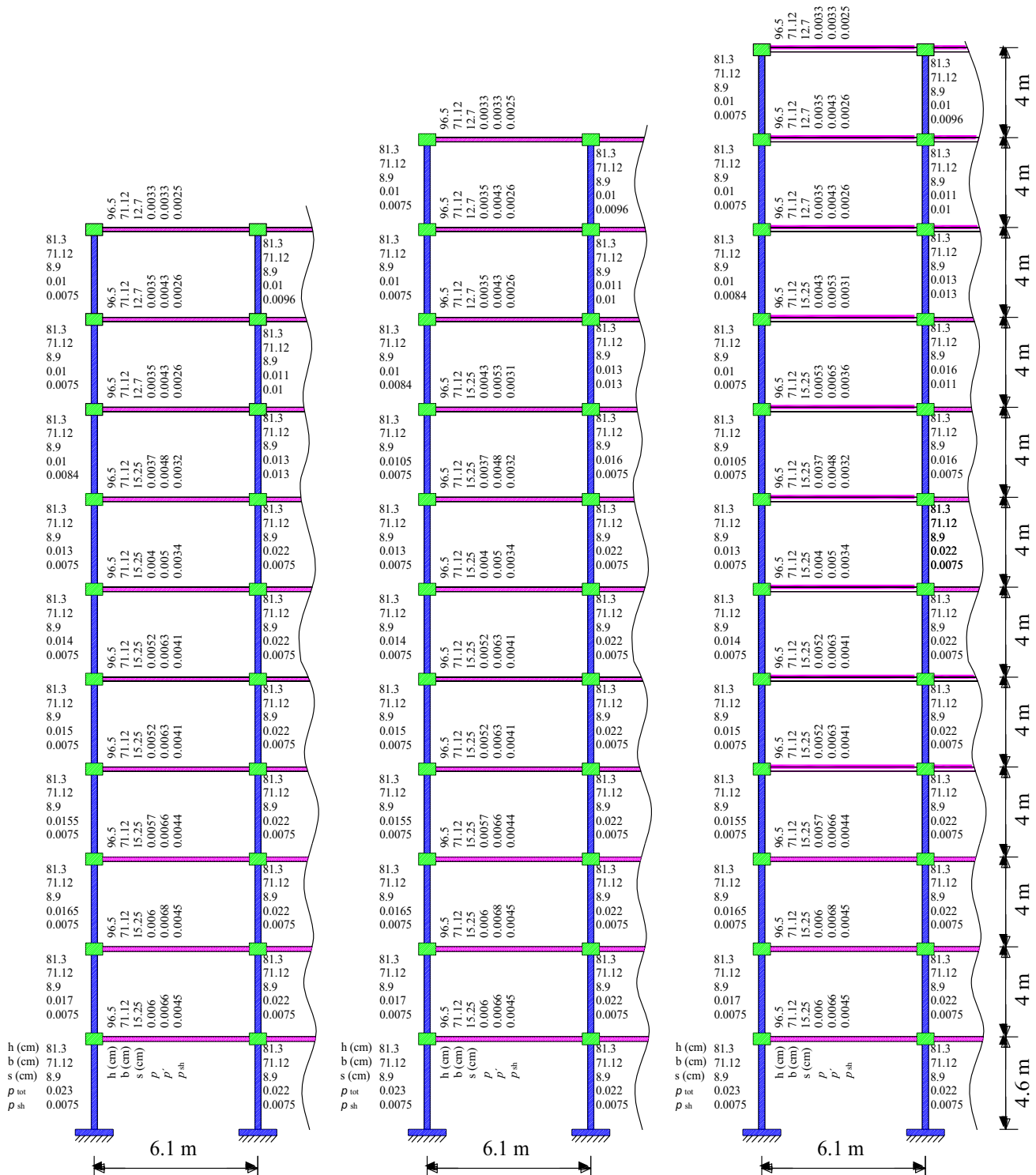


Fig. 6 Structural documentation of the 10-Story, 11-Story, and 12-Story RC MRFs

in pink color in Fig. 8). M-IDAs can be used to estimate the seismic performance levels of the structures assuming a different threshold of  $IDR_{max}$  introduced by seismic provisions. Therefore, the analytical procedure presented in Fig. 9 depicts four main parts used for preparing prediction

models. The first part in the blue color is the modeling and validation of RC MRFs using ETABS and Opensees [52] softwares (see Sect. 3). The green part, explains the preparation of training and testing datasets based on the  $IDR_{max}$  and M-IDAs as targets of prediction. In the red section,

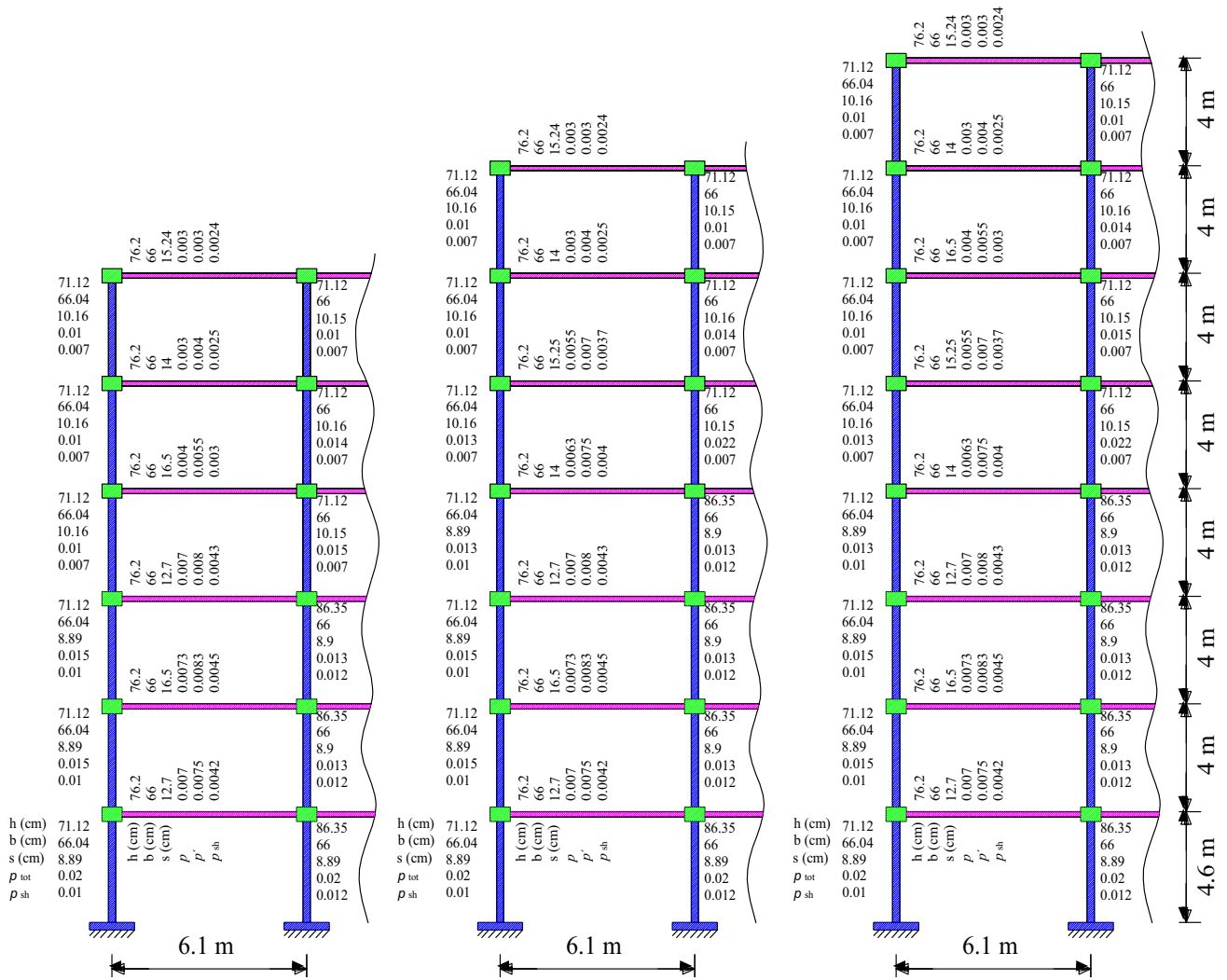


Fig. 7 Structural documentation of the 6-Story, 7-Story, and 8-Story RC MRFs

ML algorithms were implemented in Python software and improved based on some innovative methodologies for the prediction of the two aforementioned targets. After validation of predicting models, some important ML algorithms were selected for the violet part, which shows the second validation of prediction models for a new RC building to show the capability of the proposed ML-based model.

### 4.1 Data selection method

Although many features can influence the response prediction of structures, introducing all these features can reduce the speed of calculations while increase the overfitting possibility in the algorithms. Therefore, it is necessary to provide the important features while the prediction accuracy remains unchanged during the validations. To do this, different feature selection methods such as filter and wrapper methods, which contains the more suitable methods of forward feature

selection, backward feature elimination, and exhaustive feature selection, were used to achieve the importance of input features. Figure 10 presents the relative importance of seven features with higher scores achieved by trial and error using the aforementioned methods. Other features were remove since their relative importance were less than these feature. For estimating the M-IDA curve, three main features of the number of bays, fundamental period of the frame, and  $IDR_{max}$  have more scores compared to other features. On the other side, for predicting  $IDR_{max}$  as a target, five features of number of stories, weight, fundamental period of the frame, number of bays, and  $Sa(T_1)$  have scored more than 10%. According to Fig. 10, these seven features were selected in the training and testing datasets for prediction models.

It is noteworthy that to enhance the ability of the methods, the feature selection approaches were used simultaneously with embedded method to reduce the effects of those data points with low effects on the predictions of selected target.



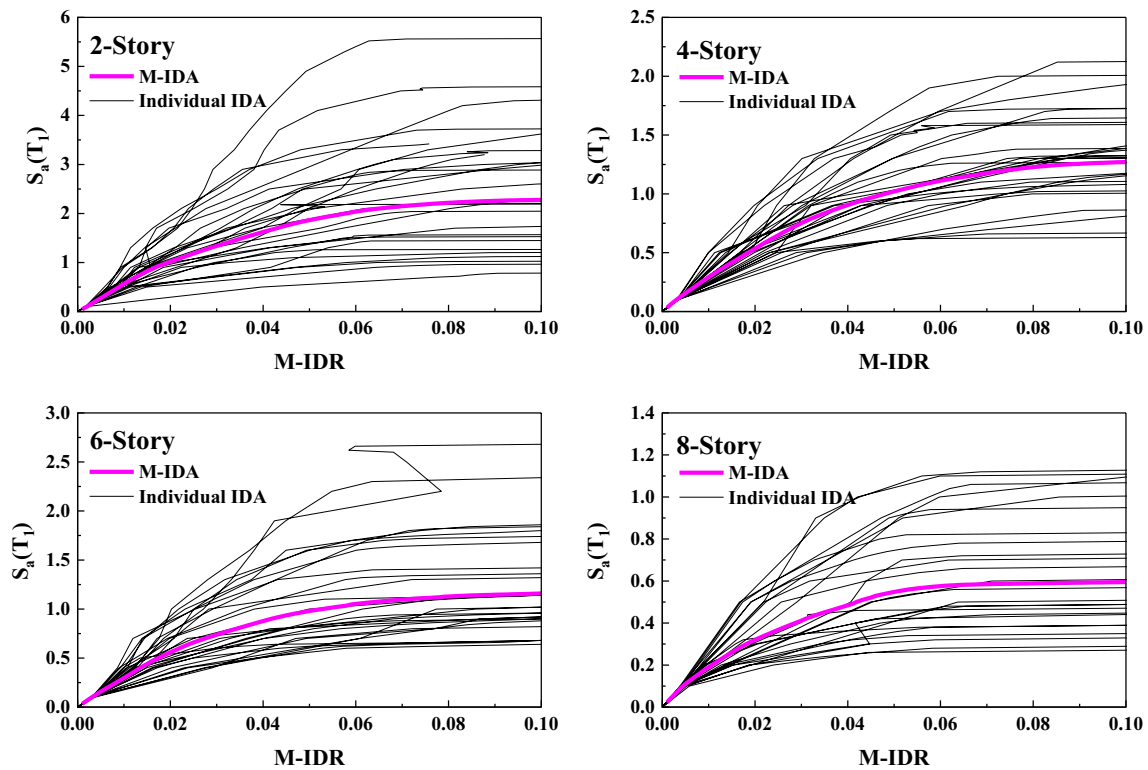


Fig. 8 IDA curves of the RC frames having three bays with 6.1 m including NP records

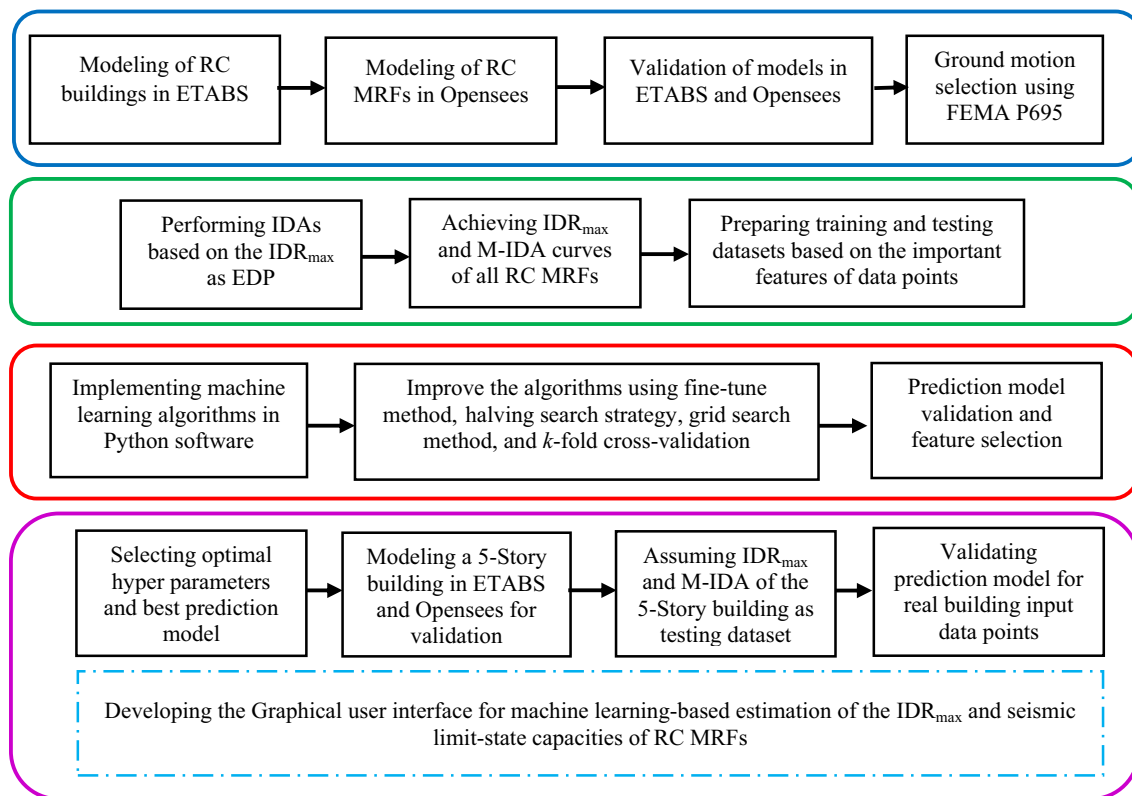


Fig. 9 Analytical procedures are used to achieve a high accuracy model

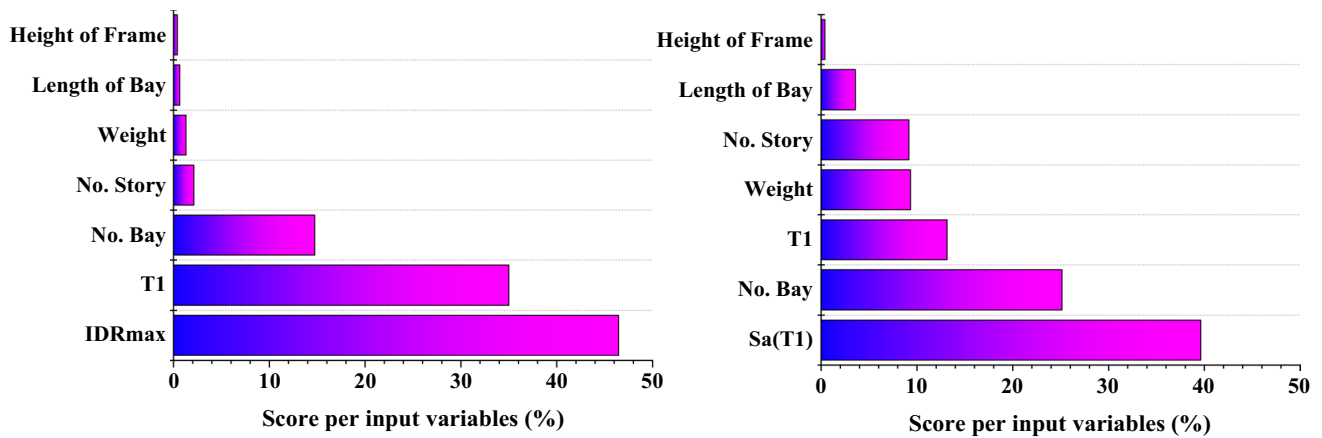


Fig. 10 Relative importance achieved by trial and error for predicting M-IDA curve (left) and IDR<sub>max</sub> (right) as targets of models

In other words, the developed embedded method reduces the number of data points for reasonable computational cost while increases the capability of ML algorithms and prevents the overfitting problem, which is the most important issue in the performance of models. Therefore, all ML methods

improved based on the developed embedded method in purpose of increasing their ability.

To compare the reliability and capability of the aforementioned ML algorithms, the statistical metrics presented in Table 1 were used. The coefficient of determination, R<sup>2</sup>, is widely used for presenting the accuracy of prediction and

Table 1 Statistical metrics used for evaluating the ML models

| Abbreviation    | Formula   | Description  |
|-----------------|---|--|
| R <sup>2</sup>  | $R^2 = 1 - \frac{\sum_{i=1}^n (\text{Actual}_i - \text{Predicted}_i)^2}{\sum_{i=1}^n (\text{Actual}_i - \text{Actual}_{\text{avg}})^2}$ | R squared  |
| MSE             | $MSE = \frac{1}{n} \sum_{i=1}^n (\text{Actual}_i - \text{Predicted}_i)^2$   | Mean squared error   |
| RMSE            | $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{Actual}_i - \text{Predicted}_i)^2}$   | Root mean squared error  |
| MAE             | $MAE = \frac{1}{n} \sum_{i=1}^n  \text{Actual}_i - \text{Predicted}_i $   | Mean absolute error  |
| MARE            | $MARE = \frac{1}{n} \sum_{i=1}^n \left  \frac{\text{Actual}_i - \text{Predicted}_i}{\text{Actual}_i} \right $                           | Mean absolute relative error   |
| MSRE            | $MSRE = \frac{1}{n} \sum_{i=1}^n \left  \frac{\text{Actual}_i - \text{Predicted}_i}{\text{Actual}_i} \right ^2$                         | Mean square relative error   |
| RMSRE           | $RMSRE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{\text{Actual}_i - \text{Predicted}_i}{\text{Actual}_i} \right)^2}$                 | Root mean squared relative error   |
| RRMSE           | $RRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\text{Actual}_i - \text{Predicted}_i)^2}}{\sum_{i=1}^n \text{Actual}_i} \times 100$      | Relative root mean square error  |
| MBE             | $MBE = \frac{1}{n} \sum_{i=1}^n (\text{Actual}_i - \text{Predicted}_i)$   | Mean bias error  |
| erMAX           | $erMAX = \max \left( \left  \frac{\text{Actual}_i - \text{Predicted}_i}{\text{Actual}_i} \right  \right)$                               | Maximum absolute relative error  |
| SD              | $SD = \sqrt{\frac{\sum (X_i - \text{Arithmetic mean})^2}{\text{total number}}}$   | Standard deviation of the difference between actual and predicted values |
| t-stat          | $t\text{-stat} = \sqrt{\frac{(n-1)MBE^2}{RMSE^2 - MBE^2}}$  | t-Statistic  |
| U <sub>95</sub> | $U_{95} = \sqrt{1.96(SD^2 + RMSE^2)}$   | Uncertainty at 95%   |

can take values between 0.0 and 1.0 (or 0.0% and 100%) to show the spreads of predicted and actual data points from the  $x=y$  line. Other metrics compare the actual and predicted values to show the capability of models for minimizing the error, which is the difference between the actual and predicted values.

Twenty ML algorithms were implemented in Python software and used as a prediction model. A sensitivity analysis was performed using the 3-Story RC frame with three bays having bay lengths of 5.0 m subjected to PL records for both models of prediction based on the  $IDR_{max}$  and  $Sa(T_1)$  as a target. Table 2 shows the comparison of statistical metrics for the performance evaluation of ML algorithms for predicting  $IDR_{max}$ . It can be seen that most ML algorithms achieved higher values of  $R^2$ , which shows the accuracy of these algorithms. In the  $IDR_{max}$  as target of testing dataset, eight methods of PLSReg, SReg, VReg, LReg, GReg, MLPReg, SVM, and LSVR had  $R^2$  values of 0.384, 0.386, 0.585, 0.350, 0.160, 0.205, 0.259, and 0.232, respectively. Although their accuracy of prediction in the training dataset was higher than approximately 90%, their performance in the testing dataset is lower than other algorithms and cannot be considered as reliable models. In addition, In the  $Sa(T_1)$  as target of testing dataset, five algorithms of LReg, PLSReg, LSVR, SReg, and GReg had  $R^2$  values of 0.775, 0.774, 0.743, 0.614, and 0.313, respectively. Therefore, these algorithms can be considered as

not reliable models that cannot achieve  $R^2$  values higher than 0.77. Comparing the metrics can provide a good information about the capability of the models and their power for estimating the targets. These tables also can be used for selecting the best ML methods. To better compare the metrics, the score marker were used, which provides the number from 1 to 20 for ranking the ML methods for each of the metrics. Then, in each ML methods, the scores of each metrics were determined to compare the capability of them. According to results of Table 2, the BReg, HistGBR, ETReg, RF, ERTReg, GBM, and XGBoost methods achieved scores of 49, 49, 80, 82, 83, 86, and 98, respectively, which are introduced as best methods. Moreover, the methods of PLSReg, LReg, NuSVR, LSVR, MLPReg, GReg, and SVM had the scores of 175, 176, 190, 199, 212, 219, and 243, respectively, in the end of ranking list.

According to results of Table 3, the ANNs, HistGBR, XGBoost, RF, NuSVR, BReg, and ETReg methods achieved scores of 49, 49, 66, 73, 81, 86, and 93, respectively, which are introduced as best models, while the methods of VReg, PLSReg, LReg, LSVR, SReg, and GReg with scores of 190, 215, 222, 236, 244, and 250, respectively, are introduced as weak prediction models. The statistical indicators used for calculating the error of methods depend on the actual and predicted values; therefore, the higher value of the error shows the dispersion of the predicted values. Although the SVM method had

**Table 2** Comparison of statistical metrics assuming the 3-Story RC frame with three bays having bay lengths of 5.0 m as test data for predicting  $IDR_{max}$

| Model       | $R^2$ | MSE   | RMSE | MAE  | MARE | MSRE  | RMSRE | RRMSE | MBE   | erMAX | SD   | $t$ -stat | $U_{95}$ |
|-------------|-------|-------|------|------|------|-------|-------|-------|-------|-------|------|-----------|----------|
| XGBoost     | 0.818 | 2.58  | 1.61 | 1.24 | 0.52 | 0.86  | 0.93  | 0.23  | 0.20  | 2.72  | 0.01 | 0.41      | 0.04     |
| RF          | 0.901 | 3.42  | 1.85 | 1.35 | 0.21 | 0.08  | 0.28  | 0.26  | -0.12 | 0.73  | 0.01 | 0.21      | 0.04     |
| BReg        | 0.898 | 2.06  | 1.44 | 1.04 | 0.16 | 0.04  | 0.21  | 0.23  | 0.50  | 0.52  | 0.01 | 1.24      | 0.03     |
| ETReg       | 0.890 | 2.25  | 1.50 | 1.10 | 0.18 | 0.05  | 0.23  | 0.26  | 0.58  | 0.60  | 0.01 | 1.38      | 0.04     |
| GBM         | 0.818 | 3.03  | 1.74 | 1.34 | 0.28 | 0.28  | 0.52  | 0.21  | 0.08  | 2.28  | 0.01 | 0.14      | 0.04     |
| HistGBR     | 0.888 | 2.28  | 1.51 | 1.07 | 0.15 | 0.04  | 0.20  | 0.20  | 0.27  | 0.48  | 0.01 | 0.60      | 0.03     |
| HistGBR(FN) | 0.830 | 2.76  | 1.66 | 1.24 | 0.48 | 0.63  | 0.79  | 0.36  | 1.34  | 3.43  | 0.01 | 4.47      | 0.04     |
| AdaBoost    | 0.864 | 5.0   | 2.24 | 1.74 | 0.47 | 0.39  | 0.62  | 0.38  | 0.05  | 1.51  | 0.02 | 0.07      | 0.06     |
| KNR         | 0.795 | 3.17  | 1.78 | 1.17 | 0.16 | 0.05  | 0.21  | 0.24  | 0.68  | 0.57  | 0.01 | 1.37      | 0.04     |
| PLSReg      | 0.384 | 6.61  | 2.57 | 2.07 | 1.11 | 5.09  | 2.26  | 0.35  | 0.07  | 6.72  | 0.02 | 0.09      | 0.06     |
| SReg        | 0.386 | 6.60  | 2.57 | 1.89 | 0.33 | 0.23  | 0.48  | 0.38  | 0.10  | 1.47  | 0.02 | 0.13      | 0.06     |
| VReg        | 0.585 | 3.14  | 1.77 | 1.38 | 0.52 | 0.86  | 0.93  | 0.23  | 0.20  | 2.72  | 0.01 | 0.37      | 0.04     |
| LReg        | 0.350 | 6.75  | 2.60 | 2.05 | 0.93 | 3.32  | 1.82  | 0.32  | 0.08  | 5.42  | 0.02 | 0.10      | 0.06     |
| GReg        | 0.160 | 7.67  | 2.77 | 2.36 | 1.98 | 19.36 | 4.40  | 0.50  | -0.21 | 13.19 | 0.02 | 0.25      | 0.07     |
| MLPReg      | 0.205 | 8.00  | 2.83 | 2.22 | 0.73 | 4.70  | 2.17  | 0.37  | -0.78 | 10.49 | 0.01 | 0.96      | 0.06     |
| SVM         | 0.259 | 11.38 | 3.37 | 2.90 | 3.43 | 61.48 | 7.84  | 0.58  | 0.58  | 23.54 | 0.03 | 0.58      | 0.09     |
| NuSVR       | 0.800 | 5.37  | 2.32 | 1.90 | 0.78 | 2.21  | 1.49  | 0.32  | 0.64  | 5.95  | 0.02 | 0.96      | 0.06     |
| LSVR        | 0.232 | 9.12  | 3.02 | 2.20 | 0.56 | 0.80  | 0.89  | 0.39  | 0.30  | 2.93  | 0.02 | 0.33      | 0.07     |
| ERTReg      | 0.894 | 2.54  | 1.59 | 1.18 | 0.15 | 0.04  | 0.21  | 0.23  | 0.53  | 0.53  | 0.01 | 1.16      | 0.04     |
| ANNs        | 0.863 | 3.22  | 1.79 | 1.35 | 0.36 | 0.37  | 0.61  | 0.26  | 0.99  | 2.18  | 0.01 | 2.21      | 0.04     |

**Table 3** Comparison of statistical metrics assuming the 3-Story RC frame with three bays having bay lengths of 5.0 m as test data for predicting  $Sa(T_1)$ 

| Model       | $R^2$ | MSE   | RMSE  | MAE   | MARE | MSRE  | RMSRE | RRMSE | MBE   | erMAX | SD   | $t$ -stat | $U_{95}$ |
|-------------|-------|-------|-------|-------|------|-------|-------|-------|-------|-------|------|-----------|----------|
| XGBoost     | 0.984 | 1.37  | 3.70  | 3.29  | 0.04 | 0.005 | 0.07  | 0.02  | -1.06 | 0.32  | 0.02 | 0.99      | 0.08     |
| RF          | 0.953 | 4.01  | 6.33  | 5.59  | 0.05 | 0.003 | 0.05  | 0.03  | -2.36 | 0.06  | 0.01 | 1.33      | 0.13     |
| BReg        | 0.952 | 4.13  | 6.42  | 5.68  | 0.05 | 0.003 | 0.05  | 0.03  | -2.42 | 0.06  | 0.01 | 1.35      | 0.13     |
| ETReg       | 0.966 | 2.92  | 5.41  | 5.09  | 0.06 | 0.004 | 0.06  | 0.04  | -3.00 | 0.08  | 0.02 | 2.21      | 0.11     |
| GBM         | 0.951 | 4.16  | 6.45  | 5.92  | 0.07 | 0.011 | 0.10  | 0.04  | -1.33 | 0.44  | 0.03 | 0.70      | 0.14     |
| HistGBR     | 0.984 | 1.33  | 3.64  | 3.32  | 0.03 | 0.001 | 0.04  | 0.02  | -1.31 | 0.08  | 0.01 | 1.28      | 0.07     |
| HistGBR(FN) | 0.986 | 1.20  | 3.46  | 2.99  | 0.11 | 0.018 | 0.13  | 0.06  | -4.35 | 0.44  | 0.02 | -         | 0.08     |
| AdaBoost    | 0.919 | 6.94  | 8.33  | 7.40  | 0.06 | 0.003 | 0.06  | 0.04  | -2.50 | 0.09  | 0.02 | 1.04      | 0.17     |
| KNR         | 0.879 | 10.32 | 10.16 | 9.53  | 0.12 | 0.023 | 0.15  | 0.07  | -3.95 | 0.54  | 0.03 | 1.40      | 0.21     |
| PLSReg      | 0.774 | 19.27 | 13.88 | 11.36 | 0.99 | 4.876 | 2.21  | 0.25  | 13.47 | 9.49  | 0.15 | 13.24     | 0.40     |
| SReg        | 0.614 | 32.97 | 18.16 | 16.19 | 1.06 | 5.385 | 2.32  | 0.27  | 15.50 | 10.01 | 0.15 | 5.44      | 0.46     |
| VReg        | 0.907 | 7.90  | 8.89  | 8.09  | 0.49 | 1.225 | 1.11  | 0.12  | 5.67  | 4.77  | 0.08 | 2.74      | 0.24     |
| LReg        | 0.775 | 19.22 | 13.86 | 11.30 | 1.00 | 4.927 | 2.22  | 0.25  | 13.66 | 9.54  | 0.15 | 19.26     | 0.40     |
| GReg        | 0.313 | 50.71 | 22.52 | 20.82 | 1.09 | 5.880 | 2.42  | 0.30  | 13.60 | 10.43 | 0.17 | 2.51      | 0.56     |
| MLPReg      | 0.982 | 1.50  | 3.88  | 3.19  | 0.13 | 0.022 | 0.15  | 0.06  | -4.08 | 0.22  | 0.03 | -         | 0.09     |
| SVM         | 0.987 | 1.14  | 3.38  | 2.93  | 0.22 | 0.266 | 0.52  | 0.05  | -0.70 | 2.34  | 0.05 | 0.70      | 0.11     |
| NuSVR       | 0.993 | 0.57  | 2.39  | 2.19  | 0.07 | 0.007 | 0.08  | 0.04  | -2.68 | 0.18  | 0.01 | -         | 0.05     |
| LSVR        | 0.743 | 21.94 | 14.81 | 12.40 | 1.05 | 5.348 | 2.31  | 0.26  | 14.82 | 9.92  | 0.15 | -         | 0.42     |
| ERTReg      | 0.947 | 4.51  | 6.72  | 6.22  | 0.06 | 0.004 | 0.07  | 0.04  | -2.92 | 0.10  | 0.02 | 1.60      | 0.14     |
| ANNs        | 0.999 | 0.09  | 0.94  | 0.85  | 0.04 | 0.014 | 0.12  | 0.01  | 0.10  | 0.57  | 0.01 | 0.36      | 0.03     |

lower performance for predicting  $IDR_{max}$  of the 3-Story RC frame, the SVM method achieved the  $R^2$  value of 0.987 for predicting  $Sa(T_1)$  that proves the acceptable performance of this method.

## 5 Performance of prediction models

The most important part of the prediction models is to prepare the datasets according to the important features. The seven important features related to each type of prediction (i.e.,  $Sa(T_1)$  or  $IDR_{max}$ ) was plotted in Fig. 10. According to these targets, the training dataset contained 92,400 data points achieved by performing IDAs. In other words, 92,400 nonlinear time history analyses were done based on increasing the intensity measures (i.e., IDA) to prepare the large database for prediction. After preparing suitable datasets, the selected ML algorithms with higher accuracy of prediction (see Tables 2 and 3) were used for seismic response prediction models. Figures 11 and 12 present prediction results of  $IDR_{max}$  for the 6-Story and 8-Story RC MRFs assuming five types of bays including PL records. It should be noted that the selected RC MRFs were removed from training datasets during the prediction. For the 6-Story RC MRFs with one-, two-, three-, four-, and five-bays, the ML algorithms of HistGBR, ANNs, and BReg had higher accuracy of prediction values of 90.2%, 93.5%, 94%, 95.4%, and 96.3%,

respectively. For the 8-Story RC MRFs with one-, two-, three-, four-, and five-bays, the ML algorithms of ETReg, BReg, and ANNs had higher accuracy of prediction values of 93.8%, 94.3%, 93.4%, 95%, and 95.3%, respectively. It can be seen that in all results, the algorithms had the most precise prediction for  $IDR_{max}$  of lower than 4.0% due to the points near the blue lines. Therefore, the mentioned algorithms can be used as a precise prediction model for  $IDR_{max}$  lower than 4.0% in all types of RC MRFs.

To present the estimation accuracy of M-IDA curve models, only having higher values of  $R^2$  is not enough due to the relations between the values of before and after data points. Therefore, the best way to present the power of the algorithm is to plot both actual and predicted curves. Figures 13 and 14 show the predicted M-IDAs versus the actual M-IDA curve of the 3-Story and 7-Story RC MRFs having five types of bays subjected to PL records. The two most precise predicted M-IDAs were plotted that show the accuracy of the prediction models used in this study and can be used as a preliminary prediction of M-IDA curves of RC MRFs.

## 6 Generality of prediction models

In Sect. 5, the capability of ML algorithms for predicting the  $IDR_{max}$  and  $Sa(T_1)$  of the aforementioned RC frames was presented. To present the overall accuracy of the proposed

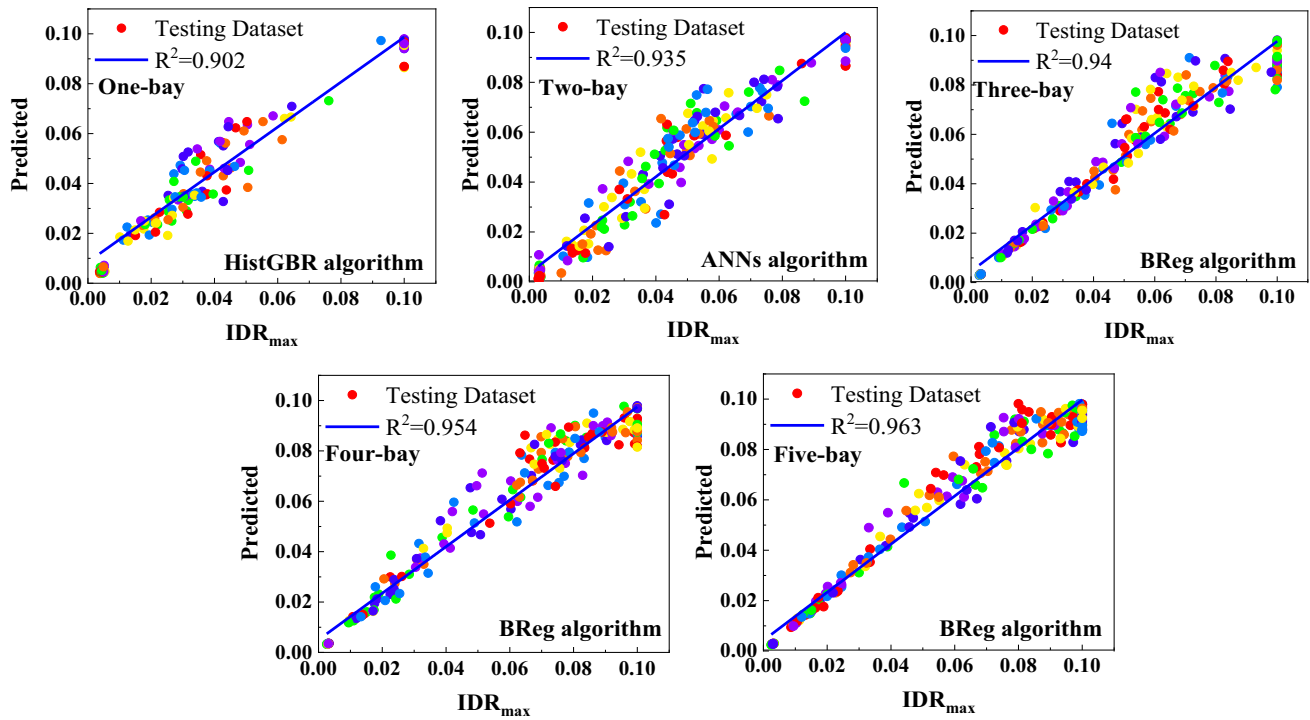


Fig. 11 IDR<sub>max</sub> prediction results for the 6-Story RC MRFs as testing datasets including PL records

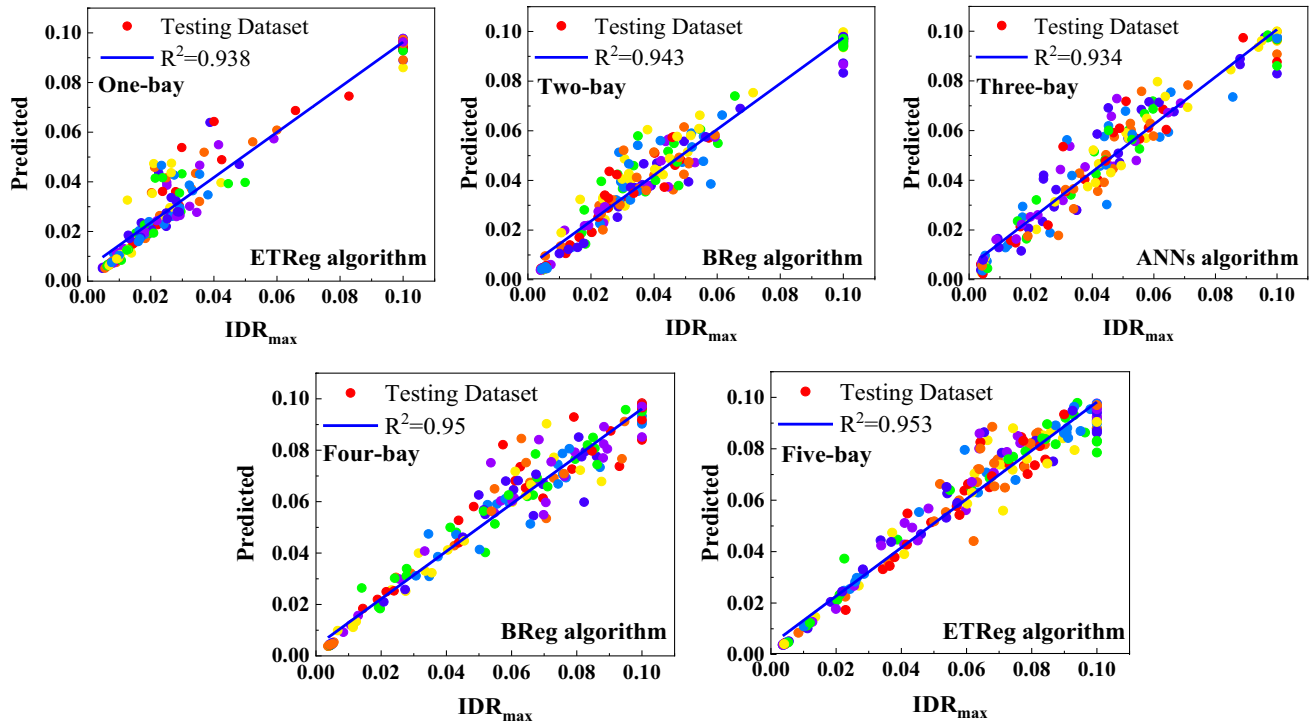
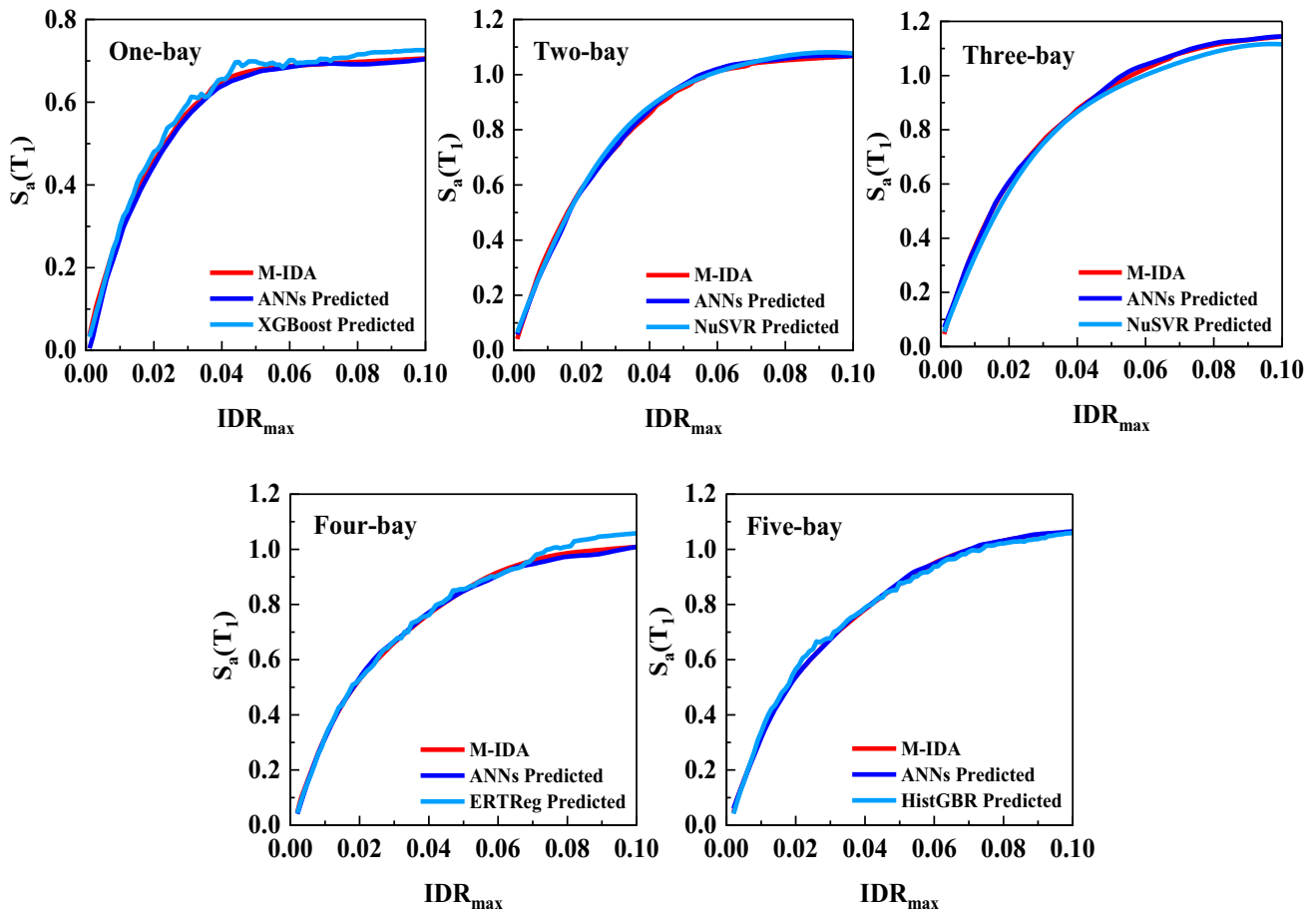


Fig. 12 IDR<sub>max</sub> prediction results for the 8-Story RC MRFs as testing datasets including PL records



**Fig. 13** Predicted M-IDAs of the 3-Story RC MRFs as testing datasets including records

ML-based prediction of  $IDR_{max}$  and  $S_a(T_1)$  as a target for the M-IDA curve, four case study RC buildings with different structural parameters were assumed to show the reliability and applicability of prediction models. Figure 15 presents the structural plan and documentation of beams and columns of a five-Story RC frame that was used for the performance evaluation of prediction models. It should be added that the testing dataset prepared for this RC frame should have same important features as the training dataset for prediction models (see Fig. 10). Therefore, the selected RC frame was modeled in ETABS and Opensees [52] softwares, and IDAs were performed based on the targets of  $S_a(T_1)$  and  $IDR_{max}$  including assumed seismic records. The results of the analysis were prepared as a testing dataset; then, trained prediction models were used to estimate  $IDR_{max}$  and  $S_a(T_1)$  as a target.

Given that it is not possible to have an experimental sample to validate prediction models, to challenge the ability of proposed ML-based models, four cases of selected RC buildings assuming different input features were assumed. In Case A, the bay length of the five-Story RC frame was selected as equal to 6.5 m. In Case B, the bay length and

story elevation of the five-Story RC frame were selected equal to 6.5 m and 3.8 m, respectively. For Case C and D, the weight of the five-Story RC frame was reduced by 10% and 20%, respectively, compared to the aforementioned loads assumed in Sect. 3, while the bay length and story elevation were selected equal to 6.5 m and 3.8 m, respectively. These four cases have different input features to challenge the possibility of using proposed ML-based models for any type of RC frame including two record subsets. The fundamental periods of Case A, Case B, Case C, and Case D were equal to 1.351, 1.291, 1.225, and 1.156, respectively. Therefore, all input features of the assumed cases are different from the training models. Figure 16 presents the comparison of  $R^2$  for ML algorithms to predict  $IDR_{max}$  of the five-Story RC frames assuming PL records. Four algorithms of BReg, ETRReg, ERTReg, and ANNs had higher values of prediction accuracy equal to 95.7%, 93.19%, 90.27%, and 90%, respectively, for the prediction of  $IDR_{max}$  in Case A, and had higher values of prediction accuracy equal to 92.78%, 90.31%, 87.85%, and 90.1%, respectively, for prediction of  $IDR_{max}$  in Case B. Moreover, in Case C, the ANNs and BReg algorithms achieved a prediction accuracy of 92.9%

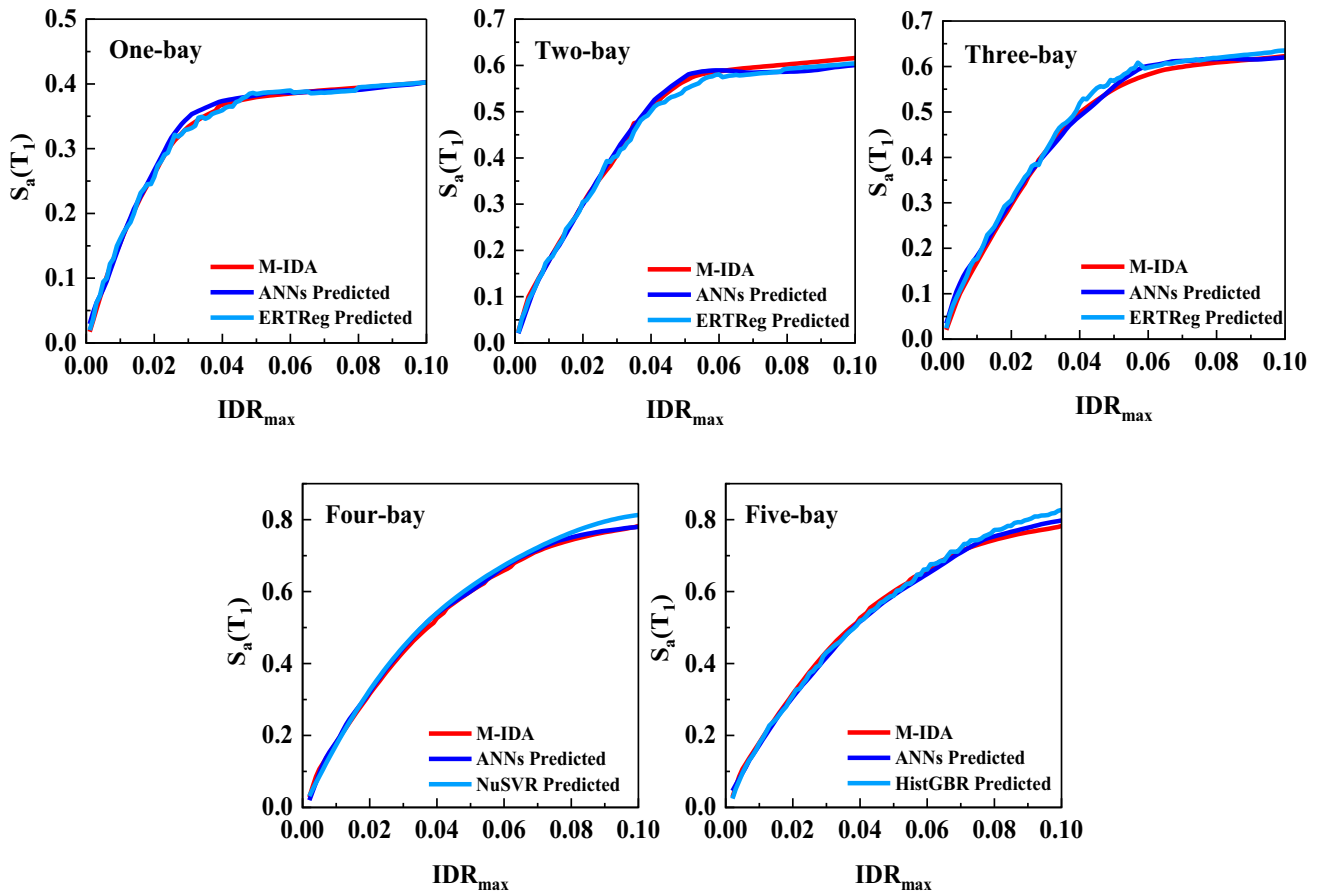


Fig. 14 Predicted M-IDs of the 7-Story RC MRFs as testing datasets including PL records

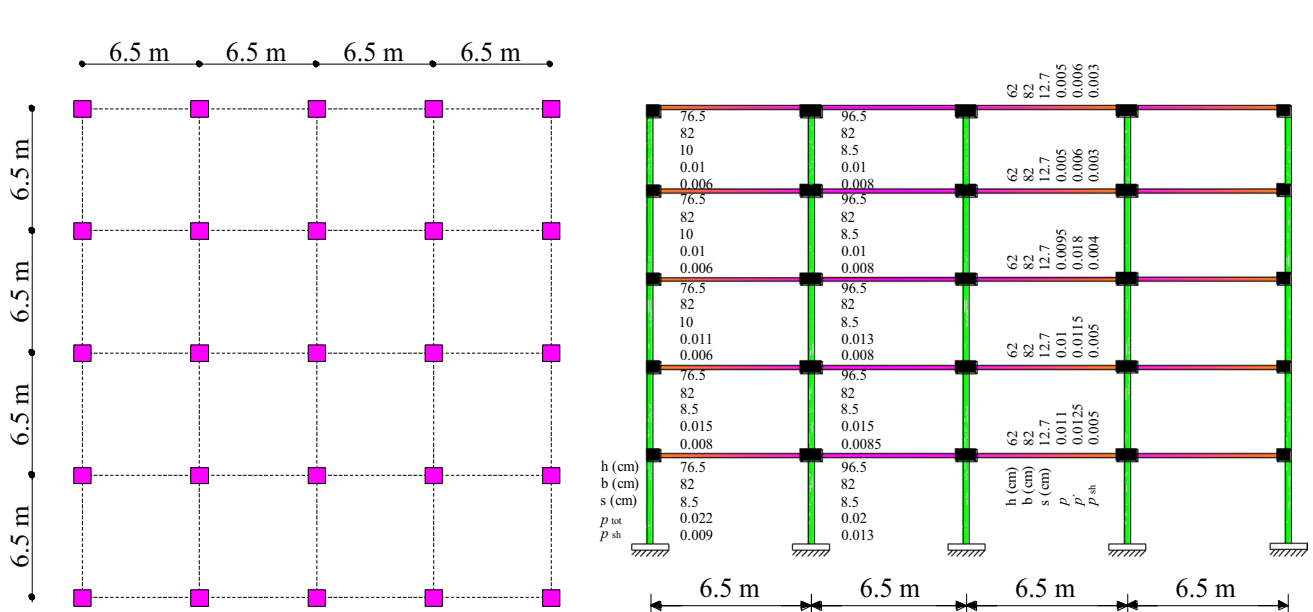


Fig. 15 A five-Story RC frame is used for the performance evaluation of trained prediction models

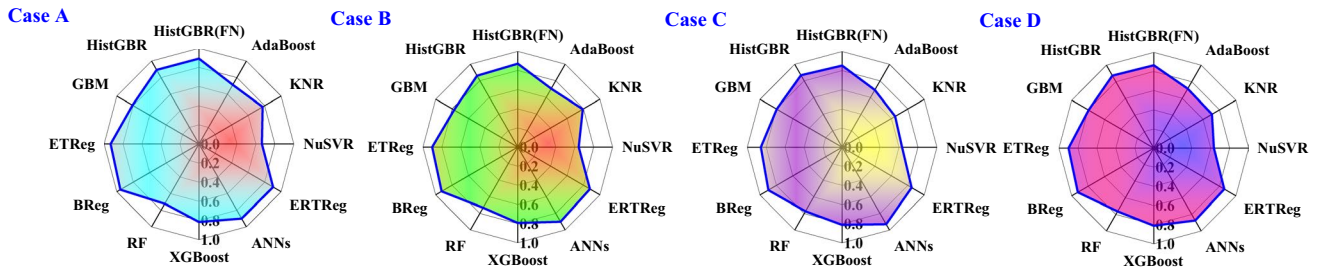


Fig. 16 Pie charts of ML algorithms for predicting  $IDR_{max}$  of the five-Story RC frames assuming PL records

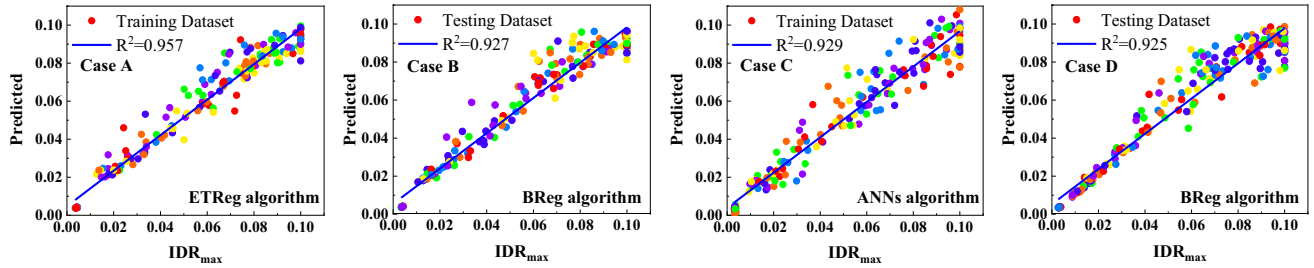


Fig. 17 Predicted  $IDR_{max}$  of the five-Story RC frames in the best ML algorithm including PL records

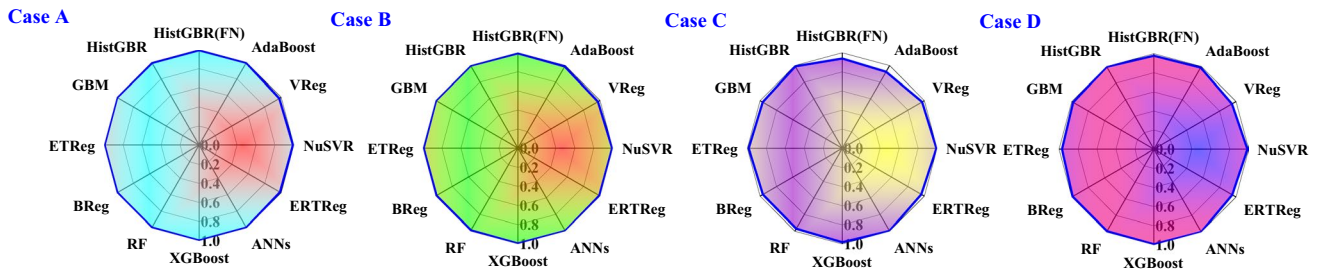


Fig. 18 Pie charts of ML methods for estimating the M-IDA curve of the five-Story RC frames assuming PL records

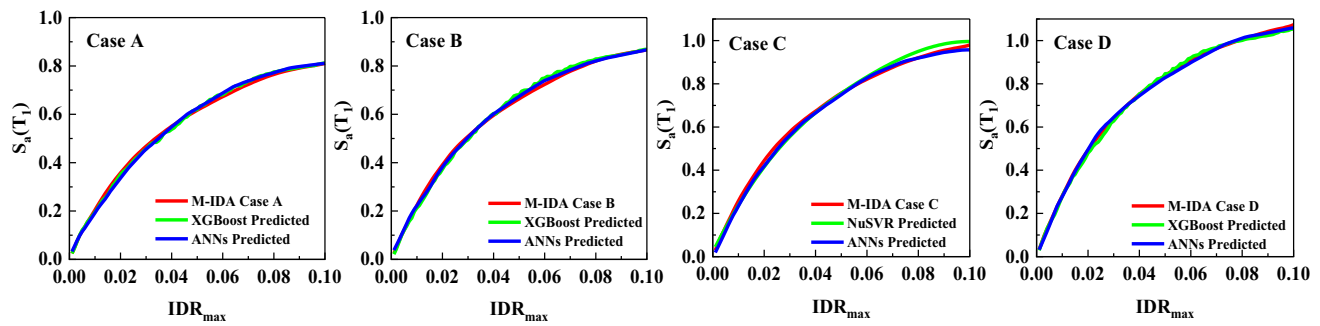
and 89.76%, respectively, while in Case D, the BReg, ETReg, and ANNs algorithms had a prediction accuracy of 92.5%, 89.93%, and 87.32%, respectively. Figure 17 depicts the scatter plots of predicted  $IDR_{max}$  of four cases of the five-Story RC frames in the best ML algorithm including PL records. It should be noted that similar results were observed for NP records, while results regarding PL records were presented for brevity.

Figure 18 presents the pie charts of ML-based models for estimating the M-IDA curve of the five-Story RC frames assuming PL records. ML methods achieved  $R^2$  values higher than 0.97 for predicting testing datasets of four cases. Although the pie charts show the highest values of the predicted M-IDA curve with  $R^2$  of more than 0.97, some of the ML algorithms cannot fit the actual M-IDA curve of RC frames. Therefore, ML algorithms were improved to

achieve the best fitting curves. Figure 19 presents the fitted predicted M-IDAs by improved ML algorithms. The ANNs and XGBoost algorithms had the best fitting curves and can be considered the most reliable prediction models.

To determine the seismic performance levels of the five-Story RC frames, the structural performance levels that were defined based on the allowable  $IDR_{max}$  values of 1.0%, 2.0%, and 4.0% corresponding to Immediate Occupancy (IO), Life Safety (LS), and Collapse Prevention (CP) performance levels, respectively, were assumed. It is noteworthy that the limit states were described according to the Table C1–3 in FEMA 356 [63] for limiting the damages states of primary structural elements of the lateral force-resisting system. According to allowable performance levels, Table 4 presents the actual values achieved by M-IDAs of the RC frames and those were predicted by improved ML algorithms. According to Table 4, the





**Fig. 19** Predicted M-IDA curves of the five-story RC frames in the best ML methods including PL records

**Table 4** Predicted seismic performance levels of the five-story RC frames based on the M-IDA curves including PL records

| M-IDA curve       | Case A |       |       | Case B |       |       |
|-------------------|--------|-------|-------|--------|-------|-------|
|                   | IO     | LS    | CP    | IO     | LS    | CP    |
| Actual value      | 0.205  | 0.357 | 0.550 | 0.224  | 0.390 | 0.595 |
| XGBoost predicted | 0.197  | 0.351 | 0.537 | 0.213  | 0.378 | 0.603 |
| ANNs predicted    | 0.193  | 0.346 | 0.548 | 0.222  | 0.383 | 0.601 |
| M-IDA curve       | Case C |       |       | Case D |       |       |
|                   | IO     | LS    | CP    | IO     | LS    | CP    |
| Actual value      | 0.257  | 0.444 | 0.673 | 0.281  | 0.492 | 0.751 |
| XGBoost predicted | –      | –     | –     | 0.278  | 0.479 | 0.753 |
| NuSVR predicted   | 0.237  | 0.415 | 0.665 | –      | –     | –     |
| ANNs predicted    | 0.244  | 0.429 | 0.669 | 0.279  | 0.495 | 0.748 |

predicted values in all performance levels are very close to the actual values; thus, the prediction models have the ability of reliable prediction and can be used by researchers for predicting RC frames.

## 7 Graphical user interface

The preliminary estimation of the performance levels can widely help designers to know about the weakness of the designed buildings, therefore, they can use the results for vulnerability assessments of structures. To prepare for better accessibility of the results of this research, Graphical User Interface (GUI) was introduced to receive input parameters related to the RC frame and seismic limitation of performance levels and provide the predicted  $S_a(T_1)$  regarding the seismic limit-state performance levels of RC MRFs prescribed by FEMA356 [63]. It should be noted that the reliability of prediction models was discussed in Sect. 6, and the introduced GUI can plot the predicted

ML-based M-IDA curve while mitigating the need for complex modeling and analyses. It is noteworthy that the input parameters can be easily achieved for the assumed structure, and in addition, for calculating the period of the structure, the formulas that have been provided by the seismic provisions (e.g., ASCE 07-16 [48]) can be used.

## 8 Conclusions

Recent studies confirm that complex modeling and analysis should be performed to determine seismic responses and seismic performance levels of RC structures, while the most of analyses are time-consuming and need to be done by high-speed computer systems. In addition, the unpredictable nature of seismic events is another factor that affects seismic performance achievement. To overcome this issue, this research proposed ML-based prediction models to estimate the  $IDR_{max}$  and  $S_a(T_1)$  for the M-IDA curve of the RC frames. The analysis results can be summarized as follows:

**Fig. 20** GUI introduced for predicting the seismic limit-state capacity of RC MRFs

Seismic Response and performance assessment of RC MRFs

Developed by Farzin Kazemi, Neda Asgarkhani and Robert Jankowski  
Gdansk University of Technology, Gdansk, Poland

Input parameters

|                          |                      |                          |
|--------------------------|----------------------|--------------------------|
| Number of Story          | <input type="text"/> | N                        |
| Fundamental Period       | <input type="text"/> | Sec                      |
| Record Dataset           | <input type="text"/> | Puls Like(1), No Puls(2) |
| Story Weight             | <input type="text"/> | ton                      |
| Story Height             | <input type="text"/> | m                        |
| Number of Bay            | <input type="text"/> | N                        |
| Bay Length               | <input type="text"/> | m                        |
| Dead Load                | <input type="text"/> | kN/m2                    |
| Live Load                | <input type="text"/> | kN/m2                    |
| Immediate Occupancy (IO) | <input type="text"/> | Drift (FEMA356)          |
| Life Safety (LS)         | <input type="text"/> | Drift (FEMA356)          |
| Collapse Prevention (CP) | <input type="text"/> | Drift (FEMA356)          |
| Total Collapse (TC)      | <input type="text"/> | Drift (FEMA356)          |

Prediction Result

|                                   |                      |   |
|-----------------------------------|----------------------|---|
| Sa(T1) in performance level of IO | <input type="text"/> | g |
| Sa(T1) in performance level of LS | <input type="text"/> | g |
| Sa(T1) in performance level of CP | <input type="text"/> | g |
| Sa(T1) in performance level of TC | <input type="text"/> | g |

Predict Default Plot Cancel

- Assuming  $IDR_{max}$  as the target of prediction, eight algorithms of PLSReg, SReg, VReg, LReg, GReg, MLPReg, SVM, and LSVR had lower  $R^2$  values (i.e., less than 65%) and cannot be used as prediction models. On the other hand, eight algorithms of KNR, PLSReg, SReg, LReg, GReg, MLPReg, SVM, and LSVR had lower  $R^2$  values (i.e., less than 77%) for predicting  $Sa(T_1)$  as a target. In addition, ML algorithms had the precise prediction values located exactly in the  $x=y$  line, assuming allowable  $IDR_{max}$  of lower than 4.0%, that shows the ability of the proposed methods for estimating  $IDR_{max}$  in all RC MRFs.
- Considering the curve plotting ability that improved in ML methods based on the allowable performance levels (i.e.,  $IDR_{max}$  values of 1.0%, 2.0%, and 4.0%), three algorithms of the XGBoost, ANNs, and NuSVR can predict the seismic performance levels of the five-story RC frame using the predicted M-IDA curves. Therefore, they can be considered as proposed prediction models for any type of RC frame.
- Four case study RC buildings were assumed to check the reliability of prediction models. In Case A, the BReg, ETReg, ERTReg, and ANNs algorithms predicted the  $IDR_{max}$  with the accuracy of 95.7%, 93.19%, 90.27%, and 90%, respectively, and in Case B, the accuracy of 92.78%, 90.31%, 87.85%, and 90.1%, respectively, were achieved by prediction models. In Case C, the ANNs and BReg algorithms with the accuracy of 92.9% and 89.76%, respectively, in Case D, the BReg, ETReg, and ANNs algorithms with the accuracy of 92.5%, 89.93%, and 87.32%, respectively, can be considered as best models of prediction.

- Graphical User Interface (GUI) was proposed for preliminary estimation of the seismic performance levels of RC frames based on the main important features that can be introduced as input parameters. In addition, the GUI can be able to plot the predicted M-IDA curve regarding both seismic events and facilitate the seismic vulnerability assessment of RC buildings. Moreover, there is no limit for introducing the thresholds of the allowable  $IDR_{max}$ , and the users can find the prediction results for the selected  $IDR_{max}$ .
- For operating the GUI, (a) receives the main important structural features that affects the seismic response and seismic limit-state capacities, (b) receives the selected  $IDR_{max}$  defined by user (e.g., four main  $IDR_{max}$  were showed in Fig. 20), (c) predicts the M-IDA curve of introduced RC frames, and (d) presents the  $Sa(T_1)$  corresponding to the selected  $IDR_{max}$ .

**Author contributions** All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by FK, NA, and RJ. The first draft of the manuscript was written by FK and NA, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. FK: writing—original draft preparation, conceptualization, software, analysis, methodology, modeling. NA: writing—original draft preparation, modeling, software, analysis, and investigation. RJ: writing—review and editing, supervision.

**Funding** The authors did not receive support from any organization for the submitted work. The authors have no relevant financial or non-financial interests to disclose.

**Data availability** Data will be made available on request.

## Declarations

**Conflict of interest** The authors declare that there is no conflict of interest with relation to the paper Machine learning-based seismic probabilistic prediction of reinforced concrete buildings submitted for publication in *Archives of Civil and Mechanical Engineering*.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Kaya Y, Safak E. Real-time analysis and interpretation of continuous data from structural health monitoring (SHM) systems. *Bull Earthq Eng*. 2015;13(3):917–34.
2. Ngeljaratan L, Moustafa MA. Structural health monitoring and seismic response assessment of bridge structures using target-tracking digital image correlation. *Eng Struct*. 2020;213: 110551.
3. Manguri A, Saeed N, Kazemi F, Szczepanski M, Jankowski R. Optimum number of actuators to minimize the cross-sectional area of prestressable cable and truss structures. *Structures*. 2023;47:2501–14.
4. Kazemi F, Jankowski R. Enhancing seismic performance of rigid and semi-rigid connections equipped with SMA bolts incorporating nonlinear soil-structure interaction. *Eng Struct*. 2023;274: 114896.
5. Kazemi F, Asgarkhani N, Jankowski R. Probabilistic assessment of SMRFs with infill masonry walls incorporating nonlinear soil-structure interaction. *Bull Earthq Eng*. 2023;21:1–32.
6. Kazemi F, Mohebi B, Yakhchalian M. Evaluation of the P-delta effect on collapse capacity of adjacent structures subjected to far-field ground motions. *Civil Eng J*. 2018;4(5):1066. <https://doi.org/10.28991/cej-0309156>.
7. Mohebi B, Kazemi F, Yakhchalian M. Investigating the P-Delta effects on the seismic collapse capacity of adjacent structures. In: 16th European conference on earthquake engineering (16ECEE), 18–21, June, Thessaloniki, Greece. 2018.
8. 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.
9. Kazemi F, Mohebi B, Yakhchalian M. Predicting the seismic collapse capacity of adjacent structures prone to pounding. *Can J Civ Eng*. 2020;47(6):663–77.
10. Kazemi F, Mohebi B, Jankowski R. Predicting the seismic collapse capacity of adjacent SMRFs retrofitted with fluid viscous dampers in pounding condition. *Mech Syst Signal Process*. 2021;161: 107939.
11. Asgarkhani N, Kazemi F, Jankowski R. Optimal retrofit strategy using viscous dampers between adjacent RC and SMRFs prone to earthquake-induced pounding. *Arch Civ Mech Eng*. 2023;23(1):1–26.
12. Kabir MAB, Hasan AS, Billah AM. Failure mode identification of column base plate connection using data-driven machine learning techniques. *Eng Struct*. 2021;240: 112389.
13. Mangalathu S, Jeon JS. Stripe-based fragility analysis of multi-span concrete bridge classes using machine learning techniques. *Earthq Eng Struct Dynam*. 2019;48(11):1238–55.
14. Nguyen HD, LaFave JM, Lee YJ, Shin M. Rapid seismic damage-state assessment of steel moment frames using machine learning. *Eng Struct*. 2022;252: 113737.
15. Wu ZN, Han XL, He A, Cai YF, Ji J. Machine learning-based adaptive degradation model for RC beams. *Eng Struct*. 2022;253: 113817.
16. Hastie T, Tibshirani R, Friedman J. Overview of supervised learning. In: *The elements of statistical learning*. New York: Springer; 2009. p. 9–41.
17. Yazdanpanah O, Dolatshahi KM, Moammer O. Rapid seismic fragility curves assessment of eccentrically braced frames through an output-only nonmodel-based procedure and machine learning techniques. *Eng Struct*. 2023;278: 115290.
18. Kazemi F, Asgarkhani N, Jankowski R. Predicting seismic response of SMRFs founded on different soil types using machine learning techniques. *Eng Struct*. 2023;274: 114953.

19. Huang CS, Hung SL, Wen CM, Tu TT. A neural network approach for structural identification and diagnosis of a building from seismic response data. *Earthq Eng Struct Dynam.* 2003;32(2):187–206.
20. Yinfeng D, Yingmin L, Ming L, Mingui X. Nonlinear structural response prediction based on support vector machines. *J Sound Vib.* 2008;311(3–5):886–97.
21. Lagaros ND, Papadrakakis M. Neural network based prediction schemes of the non-linear seismic response of 3D buildings. *Adv Eng Softw.* 2012;44(1):92–115.
22. De Lautour OR, Omenzetter P. Damage classification and estimation in experimental structures using time series analysis and pattern recognition. *Mech Syst Signal Process.* 2010;24(5):1556–69.
23. Worden K, Green PL. A machine learning approach to nonlinear modal analysis. *Mech Syst Signal Process.* 2017;84:34–53.
24. Kiani J, Camp C, Pezeshk S. On the application of machine learning techniques to derive seismic fragility curves. *Comput Struct.* 2019;218:108–22.
25. Nguyen NV, Nguyen HD, Dao ND. Machine learning models for predicting maximum displacement of triple pendulum isolation systems. *Structures.* 2022;36:404–15.
26. Oh BK, Glisic B, Park SW, Park HS. Neural network-based seismic response prediction model for building structures using artificial earthquakes. *J Sound Vib.* 2020;468: 115109.
27. Luo H, Paal SG. Artificial intelligence-enhanced seismic response prediction of reinforced concrete frames. *Adv Eng Inform.* 2022;52: 101568.
28. Gholizadeh R, Amiri GG, Mohebi B. An alternative approach to a harmony search algorithm for a construction site layout problem. *Can J Civ Eng.* 2010;37(12):1560–71.
29. Todorov B, Billah AM. Machine learning driven seismic performance limit state identification for performance-based seismic design of bridge piers. *Eng Struct.* 2022;255: 113919.
30. Dehestani A, Kazemi F, Abdi R, Nitka M. Prediction of fracture toughness in fibre-reinforced concrete, mortar, and rocks using various machine learning techniques. *Eng Fract Mech.* 2022;276: 108914.
31. Adibimanesh B, Polesek-Karczewska S, Bagherzadeh F, Szczuko P, Shafighfar T. Energy consumption optimization in wastewater treatment plants: machine learning for monitoring incineration of sewage sludge. *Sustain Energy Technol Assess.* 2023;56: 103040.
32. Shafighfar T, Bagherzadeh F, Rizi RA, Yoo DY. Data-driven compressive strength prediction of steel fiber reinforced concrete (SFRC) subjected to elevated temperatures using stacked machine learning algorithms. *J Market Res.* 2022;21:3777–94.
33. Kazemi F, Asgarkhani N, Jankowski R. Machine learning-based seismic fragility and seismic vulnerability assessment of reinforced concrete structures. *Soil Dyn Earthq Eng.* 2023;166: 107761.
34. Kazemi F, Jankowski R. Machine learning-based prediction of seismic limit-state capacity of steel moment-resisting frames considering soil-structure interaction. *Comput Struct.* 2023;274: 106886.
35. Pal M. Random forest classifier for remote sensing classification. *Int J Remote Sens.* 2005;26(1):217–22.
36. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn.* 2006;63(1):3–42.
37. Louppe G, Geurts P. Ensembles on random patches. In: *Joint European conference on machine learning and knowledge discovery in databases.* Berlin: Springer; 2012. p. 346–61.
38. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci.* 1997;55(1):119–39.
39. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *Ann Stat.* 2000;28(2):337–407.
40. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat.* 2001;29:1189–232.
41. Cortes C, Vapnik V. Support vector machine. *Mach Learn.* 1995;20(3):273–97.
42. Chang CC, Lin CJ. Training v-support vector regression: theory and algorithms. *Neural Comput.* 2002;14(8):1959–77.
43. Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. Support vector regression machines. *Adv Neural Inf Process Syst.* 1996;9.
44. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai.* 1995;14(2):1137–45.
45. McCullagh P, Nelder JA. *Generalized linear models.* London: Chapman and Hall; 1989.
46. Wolpert DH. Stacked generalization. *Neural Netw.* 1992;5(2):241–59.
47. Höskuldsson A. PLS regression methods. *J Chemometr.* 1988;2(3):211–28.
48. ASCE 7–16. *Minimum design loads and associated criteria for buildings and other structures.* American Society of Civil Engineers; 2017.
49. United States Geological Survey. <https://www.usgs.gov/programs/earthquake-hazards/hazards>. Accessed 03 Mar 2022.
50. Haselton CB, Deierlein GG. Assessing seismic collapse safety of modern reinforced concrete frame buildings. *PEER Report.* 2007;8.
51. 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. *J Build Eng.* 2021;33: 101847.
52. McKenna F, Fenves GL, Filippou FC, Scott MH. *Open system for earthquake engineering simulation (OpenSees).* Berkeley, Pacific Earthquake Engineering Research Center, University of California. 2016. <http://OpenSees.berkeley.edu>. Accessed 21 Oct 2022.
53. Asgarkhani N, Yakhchalian M, Mohebi B. Evaluation of approximate methods for estimating residual drift demands in BRBFs. *Eng Struct.* 2020;224: 110849.
54. Yakhchalian M, Asgarkhani N, Yakhchalian M. Evaluation of deflection amplification factor for steel buckling restrained braced frames. *J Build Eng.* 2020;30: 101228.
55. Yakhchalian M, Yakhchalian M, Asgarkhani N. An advanced intensity measure for residual drift assessment of steel BRB frames. *Bull Earthq Eng.* 2021;19(4):1931–55.
56. 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. *Mech Syst Signal Process.* 2022;173: 109038.
57. Kazemi F, Asgarkhani N, Manguri A, Jankowski R. Investigating an optimal computational strategy to retrofit buildings with implementing viscous dampers. *Int Conf Comput Sci ICCS Proc.* 2022. [https://doi.org/10.1007/978-3-031-08754-7\\_25](https://doi.org/10.1007/978-3-031-08754-7_25).
58. Kazemi F, Jankowski R. Seismic performance evaluation of steel buckling-restrained braced frames including SMA materials. *J Constr Steel Res.* 2023;201: 107750.
59. Ibarra LF, Medina RA, Krawinkler H. Hysteretic models that incorporate strength and stiffness deterioration. *Earthq Eng Struct Dynam.* 2005;34(12):1489–511.
60. 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.
61. Federal Emergency Management Agency (FEMA P695). *Quantification of building seismic performance factors.* US Department of Homeland Security, FEMA. 2009.

62. MATLAB/Simulink as a Technical Computing Language. Engineering computations and modeling in MATLAB. 2018.
63. FEMA-356. Prestandard and commentary for the seismic rehabilitation of buildings. Washington, DC: Federal Emergency Management Agency. 2000.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.