Structures 38 (2022) 157–167

Contents lists available at ScienceDirect

Structures

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

# Probabilistic seismic analysis of bridges through machine learning approaches

Farahnaz Soleimani

*Department of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, United States*

ARTICLE INFO

*Keywords:*
Probabilistic seismic demand model
Performance-based analysis
Machine learning
Concrete box-girder bridge
Seismic performance

ABSTRACT

Probabilistic seismic demands of bridge components such as bridge column and deck are conventionally expressed as a power-law function of a single ground motion intensity measure. This unidimensional probabilistic seismic demand model (PSDM) was introduced more than two decades ago, and since then, it was commonly used to estimate seismic demands. Over the recent years, an extensive body of research has been evolved to propose alternative PSDMs, but none has been proved to be dominantly superior over other approaches. There yet remains a milestone to enrich predictions provided by PSDMs and expanding their application beyond certain methodology, particular functional form, and corresponding assumptions on the distribution of the demands. Given the advancements in computational technologies which lead to the growth of diverse analytically-driven data, machine learning (ML) approaches have a tremendous potential to revolutionize predictions of seismic demands. This study presents a comprehensive appraisal of ML-based PSDMs to further expand the research advances in this domain and leverage the efficiency and advantages that ML methods offer compared to the unidimensional model. To this end, the efficiency of a variety of parametric and non-parametric ML algorithms with different degrees of flexibility are explored to estimate the demands associated with the primary bridge components. Moreover, by applying ML-based variable selection techniques, this study assesses the level of influence of the random variables on the generated PSDMs. These variables are used for the treatment of inherent uncertainties in material, geometric, structural, and ground motion parameters. As part of the appraisal, a ranking is provided for the investigated 39 models, such as Generalized Linear Models, Multi-order regressions, Bagging and Boosting, and Kernel-based models, according to their statistical performance in estimating the individual demands.

## 1. Introduction

The efficiency of probabilistic seismic demand models (PSDMs) governs the overall outcome of seismic performance analysis of bridges [16,15] in a postulated seismic hazard scenario and subsequent post-hazard decision makings as these demand models are typically used to derive analytical fragility curves and cover various sources of aleatory and epistemic uncertainties [39,2,5,34]. However, developing an efficient and practical framework for PSDM is challenging.

PSDM, initially formulated by Shome et al. in 1998, provides an estimation of the median value of the seismic demand ($\mu_{SD}$) as a power-law function of a ground motion intensity measure (IM) as displayed in Eq. (1) in which a and b are the regression coefficients. Based upon the lognormality assumption of the seismic demands [30,12,4], the demand models are commonly expressed in a transformed natural logarithmic space as shown in Eq. (2).

$$\mu_{SD} = a(IM)^{b} \tag{1}$$

$$\ln(\mu_{SD}) = \ln(a) + b.\ln(IM) \tag{2}$$

Although the conventional model is simple to implement, it can be improved in multiple aspects (e.g., boosting prediction power of the model, incorporating additional predictors from uncertain parameters, such as ground motion and structural-related characteristics into the formulation) particularly by applying thriving machine learning (ML) algorithms.

Over the last two decades, a substantial body of research, predominantly evolved around the concept of metamodels, has been conducted to extend the widely used univariate PSDM to multivariate regression. For example, Ghosh et al. [8] considered the classical 2nd order polynomial response surface model (PRSM), multivariate adaptive regression splines (MARS) [7], and radial basis function (RBF) [11] and 11

*E-mail address:* soleimani@gatech.edu.

https://doi.org/10.1016/j.istruc.2022.02.006
Received 21 August 2021; Received in revised form 8 January 2022; Accepted 2 February 2022
Available online 9 February 2022

bridge modeling parameters, IM-related input variables, and deterioration affected parameters as the input variables in the demand models. Kameshwar and Padgett [14] applied PRSM, adaptive basis function construction (ABFC) [13], and RBF with 9 input parameters including the material and geometric properties of bridges. Most of the previous studies [24,27,28,29,25] that used PRSM adapted the 2nd order type while the application of the polynomial orders higher than 2 has received less attention. These studies investigated horizontally curved steel I-girder bridges and steel-plate-girder bridges and included peak ground acceleration (PGA) and bridge geometries as the potential input variables.

Despite the growing body of research in developing improved demands models, none of the proposed methodologies has been proved to be dominantly superior over other approaches, and further research is required to tackle the remaining challenges. The lack of flexibility of the original unidimensional form to incorporate various sources of uncertainties could impact the reliability of the estimation of the demands. Besides, the attempts in including more random variables in PSDMs have led to overly complex models, computationally expensive, and overfitted models. Another shortcoming in the growing body of research aimed at improving the estimation of seismic demands of bridge components is the lack of systematic appraisal of different multi-parameter predictive models. As shown in the brief review here, previous studies are typically limited to regional-based bridge attributes, limited numbers of selective features corresponding to structural characteristics, and arbitrary functional forms. Furthermore, although previous efforts provide valuable insight into evaluating the importance of different inputs on estimating seismic demands of bridges, they have not typically considered systematic and robust approaches such as the ML-based variable selection techniques when dealing with large data.

Moreover, although the application of modern ML methods in different fields of seismic hazard assessment has received considerable attention in recent years, their application in the context of PSDMs is rather limited as most of the reviewed literature focused on 1st and 2nd order regression models. As opposed to traditional approaches, ML-based models have the potential to capture the complex relationships between the input variables and the seismic demands without being restricted to relatively simple functional forms and prior assumption on the distribution of parameters as is the case in the conventional unidimensional model. Thus, the application of ML approaches in the performance-based analysis of bridges is appealing [38].

In summary, this study aims to address three main challenges in building PSDMs as highlighted in the brief state-of-the-art review: I) application of modern ML algorithms in building PSDMs, II) appraisal of different parametric and nonparametric ML algorithms in PSDMs application to identify the most viable approach, and III) identification of the optimal set of input variables as the most influential predictors.

In order to address the first gap in knowledge, the current study presents an ML-based framework to develop PSDMs that leads to a more reliable fragility, risk and resilience estimation of bridges. Furthermore, this study provides a systematic appraisal of several well-established parametric and nonparametric ML-based models to balance the trade-off between model interpretability and prediction performance. The considered ML algorithms include a variety of linear and nonlinear methods such as multiple linear regression, polynomial regression, tree-based approaches, and Kernel-based regressions. Although recent years have seen an increasing number of studies in applying more methodical approaches, the application of advanced and robust ML methods such as boosted tree algorithms and Gaussian Kernel in developing PSDMs requires further investigation.

In order to establish an efficient ML-based framework, it is essential to identify the optimum set of independent variables for the predictive models. As part of the formulation of PSDMs, this study evaluates the most influential variables that control the seismic response of bridges, considering an extensive list of bridge modeling parameters [32,37]. The overall goal of variable selection in this study is to build the best possible model by removing extraneous variables without sacrificing accuracy. The redundant variables contribute to model complexity and overfitting that alter the true relationship of interest. Removing this irrelevant information can improve the model fit, reduce the computation time, and make the model more interpretable. Besides, the benefit of comparing the results from different variable selection methods in this study is to overcome the drawbacks of individual techniques. For example, forward selection does not always provide the best set of features since this method does not run through every single combination of features because of large computation time, which could also lead to a model with high multicollinearity [21]. In that case, Least Absolute Shrinkage and Selection Operator (LASSO) or Random Forest may provide better results [36].

The following sections are organized as follows: Section 2 describes the analytical modeling and seismic analysis of bridges; Section 3 provides an overview of the theoretical aspects of the applied ML algorithms, while more details are provided in Appendix B; Section 4 presents a comparison of the model performance developed using different algorithms; and Section 5 summarizes the significance of the study and the key findings.

## 2. Analytical modeling and analysis

### 2.1. Description of bridge characteristics

For the purpose of this study, a class of multi-span concrete box-girder bridges with the characteristics of bridges located in California is considered. The numerical three-dimensional models of the bridges are built in Opensees [19] which is a popular software in the seismic vulnerability assessments of bridges. The finite element models of bridges are created considering the nonlinearities in structural and material components. A "spine" modeling approach is used to represent the bridge superstructure as a single beam-column frame. Then, the substructure is built as a combination of frame and spring elements.

As depicted in Fig. 1, the column model, connected to the superstructure elastic deck elements with rigid links, is composed of fiber cross-section beam-column elements including unconfined and confined concrete, and steel reinforcements [17,20,6]. The nonlinear behavior of the columns is captured by nonlinear displacement-based elements [33]. The class of tall bridges [35] is considered for the case study with column height ratios ($H_{ratio}$) ranging from low to high (slightly tall: $1.5 \le H_{ratio} < 2.5$, moderately tall: $2.5 \le H_{ratio} < 3.5$, extremely tall: $3.5 \le H_{ratio} < 4.5$) which is derived as the ratio of the average column height in a tall bridge divided by the average column height of the representative bridge with normal column height of a bridge with regular-sized columns (~7.5 m) [18]. The columns are more often constructed within the slight and moderate range, while the extremely tall ones are typically used in mountainous areas or span precarious regions such as deep ravines between mountaintops, where other bridge types may not be a viable option. The deformation of the column footing is captured using translational and rotational springs that are attached to the columns' base. In the three-dimensional bridge models, the foundation provides a means to transfer service loads from the structure to the underlying soil. Elastic translational and rotational springs are used to model foundations and are modeled using simple linear springs. These springs are assigned to zero length elements located at the base of the bridge column, as shown in Fig. 1, to capture the longitudinal and transverse movements of the foundation system. The spring stiffnesses obtained from previous works (e.g., Ramanathan [26] in which the foundation system with different soil profiles was analyzed in LPILE) are used in this study.

Three and four spans are randomly assigned to the simulations to have an equal portion of each, while similarly, spread and pile foundations are randomly assigned to an equal number of simulations for both piers and abutments. Furthermore, column cross-section (circular and rectangular), soil type (clay and sand), and girder type (reinforced

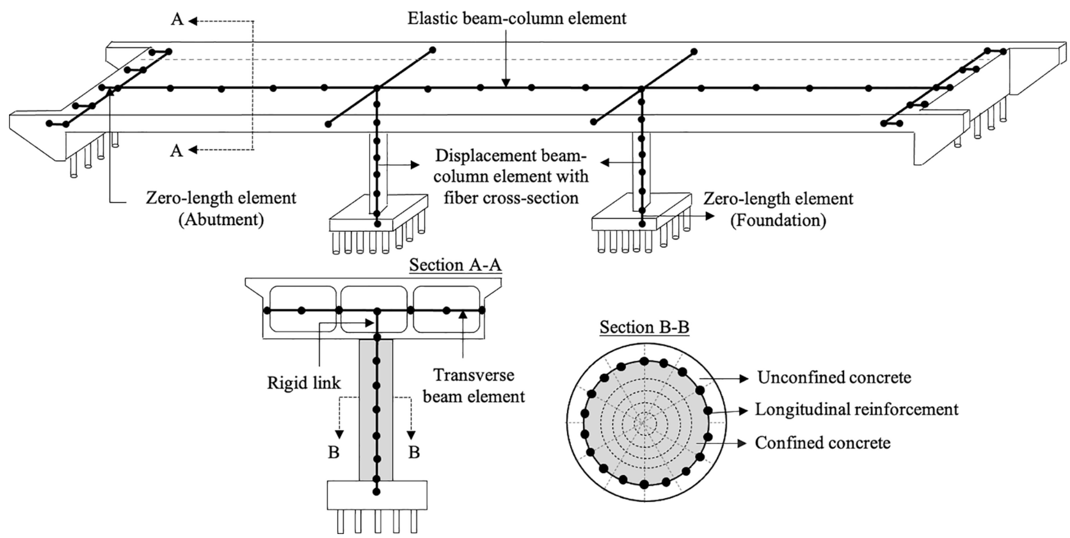


**Fig. 1.** Schematic layout for the analytical modeling of a three-span box-girder concrete bridge.

and prestressed) are the additional bridge attributes that have been randomly assigned. The associated abutment elements are represented by spring elements, and rayleigh damping with a critical ratio of 5% is employed. This study covers various sources of uncertainties corresponding to the superstructure, foundation, and substructure in addition to the general modeling parameters. To create the bridge samples for the seismic risk analysis (next section), the value of each modeling parameter is randomly sampled from the distribution provided in Table A.1 (see Appendix A). The distributions have been extracted from the review of existing bridge drawings. A detailed explanation for the modeling of individual components of bridges and the variability in their parameters can be found in previous relevant works [22,26,31].

### 2.2. Seismic performance analysis

Conducting a seismic risk assessment of bridges entails performing a particular course of actions in which the modeling parameters and ground motions are systematically varied. In the initial phase, according to the Latin Hypercube sampling technique [1], representative bridge models are created by randomly sampling across the uncertain parameters listed in Table A.1. Compared to pure random sampling, Latin Hypercube covers the probability space of the random variables and reduces the chance of unreasonable combinations which cause convergence issues when the NLTHA is performed in OpenSees and requires repeating the resampling process. In the next phase, the bridge samples are randomly paired with an equal number of ground motions. Baker's series of ground motions [3] is then applied to the generated bridge samples with scale factors of 1 and 2, resulting in a total of 320 excitations. The ground motions are scaled by a factor of two as recommended by previous studies (e.g., [26]) on the analytical seismic evaluation of bridges to have sufficient response values of IMs higher than the highest design level in California. Performing seismic analysis by applying 320 ground motions on the three classes of tall bridges leads to a total NLTHA of 960 combinations of bridge samples and ground motions. These excitations are composed of longitudinal and orthogonal components which in the process of simulations are set to be randomly oriented to the longitudinal and transverse directions of the bridge models.

Next, NLTHA is performed on each bridge sample to estimate the seismic demand of the various bridge components. The key monitored engineering demand parameters (EDPs) that are predicted by ML-based models (as the models' outputs $y_i$) are the commonly captured EDPs in the seismic analysis of bridges. Table 1 provides the list of these EDPs and their assigned notations that are used hereafter to simplify the presentation and discussion of the results. These parameters correspond to the maximum responses captured for each excitation. Curvature ductility is recorded for different locations along the column height and eventually the maximum value is used for developing the demand model. For the considered case study of concrete box-girder bridges, the bridge column has a connection at the column base close to fixity, and high curvature ductility was noted at the regions of the column close to the superstructure due to the large moment and shear transfer.

**Table 1**
Recorded key engineering demand parameters.

| Regression variable | Seismic demand | Engineering demand parameters |
|---|---|---|
| $y_1$ | $\ln(\rho_c)$ | Column curvature ductility |
| $y_2$ | $\ln(\delta_d)$ | Deck displacement |
| $y_3$ | $\ln(\delta_f)$ | Foundation translational displacement |
| $y_4$ | $\ln(\theta_f)$ | Foundation rotation |
| $y_5$ | $\ln(\delta_a)$ | Active abutment displacement |
| $y_6$ | $\ln(\delta_p)$ | Passive abutment displacement |
| $y_7$ | $\ln(\delta_t)$ | Transverse abutment displacement |

This study focuses on the PSDMs that are typically used to generate fragility curves of box-girder bridges [22,26,31]. Therefore, the key engineering demand parameters (EDPs) (primary and secondary) listed in Table 1 are considered. The notations used in Table 1 are used hereafter to facilitate the presentation. The corresponding damages to these components directly map into the bridge system-level damage states and have a significant contribution in defining the bridge system fragility. The primary EDPs of the considered class of bridges include the column and abutment demands which impact the load-carrying capacity and overall stability of the bridge structure. Four levels of damage (i.e., minor, moderate, extensive, and complete) are typically considered in the seismic performance assessment of highway bridges and generating fragility curves. The extensive and complete damage of the components corresponding to the primary EDPs could cause closure of the bridge. However, significant damages corresponding to the secondary EDPs can cause traffic restrictions that are applied to repair the component but will not lead to the entire road closure.

## 3. Machine-learning framework for the performance analysis

To add to the growing body of literature and in the light of enhancing PSDMs, this study applies emerging ML algorithms to provide a comprehensive appraisal of alternative functional forms to estimate bridge seismic demands. This includes multiple linear regression, polynomial regression, decision tree regression, and Gaussian Kernel. A general overview of these algorithms is provided in this section, while more detailed explanations with the corresponding formulations are provided in Appendix B. Interested readers are encouraged to refer to relevant statistical resources (e.g., [9]) for more details regarding the applied ML approaches.

Using ML approaches, predictive models are fitted to find the relationship between the captured peak values of EDPs, obtained from each seismic simulation (explained in the previous section), and the random input variables. Table 2 provides the list of feature candidates, corresponding to the structural modeling parameters and ground motion characteristics, that will be considered to formulate ML-based models to estimate the EDPs in Table 1. For developing the ML-based models, this study adopts the four common hazard computable IMs, including PGA, Sa(0.2 s), Sa(0.3 s), and Sa(1.0 s) (i.e., spectral acceleration at 1.0 s), for the seismic demands of highway bridges in the risk assessment software package [10,26]. In the conventional approach, Eq. (2) were used to express the mean captured EDPs in terms of a single parameter related to the ground motion IM. Fig. 2 demonstrates the general framework for developing ML-based PSDMs for the bridge components.

This study investigates the application of a variety of ML algorithms selected from parametric and non-parametric approaches that cover linear and nonlinear models with different degrees of flexibility [46–48]. Exploring the linear models, this study develops PSDMs using multi-parameters linear regression (MLR), forward stepwise regression (FSR), LASSO, principal component regression (PCR), and partial least squares regression (PLSR). As shown in Eq. (3), MLR is considered as an extension of univariate linear regression (Eq. (2)). MLR expands the explanatory input variables ($x_j$) to $m = 34$ to explain the response variable ($y_i$ in Table 1) using multiple regression coefficients $\alpha_0$ and $\alpha_j$.

$$y_i = \alpha_0 + \sum_{j=1}^{m=34} \alpha_j x_j \tag{3}$$

Increasing the number of input variables in the PSDMs could improve the estimation of responses, though it could increase the model complexity and the chance of overfitting. In order to tackle this challenge and find the optimum number of input variables for the PSDMs,

**Table 2**
The list of input variables for the ML-based models.

| Input variables | Seismic analysis characteristic | Input variables | Seismic analysis characteristic |
|---|---|---|---|
| $x_1$ | Soil type | $x_{18}$ | Foundation rotational stiffness(translational direction) |
| $x_2$ | Girder type | $x_{19}$ | Foundation rotational stiffness(longitudinal direction) |
| $x_3$ | Column cross section shape | $x_{20}$ | Foundation translational stiffness |
| $x_4$ | Abutment type | $x_{21}$ | Concrete strength |
| $x_5$ | Footing type | $x_{22}$ | Reinforcement strength |
| $x_6$ | Fixity type | $x_{23}$ | Mass |
| $x_7$ | Direction of applied ground motion | $x_{24}$ | Damping |
| $x_8$ | Number of spans | $x_{25}$ | Span ratio |
| $x_9$ | Span length | $x_{26}$ | Column height ratio |
| $x_{10}$ | Column height | $x_{27}$ | Ground motion dt |
| $x_{11}$ | Deck width | $x_{28}$ | PGA |
| $x_{12}$ | Number of cells in the box girder | $x_{29}$ | Sa (1.0 s) |
| $x_{13}$ | Girder space | $x_{30}$ | Sa (0.2 s) |
| $x_{14}$ | Top flange thickness | $x_{31}$ | Sa (0.3 s) |
| $x_{15}$ | Superstructure Depth | $x_{32}$ | Mw |
| $x_{16}$ | Reinforcement ratio | $x_{33}$ | R |
| $x_{17}$ | Abutment height | $x_{34}$ | Vs30 |

*Sa represents spectral acceleration and Vs30 represents shear-wave velocity averaged over the 30 m depth of soil.

variable selection (VS) techniques including FSR and LASSO are applied in this study. FSR involves a recursive process, while LASSO uses a regularization term (Appendix B, Eq. B.2). Using these techniques, the key influential input variables are identified, and new predictive models with reduced dimensionality compared to the MLR model are created by eliminating the extraneous variables and merely including the influential variables in the model. Furthermore, this study applies the PCR and PLSR to develop low-dimensional PSDMs based on principal component analysis (PCA). However, opposed to the reduced models generated by VS techniques that use a subset of input variables, the entire list is used by PCR and PLSR to build PSDMs.

Beyond the aforementioned linear models, several polynomial regression (PR) models (listed in Table 3) including 2nd, 3rd, and 4th degrees are considered to develop PSDMs [40]. The corresponding formulations of these PR models are provided in Appendix B, Equations B.2 to B.17. PR predicts the response variable as a function of the $p^{th}$ degree polynomial of the input variables. A general p-degree polynomial model using a single input variable and a 2nd order polynomial using multiple input variables are expressed in Eq. (4) and Eq. (5), respectively.

$$y_i = \alpha_0 + \sum_{q=1}^{p} \alpha_q x^q \tag{4}$$

$$y_i = \alpha_0 + \sum_{j=1}^{m} \alpha_j x_j + \sum_{j=1}^{m} \alpha_{jj} x_j^2 + \sum_{j=1}^{m} \sum_{l=2, l>j}^{m} \alpha_{jl} x_j x_l \tag{5}$$

Moreover, nonlinear PSDMs are formed using ensemble learning algorithms (including Bagging (BG), Least-Squares Boosting (LSB), and Random Forest (RF)) and Gaussian Kernel (GK) method. The ensemble learning approaches train and combine multiple decision trees to form a stronger predictive model following different strategies (see Appendix B) [41–43]. Compared to a single decision tree, the ensemble method provides more accurate and robust models by reducing bias and variance. Contrary to the tree-based approaches, the GK algorithm is a nonlinear regression technique that does not involve an iterative learning process [30]. The Kernel model predicts the response value $y^*$ at a query data point $x^*$ as depicted in Eqs. (6) and (7) using Kernel weight values and the distance to a set of neighboring locations $x_k$ with a bandwidth $\sigma$.

$$y_i^* = \frac{\sum_{k=1}^{q} (w(x^*, x_k) y_k)}{\sum_{k=1}^{q} w(x^*, x_k)} \tag{6}$$

$$w(x^*, x_k) = \exp\left(-\frac{(x^* - x_k)^2}{2\sigma^2}\right) \tag{7}$$

### 3.1. Training, validation, and testing of the developed models

This study uses 85% of the available data (often known as training-validation set) to train the model and validate the trained models while tuning the hyperparameters in order to have an unbiased fitness assessment. This study optimizes hyperparameters of the ML algorithms while training the models, using a grid-search method, by minimizing mean squared errors (MSEs). In this process, 10-fold cross-validation, which is a common practice in machine learning to avoid overfitting, is used to randomly partition data into similar sized 10 subsets with an equal number of data points in each subset [9]. In an iterative process, the model is trained using 9 subsets and is validated on the 10th subset. This process is iterated 10 times. The prediction accuracy and the error term are computed in each iteration, and the average values across all folds are reported as the final values.

The remaining 15% of the data (often known as test data) is used to evaluate the performance of the final tuned models. The evaluation is based on two measures, namely prediction accuracies and the mean squared error (MSE). Therefore, the model's performance is always evaluated on an independent subset that is not used for training and

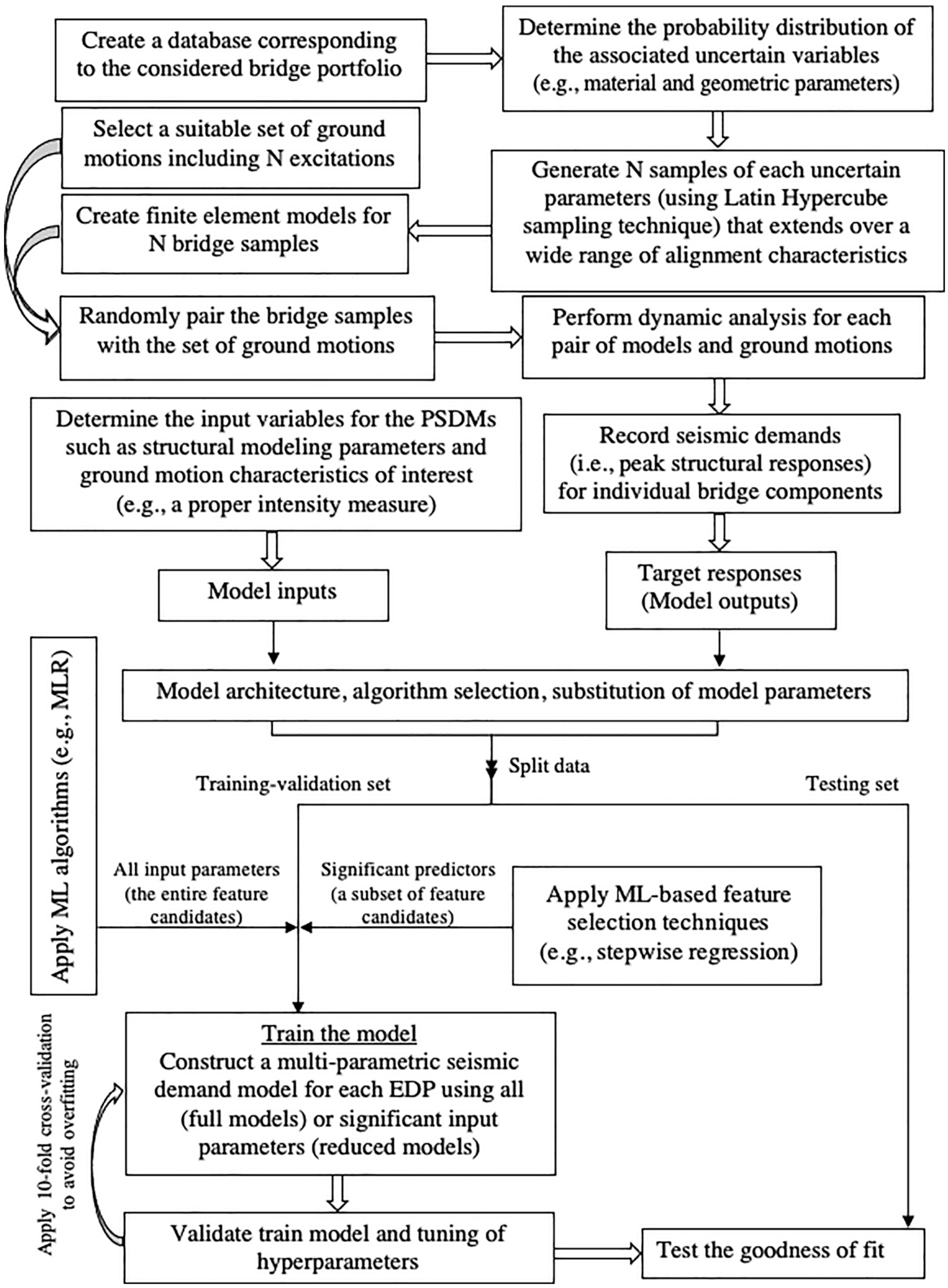


**Fig. 2.** Machine-learning framework for the development of PSDMs.

tuning the model in order to reduce the variance of the fitted model.

## 4. Performance-based development of seismic demand models

### 4.1. Performance comparison of the Machine-learning-driven PSDMs

Two groups of models are considered in this study in terms of the input variables. The list of 39 models, explored in this study, is provided in Table 4. One group is called the full model that includes the entire list of input parameters to develop predictive models, and the other is called reduced dimensionality model that uses a subset of input variables. For the reduced models (e.g., M15), the forward stepwise technique is embedded in the corresponding algorithms. For the reduced dimensionality models, RD is added as the suffix to the model's name in Table 4. For example, BG is used for the Bagging algorithm that generates a full model, while BG-RD is assigned to M7 which is a reduced dimensionality model developed using Bagging algorithm.

Table 4 presents the prediction accuracy and MSE of ML-based models to predict the column curvature ductility, $\ln(\rho_c)$. Also, Fig. 3 shows the variation of MSE across different ML-based models. Similar results for other EDPs are provided in Appendix C (Figs. C1-C5). The following remarks highlight the key findings:

As observed in Fig. 3, the linear models, and particularly the PCR, were ranked lower than the other ML-based models. This implies that to develop the PSDMs for tall concrete box-girder bridges the linear models are not able to appropriately capture the complexity in the data. Among the linear models, MLR provided lower MSEs compared to the other approaches. Besides, PCR was found as the lowest in the ranking in all cases.

Among the polynomial models, those involving the interactive terms such as the multi-degree polynomial models (MLP1, MLP2, MLP3, and MLP4), quadratic model (QPR), and the 2nd degree interactive model

**Table 3**
The list of considered polynomial algorithms (see Appendix B for formulations).

| Annotation | Algorithms | Terms | Type |
|---|---|---|---|
| PR2 | 2nd degree PR | Squared | Full list of input variables |
| PR2-RD | 2nd degree PR | Squared | Reduced dimensionality |
| PR3 | 3rd degree PR | 3rd power | Full list of input variables |
| PR3-RD | 3rd degree PR | 3rd power | Reduced dimensionality |
| PR4 | 4th degree PR | 4th power | Full list of input variables |
| PR4-RD | 4th degree PR | 4th power | Reduced dimensionality |
| PR2INT | 2nd degree PR | Interactive | Reduced dimensionality |
| QPR | Quadratic PR | 1st degree, Interactive, Squared | Reduced dimensionality |
| PLR2 | 2nd degree PR | 1st degree, Squared | Full list of input variables |
| PLR2-RD | 2nd degree PR | 1st degree, Squared | Reduced dimensionality |
| PLR3 | 3rd degree PR | 1st degree, 3rd power | Full list of input variables |
| PLR3-RD | 3rd degree PR | 1st degree, 3rd power | Reduced dimensionality |
| PLR4 | 4th degree PR | 1st degree, 4th power | Full list of input variables |
| PLR4-RD | 4th degree PR | 1st degree, 4th power | Reduced dimensionality |
| PLR23 | 2nd and 3rd degree PR | 1st degree, Squared, 3rd power | Full list of input variables |
| PLR23-RD | 2nd and 3rd degree PR | 1st degree, Squared, 3rd power | Reduced dimensionality |
| PLR234 | 2nd, 3rd, 4th degree PR | 1st degree, Squared, 3rd and 4th power | Full list of input variables |
| PLR234-RD | 2nd, 3rd, 4th degree PR | 1st degree, Squared, 3rd and 4th power | Reduced dimensionality |
| PLR24 | 2nd and 4th degree PR | 1st degree, Squared and 4th power | Full list of input variables |
| PLR24-RD | 2nd and 4th degree PR | 1st degree, Squared and 4th power | Reduced dimensionality |
| PLR34 | 3rd and 4th degree PR | 1st degree, 3rd and 4th power | Full list of input variables |
| PLR34-RD | 3rd and 4th degree PR | 1st degree, 3rd and 4th power | Reduced dimensionality |
| MLP1 | Multi degree PR | (see appendix B) | Reduced dimensionality |
| MLP2 | Multi degree PR | (see appendix B) | Reduced dimensionality |
| MLP3 | Multi degree PR | (see appendix B) | Reduced dimensionality |
| MLP4 | Multi degree PR | (see appendix B) | Reduced dimensionality |

(PR2INT) were found more suitable than the other configurations. Opposed to this observation regarding the column curvature ductility, the polynomial models without the interactive terms, more specifically PLR2, PLR3, PLR4, PLR23, PLR24, and PLR234 were the highest-ranked PR models for the seismic demands corresponding to the deck, foundation, and abutment. In terms of the pure polynomial models (PR2, PR3, and PR4), the lower degrees performed better than the higher degrees in all cases.

For the column curvature ductility, the ensemble learning methods (BG, LSB, RF) produced the most efficient predictive models by having the lowest MSEs and highest accuracies among the considered ML algorithms. Besides, their reduced models presented comparable performance to their full models. For example, the prediction accuracy of M6 and M7 were 0.858 and 0.845 (Table 5). Contrary to the Boosting algorithm, Bagging and random forest developed the best full and reduced models.

In order to predict the column curvature ductility, the GK developed the best-performed model. In fact, the full model derived by GK was

**Table 4**
Performance comparison of the ML-based PSDMs for $\mathbf{ln}(\rho_c)$.

| Models | Algorithm Annotation | Considered input variables | Prediction accuracy | MSE |
|---|---|---|---|---|
| M1 | MLR | Full list | 0.751 | 0.608 |
| M2 | FSR | Reduced | 0.747 | 0.658 |
| M3 | LASSO | Reduced | 0.751 | 0.770 |
| M4 | PLSR | Full list | 0.747 | 0.710 |
| M5 | PCR | Full list | 0.714 | 0.816 |
| M6 | BG | Full list | 0.858 | 0.197 |
| M7 | BG-RD | Reduced (insight from VS) | 0.845 | 0.234 |
| M8 | LSB | Full list | 0.870 | 0.167 |
| M9 | LSB-RD | Reduced (insight from VS) | 0.844 | 0.239 |
| M10 | RF | Full list | 0.852 | 0.213 |
| M11 | RF-RD | Reduced (insight from VS) | 0.832 | 0.275 |
| M12 | GK | Full list | 0.905 | 0.089 |
| M13 | GK-RD | Reduced (insight from VS) | 0.757 | 0.580 |
| M14 | PR2 | Full list | 0.725 | 0.739 |
| M15 | PR2-RD | Reduced (embedded VS) | 0.712 | 0.815 |
| M16 | PR3 | Full list | 0.286 | 0.804 |
| M17 | PR3-RD | Reduced (embedded VS) | 0.700 | 0.885 |
| M18 | PR4 | Full list | 0.690 | 0.940 |
| M19 | PR4-RD | Reduced (embedded VS) | 0.677 | 1.021 |
| M20 | PR2INT | Reduced (insight from VS) | 0.757 | 0.579 |
| M21 | QPR | Reduced (insight from VS) | 0.761 | 0.560 |
| M22 | PLR2 | Full list | 0.754 | 0.595 |
| M23 | PLR2-RD | Reduced (embedded VS) | 0.745 | 0.639 |
| M24 | PLR3 | Full list | 0.752 | 0.603 |
| M25 | PLR3-RD | Reduced (embedded VS) | 0.742 | 0.650 |
| M26 | PLR4 | Full list | 0.750 | 0.615 |
| M27 | PLR4-RD | Reduced (embedded VS) | 0.748 | 0.623 |
| M28 | PLR23 | Full list | 0.754 | 0.595 |
| M29 | PLR23-RD | Reduced (embedded VS) | 0.745 | 0.639 |
| M30 | PLR234 | Full list | 0.754 | 0.594 |
| M31 | PLR234-RD | Reduced (embedded VS) | 0.745 | 0.639 |
| M32 | PLR24 | Full list | 0.741 | 0.659 |
| M33 | PLR24-RD | Reduced (embedded VS) | 0.711 | 0.818 |
| M34 | PLR34 | Full list | 0.743 | 0.646 |
| M35 | PLR34-RD | Reduced (insight from VS) | 0.720 | 0.769 |
| M36 | MLP1 | Reduced (insight from VS) | 0.765 | 0.541 |
| M37 | MLP2 | Reduced (insight from VS) | 0.769 | 0.524 |
| M38 | MLP3 | Reduced (insight from VS) | 0.766 | 0.537 |
| M39 | MLP4 | Reduced (insight from VS) | 0.781 | 0.471 |

found among the highest-ranked models (e.g., for the deck displacement, GK ranked the 3rd best model) in case of predicting the seismic demands of the deck, foundation, and abutment. However, the reduced GK model (GK-RD) did not exhibit good performance in estimating most of the EDPs. The only exception is the case for estimating the active and transverse displacement of the abutments for which the GK-RD model had a close MSE value to the full GK model.

### 4.2. Recommended ML-based PSDMs

The 10 best ML-based models that are recommended for each seismic

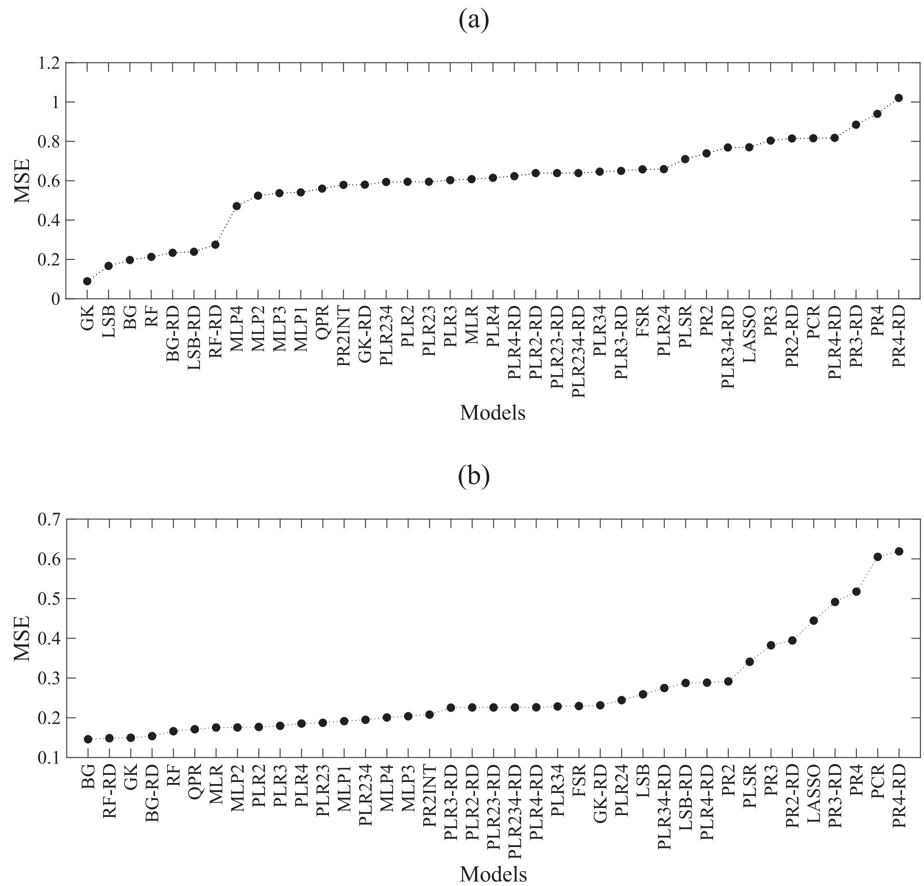


**Fig. 3.** Model performance comparison for predicting (a) $\ln(\rho_c)$ and (b) $\ln(\delta_d)$.

**Table 5**
The list of 10 best-performed ML-based models for each seismic demand model.

| $\ln(\rho_c)$ | | | | | $\ln(\delta_a)$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Model | Algorithm | Type | Number of input variables | Accuracy (%) | Model | Algorithm | Type | Number of input variables | Accuracy (%) |
| M12 | GK | Full | 34 | 90.46 | M12 | GK | Full | 34 | 82.15 |
| M8 | LSB | Full | 34 | 86.96 | M7 | BG-RD | Reduced | 7 | 76.99 |
| M6 | BG | Full | 34 | 85.82 | M8 | LSB | Full | 34 | 76.74 |
| M10 | RF | Full | 34 | 85.25 | M6 | BG | Full | 34 | 76.41 |
| M7 | BG-RD | Reduced | 6 | 84.55 | M11 | RF-RD | Reduced | 7 | 75.92 |
| M9 | LSB-RD | Reduced | 6 | 84.39 | M10 | RF | Full | 34 | 75.33 |
| M11 | RF-RD | Reduced | 6 | 83.24 | M13 | GK-RD | Reduced | 7 | 71.34 |
| M39 | MLP4 | Reduced | 6 | 78.09 | M24 | PLR3 | Full | 34 | 71.19 |
| M37 | MLP2 | Reduced | 6 | 76.88 | M26 | PLR4 | Full | 34 | 71.18 |
| M38 | MLP3 | Reduced | 6 | 76.59 | M22 | PLR2 | Full | 34 | 70.93 |

demand are provided in Table 5 and Table D.1-D.3 (see Appendix D). As observed the ranking varies based on the seismic demand of interest. In general, the ensemble learning methods, Gaussian Kernel, and some of the polynomial models outperformed the linear model which indicates the nonlinearity in the relationship between the probabilistic seismic demands and the random variables. As the order of polynomial increases, the variance in the estimators also increases, and hence, not all polynomial models perform well. More particularly, the Gaussian Kernel and random forest were found as the superior ML algorithms in all cases. Although the full and reduced models developed by random forest performed well, the full Gaussian Kernel model performed better than the reduced model.

### 4.3. Comparison with conventional model

In order to compare the developed ML-based models with the conventional form of PSDM, models M40 and M41 are generated using respectively PGA and Sa(1.0 s) as the ground motion IM for the input variable in the regression model. As stated in the *Introduction*, researchers used various IMs in their proposed PSDMs due to the lack of consensus on a single IM for predicting the bridge demands. The pioneering works investigated different IMs in developing PSDMs with respect to the metrics such as efficiency, practicality, sufficiency, and proficiency. Several studies (e.g., [23,26] tested IMs for portfolios of bridges that are found in HAZUS-MH and indicated that PGA and Sa(1.0 s) are optimal IMs for the purpose of probabilistic seismic demand analysis. Since this study is focused on typical concrete box-girder highway bridges, PGA and Sa(1.0 s) are adopted here.

According to the performance evaluation of the models presented in Fig. 4 and Fig. E.1 (see Appendix E), the ML-driven methods offer significant improvements in the predictive model performance. More specifically, to estimate the column curvature ductility, the ML-based

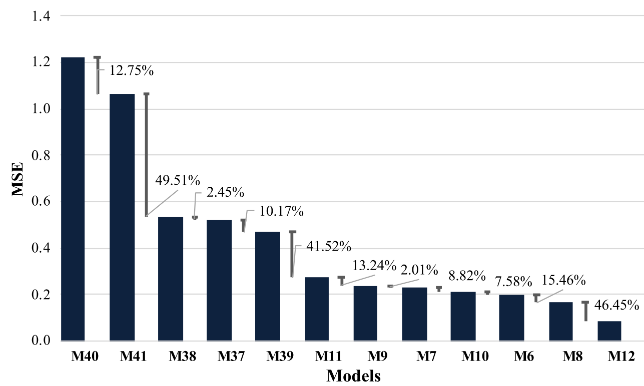


**Fig. 4.** MSE comparison for the PSDM of $\ln(\rho_c)$

models have at least 49% lower MSEs than the conventional PSDMs (Fig. 4). Comparing the conventional model with the 10 best-performed ML-based models to predict other seismic demands, Fig. E.1 indicates 71–78%, 60–69%, and 64–93% reduction in the MSEs respective to the responses of the deck (a), foundation (b, c), and abutment (d, e, f). This implies that the ML-based models could better capture the complexity of the seismic demands. Besides improving the MSEs, the ML-based models propose more reliable predictions by incorporating significant sources of uncertainties (including material, geometric, structural, and ground motion) into the model.

Fig. 5 displays the correlation between the predicted median seismic demands using PSDMs and the targeted responses that were obtained from the finite element dynamic analysis of bridges. As shown in Fig. 5.a, the ML-based results indicate that the predictions and target responses are mostly centered around the 1:1 line resulting in high values of R. In the case of predictions provided by the unidimensional model, differences between the predicted responses and the target values appear to increase over the small and large responses. The higher R values of predicted responses using ML-based models and the lower MSEs indicate that the ML-based PSDMs are more generalizable than the unidimensional regression models.

### 4.4. Machine-learning-based identification of significant variables

Increasing the number of input variables in a regression model boosts model performance up to a specific point after which the performance remains constant or decreases as a result of the increasing complexity of the model (Fig. 6). Among the applied linear ML algorithms, only a few (FSR and LASSO) develop regression models based upon VS techniques that can identify the input variables with the highest influence on predicting the seismic demand of bridges. Consequently, these approaches generate low dimensional models by eliminating the insignificant input variables from the regression model to minimize the MSE. Among the nonlinear approaches, RF provides a measure of the level of importance of each input variable in predicting the response. For example, Fig. 7 displays a comparison of the importance levels of the modeling and structural characteristics in predicting the foundation rotation.

In terms of evaluating the importance level of investigated input variables for the predictive models, different strategies are followed by the applied VS techniques. The FSR starts with zero input variables, and then, each variable is added to the model to determine its associated p-value. The variables are added to the model in the order that they have the smallest p-value or cause the highest increase in $R^2$. This procedure is repeated until all features with significant p-values are added to the model and those variables with the lowest p-values are excluded. In Fig. 6(a), the sequence of the optimum variables is Sa (1.0 s), Foundation translational stiffness, Column height ratio, Reinforcement ratio, and PGA. In Fig. 6(b), the sequence of the optimum variables is Sa (1.0 s), Foundation rotational stiffness (longitudinal direction), Foundation rotational stiffness (translational direction), and Reinforcement ratio.

LASSO and FSR perform the selection of variables in conjunction with creating the predictive model. More specifically, LASSO adds a regularization term to the objective function that eventually induces some coefficients to be equal to zero. Thereby, the input variables with non-zero coefficients are counted as the influential variables. Although FSR and LASSO provide the list of influential variables, they do not provide insight regarding the level of contribution of each variable in predicting the response. However, RF provides a measure of the level of importance of each variable in predicting the response. This measure is computed based on the sum of changes in the MSE of the predictions due to the splits of a particular variable in the growth of regression trees. This measure enables researchers to identify influential variables with the highest level of importance and also understand whether a particular influential variable has more/less impact on the response than the other influential variables.

The accuracy and MSEs of the regression models developed by these three VS techniques were compared to determine the most efficient set of variables. It was noted that LASSO involved more variables to develop the model, whereas the other two techniques generated models with comparable performance and significantly fewer variables. As an example, LASSO, FSR and random forest developed models for the column curvature ductility with prediction accuracy around 75%, 75%, 83% and identified 27, 4 and 4 significant variables, respectively.

The summary of the identified influential variables with the highest prediction power for the investigated seismic demands is provided in Table 6. For example, for predicting the column curvature ductility, the structural modeling parameters including the span length, column height, superstructure depth, reinforcement ratio, and foundation rotational stiffness were identified as the significant variables. This can

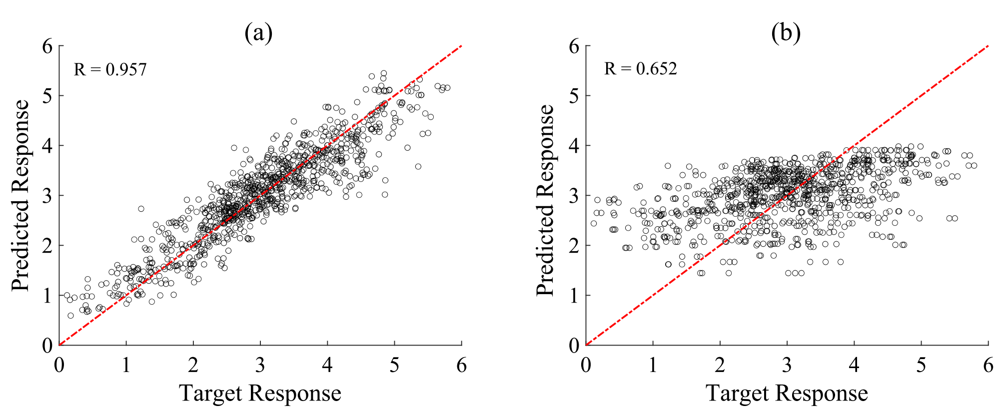


**Fig. 5.** Comparison of the target response computed by finite element analysis and the predicted seismic demands ($y_1$) using (a) GK-based PSDM (M12) and (b) the unidimensional regression models (M40).

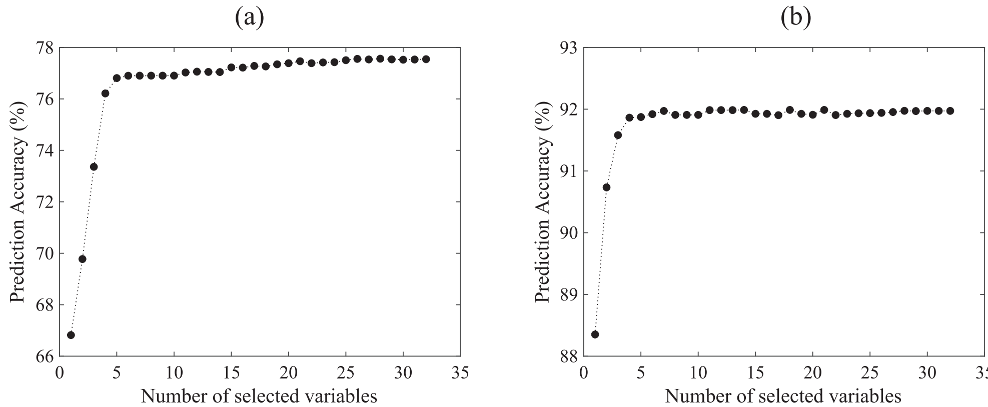


**Fig. 6.** Variation of the prediction accuracy by changing the number of input variables in the FSR models for (a) $\ln(\delta_f)$ (with 5 optimum variables) and (b) $\ln(\theta_f)$ (with 4 optimum variables) – adding more variables beyond the optimum number does not significantly improve accuracy.

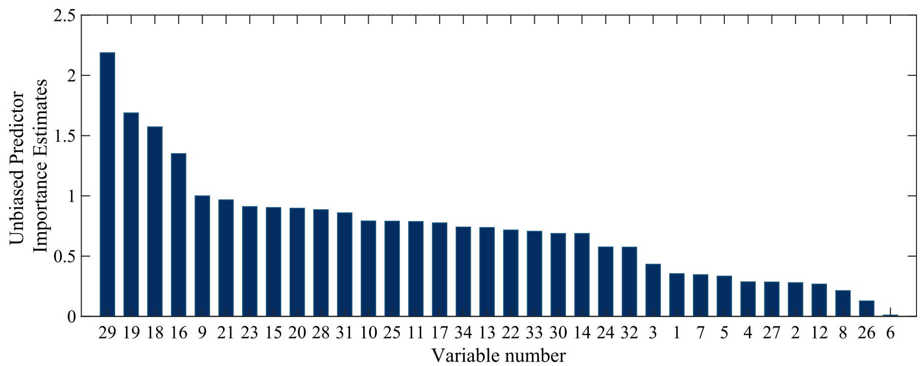


**Fig. 7.** Comparison of the level of importance (obtained from Random Forest) of the input variables in developing the regression model corresponding to the response variable $\ln(\theta_f)$.

**Table 6**
Summary of the identification of significant variables.

| Seismic demand | Significant variables |
|---|---|
| $\ln(\rho_c)$ | Span length, Column height, Superstructure Depth, Reinforcement ratio, Foundation rotational stiffness (translational direction), Sa (1.0 s) |
| $\ln(\delta_d)$ | Span length, Column height, Deck width, Reinforcement ratio, Column height ratio, PGA, Sa (1.0 s) |
| $\ln(\delta_f)$ | Reinforcement ratio, Foundation translational stiffness, Column height ratio, PGA, Sa (1.0 s) |
| $\ln(\theta_f)$ | Reinforcement ratio, Foundation rotational stiffness (translational direction), Foundation rotational stiffness (longitudinal direction), Sa (1.0 s) |
| $\ln(\delta_a)$ | Soil type, Superstructure Depth, Abutment height, Span ratio, Column height ratio, PGA, Sa (1.0 s) |
| $\ln(\delta_p)$ | Span length, Superstructure Depth, Abutment height, Foundation rotational stiffness (translational direction), Foundation rotational stiffness (longitudinal direction), Damping, Sa (1.0 s), Sa (0.3 s) |
| $\ln(\delta_t)$ | Span length, Superstructure Depth, Column height ratio, PGA, Sa (1.0 s) |

be explained in terms of the relationship between these identified influential parameters and the columns' structural characteristics since the strength and stiffness of the bridge columns are a function of the column height and reinforcement ratios. Besides, the high influence of span length on developing demand models could be associated with the direct impact of the bridge span length on the bridge system mass and the flexibility of the bridge superstructure. Moreover, it was observed that the abutment height and soil type significantly influence abutment demands because the force–displacement relationship of the abutments depends on these parameters. Besides, Sa(1.0 s) was found as the most important ground motion characteristic in the predictive model and was noted as the highest-ranked input variable for all seismic demands (e.g., Fig. 7). The higher importance of Sa(1.0 s) compared to the other IMs could be attributed to the mean fundamental period of the studied bridges which was 1.26 sec. The identified influential variables were used to develop most of the reduced dimensionality ML-based models.

## 5. Conclusions

Towards enhancing the reliability of fragility and resilience assessment of highway bridges, this study sought to improve probabilistic seismic demand models of bridge components. To this end, this study presented a systematic appraisal of a variety of ML algorithms with different degrees of interpretability and theoretical complexity. Thereby, the efficiency of 39 parametric and nonparametric ML algorithms, including multiparameter linear and multi-order regressions, ensemble learning methods, and Kernel-based algorithm, were explored. The models were evaluated in full and reduced formats by embedding ML-based variable selection techniques.

Although recent years have seen an increasing number of studies in applying more methodical approaches for improving bridge PSDMs, most of them focused on 1st order multi-parameter linear regression and 2nd order polynomial regression while the application of advanced and robust ML methods such as boosted tree algorithms and Gaussian Kernel requires further investigation. In addition, previous attempts were typically limited to regional-based bridge attributes and selective input variables corresponding to structural characteristics. Also, arbitrary fixed functional forms of PSDMs were considered, and the identification of influential parameters has been rarely conducted in previous attempts. Consequently, the produced PSDMs could include redundant variables that contribute to model complexity and overfitting which could alter the true relationship between the input parameters and the predicted seismic demands.

The proposed ML-based approaches by this study have several advantages compared with the traditional development of PSDMs. One of the main advantages of ML-based models is that they allow tracking the complex relationships between input variables and EDPs without being restricted to relatively simple functional forms as is the case in the conventional approaches. Besides, the implemented ML algorithms not only incorporate various random variables into the predictive model for the uncertainty treatment of material, geometric, structural, and ground motion parameters but also identify the most influential variables in predicting the demands to help to develop efficient and reliable models. Neglecting the uncertainty in the influential variables could lead to unreliable estimation of the seismic demands of the bridge components, while extraneous inputs increase model complexity and the chance of overfitting. In general, implementing variable selection techniques to find the optimum number of input variables improves the model fit, reduces the computation time, makes the model more interpretable, and could also be beneficial to improve the bridge database by giving insight on where to invest resources for updating the input parameters. In summary the findings of this study can be considered under two main categories based on the scope of applications. These categories include improving the prediction power of the seismic demand models and identifying the most influential predictors in the models. Improving the prediction power of PSDMs enhances the reliability of fragility, risk and resilience estimation of bridges and identifying influential parameters reduces computation time and complexity. The following remarks elaborate these findings:

1) Improved predictive models for the seismic demands:
   - Comparison of the efficiency of the proposed ML-based models against the conventional probabilistic seismic demand model (PSDM) showed a significant performance improvement in terms of prediction power when the ML-based models were used. Overall, the findings revealed that the proposed ML-based demand models provide a more reliable prediction of the different seismic demands of bridges and can be used in future studies that conduct performance-based assessments such as the formulation of more robust fragilities and reliability-based assessments.
   - Moreover, the ML-based models were ranked according to their prediction performance. It was observed that, due to the nonlinearity in the relationship between the EDPs and input variables, the ensemble learning methods, Gaussian Kernel, and some of the polynomial models outperformed the linear models. However, in most cases, the models generated by these two algorithms had lower MSEs and higher prediction accuracies than those produced by the polynomial models.
   - In general, the results suggest that Gaussian Kernel and Random Forest provide promising ML-based approaches for developing bridge PSDMs. In particular, the full Gaussian Kernel model and the full and reduced Random Forest models performed well in predicting all investigated EDPs.
   - It was notable that the performance of most of the reduced models, which could be more desirable for practical applications, was comparable to the performance of their corresponding full models.
2) Identification of influential predictors:
   - Furthermore, by applying ML variable selection techniques (LASSO, forward stepwise regression, and random forest), the most influential input variables to explain the seismic demands were identified. It was noted that increasing the number of input variables boosts model performance up to a specific point after which the performance remains constant or decreases as a result of the increasing complexity of the model.
   - EDPs associated with the column curvature ductility, the movements of the deck, foundation, and abutment were considered. The most influential variables for each EDP were summarized that are useful for future studies focused on performance-based assessments of bridge systems.
   - In particular, the spectral acceleration at 1.0 sec (Sa(1.0 s)) was identified as the most influential input variable in the context of the considered seismic demands and bridge classes. This highlights the importance of an adequate estimation of intensity measures in the seismic assessment of bridges.
   - The results indicated that for the column curvature ductility, the structural modeling parameters such as the span length, column height, superstructure depth, reinforcement ratio, and foundation rotational stiffness showed the highest prediction power. This can be explained in terms of the relationship between these identified influential parameters and the columns' structural characteristics since the strength and stiffness of the bridge columns are a function of the column height and reinforcement ratios. Besides, the high influence of span length on developing demand models could be associated with the direct impact of the bridge span length on the bridge system mass and the flexibility of the bridge superstructure. Moreover, it was observed that the abutment height and soil type significantly influence abutment demands because the force–displacement relationship of the abutments depends on these parameters.

Although the results and methodologies presented in this study pave the path towards the application of ML approaches to have a more reliable and efficient estimation of the demands during seismic events, the limitations of this study can be addressed in future works. While this study focused on the class of concrete box-girder bridges as the majority of highway bridges in the West Coast of the U.S., future work will investigate if similar conclusions about the choice of best-performed ML algorithms and influential variables apply to other bridge types (e.g., T-girder, I-girder, slab). While this study revealed the validity of a broad category of ML approaches to improve bridge PSDMs, the forthcoming study will inspect whether different emerging types of neural network [45] approaches have superior predictive capabilities as the applied approaches in this study.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.istruc.2022.02.006.

## References

[1] Ayyub BM, Lai KL. Structural reliability assessment using latin hypercube sampling. Struct Safety Reliab 1989:1177–84.
[2] Ang AH-S, De Leon D. Modeling and analysis of uncertainties for risk-informed decisions in infrastructures engineering. Struct Infrastruct Eng 2005;1(1):19–31.
[3] Baker, J. W., Lin, T., Shahi, S. K., & Jayaram, N. (2011). New ground motion selection procedures and selected motions for the PEER transportation research program. Pacific Earthquake Engineering Research Center.
[4] Cornell CA, Krawinkler H. Progress and challenges in seismic performance assessment. PEER Ctr. News 2000;3(2):1–2.
[5] Kiureghian AD, Ditlevsen O. Aleatory or epistemic? Does it matter? Struct Saf 2009;31(2):105–12.
[6] Filippou FC, Popov EP, Bertero VV. Effects of bond deterioration on hysteretic behavior of reinforced concrete joints. Washington DC: SAC Joint Venture; 1983.
[7] Friedman JH. Multivariate adaptive regression splines. Ann Stat 1991:1–67.
[8] Ghosh J, Padgett JE, Dueñas-Osorio L. Surrogate modeling and failure surface visualization for efficient seismic vulnerability assessment of highway bridges. Probab Eng Mech 2013;34:189–99.
[9] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media; 2009.

[10] HAZUS-MH. Multi-Hazard Loss Estimation Methodology: Earthquake Model HAZUS-MH MR5 Technical Manual. Washington DC: Federal Emergency Management Agency; 2011.
[11] Huang D-S. Radial basis probabilistic neural networks: Model and application. Int J Pattern Recognit Artif Intell 1999;13(07):1083–101.
[12] Hwang H, Jernigan JB, Lin Y-W. Evaluation of seismic damage to Memphis bridges and highway systems. J Bridge Eng 2000;5(4):322–30.
[13] Jekabsons G, Zhang Y. Adaptive basis function construction: an approach for adaptive building of sparse polynomial regression models. Machine Learning 2010;1(10):127–55.
[14] Kameshwar S, Padgett JE. Response and fragility assessment of bridge columns subjected to barge-bridge collision and scour. Eng Struct 2018;168:308–19.
[15] Lin L, Naumoski N, Saatcioglu M, Foo S. Improved intensity measures for probabilistic seismic demand analysis. Part 1: development of improved intensity measures. Can J Civ Eng 2011;38(1):79–88.
[16] Mackie K, Stojadinović B. Probabilistic seismic demand model for California highway bridges. J Bridge Eng 2001;6(6):468–81.
[17] Mander JB, Priestley MJN, Park R. Theoretical stress-strain model for confined concrete. J Struct Eng ASCE 1988;114(8):1804–26.
[18] Mangalathu S, Choi E, Park HC, Jeon J-S. Probabilistic seismic vulnerability assessment of tall horizontally curved concrete bridges in California. J Perform Constr Facil 2018;32(6):04018080. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001231.
[19] Mazzoni, S., McKenna, F., Scott, M. H., & Fenves, G. L. (2006). OpenSees command language manual. Pacific Earthquake Engineering Research (PEER) Center.
[20] Mengotto M, Pinto P. Method of analysis for cyclically loaded RC frames including changes in geometry and non-elastic behaviour of elements under combined normal force and bending. IABSE Congress Reports of the Working Commission. 1973.
[21] Miller A. Subset selection in regression. CRC Press; 2002.
[22] Nielson B. Analytical fragility curves for highway bridges in moderate seismic zones. Doctoral dissertation. Georgia Institute of Technology; 2005.
[23] Padgett JE, Nielson BG, DesRoches R. Selection of optimal intensity measures in probabilistic seismic demand models of highway bridge portfolios. Earthquake Eng Struct Dyn 2008;37(5):711–25.
[24] Pan Y, Agrawal AK, Ghosn M. Seismic fragility of continuous steel highway bridges in New York State. J Bridge Eng 2007;12(6):689–99.
[25] Park J, Towashiraporn P. Rapid seismic damage assessment of railway bridges using the response-surface statistical model. Struct Saf 2014;47:1–12.
[26] Ramanathan K. Next generation seismic fragility curves for California bridges incorporating the evolution in seismic design philosophy. Doctoral dissertation. Georgia Institute of Technology; 2012.
[27] Seo J, Linzell DG. Probabilistic vulnerability scenarios for horizontally curved steel I-girder bridges under earthquake loads. Transp Res Rec 2010;2202(1):206–11.
[28] Seo J, Linzell DG. Horizontally curved steel bridge seismic vulnerability assessment. Eng Struct 2012;34:21–32.
[29] Seo J, Linzell DG. Use of response surface metamodels to generate system level fragilities for existing curved steel bridges. Eng Struct 2013;52:642–53.
[30] Shome N, Cornell CA, Bazzurro P, Carballo JE. Earthquakes, records, and nonlinear responses. Earthquake Spectra 1998;14(3):469–500.
[31] Soleimani F. Fragility of California bridges-development of modification factors (Doctoral dissertation). Georgia Institute of Technology; 2017.
[32] Soleimani F, Vidakovic B, DesRoches R, Padgett J. Identification of the significant uncertain parameters in the seismic response of irregular bridges. Eng Struct 2017;141:356–72.
[33] Soleimani F, Mangalathu S, DesRoches R. A comparative analytical study on the fragility assessment of box-girder bridges with various column shapes. Eng Struct 2017;153:460–78.
[34] Soleimani F. Propagation and quantification of uncertainty in the vulnerability estimation of tall concrete bridges. Eng Struct 2020;202:109812.
[35] Soleimani F. Pattern recognition of the seismic demands for tall pier bridge systems. J Earthquake Eng 2021:1–19.
[36] Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. J Royal Stat Soc: Series B (Statistical Methodology) 2011;73(3):273–82.
[37] Xie Y, Zhang J, DesRoches R, Padgett JE. Seismic fragilities of single-column highway bridges with rocking column-footing. Earthquake Eng Struct Dyn 2019;48(7):843–64.
[38] Xie Y, Ebad Sichani M, Padgett JE, DesRoches R. The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. Earthquake Spectra 2020;36(4):1769–801. https://doi.org/10.1177/8755293020919419.
[39] Zhang R, Mahadevan S. Model uncertainty and Bayesian updating in reliability-based inspection. Struct Saf 2000;22(2):145–60.
[40] Ostertagová E. Modelling using polynomial regression. Procedia Engineering 2012;48:500–6.
[41] Soleimani, F. (2021, August). Analytical seismic performance and sensitivity evaluation of bridges based on random decision forest framework. In Structures (Vol. 32, pp. 329–341). Elsevier.
[42] Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics 2001;29(5):1189–232.
[43] Liaw A, Wiener M. Classification and regression by random forest. R news 2002;2(3):18–22.
[44] Weinberger KQ, Tesauro G. Metric learning for kernel regression. In: In: Artificial Intelligence and Statistics; 2007. p. 612–9.
[45] Soleimani F, Liu X. Artificial neural network application in predicting probabilistic seismic demands of bridge components. Earthquake Engineering and Structural Dynamics 2021.
[46] McCullagh P. Generalized linear models. Routledge; 2018.
[47] Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. Analytica chimica acta 1986;185:1–17.
[48] Jolliffe IT. A note on the use of principal components in regression. Journal of the Royal Statistical Society: Series C (Applied Statistics) 1982;31(3):300–3.