Journal of Building Engineering 95 (2024) 110124

Contents lists available at ScienceDirect

Journal of Building Engineering

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

Full length article

Interpretable machine learning models for displacement demand prediction in reinforced concrete buildings under pulse-like earthquakes

Giulia Angelucci [a], Giuseppe Quaranta [b,*], Fabrizio Mollaioli [a], Sashi K. Kunnath [c]

[a] Department of Structural and Geotechnical Engineering, Sapienza University of Rome, Via Gramsci 53, 00197 Rome, Italy
[b] Department of Structural and Geotechnical Engineering, Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy
[c] Department of Civil & Environmental Engineering, University of California Davis, Davis 95616, CA, USA

ARTICLE INFO

Keywords:
Engineering demand parameter
Gaussian process regression
Genetic programming
Machine learning
Pulse-like earthquake
Reinforced concrete

ABSTRACT

This work proposes a novel procedure to guide the development of machine learning models for estimating the seismic demand in existing reinforced concrete (RC) buildings. The proposed approach is organized across two scales. A large-scale (nonparametric) machine learning model is first obtained by means of Gaussian Process Regression (GPR) using all candidate building attributes and intensity measures. SHapley Additive exPlanations (SHAP) values are utilized to facilitate its interpretation and to assist the rational selection of a small subset of intensity measures, which is finally employed to develop a (symbolic) reduced-scale machine learning model by means of Genetic Programming (GP). Simplified models of archetype buildings are adopted to develop machine learning techniques at both scales, in such a way to alleviate the simulation time for preparing large datasets. Refined models representative of actual buildings are instead considered for the unbiased final assessment.

The proposed approach is applied to develop predictive machine learning models for the maximum inter-storey drift in bare frames, pilotis frames and frames with infills under pulse-like seismic ground motions. Consequently, the critical examination of the SHAP values revealed the most significant intensity measures and unfolded interesting patterns depending on the occupancy rate of the infills. Moreover, the final assessment demonstrates that this approach allows the management of a non-homogeneous building stock consisting of very diverse structural systems (i.e., spanning from existing buildings designed against gravity loads only to buildings that comply with outdated seismic codes) while providing satisfactory predictions of the seismic demand with minimum computational effort.

1. Introduction

The estimation of the seismic demand in structures subject to severe earthquakes is of utmost importance because of its pivotal role in seismic risk assessment and mitigation. Performance-Based Earthquake Engineering (PBEE) provides a valuable methodological framework in this regard: it comprises several steps, of which estimating the structural fragility is especially critical. The structural fragility is the probability of exceeding a predefined Limit State (LS), expressed as Engineering Demand Parameter (EDP), conditioned to different values of a designated ground motion Intensity Measure (IM). In order to quantify the seismic risk, this conditional probability is then convolved with the results of a probabilistic seismic hazard analysis [e.g., 1,2]. Since the

* Corresponding author.
E-mail address: giuseppe.quaranta@uniroma1.it (G. Quaranta).

https://doi.org/10.1016/j.jobe.2024.110124
Received 2 April 2024; Received in revised form 30 May 2024; Accepted 2 July 2024
Available online 10 July 2024

IM operates as an intermediate parameter between seismic hazard analysis (i.e., seismological attributes) and fragility analysis (i.e., structural attributes), it holds a central importance in the assessment and mitigation of seismic risk [e.g., 3].

The selection of the optimal IM has been widely investigated for different structural systems, such as steel frames [e.g., 4,5] and reinforced concrete (RC) frames [e.g., 6–8]. Several criteria have been also proposed to judge the suitability of alternative IMs in characterizing the dominant features of the seismic ground motion. Previous research works agree that efficiency and sufficiency are the most representative metrics to evaluate the predictive capability of an IM. Efficiency implies a relatively small dispersion of the EDP of interest for a given IM, making the selected measure a good descriptor of the structural response [e.g.9,10]. A sufficient IM ensures the dependency of EDP on the IM only, while resulting as minimally correlated as possible with respect to the other seismic ground motion features [e.g.11,12].

Initial studies in this field considered the use of scalar IMs. In particular, peak ground acceleration, peak ground velocity and spectral acceleration at the fundamental elastic period were among the first scalar IMs that were employed to derive the fragility curves. Although this is basically attributable to the immediate availability of the corresponding hazard curves, it is now well-recognized that peak ground acceleration and velocity are poor predictors of damage [e.g., 13]. Spectral acceleration at the fundamental elastic period can be a reliable predictor of the EDP for regular, elastic structures whose dynamic response is dominated by the first mode, but its accuracy for irregular or medium-tall structures is questionable because of the inability to account for the influence of higher modes and/or the period elongation due to the inelastic behavior [14].

The unsatisfactory predictive capabilities of these simple scalar IMs has motivated further efforts towards the search for more performing alternatives. A large number of research works has advanced new scalar IMs. For instance, the spectral acceleration averaged over a prescribed period range [15] was proposed to account for the period elongation of the structures under strong seismic ground motion. Many studies [e.g., 16] have shown that considering the elongation of the fundamental period provides a significantly more efficient predictor of the structural response. This finding was confirmed for existing infilled RC frames by O'Reilly [17], who concluded that the seismic response is not biased by velocity-based ground motion characteristics of the seismic record upon infill collapse when the average of spectral accelerations is employed. A growing mass of evidence is also emphasizing the importance of alternative scalar IMs that depend on velocity, power, or energy [e.g., 18–21]. The application of such IMs was found especially useful for structures subject to pulse-like seismic ground motions [e.g., 22]. This is because pulse-like earthquakes exhibit an impulsive velocity waveform facilitated by the directivity effect that reflects into the transmission of a large amount of energy to the structure in a relatively short time, whereas a more gradual energy transfer occurs in case of far-field, non-pulse-like earthquakes [e.g., 23–25].

Several research works have implemented radically different approaches in the attempt to further improve the prediction of relevant EDPs. A first group of research works has proposed the use of vector-valued IMs instead of a scalar IM. The advent of vector-valued IMs originates from the idea that a relevant EDP may be better predicted by increasing the number of parameters that describes the properties of the seismic ground motions [26]. For instance, Baker and Cornell [27] proposed a three-parameter vector-valued IM consisting of spectral shape, spectral acceleration at the fundamental elastic period, and a measure indicating the spectral acceleration value at a second period (which can be roughly assumed twice the elastic first-mode period). Zengin and Abrahamson [28] proposed a vector-valued IM for near-fault pulse-like seismic ground motion consisting of the so-called instantaneous power and the spectral acceleration at the fundamental elastic period of the structure. The vector-valued IM proposed by Theophilou et al. [29] includes the elastic first-mode period spectral acceleration and a normalized measure of the displacement response spectrum area between the elastic fundamental period and the elongated period of the structure. A second group of research works has proposed the use of step-wise linear or nonlinear models to predict the relevant EDP in place of standard linear regression models. The first attempt of using a nonlinear regression model is likely due to Travasarou and Bray [30], who explored this strategy to predict the seismically induced permanent displacements of slopes. Multi-linear and nonlinear regression models for RC buildings were developed by Vargas-Alzate et al. [31,32]. The search for more accurate predictions of the EDP by means of nonlinear relationships involving multiple IMs has evolved recently by leveraging on the ongoing developments in the field of machine learning [e.g.,33].

Indeed, predictive and generalization capabilities of machine learning algorithms are very attractive for this task. Existing applications in this field can be classified depending on the target outcome of the prediction, namely EDPs or the entire time-history responses. They can also be classified based on the type of input data, namely IMs only or the entire seismic ground motion records, possibly in combination with some building characteristics. For instance, Morfidis and Kostinakis [34] adopted neural networks (consisting of a single hidden layer) to predict the seismic response of RC buildings. The training of the neural networks was performed by using far-field and near-fault natural seismic records whereas lumped plasticity models were adopted for the buildings. Oh et al. [35] developed neural network models (consisting of two hidden layers) to predict maximum inter-storey drift and maximum displacement of RC buildings under earthquake. Synthetic seismic ground motions were adopted to train the neural networks whereas simplified shear-type linear elastic systems were considered for the buildings. Deep learning algorithms were adopted by Wen et al. [36] to predict both maximum inter-storey drift and peak floor acceleration of RC buildings under earthquake. As far as the training phase, they considered far-field, non-pulse-like seismic ground motion records and distributed plasticity models for the buildings. Demertzis et al. [37] compared the performance of several machine learning algorithms in predicting the maximum inter-storey drift of RC buildings, which response was simulated by means of lumped plasticity models. The prediction of the relevant EDP via machine learning algorithms in all these studies is accomplished through a suitable set of IMs, facilitating the calculation of the fragility curves required in PBEE [38,39]. The number of adopted IMs varies between 4 [35] and 14 [34,37]. The time-history of the seismic ground motion is instead considered by Wen et al. [36] in place of IMs for predicting the seismic response of RC buildings. Buildings' attributes were considered as input data in other research works together with the

IMs of the seismic ground motion. For instance, the elastic period for an assigned number of modes is considered by Oh et al. [35]. Number of floors, storey height, longitudinal column spacing, transverse span and seismic design intensity were taken into account by Wen et al. [36]. Demertzis et al. [37] considered the total height of building, the ratios of the base shear experienced by the walls along two horizontal orthogonal directions, and the structural eccentricity.

Although ongoing studies about machine learning techniques to predict the seismic demand are promising, they are based on a standard workflow that poses some critical issues. In fact, almost all the existing research leverages on complex nonparametric machine learning models in the attempt to maximize their accuracy, even though compact symbolic machine learning models are more transparent and easier to use. Moreover, while the preparation of relevant databases by means of refined structural models is useful to enhance the accuracy of the final predictions, it requires significant computational resources. Finally, machine learning algorithms have been tested so far on homogeneous databases consisting of structural systems with similar features. Their feasibility in simultaneously dealing with structural systems that were built according to different design principles (i.e., structures designed under gravity loads and in compliance with some seismic codes) has not yet been investigated. In this context, machine learning models have been mostly carried out to predict the seismic demand in buildings under far-field, non-pulse-like earthquakes whereas limited attention has been paid so far on pulse-like earthquakes. However, the effects of pulse-like earthquakes also deserve proper consideration because of their large damage potential [e.g., 23–25], and several structures, including recent ones, have not been specifically designed to withstand such type of seismic action.

This paper aims at addressing all these issues by proposing a new procedure for the development of machine learning techniques to predict the seismic demand in RC buildings. Ultimately, the proposed methodology is meant at achieving a satisfactory trade-off among conflicting requirements, such as good predictive capability and minimum simulation time as well as interpretability, soundness, easy and broad application of the final machine learning models. The goal of the present study is the prediction of a relevant EDP (i.e., maximum inter-storey drift) for different typologies of RC frames (i.e., bare frames, pilotis frames and frames with infills) subject to pulse-like earthquakes by taking into account a set of IMs together with a few accessible building attributes. The focus of this study is on existing RC structures, and then the considered stock for each typology ranges from buildings that were designed under gravity loads only to buildings that were provided with some seismic details.

2. Prediction of the seismic demand through machine learning techniques

2.1. Proposed approach

Machine learning algorithms can be powerful tools for estimating the EDPs in RC buildings subjected to seismic ground motion given a set of IMs. Nevertheless, their implementation presents some challenges, which can be attributed to both the final form of certain machine learning models and the computational resources needed to prepare large datasets. In fact, although several machine learning models have been implemented to carry out good estimates of significant EDPs, almost all of them cannot be formulated into easy-to-use closed-form expressions [e.g., 34–37]. The preparation of the databases for machine learning algorithms is also a critical task. In this regard, the use of refined lumped plasticity models [e.g., 34,37] and, to a lesser extent, distributed plasticity models [e.g., 36] is preferred, but it presents some inconveniences. Indeed, the effort required for performing a large number of nonlinear time-history analyses with detailed structural models can be prohibitive for conventional computational resources. This is why simplified structural models (i.e., shear-type models) have been also employed for this task [e.g., 35]. Additionally, a large number of details and their ranges of variation must be specified when refined structural models are employed for the numerical simulation of the considered buildings stock. This, in turn, can be a serious effort when preparing non-homogeneous databases consisting of RC buildings designed according to different rules (i.e., buildings designed against gravity loads only or in compliance with different seismic codes).

Therefore, an efficient machine-learning-based framework able to overcome all these issues is proposed in the present work for the development of predictive seismic demand models. Herein, machine learning models are developed through a two-scale approach. Initially, large-scale machine learning models are developed by taking into account all candidate IMs. Suitable tools are then implemented to interpret the corresponding predictive models. This step ensures the physical consistency of the predictive models, but also serves at estimating the relative importance of the candidate IMs and at disclosing useful trends/patterns with respect to the specific building configuration (viz., bare frames, pilotis frames, and frames with infills). This step facilitates the rational selection of a subset of most significant IMs, which is next employed to develop symbolic reduced-scale machine learning models. Although it is reasonable to expect a degradation of the predictive capacity, such compact models are more suitable for practical applications since they are given in a closed-form fashion and involve a fewer IMs. As far as the structural modeling approach is concerned, simplified models of archetype RC buildings based on a limited number of representative parameters are adopted to develop machine learning techniques at both scales. This alleviates the total computational effort and allows to represent the complex nonlinear behaviors of different RC buildings by means of general phenomenological models that depend on a few parameters. This also allows for circumventing the need to specify all structural details and their range of variation, which becomes highly complex or even impractical when attempting to take into account a portfolio of buildings with significantly different characteristics. Because of the inherent approximation of such simplified building models, the final reduced-scale machine learning models are assessed against the results obtained from more refined models representative of actual RC buildings.

The procedure for the development of predictive models through machine learning is outlined in Fig. 1 and has been implemented by combining different computer programs. Nonlinear time-history analyses of simplified and refined structural models are performed by means of Matlab and OpenSees, respectively, whereas all computations related to the development and interpretation of machine learning models have been accomplished using Python.

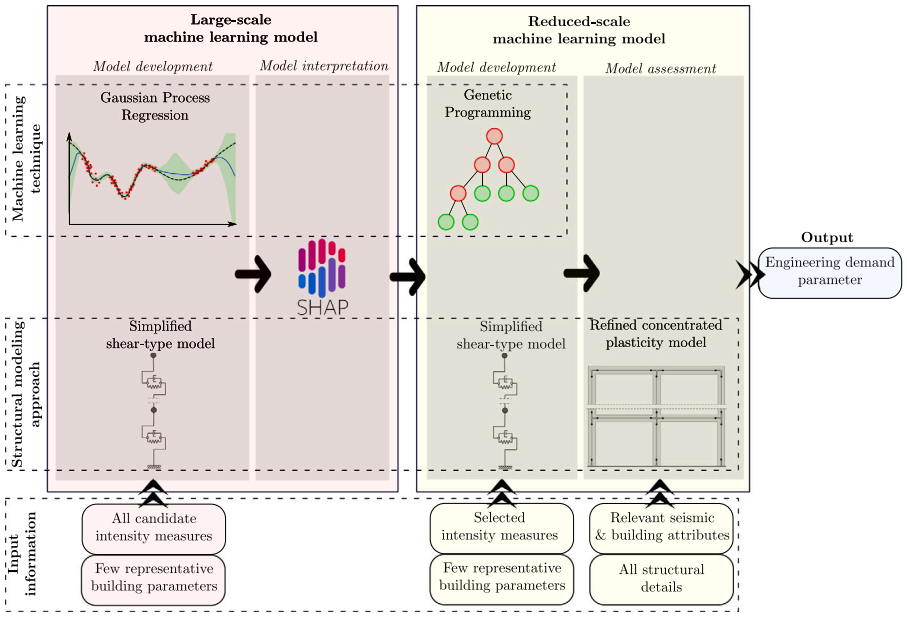

Fig. 1. Proposed workflow for the development of interpretable machine learning models to predict the seismic displacement demand.

2.2. Machine learning tools

Several techniques can be implemented to develop large-scale predictive models of the displacement demand for RC buildings under pulse-like earthquakes. Gaussian Process Regression (GPR) is herein adopted because the database size in the present study is not extremely large, and previous studies have demonstrated that it is capable of handling effectively less data compared to other machine learning techniques [e.g.,40]. However, any other suitable technique can be adopted for this task. GPR is a nonlinear, nonparametric regression algorithm based on Bayesian probability theory [41–44]. By resorting to a function-space-based formalism to provide a concise overview of the GPR technique, it is first considered the regression problem consisting of the output y of a function f given input x:

$$y = f(\mathbf{x}) + \epsilon. \tag{1}$$

Herein, y represents the EDP value and $\mathbf{x}$ is the vector collecting the selected explanatory variables. In the present study, the vector $\mathbf{x}$ is partitioned into two subsets such that $\mathbf{x} = \{\mathbf{b} \quad \mathbf{IM}\}$, where $\mathbf{b}$ is the subset of the explanatory variables that refer to the considered building while $\mathbf{IM}$ collects all the IMs of the seismic ground motion. Finally, ϵ is the so-called noise term. It is noted that both f and ϵ are uncertain. The uncertainty regarding f can be reduced by observing its output at different inputs. The noise term ϵ is a random variable and follows a Gaussian distribution such that $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$: it is always present no matter how many inputs are available, since it arises inevitably because of the error due to the use of the IMs for estimating the EDP in place of a nonlinear time-history analysis. In particular, the function f is distributed as a Gaussian process:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k\left(\mathbf{x}, \mathbf{x}'\right)\right). \tag{2}$$

A Gaussian process $\mathcal{GP}$ is a distribution over functions defined by a mean and a covariance function. The mean function $m(\mathbf{x})$ reflects the expected function value at input $\mathbf{x}$ (i.e. the average of all functions in the distribution evaluated at input $\mathbf{x}$):

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]. \tag{3}$$

It is common expedient to assume the prior mean function $m(\mathbf{x}) = 0$. The term $k\left(\mathbf{x}, \mathbf{x}'\right)$ is the covariance function and it models the dependence between the function values at different input points $\mathbf{x}$ and $\mathbf{x}'$:

$$k\left(\mathbf{x}, \mathbf{x}'\right) = \mathbb{E}\left[\left(f(\mathbf{x}) - m(\mathbf{x})\right)\left(f\left(\mathbf{x}'\right) - m\left(\mathbf{x}'\right)\right)\right]. \tag{4}$$

The function k is the kernel of the Gaussian process. Now, it is supposed that n nonlinear time-history analyses are performed for a building stock by considering a suite of seismic ground motion records. This allows to prepare a training database $D_t = \left\{\mathbf{X}_t, \mathbf{y}_t\right\}$ where $\mathbf{X}_t$ is the matrix with on each row the inputs $\mathbf{x}_i = \left\{\mathbf{b}_i \quad \mathbf{IM}_i\right\}$ for $i = 1, \ldots, n$ while $\mathbf{y}_t = \left\{\mathrm{EDP}_1 \quad \ldots \quad \mathrm{EDP}_i \quad \ldots \quad \mathrm{EDP}_n\right\}^\top$ collects the corresponding outputs. This training database serves at predicting the p outputs $\mathbf{y}^* = \left\{\mathrm{EDP}_1^* \quad \ldots \quad \mathrm{EDP}_j^* \quad \ldots \quad \mathrm{EDP}_p^*\right\}^\top$ corresponding to the matrix $\mathbf{X}^*$ with on each row the new inputs $\mathbf{x}_j^* = \left\{\mathbf{b}_j^* \quad \mathbf{IM}_j^*\right\}$ for $j = 1, \ldots, p$ by drawing f^* from the posterior

distribution φ, with $\mathbf{f}^* = \left\{ f\left(\mathbf{x}_1^*\right) \quad \dots \quad f\left(\mathbf{x}_j^*\right) \quad \dots f\left(\mathbf{x}_p^*\right) \right\}$ (i.e., $\mathbf{f}^*$ is a sample of the function values). This is a conditional multivariate Gaussian distribution $\varphi\left(\mathbf{f}^* | \mathbf{X}_t, \mathbf{y}_t, \mathbf{X}^*\right)$ whose mean and covariance matrix are:

$$\mathbf{K}\left(\mathbf{X}^*, \mathbf{X}_t\right)\left[\mathbf{K}\left(\mathbf{X}_t, \mathbf{X}_t\right) + \sigma_\epsilon^2 \mathbf{I}\right]^{-1} \mathbf{y}_t, \tag{5}$$

$$\mathbf{K}\left(\mathbf{X}^*, \mathbf{X}^*\right) - \mathbf{K}\left(\mathbf{X}^*, \mathbf{X}_t\right)\left[\mathbf{K}\left(\mathbf{X}_t, \mathbf{X}_t\right) + \sigma_\epsilon^2 \mathbf{I}\right]^{-1} \mathbf{K}\left(\mathbf{X}_t, \mathbf{X}^*\right), \tag{6}$$

respectively, where $\mathbf{K}\left(\mathbf{X}_t, \mathbf{X}_t\right)$ is the covariance matrix between all training data, $\mathbf{K}\left(\mathbf{X}^*, \mathbf{X}^*\right)$ is the covariance matrix between the new data while $\mathbf{K}\left(\mathbf{X}^*, \mathbf{X}_t\right)$ and $\mathbf{K}\left(\mathbf{X}_t, \mathbf{X}^*\right)$ are the mixed covariance matrices involving both training data and new data. All covariance matrices are built by means of the selected kernel function k. Moreover, $\mathbf{I}$ is the identity matrix. Hyperparameters optimization in GPR is performed by maximizing the log-marginal-likelihood. This automatically performs a trade-off between bias and variance, and hence avoids over-fitting phenomena. The maximum of the log-marginal-likelihood is found by means of a gradient-based optimizer and, since local maxima might occur in the search space, the optimization problem is solved repeatedly by assuming a different starting point each time to prevent the converge towards a sub-optimal solution. All starting points are determined through Latin hypercube sampling technique, and the final value of the hyperparameters corresponds to the largest value of the log-marginal-likelihood.

It is evident that GPR technique does not assume a functional form to predict the EDP. Moreover, it does not inherently provide an explicit measure of feature importance like some other machine learning techniques. This does not facilitate the appraisal of the role of building features and seismic intensity. Such limitation, in turn, does not reveal the manner in which the corresponding machine learning model works to make the final predictions of the seismic demand given the building features and the seismic intensity. It also hinders the possibility to check the consistency of the relationship between buildings' features, seismic intensity and seismic demand. There are, however, some tools that can shed a light on the inner work of machine learning models. The most adopted approaches are the SHapley Additive exPlanations (SHAP) and, to a lesser extent, the Shapley Additive Global importancE (SAGE), which are two game theoretic methods based on the Shapley value [45]. SHAP interpretation is especially appealing for the present study since it can reflect the influence of the explanatory variables in every EDP value, and can also reveal both positive and negative effects of the input explanatory variables, allowing to check the soundness of the final predictive models. Therefore, once a large-scale machine learning model has been developed by means of the GPR technique, SHAP values are calculated. It is noted that direct calculation of the exact SHAP values is time-consuming. Therefore, a simplified method known as KernelSHAP [46] is implemented to approximate the SHAP values using all available data.

Apart from being useful to understand how machine learning models work, SHAP interpretation has been also proposed to assist the selection of the most relevant explanatory variables involved into a model [47]. In this regard, recent studies have highlighted that SHAP interpretation can outperform conventional approaches for features selection [48]. Therefore, the critical examination of large-scale machine learning models developed by means of the GPR through SHAP interpretation is accomplished to reduce the initial set $\mathbf{x}$ to $\overline{\mathbf{x}} = \left\{ \mathbf{b} \quad \overline{\mathbf{IM}} \right\}$ where $\overline{\mathbf{IM}} \subset \mathbf{IM}$. This, in turn, is adopted to develop reduced-scale machine learning models by means of Genetic Programming (GP). This is a very common parametric regression technique that simulates biological breeding and Darwinian evolution [49,50] in the attempt to infer symbolic predictive models from available data. Accordingly, a population of candidate computer programs (i.e., a predictive model that includes the instructions for combining variables, constants and operators) are manipulated iteratively through a sequence of genetic operators, namely selection, crossover and mutation. This iterative procedure starts with a random population of computer programs and runs until a stopping criterion is not fulfilled. One of the most critical issues in implementing GP algorithms is the uncontrolled growth of the model complexity. This phenomenon is also known as bloat and identifies an excessive growth of the model complexity that leads to a small increment of the model accuracy over the training dataset. In contrast, it often goes with a significant worsening of the predictive capability against new data. Out of the available approaches to cope with bloat in GP [51], the strategy implemented in the present work assumes model accuracy and model complexity as conflicting criteria as proposed by Ekart and Nemeth [52]. Therefore, tournament selection operator and Pareto's nondomination concept are implemented as follows:

- a number of candidate computer programs (the comparison set) is randomly picked from the current population;
- another computer program is randomly picked from the current population;
- if the new computer program is not dominated by any individual from the comparison set, then it is selected;
- otherwise (i.e., . if the new computer program is dominated by some individuals from the comparison set), the procedure is repeated by randomly selecting another computer program from the current population.

A non-dominated solution is better than the others over some criteria but worse over the other criteria. While the root mean squared error function is considered to quantify the model accuracy, the threshold function proposed by Ekart and Nemeth [52] is adopted to measure the model complexity. Both objective functions are evaluated against the training database $\mathcal{D}_t = \left\{ \overline{\mathbf{X}}_t, \mathbf{y}_t \right\}$, where $\overline{\mathbf{X}}_t$ is the matrix with on each row the inputs $\overline{\mathbf{x}}_i = \left\{ \mathbf{b}_i \quad \overline{\mathbf{IM}}_i \right\}$. Thanks to the symbolic formulation, reduced-scale machine learning models developed by means of GP can be readily adopted to predict the outputs $\mathbf{y}^*$ corresponding to the matrix $\overline{\mathbf{X}}^*$ with on each row the new inputs $\overline{\mathbf{x}}_j^* = \left\{ \mathbf{b}_j^* \quad \overline{\mathbf{IM}}_j^* \right\}$. This facilitates their direct implementation, assessment, and interpretation.

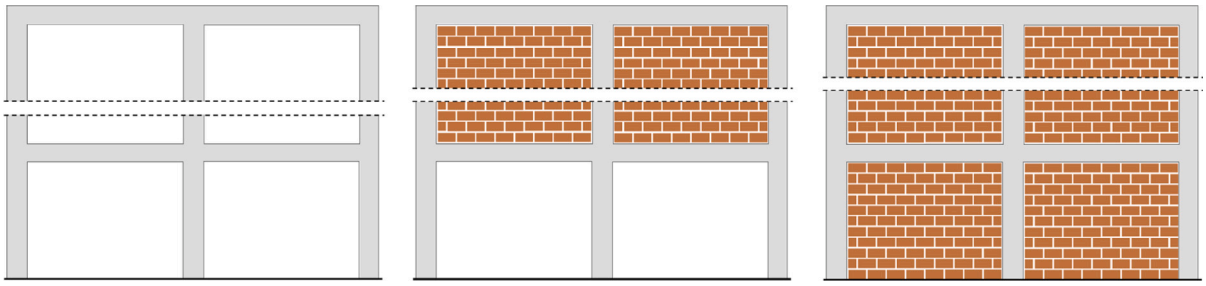

Fig. 2. Filling distribution of masonry panels within the considered RC buildings: bare frame (left); pilotis frame (center); infilled frame (right).

3. Seismic demand and intensity measures for RC buildings under pulse-like earthquakes

3.1. Numerical modeling of reinforced concrete buildings

Multi-story RC buildings representative of the majority of the existing Italian building stock are analyzed in the present work. Hence, the present study focuses on existing RC buildings designed for gravity loads only or in compliance with outdated seismic codes. The following typologies of RC buildings are analyzed (see Fig. 2): bare frames without masonry infills; pilotis frames with infill-free ground floor; infilled frames with uniformly infilled stories.

As far as the development of machine learning models is concerned, the considered RC buildings are not intended to replicate actual structures. Conversely, simplified phenomenological models of archetype RC buildings are considered in such a way to account for the huge variability of existing RC building features by means of a few, yet relevant and accessible, data. The use of simplified structural models alleviate the computational effort required to prepare the databases upon which machine learning models are developed while ensuring their application to a large and non-homogeneous building stock. Accordingly, an equivalent shear-type model (ESTM) is herein employed for this task, as shown in Fig. 3. The resulting multi-degree-of-freedom (MDOF) model consists of a discrete, stick-like system of joint masses lumped at each floor level of the structure and connected by nonlinear shear springs that simulate the lateral stiffness and strength characteristics of the structural members. The relevant backbone at each story is then defined in terms of inter-storey shear-displacement relationship by assuming that column ends are restrained against rotation. While flexural deformations are not negligible in new RC structures conforming to modern seismic codes, shear-type deformations are usually prevalent in existing RC buildings designed for gravity loads only or in compliance with outdated seismic codes. In such a case, the assumption of shear-type behavior allows to reproduce the seismic response of existing moment resisting frames with a reasonable level of approximation as pointed out by previous studies [e.g., 53,54]. The effectiveness of the ESTM model has been demonstrated in previous research works by comparing the results with those derived from more accurate nonlinear models [55].

A generalized degrading hysteretic behavior characterized by stiffness degradation, strength degradation and pinching is implemented in the ESTM to represent the inelastic cyclic response of the frames at each story [56]. The overall cyclic behavior of the constitutive model is illustrated in Fig. 3. It is governed by four parameters, namely p, α, β, and γ. The post-yielding stiffness is regulated by the strain hardening ratio p, namely the ratio between post-yield tangent and initial elastic tangent. The parameter α controls the unloading stiffness degradation (i.e., unloading degradation is negligible when α is larger than 20). The strength degradation is governed by the parameter β (if it is equal to 0, then no strength degradation due to energy dissipation occurs). Lastly, the parameter γ denotes the pinching effect due to closing cracks during reloading ($\gamma = 1$ represents a non-pinching hysteretic behavior whereas $\gamma = 0$ corresponds to the maximum pinching effect). Moreover, K_0 is the initial elastic stiffness and μ is the ductility ratio while F_y and u_y are yield strength and yield displacement, respectively. Once the base shear seismic coefficient C_y is assigned (i.e., the ratio between the maximum base shear and the total weight of the structure), a large variety of structural behaviors can be simulated in a straightforward way by tuning the model parameters p, α, β, and γ.

Refined models of actual RC frames are instead adopted for the unbiased assessment of the final reduced-scale machine learning models, as shown in Fig. 4. It is pointed out that buildings other than those adopted to develop machine learning models are considered for this task. These refined models are developed using the open-source software platform OpenSees [57]. Beam with fiber-hinges element objects are used to simulate the response of beams and columns. Since these structural members are expected to experience nonlinear deformations at beam–column connections, plasticity is concentrated over specified hinge lengths of the element ends, which is assumed to be equal to the cross-section height of the element itself. Beam with hinges element are actually divided into three parts: two inelastic fiber-hinges at both ends and linear-elastic region in the middle [58]. Fiber section model is adopted to directly account for uniaxial bending and axial force. Unlike other types of distributed plasticity elements where the Gauss integration points distribute along the entire element length, the beam with hinges element localizes the integration points in the hinge regions (namely, two integration points per hinge are used to represent the curvature distribution). Cracked section properties are considered in the elastic portion of the beam–column elements. Inelastic fiber section are defined at the end of the elements with constitutive relationships for both confined and unconfined concrete characterized by the uniaxial Kent–Scott–Park concrete material object (`Concrete01` of the OpenSees library) with degraded linear unloading/reloading stiffness and no tensile

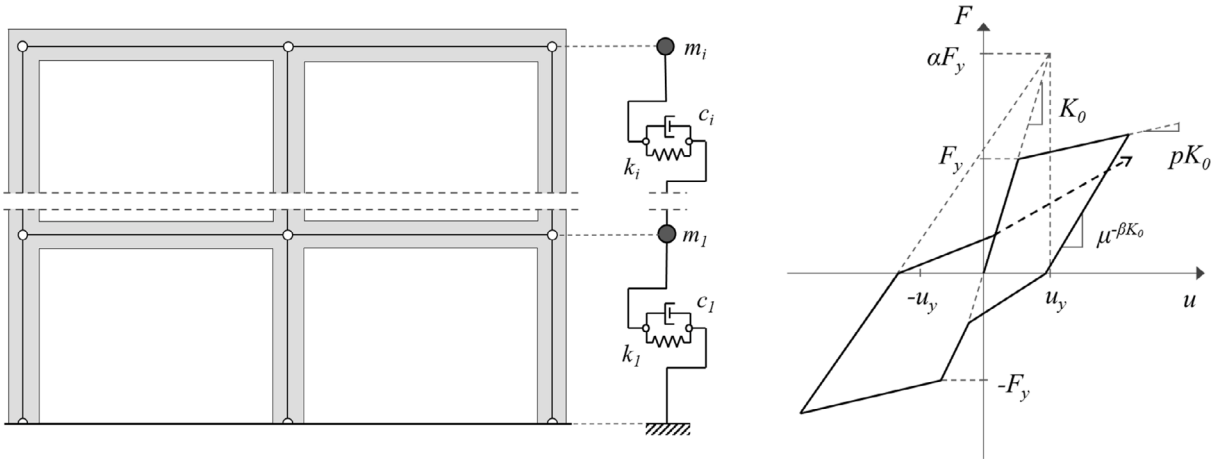

Fig. 3. Simplified numerical model of multi-storey RC frames: equivalent shear-type model and generalized degrading hysteretic behavior.

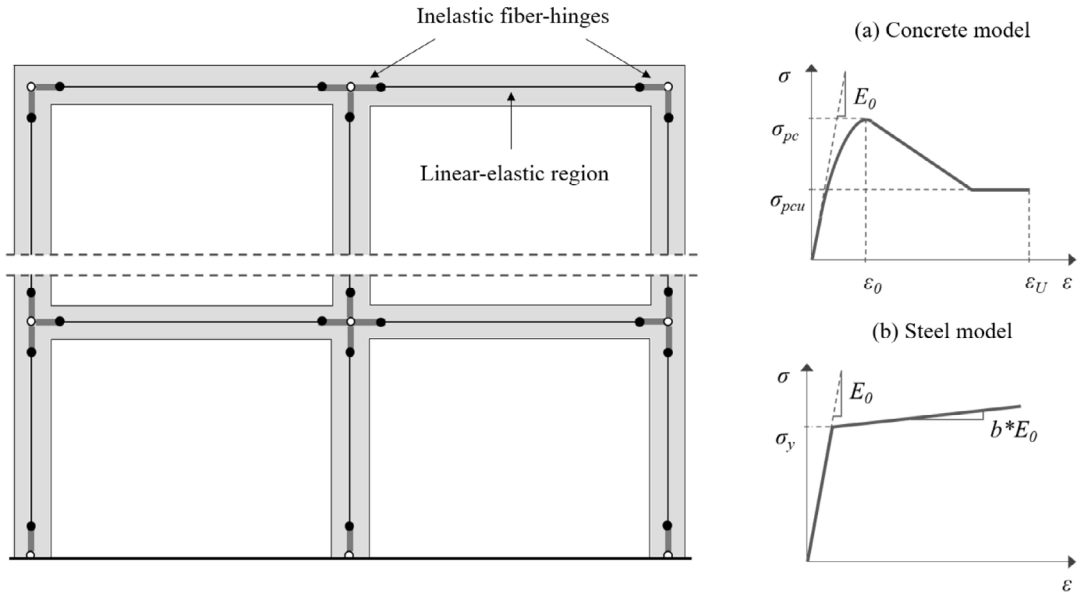

Fig. 4. Refined numerical model of multi-storey RC frames: concentrated plasticity model and stress–strain relationships for steel and concrete.

strength. A simple uniaxial bilinear model has been selected for the longitudinal reinforcement fibers (`Steel01` of the OpenSees library). The structural masses are concentrated at the nodes and a Rayleigh damping model is implemented.

The equivalent diagonal strut approach is employed in this study to simulate the in-plane seismic response of masonry infills within, both, simplified and refined RC frame models. Although such a macro-modeling approach cannot simulate accurately the complex response of structures under cyclic loading, it aligns with the scopes of the present study since it provides a simple, yet efficient, tool to describe the contribution of the masonry panels to the global behavior of RC buildings. Each infill panel is thus simulated using a single diagonal compression-only strut element. The in-plane seismic response of the infill panel is defined by a multi-linear force–displacement backbone consistent with the constitutive model by Decanini et al. [59], as shown in Fig. 5.

The first branch of the lateral force–displacement skeleton curve represents the initial uncracked stiffness K_0 up to the onset of cracking at F_{mf}, and it is determined according to simple mechanical considerations. The second branch corresponds to the post-cracking phase up to the development of the peak strength at F_{mfc}. The stiffness at complete cracking K_{mfc} and the peak strength of the equivalent strut are computed following the procedure proposed by Decanini and Fantin [60], which takes into account multiple failure mechanisms (i.e., bed-joint sliding failure mode, diagonal tension failure mode, diagonal compression failure mode and corner compression failure mode). Lastly, the descending branch of the post-peak strength deteriorates linearly until reaching zero residual infill strength at the ultimate displacement. Drift limits corresponding to cracking u_f, peak u_{fc}, and collapse u_r are determined following the guidelines provided by Liberatore et al. [61].

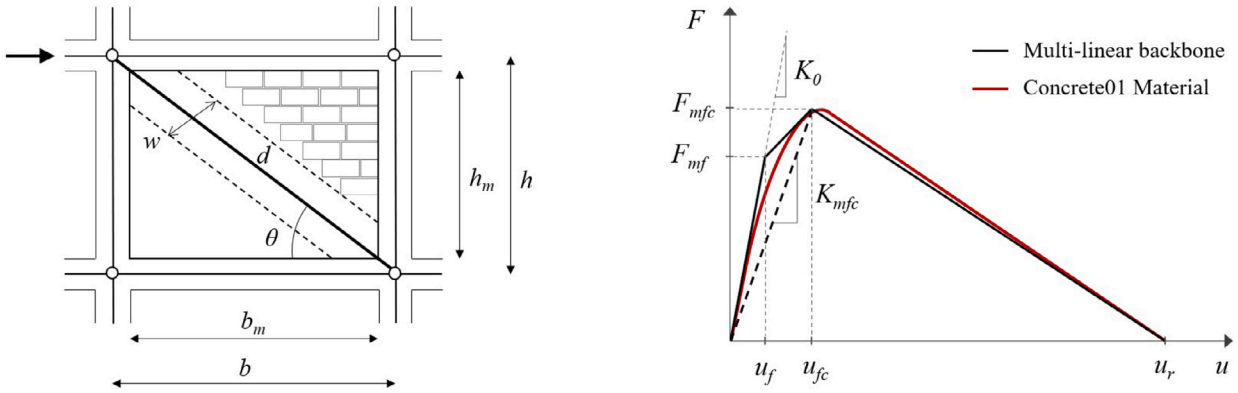

Fig. 5. Macro-model for the in-plane masonry infill response simulation.

Table 1
List of candidate IMs adopted to develop predictive machine learning models of IDI_{max}.

IM	Definition
Peak ground acceleration, PGA	$\max_t\{\ddot{u}_g(t)\}$
Arias intensity, I_A	$(\pi/2g)\int_0^{t_f}\ddot{u}_g^2(t)\mathrm{d}t$
Spectral acceleration at T_1, S_{pa,T_1}	$S_{pa}(T_1)$
Spectral seismic coefficient at T_1, C_{y,T_1}	$C_y(T_1)$
Average spectral acceleration, $S_{a,avg}$	$\left[\prod_{\ell=1}^{q}S_{pa}(T_\ell)\right]^{1/q}$
Peak ground velocity, PGV	$\max_t\{\dot{u}_g(t)\}$
Housner intensity, I_H	$\int_{0.1}^{2.5}S_{pv}(T)\,\mathrm{d}T$
Spectral velocity at T_1, S_{pv,T_1}	$S_{pv}(T_1)$
Spectral relative input energy at T_1, E_{ir,T_1}	$-\int_0^{t_f}\ddot{u}_g(t)\,\dot{u}_r(t)\,\mathrm{d}t$
Relative input equivalent velocity spectrum intensity, $VE_{ir}SI$	$\int_{0.1}^{3.0}\sqrt{2E_{ir}(T)}\mathrm{d}T$
Modified relative input equivalent velocity spectrum intensity, $MVE_{ir}SI$	$\int_{0.2T_1}^{2T_1}\sqrt{2E_{ir}(T)}\mathrm{d}T$
Modified damping equivalent velocity spectrum intensity, MVE_dSI	$\int_{0.2T_1}^{2T_1}\sqrt{2E_d(T)}\mathrm{d}T$
Modified effective cyclic energy, ECE	$\max_t\int_t^{t+\Delta t}\left(E_h(t)+E_d(t)\right)\mathrm{d}t$ for $\dot{u}_g(t)=\dot{u}_g(t+\Delta t)=0$
Peak ground displacement, PGD	$\max_t\{u_g(t)\}$
Spectral maximum displacement at T_1, u_{max,T_1}	$u_{max}(T_1)$
Effective peak of input energy, ΔE_{ir}	$\max_t\{\Delta E_{ir}(t)\}$
Effective peak of dissipated energy, ΔE_{h+d}	$\max_t\{\Delta\left(E_h(t)+E_d(t)\right)\}$
Effective peak of hysteretic energy, ΔE_h	$\max_t\{\Delta E_h(t)\}$
Peak of input power, $P_{E_{ir}}$	$\max_t\{P_{E_{ir}}(t)\}$
Peak of dissipated power, $P_{E_{h+d}}$	$\max_t\{P_{E_h}(t)+P_{E_d}(t)\}$
Peak of hysteretic power, P_{E_h}	$\max_t\{P_{E_h}(t)\}$

The contribution of masonry panels within the simplified structural model of the RC frames is numerically simulated by assuming that the shear spring of the RC frame and the shear spring of the infill elements act in parallel at the same story level. A similar assumption is made to account for the contribution of masonry panels within the refined models developed via OpenSees, where the infill elements are simulated using the uniaxial Kent–Scott–Park concrete material object (`Concrete01` of the OpenSees library) as suggested by Noh et al. [62] and Di Domenico et al. [63]. Since this material model provides a parabolic stress–strain response up to the maximum compressive strength, the actual deformation at peak strength is reduced by half of its actual value to obtain a realistic estimate of the elastic stiffness (i.e., tangent line of the first linear branch of the response envelope).

3.2. Engineering demand parameter and intensity measures

The maximum inter-storey drift ratio IDI_{max} is the target EDP. The list of candidate IMs that will be considered for its prediction through machine learning models is provided in Table 1. Herein, t stands for time and Δt for the time interval, T is the period, g is the gravity acceleration, and t_f is the total duration of the seismic ground motion whereas $\ddot{u}_g(t)$, $\dot{u}_g(t)$ and $u_g(t)$ are the seismic ground acceleration, velocity and displacement, respectively. The relative velocity time history is denoted as $\dot{u}_r(t)$ while S_{pa} and S_{pv} are the pseudo-acceleration and the pseudo-velocity spectrum, respectively. On the one hand, E_{ir}, E_d and E_h are the relative input energy, the total dissipated energy and the hysteretic energy, respectively. On the other hand, $P_{E_{ir}}$, P_{E_d} and P_{E_h} are the relative input power, the total dissipated power and the hysteretic power, respectively. Finally, T_ℓ (with $\ell=1,\ldots,q$) is the ℓth elastic period of the structure (T_1 being the fundamental period). A damping ratio of 5% is considered.

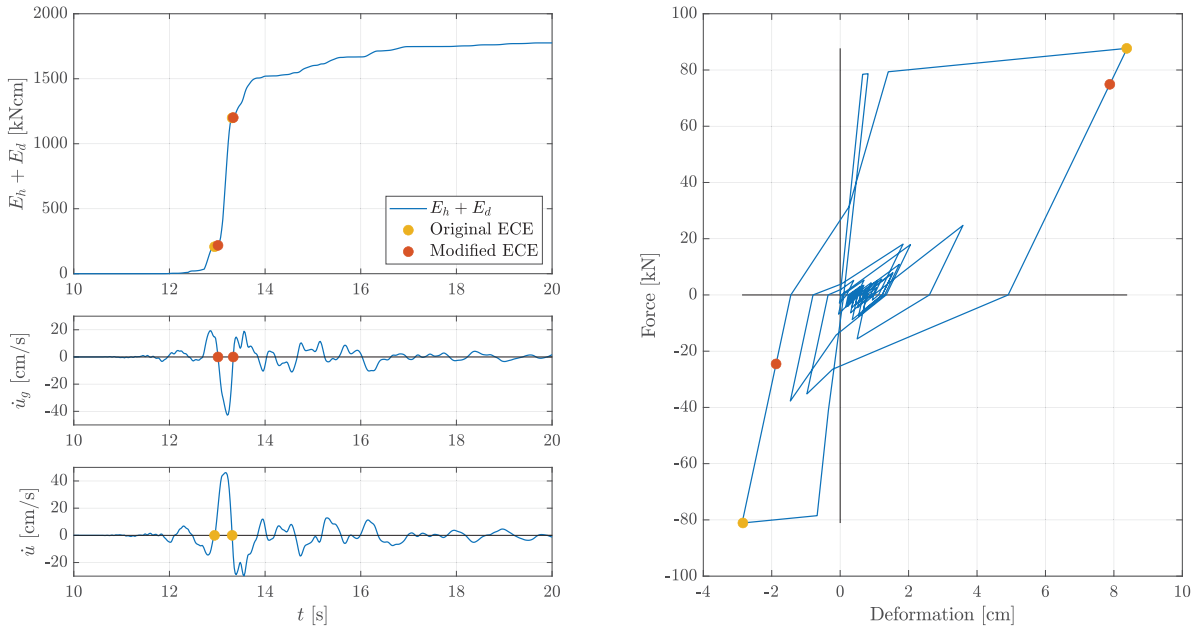

Fig. 6. Comparison between original and modified ECE.

The inventory consists of 21 candidate IMs. It encompasses both non-structure-specific and structure-specific IMs as well as acceleration-, velocity-, displacement-, energy-, and power-based IMs. This inventory of candidate IMs has been prepared taking into account previous studies [e.g., 6,33,64], but some novelties are implemented in the present work.

The Effective Cyclic Energy (ECE) considered in Table 1 is a modified version of the original formulation by Kalkan and Kunnath [65], who recognized that the peak-to-peak dissipated energy demand imposed on a single-degree-of-freedom (SDOF) system over an effective ground motion duration is equivalent to the time interval between two zero-crossings of the velocity time response. This aligns with the concept of Momentary Input Energy (MIE) developed by Hori and Inoue [66] and extended recently by Fujii [67] as well as Fujii and Murakami [68]. Both effective cyclic energy and momentary energy, in fact, rely on the evidence that input energy at the corner points of a half force–deformation hysteresis loop (i.e., when velocity time response is null during reversal) is equal to the sum of damping and hysteretic energy, as the kinetic energy vanishes. Since this energy measure is defined as the peak between two consecutive zeros of the velocity time response, it is directly linked to the peak absolute cyclic deformation experienced by the structural system during the seismic ground motion. Thus, both ECE and MIE serve as structure-specific IMs of the critical seismic energy transferred by impulsive ground motions to the structure in a single effective cycle. In the present study, a modified version of the ECE is introduced to carry out an IM based on the seismic ground motion only that can be applicable effortlessly to MDOF systems. In particular, instead of considering the velocity response, the peak value between two zero-crossing points in the velocity seismic ground motion is considered. So doing, ECE only depends on the reversals of the seismic ground velocity, without explicit involvement of the actual structural response. As an example, original and modified ECE are compared in Fig. 6 considering an inelastic SDOF system subject to 2000 Yountville earthquake (Napa Fire Station #3). It is noted that the peak structural response is related to a single effective hysteretic cycle, which causes a sudden excursion of the structure into the inelastic range. It can also be observed that the proposed version of the ECE is a satisfactory proxy of the seismic demand, given it relies solely on seismic ground motion while does not depend on structural features.

Regarding seismic energy, it has been recognized that it gradually increases and reaches the peak at the end of the seismic ground motion in case of far-field earthquakes [69]. Conversely, it exhibits a more discontinuous trend in case of pulse-like near-fault earthquakes [69]. Therefore, the effective peak values of relative input energy, total dissipated energy and hysteretic energy in Table 1 correspond to their largest jumps over the total duration of the pulse-like seismic ground motion, which are denoted as $\Delta E_{ir}(t)$, $\Delta\left(E_h(t)+E_d(t)\right)$, and $\Delta E_h(t)$, respectively. Similarly, peak values of relative input power, total dissipated power and hysteretic power in Table 1 are their corresponding maxima over the total duration of the seismic ground motion, which are denoted as $P_{E_{ir}}(t)$, $P_{E_h}(t)+P_{E_d}(t)$, and $P_{E_h}(t)$, respectively. As far as the calculation of ΔE_{ir}, ΔE_{h+d}, and ΔE_h is concerned, there are no details in the relevant literature, and thus it is inferred that their detection has been done manually thus far. An automatic procedure based on the work by Tyson et al. [70] is implemented here, which consists of the following steps.

- Initially, energy values are scaled so that the maximum value is equal to one.
- The probability density estimate of the normalized energy value is next obtained via kernel density estimation (Normal kernel is adopted for this task, and the number of points is down-sampled with a scale factor equal to 10).

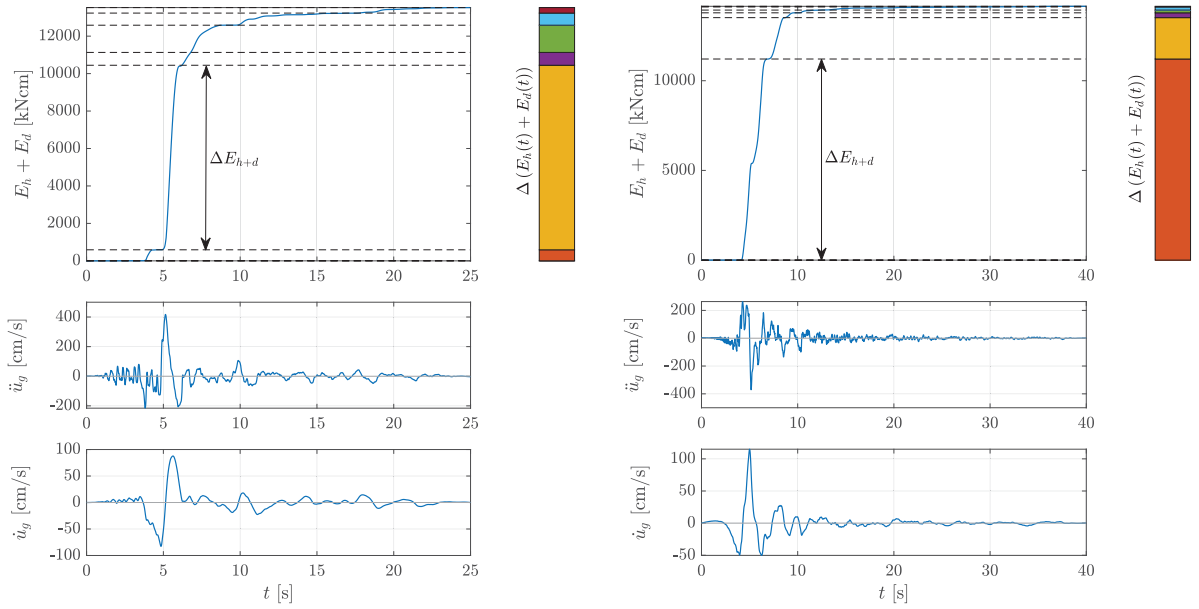

Fig. 7. Application of the proposed automatic procedure to identify the effective energy peaks.

Table 2
Hysteretic parameters for the simplified numerical models of multi-story RC frames adopted for the development of machine learning models.

C_y	p	α	β	γ
0.08	0.01	2.0	0.1	0.6
0.15	0.05	2.0	0.0	0.8
0.25	0.05	2.0	0.0	0.8

- The peaks in the probability density estimate of the energy are then identified, and those below an assigned threshold are discarded because due to noise or attributable to numerical effects (the threshold is set to 2% of the highest peak in the probability density estimate).
- Finally, all energy peaks are calculated as the difference of the energy values between two consecutive peaks of the probability density estimate. The highest energy peak (i.e., effective energy) is considered as the IM.

As an example, Fig. 7 illustrates the application of the proposed procedure for the 1994 Northridge earthquake (Newhall – W Pico Canyon Rd. station) and the 1979 Imperial Valley earthquake (EC Meloland Overpass FF station). Notably, the proposed statistic-based criterion was able to recognize a very short energy plateau occurring for the 1979 Imperial Valley earthquake, and the energy jumps preceding and following such a step have been automatically merged into a single one.

4. Numerical investigation

4.1. Buildings data

Simplified numerical models of archetype multi-story RC buildings with 2, 4, 6, 8 and 10 floor have been prepared to develop the predictive machine learning models. The ranges of the fundamental period T_1 are 0.575–1.951 s for bare frames, 0.475–0.560 s for pilotis frames, and 0.120–0.325 s for infilled frames. Three combinations of model parameters are adopted for each frame in order to represent different degrading hysteretic behaviors commonly observed in existing residential Italian RC buildings. These model parameters are listed in Table 2, where the lowest C_y value is associated with structures designed for gravity loads only [71], which were built without suitable reinforcement detailing and using materials with poor mechanical properties. The highest value of C_y denotes structures designed for seismic loads in compliance with outdated seismic codes dating back to the 90s [72].

The unbiased assessment of the final reduced-order predictive machine learning models is performed on 4-storey and a 6-storey, three-bay actual RC frames. Their overall geometries and structural details are provided in Fig. 8 and Table 3, respectively. The uncracked concrete elastic modulus is 15.5 GPa whereas the concrete compressive strength is 27 MPa. Yielding stress of the steel is 440 MPa. A Rayleigh damping model is assumed with a constant damping ratio equal to 5%. The modal analysis based on the

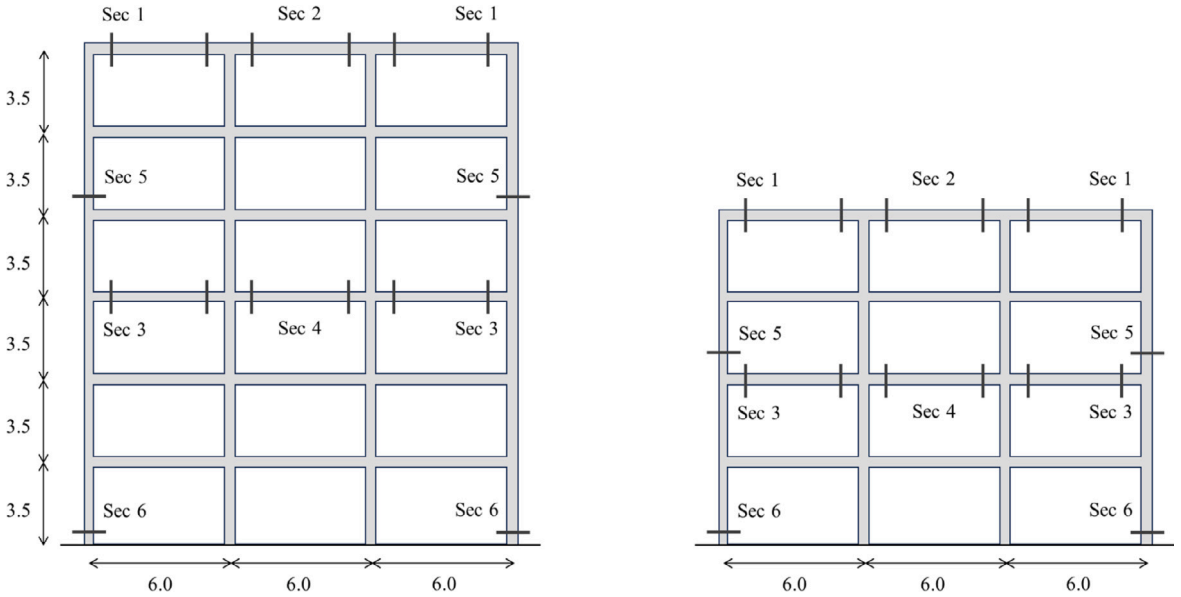

Fig. 8. Overall geometry of the refined numerical models of the multi-story RC buildings adopted for the assessment of machine learning models (units in [m]).

Table 3
Structural details of the refined numerical models of the 4-storey and a 6-storey, three-bay RC frames adopted for the assessment of machine learning models (legend: b is the width of the cross-section; h is the height of the cross-section; N_{bot}, N_{top} and N_{lat} denote the number of rebars on the top, the bottom and the lateral side of the cross-section, respectively; ϕ_{bot}, ϕ_{top} and ϕ_{lat} denote the diameter of the rebars on the top, the bottom and the lateral side of the cross-section, respectively.

Number of floors	Section	b [m]	h [m]	N_{top}	ϕ_{top} [mm]	N_{bot}	ϕ_{bot} [mm]	N_{lat}	ϕ_{lat}
4	1	0.3	0.6	3	18	3	18	–	–
	2	0.3	0.6	4	18	3	18	–	–
	3	0.3	0.6	5	18	3	18	–	–
	4	0.3	0.6	5	18	3	18	–	–
	5	0.3	0.35	4	18	4	18	1	18
	6	0.3	0.35	6	18	6	18	1	18
6	1	0.3	0.6	3	20	2	20	–	–
	2	0.3	0.6	3	20	2	20	–	–
	3	0.3	0.6	4	20	3	20	–	–
	4	0.3	0.6	4	20	3	20	–	–
	5	0.3	0.35	4	20	4	20	1	20
	6	0.3	0.35	6	20	6	20	1	20

reduced cracked stiffness of the structural members provide the following values of the fundamental period T_1: 0.967–1.165 s for bare frames; 0.674–0.715 s for pilotis frames; 0.395–0.556 s for infilled frames.

Clay masonry infill consisting of 24 cm × 12 cm × 12 cm bricks with horizontal holes are adopted for both simplified and refined models of pilotis frames with infill-free ground floor and infilled frames with uniformly infilled stories. Compressive strength and shear strength (measured by diagonal compression test) are 1.20 MPa and 0.20 MPa, respectively, while the elastic modulus is 1050 MPa. Such typology for masonry panels is very common in old existing residential Italian RC buildings and is typically categorized as a weak infill because of the significant perforation rate and the slenderness ratio resulting from the small thickness.

4.2. Seismic ground motion records

A set of 60 horizontal pulse-like seismic ground motions is selected through a procedure based on the Variational Mode Decomposition technique [23,73]. The distributions of magnitude M_w, epicentral distance D, soil preferred shear-velocity V_{s30}, and pulse period T_p within the selected database of pulse-like seismic ground motion records are shown in Fig. 9. The ranges of several seismic parameters are listed in Table 4.

4.3. Development of large-scale machine learning models

The search for the best predictive models using GPR is performed considering all the features, namely $\mathbf{x} = \{\mathbf{b} \quad \mathbf{IM}\}$ with $\mathbf{b} = \{T_1 \quad C_y\}$ and **IM** collecting all the IMs listed in Table 1. The candidate kernel functions are the following: rational quadratic

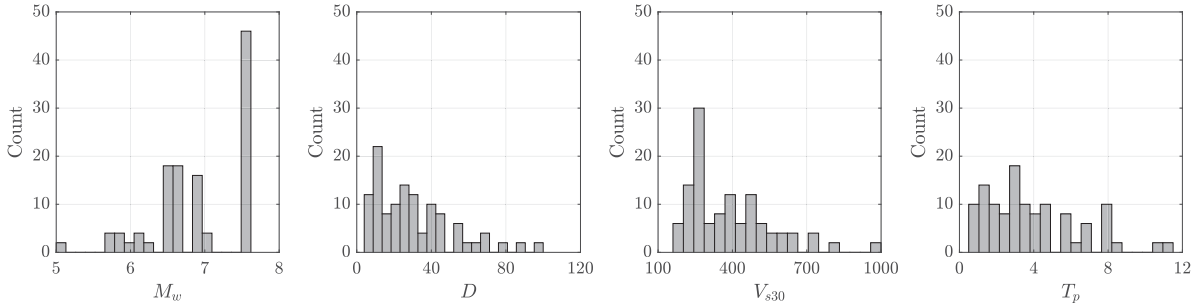

Fig. 9. Main characteristics of the selected pulse-like seismic ground motion records.

Table 4
Minimum and maximum value of several seismic parameters within the selected database of pulse-like seismic ground motion records.

Seismic parameter	Minimum value	Maximum value
Magnitude, M_w	5.0	7.5
Epicentral distance, D [km]	4.4	98.2
Soil preferred shear-velocity, V_{s30} [m/s]	186	1000
Pulse period, T_p [s]	0.7	11.3
Closest site-to-source distance, ClstD [km]	0.07	35.7
Peak ground acceleration, PGA [cm/s^2]	104.9	1245.0
Peak ground velocity, PGV [cm/s]	20.3	169.6
Arias Intensity, I_A [cm/s]	28.2	1048.7
Significant duration, t_D [s]	2.9	30.8
Housner Intensity, I_H [cm]	29.2	505.8
Maximum spectral seismic coefficient at T_1, C_{y,T_1} [g]	0.3	2.9
Maximum spectral relative input energy at T_1, E_{ir,T_1} [cm^2/s^2]	801.8	320,101.6
Seismic hazard energy factor, AE_{ir} [cm^2/s]	1492.4	342,021.1
Average spectral acceleration, $S_{a,avg}$ [cm/s^2]	62.67	1824.14
Modified effective cyclic energy, ECE [kNcm]	1.90	35,645.77

Table 5
Performance metrics of the GPR models for displacement demand prediction of RC frames.

Metric	Bare RC frames		Pilotis RC frames		RC frames with infills	
	Training dataset	Validation dataset	Training dataset	Validation dataset	Training dataset	Validation dataset
R^2	0.79	0.78	0.80	0.80	0.82	0.80
RMSE	0.31	0.33	0.30	0.32	0.34	0.36
MAE	0.23	0.23	0.23	0.24	0.23	0.25
MedAE	0.15	0.18	0.17	0.17	0.13	0.16
MAPE	0.16	0.16	0.24	0.24	0.32	0.39

kernel; radial basis function kernel; Matérn kernel (with smoothness equal to 1/2, 3/2 or 5/2). Both isotropic and anisotropic kernel variants are considered. The limited-memory Broyden–Fletcher–Goldfarb–Shanno optimization algorithm is used. It is remarked that simplified (i.e., shear-type) building models are adopted to prepare the training and validation databases needed for the development of large-scale models based on GPR (75% and 25% of the full database are considered as training and validation databases, respectively).

Table 5 and Figs. 10–12 quantify the accuracy of the predictions and the efficiency of the IMs (i.e., the dispersion of the EDP given IMs). Coefficient of determination R^2, root mean square error RMSE, mean absolute error MAE, median absolute error MedAE, and mean absolute percentage error MAPE are listed in Table 5. These statistical metrics rank among the most frequently employed in machine learning regression and altogether can provide an unbiased evaluation of their performances [74]. The comparison between actual and predicted values of the maximum inter-storey drift is displayed into Figs. 10–12. These plots also show the cumulative distribution function F for both maximum inter-storey drift values and residuals. All results into Table 5 and Figs. 10–12 are properly disaggregated (i.e., actual data and predictions for training and validation sets are examined individually).

Results in Table 5 confirm a high robustness of the GPR models, wherein the statistics remain almost constant across all the datasets. Figs. 10–12 also demonstrate an unbiased distribution of the residuals up to maximum inter-storey drift value equal to 2%. Above this threshold, GPR models underestimate the maximum inter-storey drift value. The degradation of the accuracy of predictive models based on machine learning for increasing values of the drift was previously observed by Wen et al. [36]. It can also be inferred from Table 5 and Figs. 10–12 that both accuracy and dispersion increase when moving from bare frames to pilotis frames and frames with infills. This can be explained by observing that pilotis frames and frames with infills experience lower

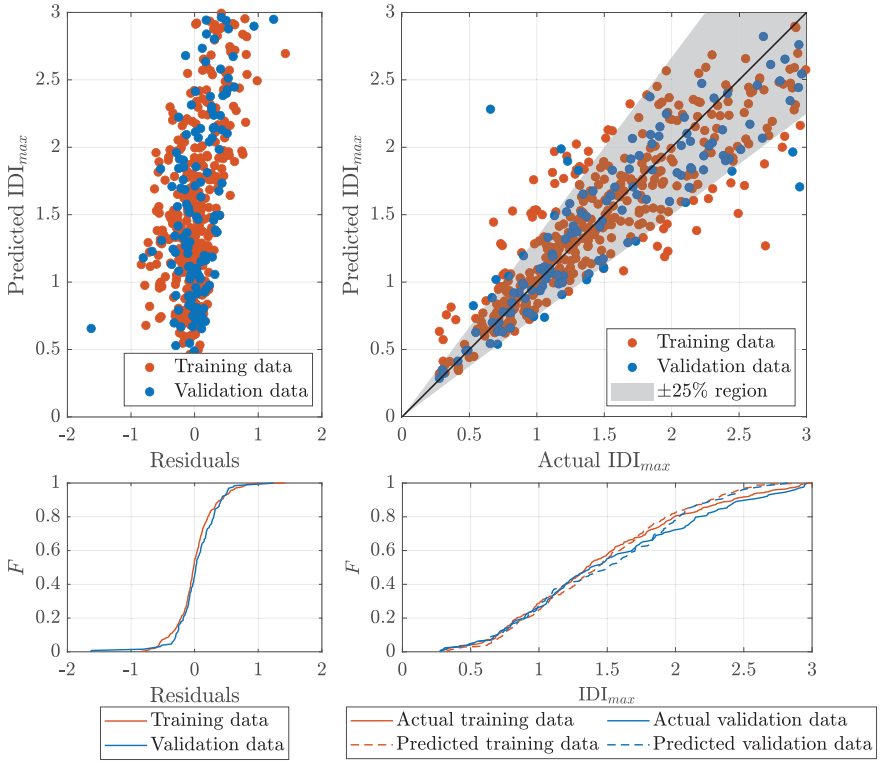

Fig. 10. Accuracy of the GPR model for predicting the maximum inter-storey drift of bare RC frames.

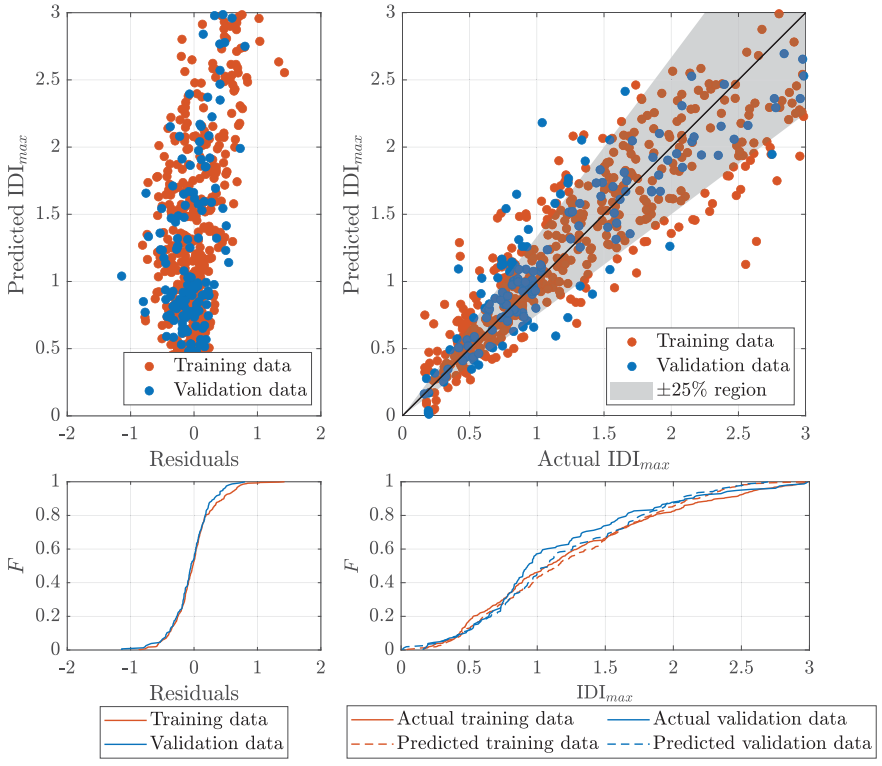

Fig. 11. Accuracy of the GPR model for predicting the maximum inter-storey drift of pilotis RC frames.

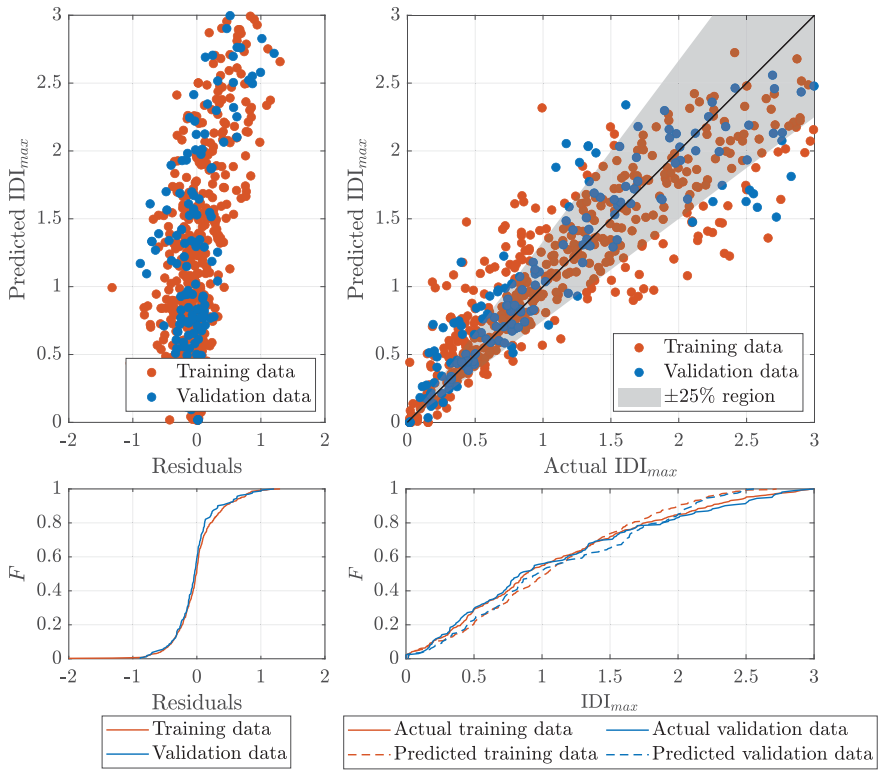

Fig. 12. Accuracy of the GPR model for predicting the maximum inter-storey drift of RC frames with infills.

displacement demand values than bare frames, thereby promoting more precise predictions on average. However, the presence of the infills complicate the dynamics of the structure, which is reflected in the increased dispersion in the predictions.

The coefficient of determination R^2 in Table 5 is about 80% for all building configurations. This outcome is in agreement with the best results obtained recently by Demertzis et al. [37] for pilotis frames under ordinary (i.e., non-pulse-like) earthquakes while is slightly lower than the results carried out for bare frames and frames with infills. This is mainly attributable to the fact that Demertzis et al. [37] considered buildings designed and built according to modern seismic codes. Conversely, the present study addresses a more challenging situation, for which buildings designed for vertical loads only and built with materials having poor mechanical characteristics are taken into account together with buildings designed in compliance with old and modern seismic codes. The recent study by Wen et al. [36] is pertinent as well for a comparative assessment of alternative approaches. They found a comparable coefficient of determination R^2 for bare frames designed in accordance with modern seismic codes and subjected to ordinary earthquakes in the range of drift values relevant for the present study.

4.4. Interpretation of large-scale machine learning models

Figs. 13–21 provide SHAP features importance plots, SHAP summary plots and features dependence plots of the GPR models developed for predicting the maximum inter-storey drift of bare frames, pilotis frames and frames with infills. The analysis of the SHAP features importance plots highlights the relative prominence of the considered explanatory variables (i.e., the importance of each feature with respect to the others), but also outlines interesting trends that demonstrate a correlation of the relative features importance with the frame filling rate (i.e., the filled surface area of the frame with respect to the total value).

A general overview of the SHAP features importance plots in Figs. 13, 16 and 19 reveals that the number of pertinent explanatory variables influencing the prediction of maximum inter-storey drift is contingent upon the rate at which the frame is filled. On average, the relative difference among the feature importance values for bare frames is less than that observed for pilotis frames and filled frames. This, in turn, implies that a smaller subset of features can explain most of the maximum inter-storey drift values for pilotis frames and filled frames while a larger number of variables is needed for bare frames.

As far as the building features are concerned, Fig. 13 shows that T_1 has a moderate relative relevance to predict the maximum inter-storey drift for bare frames. However, the larger the frame filling rate, the larger the relative importance of T_1. In fact, Fig. 16 demonstrates that T_1 is rather significant for pilotis frames compared to others features whereas it ranks among the best for frames with infills, as shown in Fig. 19. The seismic coefficient C_y is of utmost importance to estimate the maximum inter-storey drift for RC frames regardless of their filling rate since it classifies systematically among the most important explanatory variables for

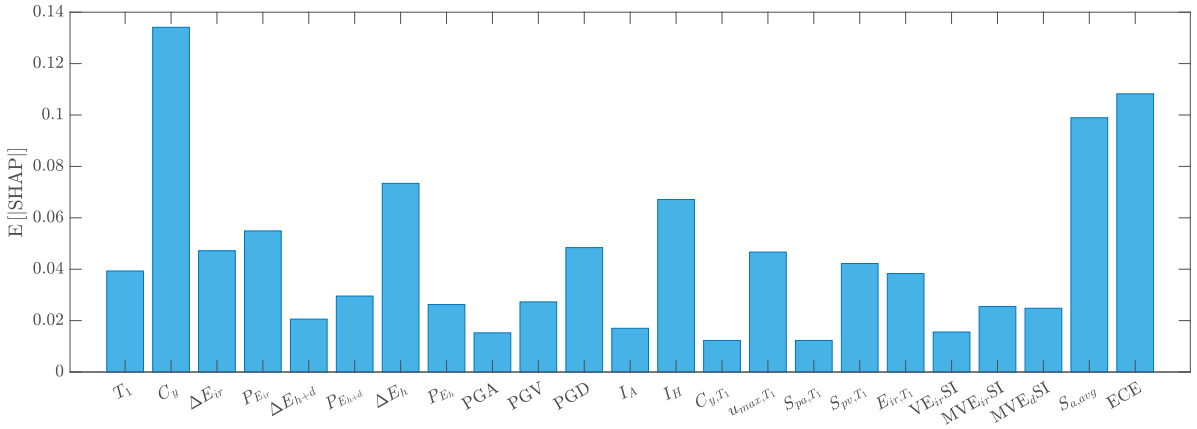

Fig. 13. SHAP features importance diagram of the GPR model for predicting the maximum inter-storey drifts of bare RC frames.

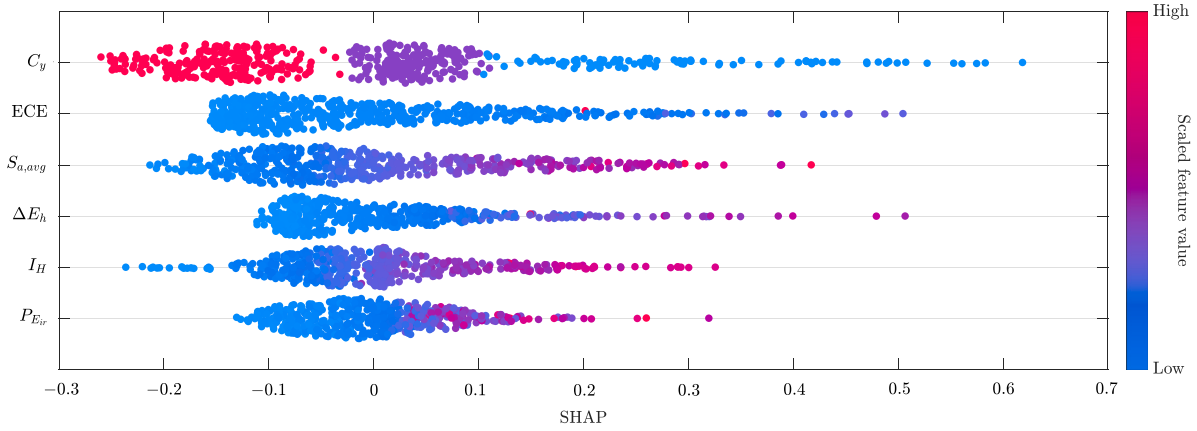

Fig. 14. SHAP summary plot of the GPR model for predicting the maximum inter-storey drifts of bare RC frames.

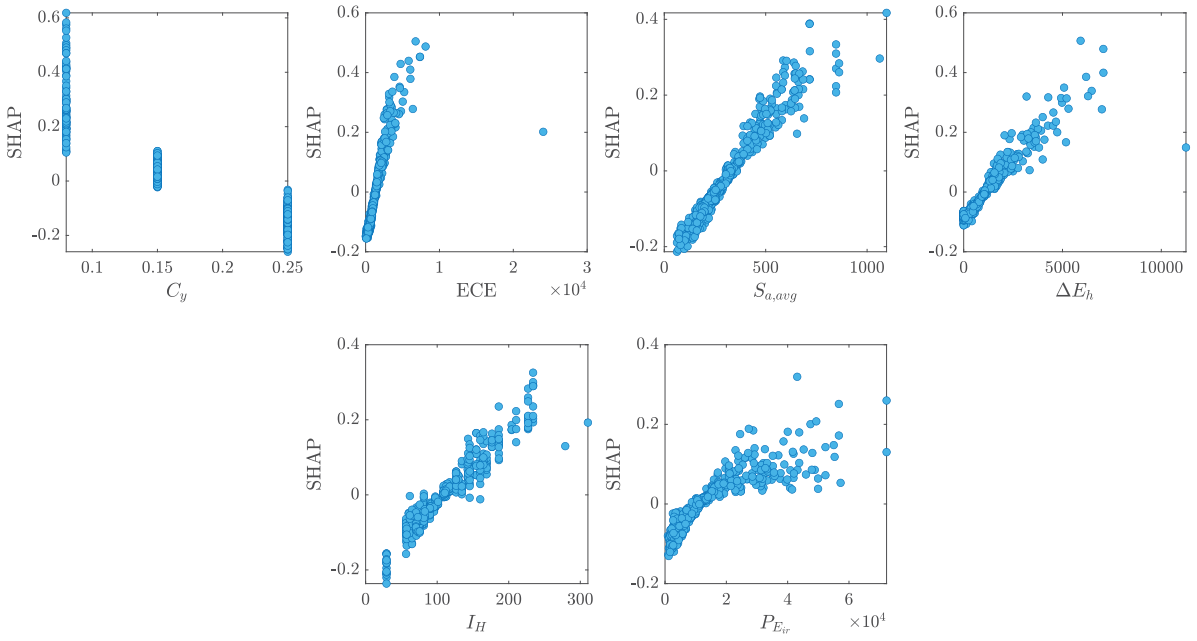

Fig. 15. Features dependence plots of the GPR model for predicting the maximum inter-storey drifts of bare RC frames.

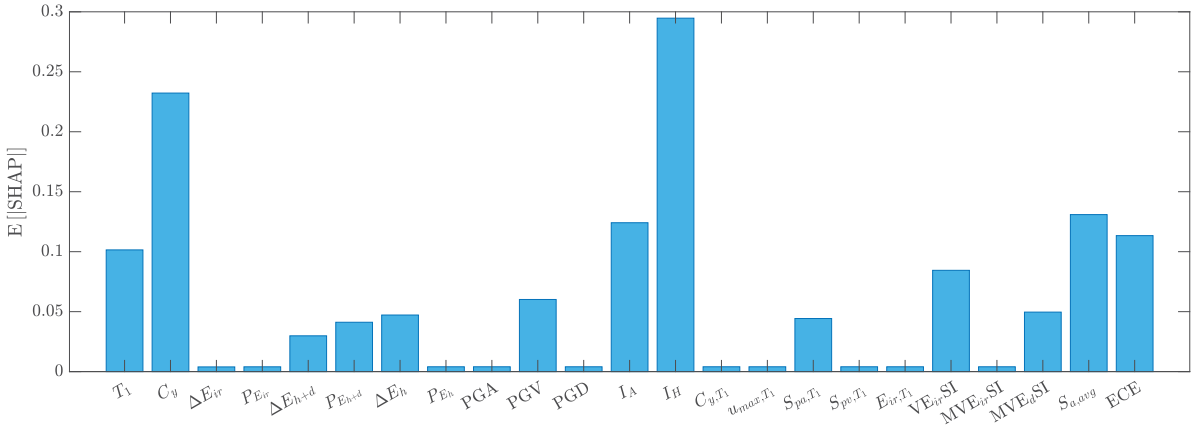

Fig. 16. SHAP features importance diagram of the GPR model for predicting the maximum inter-storey drifts of pilotis RC frames.

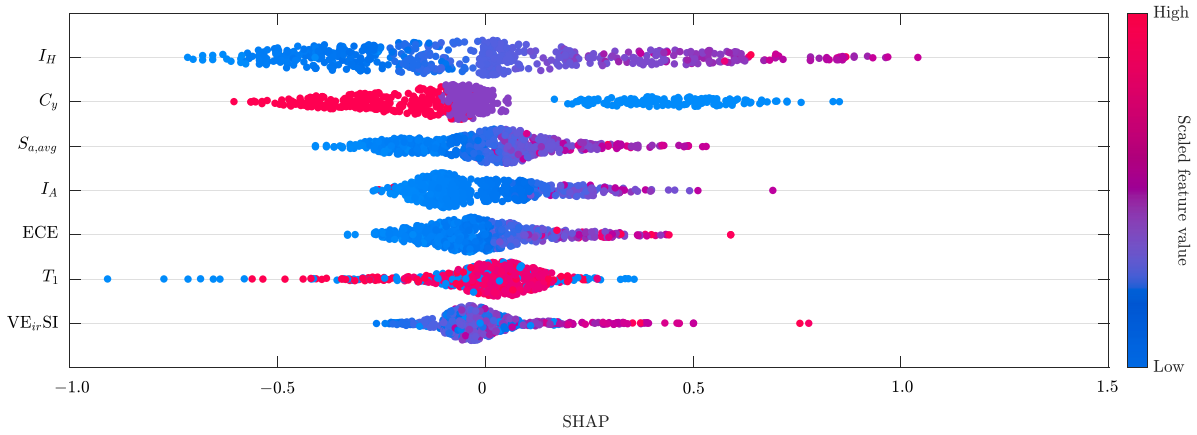

Fig. 17. SHAP summary plot of the GPR model for predicting the maximum inter-storey drifts of pilotis RC frames.

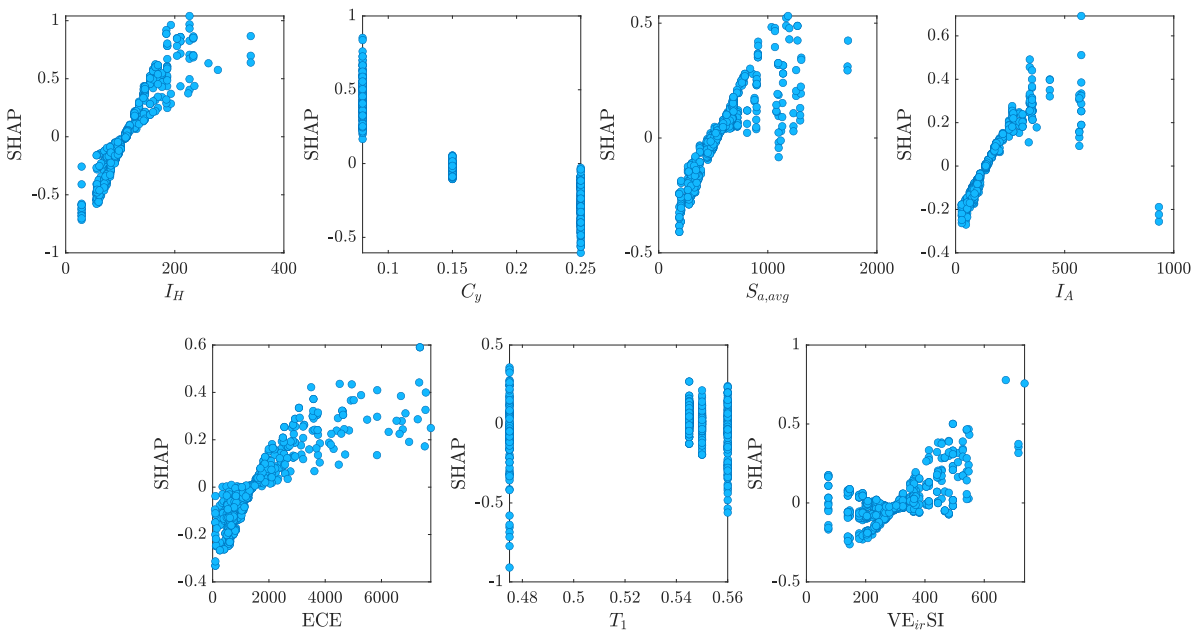

Fig. 18. Features dependence plots of the GPR model for predicting the maximum inter-storey drifts of pilotis RC frames.

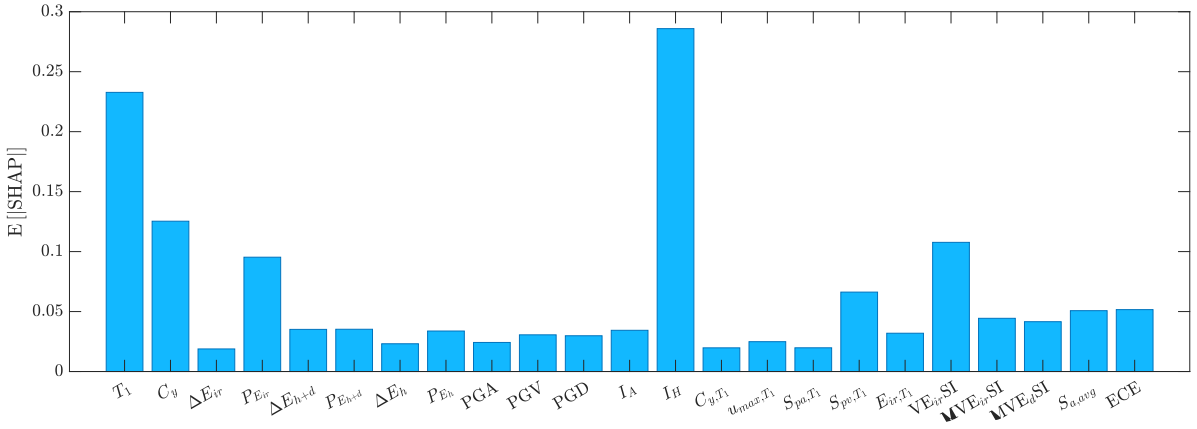

Fig. 19. SHAP features importance diagram of the GPR model for predicting the maximum inter-storey drifts of RC frames with infills.

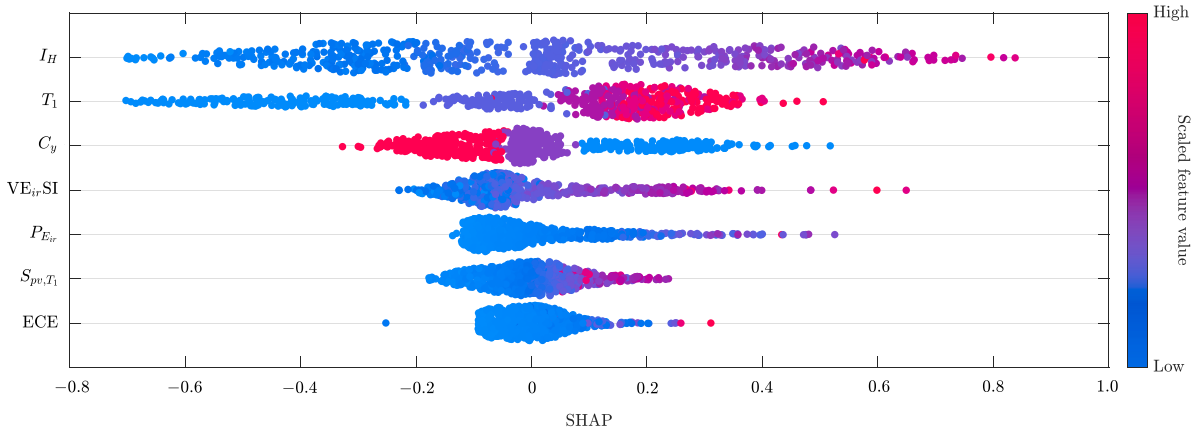

Fig. 20. SHAP summary plot of the GPR model for predicting the maximum inter-storey drifts of RC frames with infills.

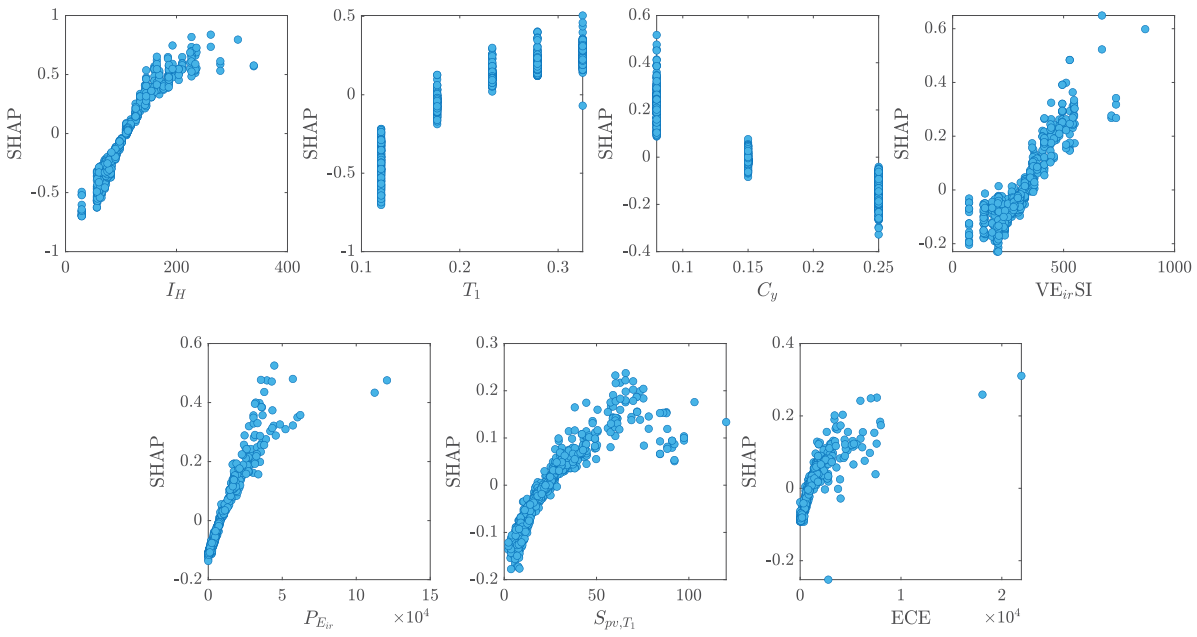

Fig. 21. Features dependence plots of the GPR model for predicting the maximum inter-storey drifts of RC frames with infills.

bare frames, pilotis frames and frames with infills, as demonstrated in Fig. 13, Fig. 16 and Fig. 19, respectively. It was found that the coefficient of determination R^2 dropped to 50% or even lower values when the seismic coefficient C_y was tentatively removed from the candidate set of building attributes. Indeed, this was an expected result: since the building stock considered in the present study includes structural systems having dissimilar behavior under seismic loads, they are expected to exhibit different values of the seismic demand even if the seismic intensity is the same. Therefore, accounting for a representative parameter of the building behavior such as the seismic coefficient C_y leads to improved predictions of the seismic demand. It is interesting to note that the relative importance of C_y and T_1 exhibit opposite trends with respect to the frame filling rate. In fact, while the relative importance of T_1 grows up for increasing values of the frame filling rate, C_y becomes less relevant compared to other features as the frame fills up.

Regarding the earthquake features, ECE, $S_{a,avg}$ and ΔE_H stand out for their relative importance in predicting the maximum inter-storey drift for bare frames, as it can be inferred from Fig. 13. In this case, I_H and $P_{E_{ir}}$ are also somewhat relevant. However, the comparison of Figs. 13, 16 and 19 demonstrates that the relative importance of ECE, $S_{a,avg}$ and ΔE_H reduces as the frames are filled. Conversely, from being an explanatory variable of moderate relative importance for bare frames, I_H becomes more important as the frame fills up. There is no evidence from Figs. 13, 16 and 19 about a clear trend for $P_{E_{ir}}$, which is somewhat important to predict the maximum inter-storey drift for bare frames and frames with infills, but it is almost negligible in case of pilotis frames. A further IM that deserves consideration for predicting the maximum inter-storey drift of RC frames is VE_{ir}SI, whose relative importance is modest for bare frames but attains a rather significant role for pilotis frames and frames with infills.

The consistency of these findings with previous studies deserves some considerations. In this regard, Fiore et al. [33] adopted another machine learning technique, namely Evolutionary Polynomial Regression (EPR) technique, to predict the maximum inter-storey drift of a bare RC frame under different pulse-like earthquakes taking into account a smaller number of IMs. The present study shares with the one by Fiore et al. [33] the following IMs: PGA; PGV; PGD; I_A; I_H; S_{pa,T_1}; $\mathrm{E}_{ir,T1}$; VE_{ir}SI; MVE_{ir}SI. The analysis reported by Fiore et al. [33] reveals that I_H carries the utmost significance within such subset of IMs. The symbolic expressions developed by Fiore et al. [33] also highlight that PGD and VE_{ir}SI are somewhat important regarding this specific set of IMs. These results obtained by Fiore et al. [33] are in fully agreement with the SHAP features importance plot in Fig. 13, where it can be observed that I_H, PGD and VE_{ir}SI rank at the first three positions, respectively, as far as such subset of IMs is concerned. Interestingly, previous implementation of different machine learning models to predict the seismic demand in RC buildings under far-field earthquakes have further confirmed the important role of I_H over alternative IMs [34,37]. Hence, it seems that there exists a general consensus among different machine learning models about the prominence of I_H to estimate the seismic displacement demand for different configurations of RC buildings, irrespective of whether the earthquake is pulse-like or not.

SHAP summary plots in Figs. 14, 17 and 20, as well as features dependence plots in Figs. 15, 18 and 21 are limited to the five most important IMs (according to the SHAP features importance plots in Figs. 13, 16 and 19). They display the SHAP values for each data point, with the color of the dots in Figs. 14, 17 and 20 representing the normalized actual data point value (red indicating high values and blue indicating small ones). Positive SHAP values signify that the feature tends to positively influence the prediction of the output, meaning the maximum inter-storey drift increases. These plots are useful to check the physical consistency of the GPR models with respect to the most influential explanatory variables. As expected, the lower C_y, the larger the maximum inter-storey drift. On the contrary, the larger the IM (whatever it is), the larger the maximum inter-storey drift. It is noted that colormap interpretation in Figs. 14, 17 and 20 is not always immediate for a few IMs because of some outliers that prevent obtaining a uniform color scaling. These outliers are evident in Figs. 15, 18 and 21, which further confirm the expected general direct relationships between IMs and maximum inter-storey drift. Within this framework, the impact of T_1 on the maximum inter-storey drift predicted by the GPR models is rather peculiar. While the relative influence of the period T_1 is not very large for bare frames, it has a moderate impact for pilotis frames, but without a clear relationship with the maximum inter-storey drift as it can be inferred from Fig. 17. The fundamental period T_1 seems especially important to predict the displacement demand for frames with infills: in such case, according to Fig. 20, the larger is T_1, the larger the maximum inter-storey drift.

4.5. Development and assessment of reduced-scale machine learning models

The search for the reduced-scale predictive models of the maximum inter-storey drift in RC frames via GP is performed by adopting a tree-based representation of the candidate solutions. The initial population is generated according to the ramped half-and-half method. This initial population is then manipulated iteratively thorough the following combination of genetic operators: tournament selection (with a tournament size equal to 10), subtree crossover (with crossover rate equal to 0.90, where functions and leaves are selected as crossover point 90% and 10% of the times, respectively), point mutation (with mutation rate equal to 0.15), and reproduction (by duplicating 5 candidate solutions within the current population to the next without changes). A constant population size equal to 2000 and a maximum number of iterations equal to 100 are assumed. Common arithmetic, logarithmic, exponential and power operators are selected as candidates for the model formulation (no logic operators are employed). Following the examination of the large-scale models developed through GPR by means of SHAP interpretation in Figs. 13–21, the search for the best predictive models using GP is constrained to a limited number of most important features, namely $\overline{\mathbf{x}} = \left\{ \mathbf{b} \quad \overline{\mathbf{IM}} \right\}$ with $\overline{\mathbf{IM}} = \left\{ I_H \quad \mathrm{VE}_{ir}\mathrm{SI} \quad S_{a,avg} \quad \mathrm{ECE} \right\}$. This is because they turn out to be the most important across all building configurations. The development of reduced-scale models based on GP algorithm is performed by means of the same training and validation databases adopted previously for the elaboration of large-scale models via GPR, which were prepared using simplified (i.e., shear-type) building models.

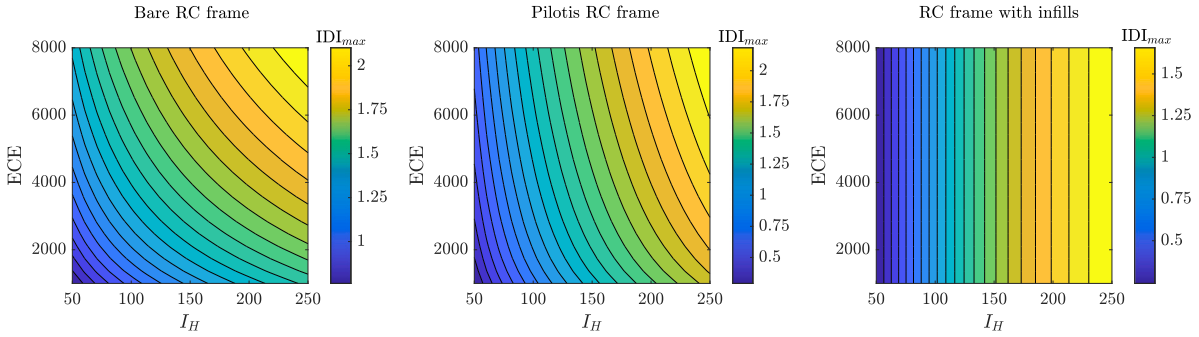

Fig. 22. Sensitivity of the GP models for predicting the maximum inter-storey drifts of RC frames using refined structural models.

Consequently, the following predictive models for the maximum inter-storey drift IDI$_{max}$ have been obtained by means of GP:

$$\mathrm{IDI}_{max} = -1.08 + 0.02\log\left(-0.15 I_H\left(-\mathrm{ECE} - 2.15 I_H + 0.76\sqrt[3]{S_{a,avg}^2} + 0.58 S_{a,avg} - \frac{S_{a,avg}}{C_y} + \frac{1.56 S_{a,avg}}{T_1}\right)\right)^2 \text{ (bare RC frames)}, \tag{7}$$

$$\mathrm{IDI}_{max} = -1.65 + 0.09\sqrt{0.94\sqrt[3]{I_H} + 1.98 I_H} + 0.09\log I_H + 0.05\sqrt{\frac{\sqrt{\mathrm{ECE}}\log(-1.43 + \log I_H)}{C_y}} \text{ (pilotis RC frames)}, \tag{8}$$

$$\mathrm{IDI}_{max} = -2.14 + 0.76\log\left(15 + 1.32T + \frac{0.01\sqrt[3]{I_H^8}\left(0.46\sqrt[3]{T_1^4} - \sqrt[3]{I_H T_1^4}\right)^2}{C_y}\right) \text{ (RC frames with infills)}, \tag{9}$$

under the condition $0.3\% \le \mathrm{IDI}_{max} \le 3\%$. The validity of Eqs. (7)–(9) is limited to the ranges of the involved parameters within the training database, and the IMs therein must be expressed using the units reported in Table 5. Eqs. (7)–(9) can be readily adopted for a rapid preliminary estimate of the maximum inter-storey drift for a broad portfolio of RC frames under pulse-like earthquakes. They require only two variables related to the examined RC building (i.e., T_1 and C_y). Moreover, they involve very few IMs. These reduced-scale (parametric) models obtained through GP are fully consistent with large-scale (nonparametric) models developed by means of GPR. In general, Eqs. (8)–(9) display a number of IMs lower than Eq. (7). This well reflects the results in Figs. 13, 16 and 19, from which is deduced that the number of most prominent variables decreases when moving from bare frames to pilotis frames and frames with infills. In particular, the predictive model for bare frames given by Eq. (7) involves all candidate explanatory variables with the only exception of VE$_{ir}$SI. This is in agreement with Fig. 13, where the SHAP features importance plot for bare frames confirms that VE$_{ir}$SI belongs to a group of IMs with similar, yet low, relevance. The predictive models for bare frames and pilotis frames share the same IMs with the exception of $S_{a,avg}$, which is missing in Eq. (8). This evidence well align with the results in Fig. 16, where it can be observed that $S_{a,avg}$ does not clearly stand out as relevant IM for pilotis frames. The fundamental period T_1 is also missing in Eq. (8) while it appears in Eq. (7). This is attributable to the fact that T_1 fulfills a certain role for pilotis frames, but its ultimate influence on the maximum inter-storey drift is not unique in such case, as it can be observed in Fig. 17. As expected, it is straightforward to infer from Eqs. (7)–(9) that the larger C_y, the lower is the seismic demand, regardless of the building typologies (i.e., bare frame, pilotis frame, or frame with infills). Both C_y and I_H are involved in all possible configurations, thereby confirming as the most critical variables in accordance with the outcomes of Figs. 13, 16 and 19. Notably, I_H is the sole IM required to predict the maximum inter-storey drift for frames with infills as per Eq. (9).

Once the final symbolic predictive models have been carried out, they are now assessed against new data obtained from refined building models. In this sense, Fig. 22 is useful to further understand the impact of I_H and ECE on the maximum inter-storey drift as predicted by Eqs. (7)–(9) for bare frames (T_1 = 1.165 s), pilotis frame (T_1 = 0.715 s) and frame with infills (T_1 = 0.556 s). It is noted that T_1 for pilotis frame and frame with infills lies outside the range considered for the development of machine learning models. This is irrelevant for the pilotis frame because the corresponding model given by Eq. (8) does not depend on T_1. In case of frame with infills, Eq. (9) is applied by replacing the actual value of T_1 with the maximum value within the range considered for the development of machine learning models (T_1 = 0.3250 s). Fig. 22 shows that, while both IMs equally contribute to the maximum inter-storey drift for bare frames, ECE becomes less important in case of pilotis frames and is no longer relevant for frames with infills. This evidence further confirms the consistency between reduced-scale and large-scale machine learning models for bare frames, pilotis frames and frames with infills. As expected, the colorbars in Fig. 22 also displays that the maximum inter-storey drift predicted by Eqs. (7)–(9) reduces when moving from bare frames to pilotis frames and frames with infills for assigned values of I_H and ECE.

Table 6
Performance metrics of the GP models for displacement demand prediction of RC frames using refined structural models.

Metric	Bare RC frames	Pilotis RC frames	RC frames with infills
R^2	0.82	0.73	0.71
RMSE	0.25	0.26	0.33
MAE	0.18	0.19	0.22
MedAE	0.13	0.13	0.15
MAPE	0.18	0.17	0.21

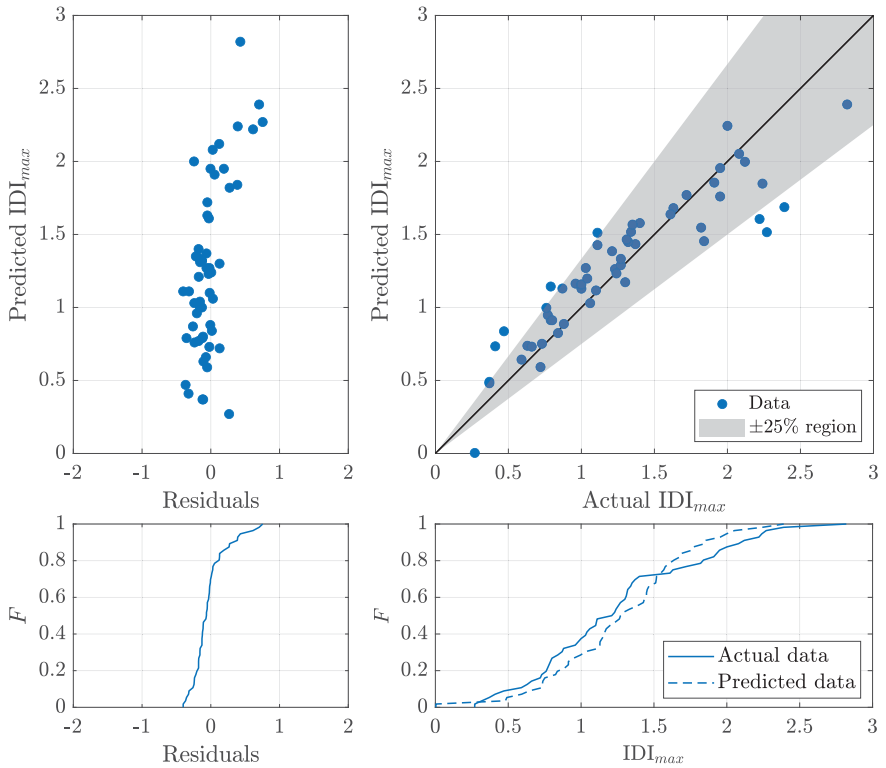

Fig. 23. Accuracy of the GP models for predicting the maximum inter-storey drifts of bare RC frames using refined structural models.

The assessment of symbolic predictive models against new data generated from refined structural models is especially important to quantify the approximation error due to the use of simplified structural models. Table 6 and Figs. 23–25 provide a comprehensive view about the accuracy of the predictive models given by Eqs. (7)–(9). The reduction of R^2 from about 80% in Table 5 to 75% in Table 6 is attributable to two concurrent factors. On the one hand, the huge reduction in dimension of the machine learning models facilitates their implementation in practical applications, but it unavoidably leads to a deterioration in accuracy, though the impact is limited thanks to the careful features selection driven by SHAP interpretation. Moreover, simplified structural models make easy the databases preparation (in terms of both simulation time and buildings data assignment), but approximation errors are expected when they are considered in place of refined structural models.

While Table 6 and Figs. 23–25 provide evidence about the efficiency of the involved IMs, it is mandatory to assess the predictive models in terms of sufficiency. In this regard, Fig. 26 confirms that residuals and seismological parameters are fully uncorrelated each other since the slope coefficient of the corresponding linear regressions is about zero.

5. Conclusions

This study has introduced a novel approach for assisting the development of machine learning models to estimate the seismic demand in reinforced concrete (RC) buildings. The proposed approach operates on two distinct scales. Initially, a large-scale (nonparametric) machine learning model is obtained by means of Gaussian Process Regression (GPR) taking into account essential building information and relevant intensity measures (IMs). Next, SHapley Additive exPlanations (SHAP) values are employed to enhance interpretability and support the selection of a reduced set of intensity measures. This selected subset of features is then

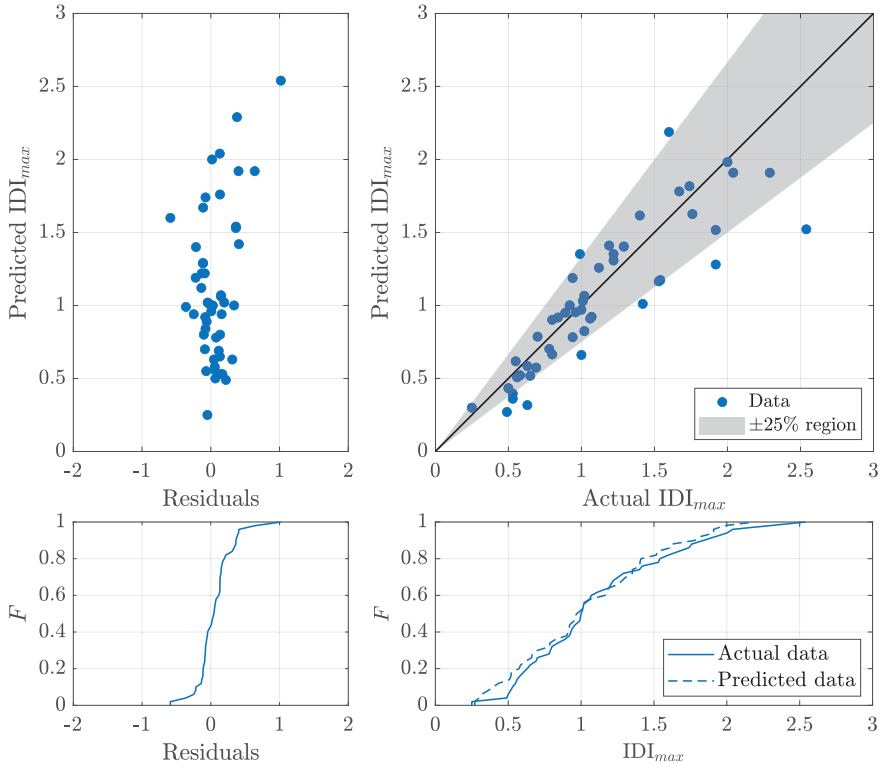

Fig. 24. Accuracy of the GP models for predicting the maximum inter-storey drifts of pilotis RC frames using refined structural models.

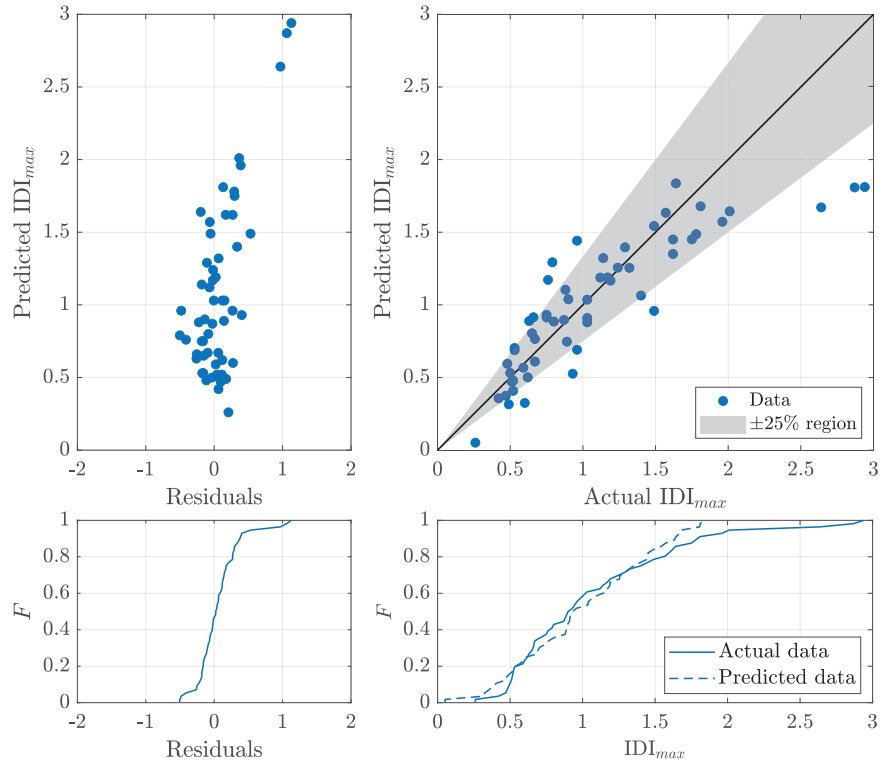

Fig. 25. Accuracy of the GP models for predicting the maximum inter-storey drifts of RC frames with infills using refined structural models.

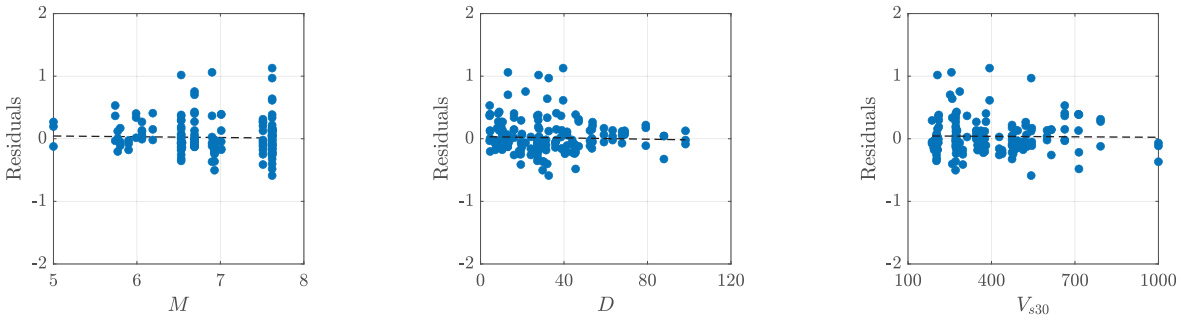

Fig. 26. Sufficiency of the GP models for predicting the maximum inter-storey drifts of RC frames using refined structural models (the dashed line indicates the linear regression).

used to carry out the final (symbolic) reduced-scale machine learning model through Genetic Programming (GP). Simplified building models are utilized to prepare the data needed to develop predictive machine learning models at both scales, and refined building models are adopted for the final evaluations.

Predictive machine learning models for the maximum inter-storey drift in bare frames, pilotis frames, and frames with infills under pulse-like seismic ground motion are proposed utilizing this approach. The accuracy of the final predictive models is suitable for preliminary evaluations and applications at regional scale, whereas the independence of the predictions from other seismological parameters confirmed their sufficiency.

Notably, the proposed symbolic expressions to predict the displacement demand of RC frames under pulse-like earthquakes are ready-to-use in future applications. As far as the building attributes are concerned, only the fundamental period and the base shear seismic coefficient are required, and both are easily accessible. Several empirical formulations exist to predict the fundamental period of RC frames from the height only [e.g., 75] whereas numerical values of the base shear seismic coefficient can be retrieved from existing literature and they mainly depend on the construction year [e.g., 71]. A limitation is due to the fact that a ground motion prediction equation for ECE is not available yet. Moreover, it is understood that the symbolic expressions reported in the present study apply if and only if the input explanatory variables (i.e., building attributes and seismic intensity measures) fall within the range of values of the adopted database.

The critical examination of the predictive machine learning models across both scales has demonstrated that Housner intensity stands out for its importance for all building configurations. This evidence has been obtained for RC buildings designed against gravity loads only or in compliance with outdated seismic codes when subjected to pulse-like earthquakes, but well agrees with previous predictive machine learning models developed for ordinary (i.e., non-pulse-like) earthquakes, thereby suggesting that Housner intensity can be designated as general, most representative IMs for all seismic scenarios.

CRediT authorship contribution statement

Giulia Angelucci: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Giuseppe Quaranta:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Fabrizio Mollaioli:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Sashi K. Kunnath:** Writing – review & editing, Methodology, Conceptualization.

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 article.

Data availability

Data will be made available on request.

Acknowledgments

Giulia Angelucci gratefully acknowledges the financial support of the project n. CN1221844D08208F from Sapienza University of Rome under the umbrella of the national program PNRR – CN1 – Spoke 5 (D. D. n. 1031, 17-06-2022). The work of Giuseppe Quaranta and Fabrizio Mollaioli is framed within the project "Artificial Intelligence for SUstainable seismic risk reduction of STructures (AI-SUST)" (project code: 2022LEFKHS) funded by European Union – NextGeneration EU through the PRIN 2022 program of the Italian Ministry of University and Research (MUR) (D. D. n. 104, 02-02-2022). This work reflects only the authors' views and opinions whereas the MUR cannot be considered responsible for them.

References

[1] C.A. Cornell, F. Jalayer, R.O. Hamburger, D.A. Foutch, Probabilistic basis for 2000 SAC federal emergency management agency steel moment frame guidelines, J. Struct. Eng. 128 (4) (2002) 526–533.

[2] J. Moehle, G.G. Deierlein, A framework methodology for performance-based earthquake engineering, in: 13th World Conference on Earthquake Engineering, vol. 679, WCEE Vancouver, 2004, p. 12.

[3] H. Ebrahimian, F. Jalayer, Selection of seismic intensity measures for prescribed limit states using alternative nonlinear dynamic analysis methods, Earthq. Eng. Struct. Dyn. 50 (5) (2021) 1235–1250.

[4] N. Luco, C.A. Cornell, Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions, Earthq. Spectra 23 (2) (2007) 357–392.

[5] H.D. Nguyen, M. Shin, J.M. LaFave, Optimal intensity measures for probabilistic seismic demand models of steel moment frames, J. Build. Eng. 65 (2023) 105629.

[6] F. Mollaioli, A. Lucchini, Y. Cheng, G. Monti, Intensity measures for the seismic response prediction of base-isolated buildings, Bull. Earthq. Eng. 11 (2013) 1841–1866.

[7] J. Donaire-Ávila, F. Mollaioli, A. Lucchini, A. Benavent-Climent, Intensity measures for the seismic response prediction of mid-rise buildings with hysteretic dampers, Eng. Struct. 102 (2015) 278–295.

[8] F. Mazza, R. Labernarda, Structural and non-structural intensity measures for the assessment of base-isolated structures subjected to pulse-like near-fault earthquakes, Soil Dyn. Earthq. Eng. 96 (2017) 115–127.

[9] M. De Biasio, S. Grange, F. Dufour, F. Allain, I. Petre-Lazar, A simple and efficient intensity measure to account for nonlinear structural behavior, Earthq. Spectra 30 (4) (2014) 1403–1426.

[10] S. Ghimire, P. Guéguen, A. Astorga, Analysis of the efficiency of intensity measures from real earthquake data recorded in buildings, Soil Dyn. Earthq. Eng. 147 (2021) 106751.

[11] F. Jalayer, J. Beck, F. Zareian, Analyzing the sufficiency of alternative scalar and vector intensity measures of ground shaking based on information theory, J. Eng. Mech. 138 (3) (2012) 307–316.

[12] S.L. Dhulipala, A. Rodriguez-Marek, M.M. Flint, Importance of intensity measure sufficiency for structural seismic demand hazard analysis, in: Proceedings of the 11 Th National Conference on Earthquake Engineering, Los Angeles, CA, USA, 2018, pp. 25–29.

[13] T. Rossetto, A. Elnashai, Derivation of vulnerability functions for European-type RC structures based on observational data, Eng. Struct. 25 (10) (2003) 1241–1263.

[14] M.Z. Esteghamati, Q. Huang, Evaluating the impact of higher-mode and inelastic dynamic responses of concrete frames on the performance of seismic intensity measures, Structures 56 (2023) 105029.

[15] M. Bianchini, P. Diotallevi, J. Baker, et al., Prediction of inelastic structural response using an average of spectral accelerations, ICOSSAR09, in: 10th International Conference on Structural Safety and Reliability, vol. 1317, 2009, pp. 2164–2171.

[16] A.K. Kazantzi, D. Vamvatsikos, Intensity measure selection for vulnerability studies of building classes, Earthq. Eng. Struct. Dyn. 44 (15) (2015) 2677–2694.

[17] G.J. O'Reilly, Limitations of sa (t 1) as an intensity measure when assessing non-ductile infilled RC frame structures, Bull. Earthq. Eng. 19 (6) (2021) 2389–2417.

[18] L.D. Decanini, F. Mollaioli, An energy-based methodology for the assessment of seismic demand, Soil Dyn. Earthq. Eng. 21 (2) (2001) 113–137.

[19] F. Mollaioli, J. Donaire-Avila, A. Lucchini, A. Benavent-Climent, On the importance of energy-based parameters, in: COMPDYN 2019 7th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, Crete, Greece, 2019, pp. 24–26.

[20] S. Pathak, A. Khennane, S. Al-Deen, Power based seismic collapse criterion for ductile and non-ductile framed structures, Bull. Earthq. Eng. 18 (2020) 5983–6014.

[21] S. Kunnath, Y. Chai, Cumulative damage-based inelastic cyclic demand spectrum, Earthq. Eng. Struct. Dyn. 33 (4) (2004) 499–520.

[22] G. Angelucci, F. Mollaioli, G. Quaranta, Correlation between energy and displacement demands for infilled reinforced concrete frames, Front. Built Environ. 9 (2023) 1198478.

[23] G. Quaranta, F. Mollaioli, Analysis of near-fault pulse-like seismic signals through variational mode decomposition technique, Eng. Struct. 193 (2019) 121–135.

[24] C. Li, Z. Zuo, S. Kunnath, L. Chen, Orientation of the strongest velocity pulses and the maximum structural response to pulse-like ground motions, Soil Dyn. Earthq. Eng. 136 (2020) 106240.

[25] R. Gentile, C. Galasso, Accounting for directivity-induced pulse-like ground motions in building portfolio loss assessment, Bull. Earthq. Eng. 19 (15) (2021) 6303–6328.

[26] J.W. Baker, Probabilistic structural response assessment using vector-valued intensity measures, Earthq. Eng. Struct. Dyn. 36 (13) (2007) 1861–1883.

[27] J.W. Baker, C.A. Cornell, Vector-valued intensity measures incorporating spectral shape for prediction of structural response, J. Earthq. Eng. 12 (4) (2008) 534–554.

[28] E. Zengin, N.A. Abrahamson, A vector-valued intensity measure for near-fault ground motions, Earthq. Eng. Struct. Dyn. 49 (7) (2020) 716–734.

[29] A.-A.I. Theophilou, M.K. Chryssanthopoulos, A.J. Kappos, A vector-valued ground motion intensity measure incorporating normalized spectral area, Bull. Earthq. Eng. 15 (2017) 249–270.

[30] T. Travasarou, J. Bray, Optimal ground motion intensity measures for assessment of seismic slope displacements, in: 2003 Pacific Conference on Earthquake Engineering, Christchurch, New Zealand, Feb, 2003.

[31] Y.F. Vargas-Alzate, R. Gonzalez-Drigo, J.A. Avila-Haro, Multi-regression analysis to enhance the predictability of the seismic response of buildings, Infrastructures 7 (4) (2022) 51.

[32] Y.F. Vargas-Alzate, J.E. Hurtado, L.G. Pujades, New insights into the relationship between seismic intensity measures and nonlinear structural response, Bull. Earthq. Eng. (2022) 1–37.

[33] A. Fiore, F. Mollaioli, G. Quaranta, G.C. Marano, Seismic response prediction of reinforced concrete buildings through nonlinear combinations of intensity measures, Bull. Earthq. Eng. 16 (2018) 6047–6076.

[34] K. Morfidis, K. Kostinakis, Seismic parameters' combinations for the optimum prediction of the damage state of R/C buildings using neural networks, Adv. Eng. Softw. 106 (2017) 1–16.

[35] B.K. Oh, B. Glisic, S.W. Park, H.S. Park, Neural network-based seismic response prediction model for building structures using artificial earthquakes, J. Sound Vib. 468 (2020) 115109.

[36] W. Wen, C. Zhang, C. Zhai, Rapid seismic response prediction of RC frames based on deep learning and limited building information, Eng. Struct. 267 (2022) 114638.

[37] K. Demertzis, K. Kostinakis, K. Morfidis, L. Iliadis, An interpretable machine learning method for the prediction of R/C buildings' seismic response, J. Build. Eng. 63 (2023) 105493.

[38] C.C. Mitropoulou, M. Papadrakakis, Developing fragility curves based on neural network IDA predictions, Eng. Struct. 33 (12) (2011) 3409–3421.

[39] Y. Yan, Y. Xia, J. Yang, L. Sun, Optimal selection of scalar and vector-valued seismic intensity measures based on Gaussian process regression, Soil Dyn. Earthq. Eng. 152 (2022) 106961.

[40] E. Uncuoglu, H. Citakoglu, L. Latifoglu, S. Bayram, M. Laman, M. Ilkentapar, A.A. Oner, Comparison of neural network, Gaussian regression, support vector machine, long short-term memory, multi-gene genetic programming, and M5 trees methods for solving civil engineering problems, Appl. Soft Comput. 129 (2022) 109623.
[41] M. Seeger, Gaussian processes for machine learning, Int. J. Neural Syst. 14 (02) (2004) 69–106.
[42] C.K. Williams, C.E. Rasmussen, Gaussian Processes for Machine Learning, vol. 2, (no. 3) MIT press Cambridge, MA, 2006.
[43] C.E. Rasmussen, H. Nickisch, Gaussian processes for machine learning (GPML) toolbox, J. Mach. Learn. Res. 11 (2010) 3011–3015.
[44] R. Gentile, C. Galasso, Gaussian process regression for seismic fragility assessment of building portfolios, Struct. Saf. 87 (2020) 101980.
[45] E. Mosca, F. Szigeti, S. Tragianni, D. Gallagher, G. Groh, SHAP-based explanation methods: A review for NLP interpretability, in: Proceedings of the 29th International Conference on Computational Linguistics, 2022, pp. 4593–4603.
[46] C. Molnar, Interpretable machine learning, Lulu. com, 2020.
[47] D. Fryer, I. Strümke, H. Nguyen, Shapley values for feature selection: The good, the bad, and the axioms, Ieee Access 9 (2021) 144352–144360.
[48] W.E. Marcílio, D.M. Eler, From explanations to feature selection: Assessing SHAP values as feature selection mechanism, in: 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI, Ieee, 2020, pp. 340–347.
[49] J.R. Koza, Genetic programming as a means for programming computers by natural selection, Stat. Comput. 4 (1994) 87–112.
[50] W.B. Langdon, R. Poli, Foundations of Genetic Programming, Springer Science & Business Media, 2013.
[51] G. Quaranta, W. Lacarbonara, S.F. Masri, A review on computational intelligence for identification of nonlinear dynamical systems, Nonlinear Dynam. 99 (2) (2020) 1709–1761.
[52] A. Ekart, S.Z. Nemeth, Selection based on the pareto nondomination criterion for controlling code growth in genetic programming, Genet. Programm. Evolvable Mach. 2 (2001) 61–73.
[53] F. Mollaioli, S. Bruno, L. Decanini, R. Saragoni, Correlations between energy and displacement demands for performance-based seismic engineering, Pure Appl. Geophys. 168 (2011) 237–259.
[54] M.G. d'Aragona, M. Polese, A. Prota, Stick-IT: A simplified model for rapid estimation of IDR and PFA for existing low-rise symmetric infilled RC building typologies, Eng. Struct. 223 (2020) 111182.
[55] L.D. Decanini, F. Mollaioli, A. Mura, et al., Shear-beam model for the prediction of the response of MDOF systems subjected to severe earthquake ground shaking, in: 12th European Conference on Earthquake Engineering, vol. 400, Elsevier Science Ltd, 2002.
[56] F. Mollaioli, S. Bruno, Influence of site effects on inelastic displacement ratios for SDOF and MDOF systems, Comput. Math. Appl. 55 (2) (2008) 184–207.
[57] F. McKenna, M.H. Scott, G.L. Fenves, Nonlinear finite-element analysis software architecture using object composition, J. Comput. Civ. Eng. 24 (1) (2010) 95–107.
[58] M.H. Scott, G.L. Fenves, Plastic hinge integration methods for force-based beam–column elements, J. Struct. Eng. 132 (2) (2006) 244–252.
[59] L. Decanini, L. Liberatore, F. Mollaioli, Damage potential of the 2009 L'Aquila, Italy, earthquake, J. Earthq. Tsunami 6 (03) (2012) 1250032.
[60] L. Decanini, G. Fantin, Modelos simplificados de la mamponesteria incluida en porticos caracteristicas de rigidez y resistencia lateral en astado limite, Jornadas Argentinas de Ingenieria Estructural III 2, 817–8605.
[61] L. Liberatore, F. Noto, F. Mollaioli, P. Franchin, In-plane response of masonry infill walls: Comprehensive experimentally-based equivalent strut model for deterministic and probabilistic analysis, Eng. Struct. 167 (2018) 533–548.
[62] N.M. Noh, L. Liberatore, F. Mollaioli, S. Tesfamariam, Modelling of masonry infilled RC frames subjected to cyclic loads: State of the art review and modelling with OpenSees, Eng. Struct. 150 (2017) 599–621.
[63] M. Di Domenico, M.T. De Risi, V. Manfredi, M. Terrenzi, G. Camata, F. Mollaioli, F. Noto, P. Ricci, P. Franchin, A. Masi, et al., Modelling and seismic response analysis of Italian pre-code and low-code reinforced concrete buildings. Part II: Infilled frames, J. Earthq. Eng. 27 (6) (2023) 1534–1564.
[64] L.D. Decanini, F. Mollaioli, Formulation of elastic earthquake input energy spectra, Earthq. Eng. Struct. Dyn. 27 (12) (1998) 1503–1522.
[65] E. Kalkan, S.K. Kunnath, Effective cyclic energy as a measure of seismic demand, J. Earthq. Eng. 11 (5) (2007) 725–751.
[66] N. Hori, N. Inoue, Damaging properties of ground motions and prediction of maximum response of structures based on momentary energy response, Earthq. Eng. Struct. Dyn. 31 (9) (2002) 1657–1679.
[67] K. Fujii, Bidirectional seismic energy input to an isotropic nonlinear one-mass two-degree-of-freedom system, Buildings 11 (4) (2021) 143.
[68] K. Fujii, Y. Murakami, Bidirectional momentary energy input to a one-mass two-DOF system, in: Proceedings of the 17th World Conference on Earthquake Engineering, vol. 20, Sendai, Japan, 2020.
[69] S.A. Enderami, S.B. Beheshti-Aval, M.A. Saadeghvaziri, New energy based approach to predict seismic demands of steel moment resisting frames subjected to near-fault ground motions, Eng. Struct. 72 (2014) 182–192.
[70] C. Tyson, C. McAndrew, P. Tuma, I. Pegg, A. Sarkar, Automated nonparametric method for detection of step-like features in biological data sets, Cytomet. Part A 87 (5) (2015) 393–404.
[71] N. Jayaram, F. Mollaioli, P. Bazzurro, A. De Sortis, S. Bruno, Prediction of structural response of reinforced concrete frames subjected to earthquake ground motions, in: 9th US National and 10th Canadian Conference on Earthquake Engineering, EERI, 2010, pp. 428–437.
[72] DMLL PP 9/1/1996, Norme tecniche per il calcolo, l'esecuzione ed il collaudo delle strutture in cemento armato, normale e precompresso e per le strutture metalliche, 1996.
[73] G. Quaranta, G. Angelucci, F. Mollaioli, Near-fault earthquakes with pulse-like horizontal and vertical seismic ground motion components: Analysis and effects on elastomeric bearings, Soil Dyn. Earthq. Eng. 160 (2022) 107361.
[74] A. Botchkarev, Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology, 2018, arXiv preprint arXiv:1809.03006.
[75] M.R. Gallipoli, B. Petrovic, G. Calamita, N. Tragni, C. Scaini, C. Barnaba, M. Vona, S. Parolai, Towards specific T–H relationships: FRIBAS database for better characterization of RC and URM buildings, Bull. Earthq. Eng. 21 (4) (2023) 2281–2307.