Received: 5 January 2024 | Revised: 24 February 2024 | Accepted: 27 February 2024

DOI: 10.1002/eqe.4110

RESEARCH ARTICLE

WILEY

A multi-level approach to predict the seismic response of rigid rocking structures using artificial neural networks

Seyed Amir Banimahd[1] | **Anastasios I. Giouvanidis**[2] | **Shaghayegh Karimzadeh**[3] | **Paulo B. Lourenço**[3]

[1]Department of Civil Engineering, Faculty of Engineering, Ardakan University, Ardakan, Iran

[2]Department of Civil and Environmental Engineering, University of Auckland, Auckland, New Zealand

[3]Department of Civil Engineering, ISISE, ARISE, University of Minho, Guimarães, Portugal

Correspondence
Anastasios I. Giouvanidis, Department of Civil and Environmental Engineering, University of Auckland, Auckland, New Zealand.
Email: a.giouvanidis@auckland.ac.nz

Funding information
Fundação para a Ciência e a Tecnologia; European Research Council

Abstract

This paper explores the use of Artificial Neural Networks (ANN) for the rocking problem. The paper adopts rigid rocking blocks of different sizes and slenderness, which undergo rocking motion without sliding and bouncing when subjected to recorded earthquakes. This research focuses on the cases where the blocks overturn or safely return to their initial (rest) position at the end of the ground shaking. An ANN model is trained to efficiently categorise the response into overturning or safe rocking using the structural parameters, ground motion characteristics, and the coefficient of restitution as input. The results show the substantial contribution of velocity and frequency characteristics of the ground motion to overturning. In addition, ANN is used to predict the response amplitude and identify the most critical input variables that govern safe rocking. The analysis reveals that rocking amplitude is governed by a combination of duration, frequency, and intensity characteristics of the ground excitation. Interestingly, the maximum incremental velocity (MIV), a novel intensity measure for the rocking literature, shows a substantial correlation with the rocking amplitude. In this context, this paper proposes closed-form expressions using the most influential input variables to provide a quick, yet adequately accurate, response prediction. Finally, this study pays special attention to the contribution of the coefficient of restitution, which, in general, is less critical to the peak safe rocking response, while it becomes more important to the overturning response.

KEYWORDS
artificial neural networks, coefficient of restitution, intensity measures, machine learning, rocking, seismic signal

1 | INTRODUCTION

Rocking is regarded as a non-traditional seismic isolation technique that provides stability to the structure due to the activation of its rotational inertia.[1,2] Thus, it finds application to building contents,[3,4] masonry structures,[5–7] as well as

buildings[8] and bridges.[9–13] Despite its merits, rocking motion is highly nonlinear[14] and nonsmooth.[15] Therefore, it faces marked challenges. Rocking is particularly sensitive to characteristics of the ground excitation.[16,17] Psycharis et al.[18] illustrated the importance of peak ground velocity (*PGV*) over peak ground acceleration (*PGA*) to rocking response; similar results were later found in Sieber et al.[19] Dimitrakopoulos and Paraskeva[20] highlighted the contribution of *PGV* and a combination of *PGA* with a frequency characteristic (*PGA*/*PGV*) to rocking amplitude. Later, Petrone et al.[21] and Lachanas et al.[22] verified the importance of *PGV* and *PGA* to the overturning of slender and stocky rocking structures, respectively. Liu et al.[23] showed the influence of the combined consideration of peak and frequency characteristics of the ground excitation on rocking response. Further accreditation of velocity and frequency characteristics as optimal intensity measures (IMs) (i.e., exhibiting a strong correlation with the response) came from Pappas et al.[24] and Kavvadias et al.[25] Recently, Giouvanidis and Dimitrakopoulos[26] revealed the remarkable contribution of duration-based characteristics, such as the uniform duration t_{uni} and cumulative absolute velocity of exceedance CAV_{exc}, to rocking amplification (i.e., rocking without overturning).

Due to the stochastic nature of earthquakes[27] and the sensitivity of the dynamic rocking response to different seismic signal characteristics, most studies investigated one or two ground motion parameters, formulating scalar or bivariate IMs, respectively. However, a scalar or bivariate IM could be insufficient to represent the impact of the seismic signal on demand estimation. On the other hand, introducing multiple IMs to conventional methods, for example, regression models, can be computationally demanding. Thus, the relation between rocking response and seismic signal is not yet fully understood. Machine learning (ML) is a promising means that can reveal the underlying nonlinearity of such a relation.[28] In this context, Xie et al.[29] used ML to identify velocity characteristics as optimal and propose closed-form expressions of fragility estimates for a single-column rocking bridge. Hu et al.[30] adopted various ML methodologies, such as linear and ridge regression, decision trees, random forests, extreme gradient boosting and adaptive boosting, to propose seismic demand models of self-centring dual rocking core systems. Similarly, Gajan[31] adopted k-nearest neighbours, support vector machines, and random forests to predict the peak rotation and safety factor against the overturning of rocking structures under earthquakes. Recently, Achmet et al.[32] adopted k-nearest neighbours and support vector machines to classify rocking response into safe rocking and overturning.

The application of ANN in the rocking literature has not yet received the attention it deserves, despite being widely employed in various engineering fields,[33–37] showing their superiority as response predictors.[38–41] Therefore, this work explores the ability of ANN models to predict the relation between seismic signal and rocking response without the need to use complex mathematical models. Specifically, this work investigates the ANN model's ability to predict rocking response and, importantly, identify the critical ground motion characteristics that govern rocking behaviour. Both safe rocking and overturning are considered, while this study gives special attention to the role of the coefficient of restitution in rocking response. From a practical engineering viewpoint, using ML methodologies might be challenging. Thus, this study also proposes simple closed-form expressions with respect to the most optimal ground motion IMs for a rapid, yet adequately accurate, response prediction.

The outline of this paper is as follows. Section 2 presents the analytical dynamics of rocking motion. Section 3 introduces the engineering demand parameter that captures the seismic response of a rigid rocking block, and the intensity measures that represent different attributes of the ground excitation. Section 4 provides an overview of the adopted methodology, while Section 5 introduces the need to reduce (linear) correlations among the input variables. Section 6 provides a detailed description of the ANN architecture. Section 7 offers insights into the binary distinction between safe rocking and overturning (i.e., classification problem), whereas Section 8 focuses on the prediction of the (safe) rocking amplitude (i.e., regression problem). Both Sections 7 and 8 give special attention to the input variables that contribute the most to each problem. In addition, Section 8 proposes seismic demand models based on the most influential input variables for a quick and adequately accurate estimation of the rocking amplitude. Finally, Section 9 summarises the main outcomes of this paper.

2 | ANALYTICAL ROCKING DYNAMICS

The slender rocking block of Figure 1 is an archetypal structural model that describes the dynamics of a broad class of rocking configurations.[26] When the block is subjected to a horizontal ground excitation, it uplifts and starts rocking once the ground acceleration exceeds the minimum threshold of:

$$\left|\ddot{u}_g\right|_{\min} = g \tan \alpha \tag{1}$$

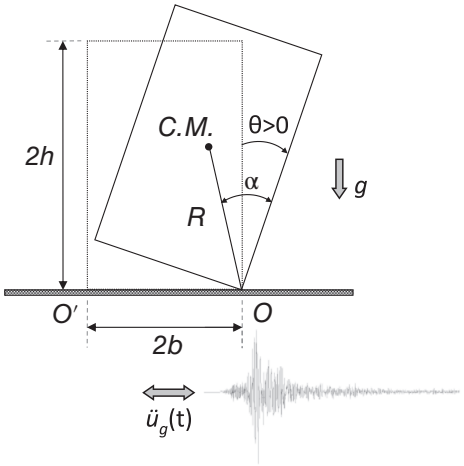

FIGURE 1 The archetypal rocking block model under a horizontal ground excitation.

where g denotes the gravitational acceleration, and α is the slenderness angle $\alpha = tan^{-1}(b/h)$ with b being the block's half-width and h its half-height (Figure 1).

After rocking initiation, moment equilibrium with respect to the pre-defined pivot (contact) points yields the equation that describes the rocking motion of the block.

$$\ddot{\theta} = -p^2\left[\sin(\alpha sgn(\theta) - \theta) + \frac{\ddot{u}_g}{g}\cos(\alpha sgn(\theta) - \theta)\right] \tag{2}$$

where $p = \sqrt{mgR/I_0}$ is the frequency parameter of the block. I_0 represents the moment of inertia of the rigid body with respect to the pivot point(s), m is the mass of the block, R is the half-diagonal distance measured from the pivot point(s) to the centre of mass, and θ is the rocking rotation. $sgn(\theta)$ is the sign function with respect to θ.

During rocking, when $\theta = 0$, impact happens, and energy is lost. The classic rocking theory[1] considers impact an instantaneous event during which no bouncing and sliding occur at the contact interface. Hence, the angular velocities before ($\dot{\theta}^-$) and after impact ($\dot{\theta}^+$) can be connected via the (dimensionless) coefficient of restitution $\dot{\theta}^+ = \eta\dot{\theta}^-$ [1] To estimate η, Housner[1] applied the conservation of the moment of momentum before and after impact, which yielded a closed-form expression with respect to the geometrical characteristics of the block. However, later studies showed that the amount of energy lost at impact is not solely a function of the block's geometry[42] but is also affected by the material of the block and the properties of the contact interface. Thus, the value of the coefficient of restitution η is usually calibrated from experiments[43] or assumed.[4] This study considers η an independent input variable and thoroughly investigates its contribution to rocking response.

3 | APPROPRIATE DEMAND PARAMETER AND INTENSITY MEASURES

For the purposes of the present study, the most appropriate engineering demand parameter that efficiently describes the response of a rigid structure undergoing pure rocking motion (i.e., no sliding and bouncing) is the absolute peak rocking rotation $|\theta|_{max}$ normalised with the slenderness angle α. A rigid rocking block remains in full contact with the ground if the ground acceleration does not exceed the minimum threshold of Equation (1). For such cases, the block sustains non-rocking (NR) with $|\theta|_{max}/\alpha = 0$. In case the ground acceleration exceeds the minimum threshold, the block commences rocking motion ($|\theta|_{max}/\alpha > 0$). By the end of the ground excitation, the block either returns to its initial (rest) position, having undergone safe rocking (SR) motion with a finite nonzero peak rotation ($|\theta|_{max}/\alpha \neq 0$), or overturns (OT) with a theoretically infinite peak rotation ($|\theta|_{max}/\alpha \to \infty$). From a numerical perspective, overturning occurs when $|\theta|_{max}/\alpha \geq \pi/2$.

In general, the response of a rigid rocking structure subjected to an earthquake ground excitation is a function of the structural characteristics p, α, the independent coefficient of restitution η, and the parameters that characterise the waveform of a ground excitation. The selection of appropriate parameters, that is, intensity measures (IMs), that can efficiently represent different attributes of a seismic waveform, such as frequency, amplitude, duration, and energy content, among

TABLE 1 List of dimensionless IMs adopted to represent different attributes of a seismic waveform.

Category	IM	Definition
Intensity-based	$IM_1 = PGA/gtan\alpha$	$PGA = \max(\lvert\ddot{u}_g(t)\rvert)$[44]
	$IM_2 = SPA/gtan\alpha$	SPA = sustained peak acceleration[44]
	$IM_3 = pPGV/gtan\alpha$	$PGV = \max(\lvert\dot{u}_g(t)\rvert)$[44]
	$IM_4 = pSPV/gtan\alpha$	SPV = sustained peak velocity[44]
	$IM_5 = p^2PGD/gtan\alpha$	$PGD = \max(\lvert u_g(t)\rvert$[44]
	$IM_6 = pMIV/gtan\alpha$	MIV = maximum incremental velocity
Frequency-based	$IM_7 = PGA/pPGV$	
	$IM_8 = PGV/pPGD$	
	$IM_9 = pT_m$	$T_m = \sum_i (C_i^2/f_i)/\sum_i C_i^2$ = mean period[45]
	$IM_{10} = \omega_D/p$	ω_D = dominant frequency (maximum of Fourier amplitude spectrum)[44]
	$IM_{11} = \omega_C/p$	ω_C = central frequency[44]
	$IM_{12} = 1/(p\sqrt{t_{uni}/v_0})$	$v_0 = N/t_{brc}$ = number of crossings per unit of (bracketed) time
Duration-based	$IM_{13} = pt_{sig}$	t_{sig} = significant duration[52]
	$IM_{14} = pt_{uni}$	t_{uni} = uniform duration[26]
	$IM_{15} = pt_{brc}$	t_{brc} = bracketed duration[53]
	$IM_{16} = pt_D$	t_D[54]
	$IM_{17} = pt_{sust}$	t_{sust} = sustained duration
Energy-based	$IM_{18} = A_{RMS}/gtan\alpha$	$A_{RMS} = \sqrt{\frac{1}{t_{tot}}\int_0^{t_{tot}} (\ddot{u}_g(t))^2 dt}$ = root mean square acceleration[44]
	$IM_{19} = pV_{RMS}/gtan\alpha$	$V_{RMS} = \sqrt{\frac{1}{t_{tot}}\int_0^{t_{tot}} (\dot{u}_g(t))^2 dt}$ = root mean square velocity[44]
	$IM_{20} = p^2D_{RMS}/gtan\alpha$	$D_{RMS} = \sqrt{\frac{1}{t_{tot}}\int_0^{t_{tot}} (u_g(t))^2 dt}$ = root mean square displacement[44]
	$IM_{21} = p^2L_e/gtan\alpha$	$L_e = PGA \cdot T_m^2$ = energetic length[46]
	$IM_{22} = pI_A/gtan\alpha$	I_A = Arias intensity[47]
	$IM_{23} = p^{1/2}I_C/(gtan\alpha)^{3/2}$	I_C = characteristic intensity[49]
	$IM_{24} = p^{5/4}I_F/gtan\alpha$	$I_F = PGV \cdot t_{sig}^{1/4}$ = Fajfar index[48]
	$IM_{25} = p^3SED/(gtan\alpha)^2$	SED = specific energy density[50]
	$IM_{26} = p^{1/3}I_{Ra}/gtan\alpha$	$I_{Ra} = PGA \cdot t_{sig}^{1/3}$ = Rindell acceleration[51]
	$IM_{27} = pI_{Rv}/(gtan\alpha)^{2/3}$	$I_{Rv} = PGV^{2/3} \cdot t_{sig}^{1/3}$ = Rindell velocity[51]
	$IM_{28} = p^{7/3}I_{Rd}/gtan\alpha$	$I_{Rd} = PGD \cdot t_{sig}^{1/3}$ = Rindell displacement[51]

others, is a challenging task. This study extends the work of Giouvanidis and Dimitrakopoulos[26] and formulates dimensionless IMs, which are both structure- and record-dependent. In Table 1, PGA, PGV, and PGD are the peak ground acceleration, velocity, and displacement,[44] respectively, while $\ddot{u}_g(t)$, $\dot{u}_g(t)$, and $u_g(t)$ are the acceleration, velocity, and displacement time histories of a ground motion. SPA and SPV are the sustained peak acceleration and velocity defined as the third highest (in absolute value) acceleration and velocity peaks with a minimum duration of $20dt$ among them, where dt denotes the time-step of the ground motion record.[44] T_m is the mean period of the ground excitation,[45] whereas ω_d, ω_c represent the dominant and central frequency of the signal.[44] A_{RMS}, V_{RMS}, and D_{RMS} are the root mean square acceleration, velocity, and displacement.[44] L_e is the energetic length scale.[46] In addition, I_A is the Arias intensity,[47] I_F is the Fajfar index,[48] and I_C is the characteristic intensity.[49] SED is the specific energy density,[50] whereas I_{Ra}, I_{Rv}, and I_{Rd} are the Riddell acceleration, velocity, and displacement intensity.[51] Also, t_{sig} is the significant duration[52] (Figure 2E), t_{brc} represents the bracketed duration[53] (Figure 2A), and t_{uni} is the uniform duration[26] (Figure 2D). Due to the importance of velocity metrics on rocking response, the sustained duration t_{sust} is defined as the time interval between the first and third peaks of the absolute ground velocity time history (Figure 2C). Zhou and Katayama[54] introduced a duration-based IM based on the characteristics of the power spectral density of the strong motion record[44] denoted as $t_{D.}$[55] Giouvanidis and Dimitrakopoulos,[26] unveiled the substantial contribution of the individual impulses of exceedance $|\Delta\dot{u}_g|$, that is, the

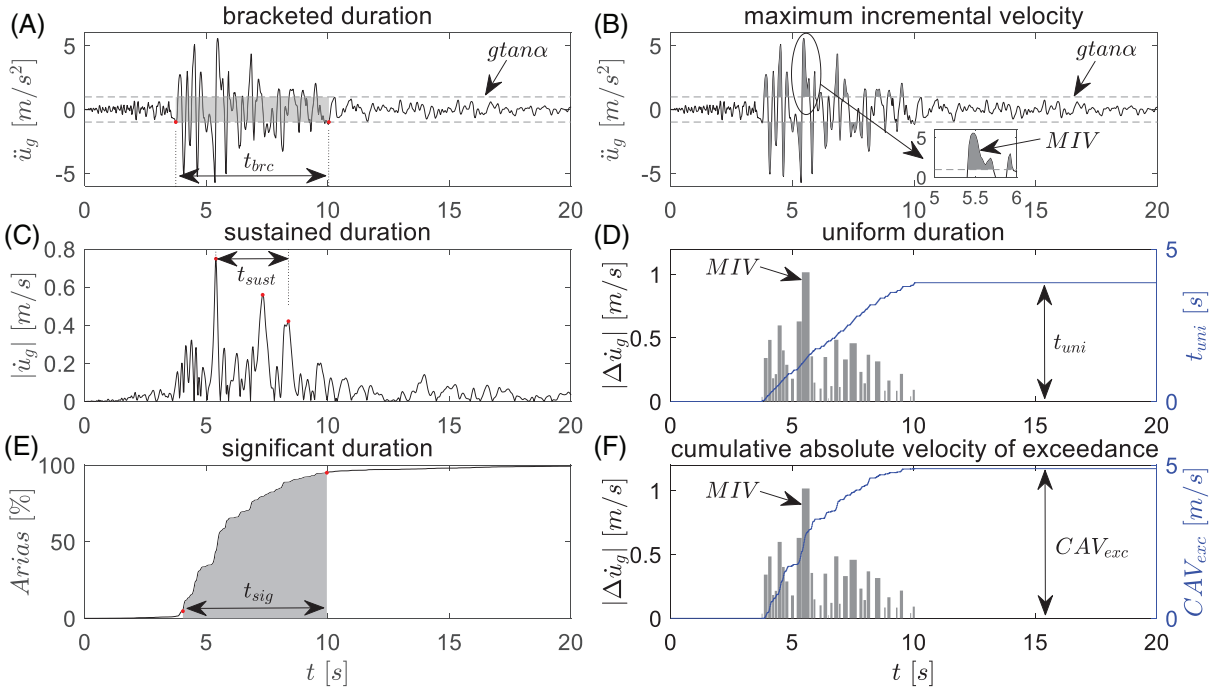

FIGURE 2 Intensity measures of Table 1 using the Northridge-01 earthquake record (Newhall fire station).

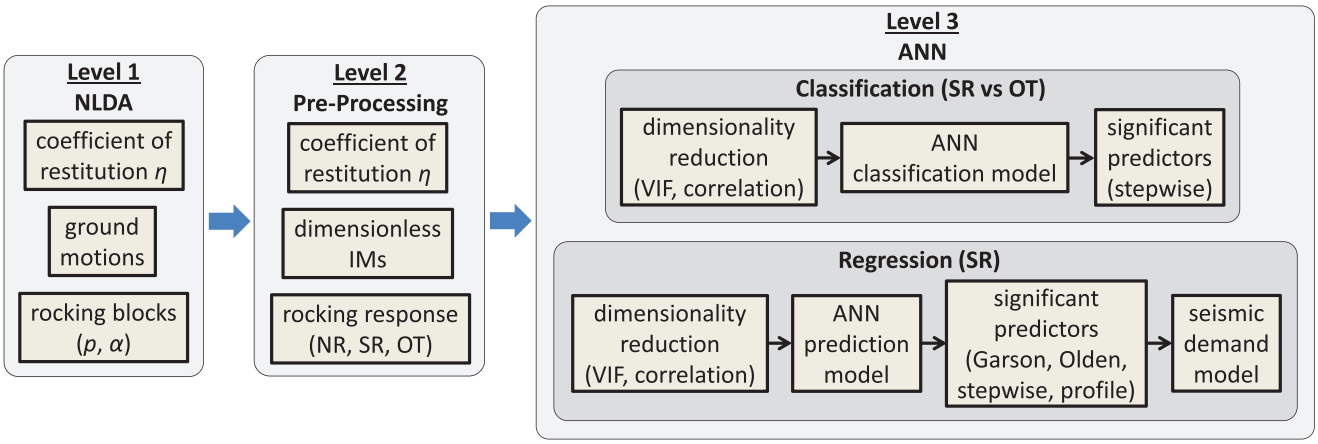

FIGURE 3 The adopted multi-level approach. (NR=non-rocking; SR=safe rocking; OT=overturning).

enclosed (shaded) areas between the accelerogram and the rocking initiation threshold of Equation (1) (Figure 2B), on the safe rocking response. The summation of those areas provides the cumulative absolute velocity of exceedance (CAV_{exc})[26] (Figure 2F). Instead, this work exploits the contribution of these areas' (i) amplitude and (ii) frequency on the response through two novel IMs for the rocking literature (IM_6 and IM_{12} in Table 1). Specifically, the maximum (absolute) area between the accelerogram and two crossing points of the minimum ground acceleration threshold represents the maximum incremental velocity (MIV)[56] (Figure 2B). To capture the distribution of the individual impulses of exceedance throughout time, a frequency-based IM considers the number of crossings (N) per unit of time during which the ground acceleration is capable of triggering rocking motion. Herein, only crossings within the bracketed duration are considered.

4 | THE ADOPTED MULTI-LEVEL APPROACH FOR THE ROCKING PROBLEM

Figure 3 illustrates the multi-level approach adopted to address the rocking problem using ANN. Level 1 consists of the nonlinear dynamic analyses (NLDA) of rocking structures when subjected to a suite of recorded earthquakes, while the

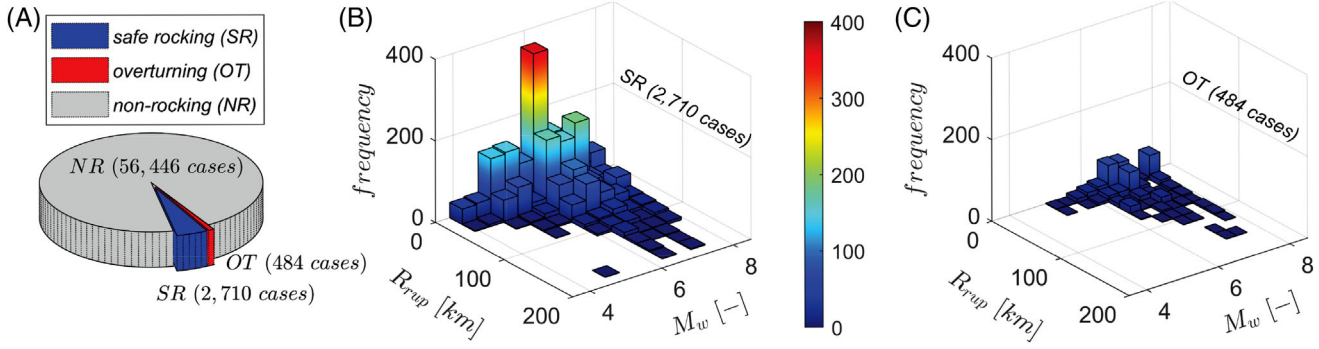

FIGURE 4 (A) Total number of safe rocking (SR), overturning (OT) and non-rocking (NR) cases after nonlinear dynamic analyses, and M_w–R_{rup} distribution with reference to the (B) safe rocking and (C) overturning cases.

other Levels are presented below. As stated, the seismic rocking response depends on the structural characteristics p, α, the coefficient of restitution η, and the parameters that characterise the ground motion waveform. This work assumes that both structural characteristics and the coefficient of restitution follow a uniform distribution. The frequency parameter of the blocks varies $0.7 \leq p \leq 4$ rad/s, yielding a variety of rocking geometries with diagonal distance $0.92 \leq 2R \leq 30$ m. The slenderness angle of the blocks lies between $0.07 \leq \alpha \leq 0.4$ rad, covering a wide range of rocking configurations with aspect ratios $2.3 \leq h/b \leq 15$, capable of representing, for example, building/lab contents, masonry façades or bridge columns. Importantly, this work investigates the contribution of the coefficient of restitution to rocking response. As stated, η is an independent input variable with values $0.72 \leq \eta \leq 0.99$ based on experimental results.[57,58] Finally, recorded earthquake ground motions from the PEER database[59] are employed for the nonlinear dynamic analyses. Each record consists of three components. The effect of the vertical component on pure rocking response is marginal[60,61]; thus, it is neglected. In addition, due to some correlation between the two horizontal components, only one (randomly selected) is used for the dynamic analyses. To induce higher levels of seismic demand, scale factors of 1, 1.5, and 2 are assigned to the accelerograms.[4,62] To preserve the kinematic features of the ground motions and avoid inducing bias in the response, larger scale factors have not been considered.[63,64] Subsequently, 59,640 response-history analyses are conducted.

Figure 4A illustrates the total number of safe rocking (SR), overturning (OT) and non-rocking (NR) cases after conducting nonlinear dynamic analyses (Level 1 in Figure 3). The NR cases are dictated by Equation (1), which is deterministically defined. Thus, in this study, this response mode is omitted. Subsequently, only the safe rocking (2710) and overturning (484) cases are considered (Level 2), which serve as the input dataset for training the ANN models for the classification and regression problems (Level 3). Figure 4B,C show how the safe rocking and overturning cases are distributed with respect to the rupture distance (R_{rup}) and moment magnitude (M_w). Most earthquake records that induce high rocking demands are characterised by R_{rup} <80 km and $M_w > 5$.

5 | DIMENSIONALITY REDUCTION AND SELECTION OF TAILORED INPUT VARIABLES

Machine learning models may produce biased results if there are linear correlations (i.e., collinearity/multicollinearity) between/among the explanatory input variables.[65] One way to detect collinearity between two variables is through the correlation matrix. In general, a Pearson correlation coefficient greater than 0.8–0.9 indicates collinearity.[65] The presence of collinearity can cause problems in the prediction since, for instance, when two IMs tend to increase or decrease simultaneously, it becomes difficult for the ML algorithm to quantify the amount each IM contributes to the response. In other words, the importance of one predictor may be concealed due to the presence of collinearity. In addition, multicollinearity, which is the existence of linear correlations among three or more input variables, may also be present. Similarly, multicollinearity may reduce the predictive power of a single input variable by the extent to which it is associated with other variables.[65] Hence, the use of simple correlation analysis is not sufficient to detect multicollinearity. Different methodologies have been proposed to identify multicollinearity, such as condition index,[66] pairwise correlation coefficient[67] and variance inflation factor (VIF).[68] This study adopts the *VIF* method to investigate potential multicollinearity. Specifically,

the *VIF* for each input variable is evaluated through linear regression analysis with the remaining input variables:

$$VIF_{IMi} = \frac{1}{1 - R^2_{IMi}} \tag{3}$$

where R^2_{IMi} is the square of the multiple correlation coefficient between IM_i and the remaining IMs. The minimum value of *VIF* is 1 when $R^2_{IMi} = 0$ and indicates a complete absence of multicollinearity. As the value of *VIF* increases, the level of multicollinearity increases. In general, *VIF* values greater than 10 indicate the presence of multicollinearity.[65,68,69] Common ways to minimise the impact of multicollinearity on the prediction are: (i) imposing a penalty for including additional variables into the model and/or (ii) combining two or more input variables into a single index, especially when they are conceptually similar.[69,70] This study adopts the former method. This pre-processing analysis is implemented through dimensionality reduction, namely by iteratively removing input variables with correlation coefficient and *VIF* value higher than the adopted thresholds.

6 | ARTIFICIAL NEURAL NETWORK ARCHITECTURE

As essential tool in machine learning, ANN is a computational modelling tool utilised to learn and model real-world complex problems for which conventional algorithms struggle. ANN is an established data processing algorithm used to identify complex nonlinear relations between a set of input and output variables. It can be purely data-driven or governed by physical laws (i.e. physics-informed ANN). The key difference between them lies in the ability of the latter to incorporate physical laws of dynamics into the learning process. Therefore, in principle, physics-informed ANN is able to provide predictions tailored to the physical restrictions that govern the particular engineering problem. Investigating the applicability of physics-informed ANN to the rocking problem merits further investigation, which is beyond the scope of the present study.

Figure 5 illustrates the architecture of a typical ANN model, which is formed by (at least) three layers, namely input layer, output layer, and hidden layer(s). The appropriate number of hidden layers is a challenging task. In general, the number of hidden layers varies based on the complexity and nature of the problem and the necessity to capture the nonlinearity between input and output variables. For most engineering problems, up to three hidden layers are considered adequate for both regression and classification. Each layer consists of neurons, which are connected and interact with each other via adjustable weights. The number of neurons in the input and output layers corresponds to the number of input and output variables, respectively. On the other hand, the number of neurons in each hidden layer varies and depends on the peculiarities of the problem. A model with too few neurons cannot properly identify and learn the underlying nonlinear patterns in the dataset, while too many neurons may cause overfitting of the model.[71,72]

The most common ANN architecture is the multi-layer feed-forward network or multi-layer perceptron (MLP) algorithm.[73] In MLP algorithms, signals are transmitted in a forward direction (i.e., forward propagation) from the input layer to the hidden layer(s) and then towards the output layer. To develop a numerical model exploiting a relation between input and output variables, the signal entering the neurons in the hidden layer(s) is obtained by multiplying the input signal (x_j) with appropriate weighting factors (w_{kj}). A bias (b_k) is also added to shift the activation function by a constant amount. The linear combination of the weighted inputs and bias is the net input (u_k) for the activation function (Figure 5).

$$u_k = \sum_{j=1}^{m} w_{kj} x_j + b_k \tag{4}$$

The activation function $f(\bullet)$ identifies complex patterns in the dataset and introduces nonlinearities to the network. The output value of the signal depends on the type of the activation function. Two basic types of activation functions are (i) linear and (ii) nonlinear. For regression problems, a nonlinear function is usually adopted. Thus, using the hyperbolic tangent sigmoid function, the output of each neuron $f(u)$ becomes[71]:

$$f(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}} \tag{5}$$

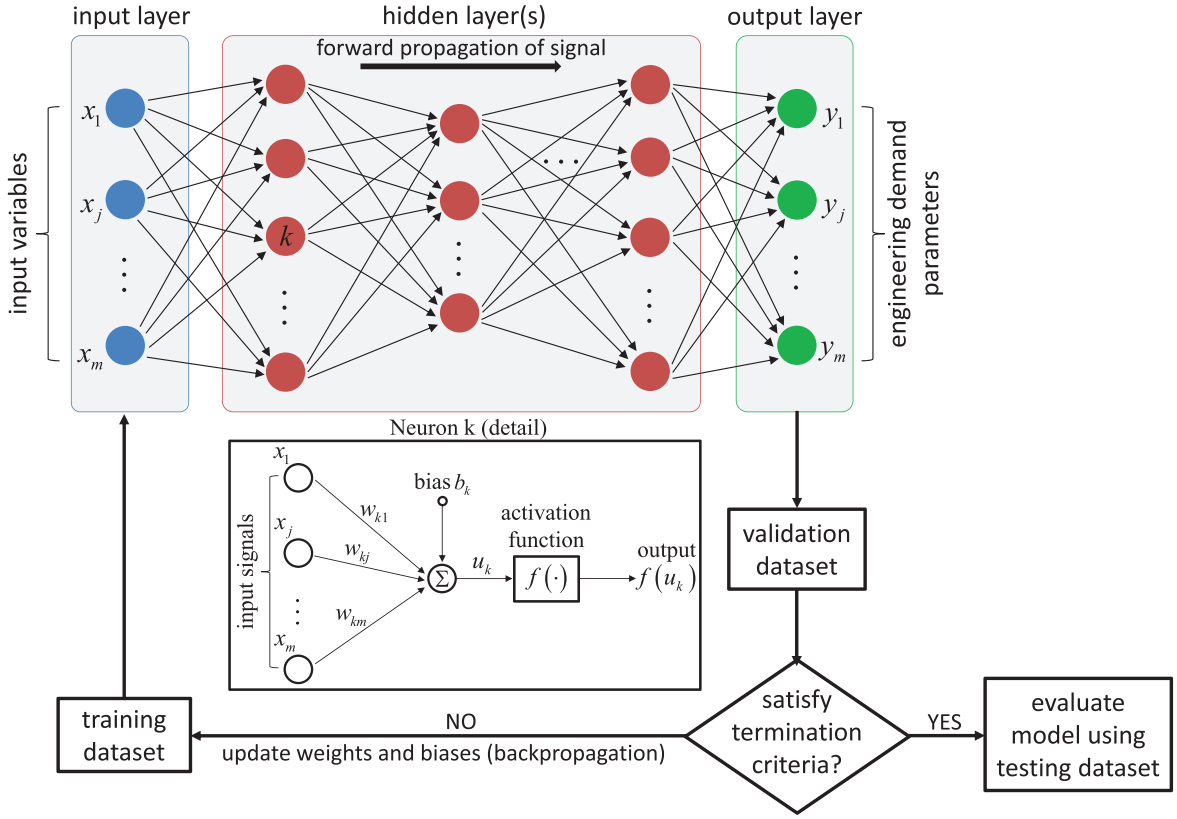

FIGURE 5 Architecture of a typical artificial neural network (ANN) and a neuron k.

For classification problems, a nonlinear activation function can also be used. This study adopts the softmax (or normalised exponential) function[74]:

$$f(u)_i = \frac{e^{u_i}}{\sum_{j=1}^{l} e^{u_j}} \tag{6}$$

where $f(u)_i$ is the i-th value of the output vector of the classification model, and l is the number of classes. Once the signal reaches the output layer, the assumed termination criterion is evaluated. The termination criterion for regression and classification problems is the error between outputs and desired targets. Minimisation of this error serves as the objective function of the problem. To solve such nonlinear least-squares problems within an iterative procedure, this paper adopts the Levenberg–Marquardt (LM) algorithm.[75,76] When minimisation of the objective function is achieved, the learning process terminates, and the ANN model is evaluated against the test dataset. Otherwise, the output signals propagate to the input layer to adjust the parameters of the model (i.e., weights and biases) using a backpropagation algorithm (Figure 5). These parameters are progressively adjusted till the termination criterion is satisfied. The ultimate goal of the training procedure is to define an optimal set of weights and biases that produces the most accurate output.

Two main challenges of machine learning algorithms, which may result in their poor performance, are (i) overfitting and (ii) shortage of input data for training.[71] When overfitting occurs, the model performs well on the training dataset. However, its performance deteriorates on the test dataset. There are several solutions to overcome overfitting.[77] This study adopts the early stopping method.[78] In this context, the whole dataset is divided into training, validation, and test datasets. The training dataset is utilised to optimise the parameters of the ANN model (i.e., weights and biases). The validation dataset, which is isolated from the training dataset, is used during the training process to evaluate the performance of the trained model. An indication of overfitting is when the error on the validation dataset starts increasing. The early stopping method identifies when the error starts increasing and terminates the training process. When the performance of the model is satisfactory on the training-validation dataset, the test dataset is used to evaluate the effectiveness of the model. Appropriate division of the input data is of primary importance to ensure a well-trained and thus generalised model. A generalised model is adequately robust to an unseen (new) dataset. To this aim, the entire dataset is partitioned

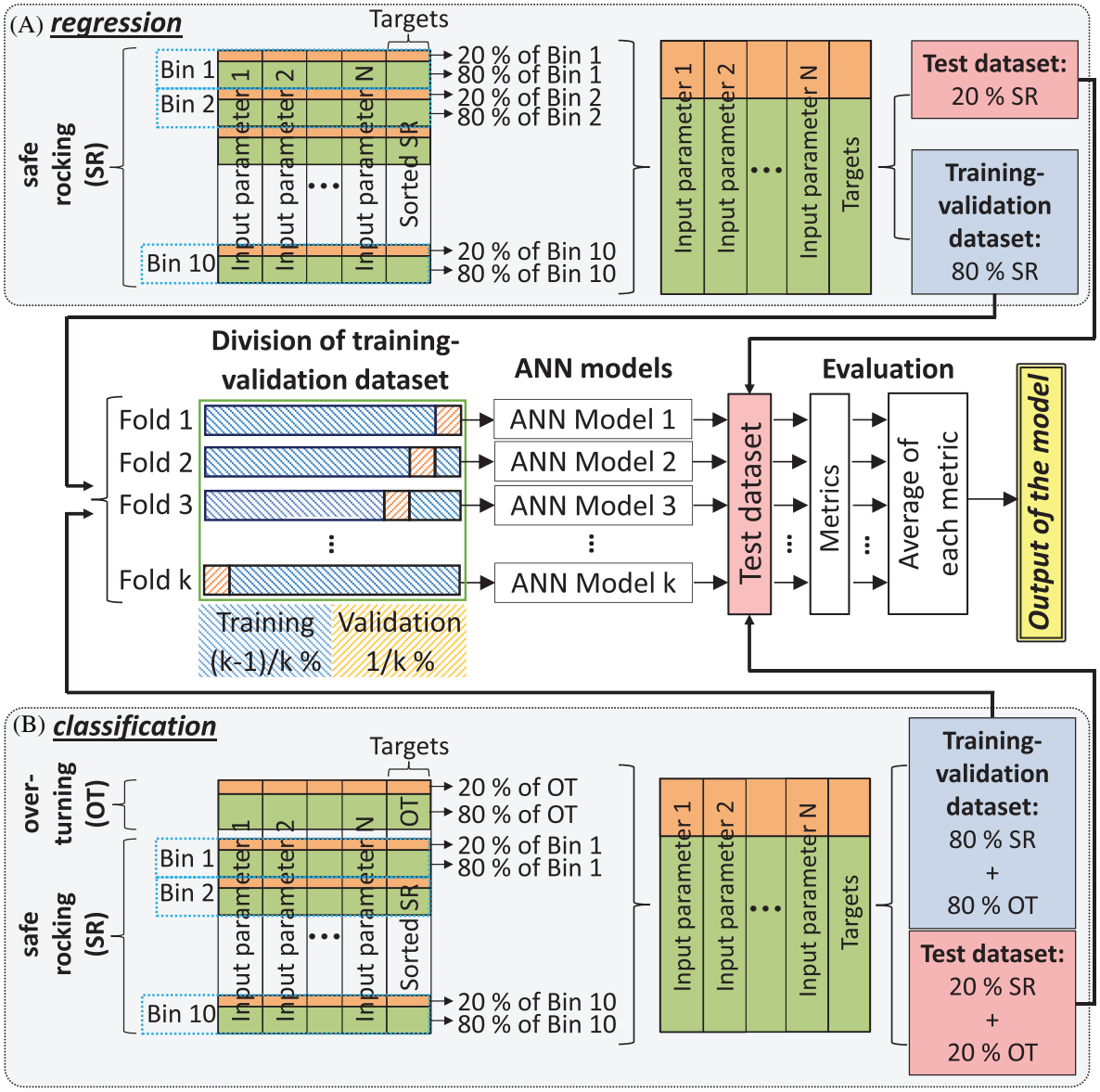

FIGURE 6 Flowchart of the adopted algorithm for the (A) regression, and (B) classification problems.

into bins of equal size, determined by sorting the target values. The training-validation dataset, which is assumed to be 80% of the whole dataset, is randomly extracted from each bin (Figure 6A). This selection guarantees well-distributed data, contributing to a more efficient training procedure. As a result, the ANN model is trained based on a wide range of target values, which facilitates the generalisation of the model.

In cases where there is a shortage of input data, the risk of losing important information on the trends between input and output variables becomes higher. As a result, the ANN model may become biased. To improve the stability of the model and minimise issues arising due to the lack of sufficient input data, cross-validation, such as the k-fold cross-validation approach, is employed.[79] The procedure commences assuming that 20% of the whole dataset serves as the test dataset. This part does not participate in the learning process. The remaining 80% is randomly divided into k groups of samples such that k-1 subsets are used for training and the remaining one for validation (Figure 6A). In this study, k equals 10.[79,80] Note that selecting the appropriate number of folds is an optimisation problem, which, however, is beyond the scope of the present study. The procedure is repeated k times so that each subset has been used as a validation dataset. Subsequently, k ANN regression models are trained based on the training-validation dataset and evaluated based on the test dataset.

A similar procedure is utilised for the classification problem (Figure 6B). The main difference between the regression and classification problem lies in that, for the regression problem, the overturning samples are omitted. On the

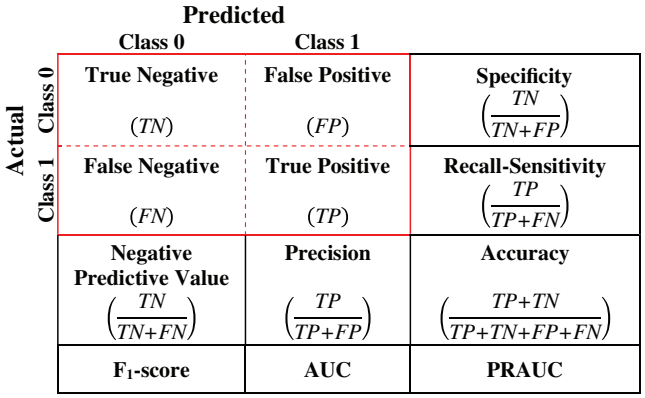

FIGURE 7 Confusion matrix and corresponding metrics for evaluating a classification model.

contrary, predicting overturning is the main objective in the classification problem; thus, the overturning samples are considered. Similarly, to construct the training-validation dataset, this work randomly selects 80% from each bin of the safe rocking samples and 80% of the overturning samples. The remaining 20% of safe rocking and overturning samples are considered as the test dataset. Again, the k-fold cross-validation approach is adopted to facilitate a more efficient training process.

6.1 | Evaluating the performance of an ANN model

6.1.1 | Regression

The performance of the ANN model is evaluated using different statistical indices. For regression problems, where the output takes a continuous value, the performance of the model can be evaluated through the root mean square error (*RMSE*), (Pearson) correlation coefficient (r_P), coefficient of determination (R^2), mean absolute error (*MAE*), and symmetric mean absolute percentage error (*sMAPE*):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2}, \quad r_P = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2 \cdot \sum_{i=1}^{n}(Y_i-\bar{Y})^2}}, \quad R^2 = 1 - \frac{\sum_{i=1}^{n}(X_i-Y_i)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$
$$MAE = \frac{1}{n}\sum_{i=1}^{n}|X_i - Y_i|, \quad sMAPE = \frac{1}{n}\frac{|X_i-Y_i|}{|X_i|+|Y_i|} \tag{7}$$

where n is the number of data points. X_i and Y_i are the actual and predicted values, respectively, whereas $\bar{X}$ and $\bar{Y}$ are the mean of the actual and predicted values.

6.1.2 | Classification

In classification problems, logistic regression as a supervised learning approach is a powerful tool for solving binary problems.[81] The output of the logistic regression algorithm is a continuous value varying between zero and unity. This value can be converted into a firm class label using a pre-defined threshold where samples with outputs higher (or equivalently lower) than the assumed threshold are categorised into a specified class. For binary classification problems, the default and most common threshold is 0.5. The performance of a classifier can be evaluated through the confusion matrix. Figure 7 illustrates a typical confusion matrix for binary problems, where the safe rocking (SR) cases are the negative class (Class 0) while the overturning (OT) cases are the positive class (Class 1). Four key values can be arranged in a confusion matrix: True Positive (*TP*) indicates the number of positive (overturning) samples which are correctly classified, False Positive (*FP*) represents the number of actual negative (safe rocking) samples which are falsely classified as positive (overturning), True Negative (*TN*) denotes the number of negative (safe rocking) samples which are correctly classified, and False Negative (*FN*) is the number of actual positive (overturning) samples which are falsely classified as negative (safe

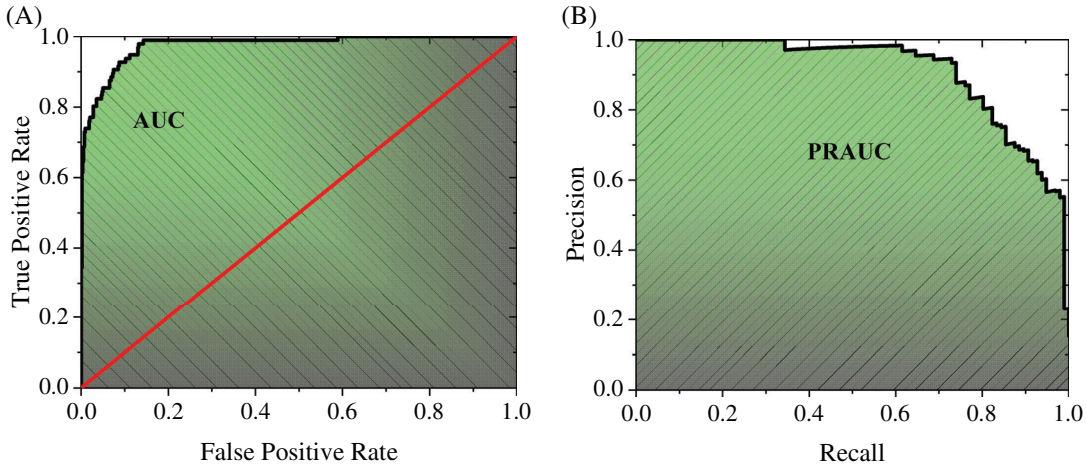

FIGURE 8 Typical (A) Receiver Operating Characteristics (ROC) curve, and (B) Precision–Recall (PR) curve.

rocking). Along with the confusion matrix, metrics such as Accuracy, Recall or Sensitivity or True Positive Rate (TPR), Precision, F_1-score, and False Positive Rate (FPR) are also used to evaluate the performance of a classifier:

$$\begin{aligned} &\text{Accuracy} = \frac{TN+TP}{TN+TP+FN+FP}, \quad \text{Recall} = \frac{TP}{TP+FN}, \quad \text{Precision} = \frac{TP}{TP+FP} \\ &\text{F}_1\text{−score} = 2\frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}, \quad \text{FPR} = \frac{FP}{TN+FP} \end{aligned} \tag{8}$$

Accuracy is the ratio of all correctly predicted samples over the total number of samples. Precision identifies the proportion of the positive samples that are correctly predicted, whereas Recall (or Sensitivity or TPR) is the fraction of the truly predicted positive samples out of all actual positive samples. Among those metrics, Accuracy may be misleading, especially for an imbalanced dataset, that is, with a highly unequal distribution of classes in the training dataset—the number of safe rocking versus overturning cases—since it does not provide information about the class of the predicted samples. Classifying the minor (overturning) class in an imbalanced dataset is very important. Thus, Recall and Precision are more effective metrics than Accuracy. However, when the importance of false positive cases is high, namely the actual safe rocking cases that are falsely predicted as overturning, Precision is a valuable metric for evaluating the model. In this study, high Precision means a low number of safe rocking cases classified as overturning. In contrast, Recall becomes appropriate when the importance of false negative cases is high, herein the actual overturning cases falsely classified as safe rocking. High Recall implies a low number of overturning cases classified as safe rocking. Therefore, in cases where both Recall and Precision need to be considered, the F_1-score becomes a more efficient metric. F_1-score represents the harmonic mean between Precision and Recall and, similarly, takes values between zero and unity. A higher F_1-score implies a better classifier with a balance between Precision and Recall.

A limitation of the evaluation metrics of Equation (8) is their sensitivity to the assumed threshold value. Particularly for imbalanced datasets, the threshold value might affect the prediction of the class of a sample. In lieu of predicting the class of a sample, the probability of a sample belonging to each class can be employed as a metric for evaluating classification models. Probabilistic measures, such as the Receiver Operator Characteristics (ROC) curve and the Precision–Recall (PR) curve, enable the prediction of the sample class and provide the predicted probability of the event across various threshold values, ranging from zero to unity. Specifically, ROC is the curve with the False Positive Rate (FPR) values on the abscissa and the True Positive Rate (TPR) values on the ordinate, where each point on the ROC curve corresponds to a different threshold value (Figure 8A). The Area Under Curve (AUC)[81] is a single metric defined as the area under the ROC curve. AUC ranges from zero to unity. When AUC = 1, the classifier perfectly distinguishes between all positive and negative class points. On the contrary, when AUC = 0, the classifier predicts all negatives as positives and all positives as negatives. For the special case of AUC = 0.5 (denoted with the red line in Figure 8A), the classifier is unable to distinguish between the two classes, meaning that its predictions become random. Thus, in general, to get better class predictions, $0.5 < \text{AUC} < 1$. The higher the AUC, the better the classifier. Since AUC provides information for both classes, it can be safely used to evaluate classification models even for imbalanced datasets. However, for severely imbalanced datasets with the negative (safe rocking) class being the major, the positive (overturning) class as the minor has less effect on the ROC curve and,

subsequently, the AUC value. Therefore, a probabilistic curve that focuses on the minor class is a more appropriate metric. To this end, Precision–Recall (PR) is a curve with the Precision values on the ordinate and the Recall values on the abscissa, considering different probability thresholds (Figure 8B). Similarly, the area under the PR curve (PRAUC) is a metric used to evaluate a classifier. A higher PRAUC value denotes a better classifier.

6.2 | Identification of significant predictors (feature importance) – Sensitivity analysis

6.2.1 | Regression

Various methods have been proposed to identify the most effective parameters (or significant predictors) within a neural network approach. This study adopts four different approaches: the Garson method,[82] the connection weight or Olden method,[83] the improved stepwise method,[84] and the profile method.[85]

- ***Garson method***

According to the Garson method, each connection weight between the hidden and output layers is partitioned into components associated with each input neuron. The relative importance factor is[82]:

$$I_{ik} = \sum_{j=1}^{L} \left(\frac{w_{ij} v_{jk}}{\sum_{r=1}^{N} w_{rj}} \right) \Bigg/ \sum_{i=1}^{N} \left[\sum_{j=1}^{L} \left(\frac{w_{ij} v_{jk}}{\sum_{r=1}^{N} w_{rj}} \right) \right] \tag{9}$$

where I_{ik} is the influence of the i-th input parameter on the k-th output. N and L are the number of neurons related to the input and hidden layers, respectively. w_{ij} is the connection weight between the i-th input neuron and j-th hidden neuron, while v_{jk} is the connection weight between the j-th hidden neuron and k-th output neuron. A limitation of the Garson method is that it uses solely the magnitude of the connection weights to calculate the importance factor.

- ***Connection weight or Olden method***

Olden and Jackson[83] introduced the connection weight or Olden method, in which both the magnitude and the direction (i.e. positive or negative) of the connection weights are considered to calculate the importance factor of each input parameter to the output. The importance factor is the sum of the products of the connection weights w_{ij} between the input and hidden layers with the connection weights v_{jk} between the hidden and output layers:

$$I_{ik} = \sum_{j=1}^{L} w_{ij} v_{jk} \tag{10}$$

- ***Improved stepwise method***

In the classical stepwise method, the effect of the step-by-step adding and/or withdrawing input parameters on the output is quantified by the mean square error (*MSE*) between the output and the target. However, such a method requires a new ANN model for each step that needs training. In contrast, the improved stepwise method utilises only a single ANN model.[84] In this approach, each input parameter is sequentially fixed at its mean value to investigate potential deviations of the *MSE*. The resulting *MSE* estimates the relative importance of each input parameter on the target, with the most important parameter having the largest *MSE* value.

- ***Profile method***

Contrary to the improved stepwise method, the profile approach[85] keeps the examined input parameter unchanged while all remaining parameters are simultaneously fixed at their minimum, first quartile, median, third quartile, and

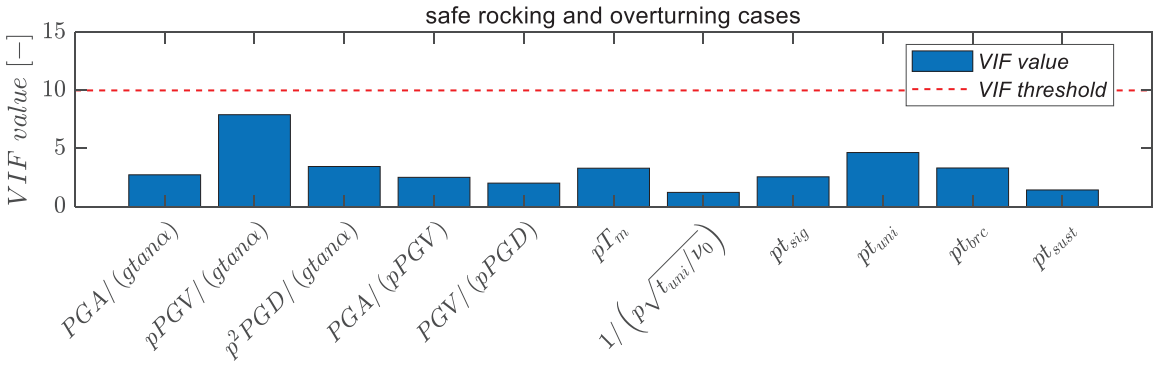

FIGURE 9 *VIF* analysis on the examined intensity measures of Table 1 for the classification problem after removing linear dependency.

maximum values, successively. Therefore, for each input parameter, five new datasets are created. For each dataset, the ANN model predicts the outputs. The variation of the predicted outputs for each input parameter reveals its importance. Specifically, the larger the variation of the output, the higher the impact of the examined input parameter on the output.

6.2.2 | Classification

To identify the most effective parameters (or significant predictors) in classification problems, the Garson, Olden, and profile methods cannot be used since they are developed based on continuous output values. Therefore, this study utilises the improved stepwise method, which can also be used for discrete output values.

7 | CLASSIFYING ROCKING RESPONSE WITH THE USE OF ANN – SAFE ROCKING VERSUS OVERTURNING AND THE ROLE OF THE COEFFICIENT OF RESTITUTION

This section generates an ANN model to classify rocking response to either (i) safe rocking, where the adopted structures undergo rocking motion without any danger of overturning, or (ii) overturning, based on characteristics of the seismic signal, structural parameters, and the coefficient of restitution.

7.1 | Dimensionality reduction for the classification of rocking response

Prior to training the ANN model, tailored input variables need to be selected from Table 1. To avoid collinearity/multicollinearity, reduce the dimensionality of the problem, and enhance the reliability of the ANN model, this section conducts correlation and *VIF* analyses, the results of which are omitted herein for brevity. Figure 9 shows the (linearly) independent input variables to serve as input dataset for the classification problem. Note that the coefficient of restitution η, despite not being considered in the correlation and *VIF* analyses due to its status as an independent input variable, is still considered during the training process of the ANN model.

7.2 | An ANN model to classify rocking response and identify the most significant predictors

There is no easy method to define the most optimal parameters to describe a neural network, for instance, the number of hidden layers and neurons for each hidden layer. This subject is an optimisation problem, which is beyond the scope of the present study. However, 'trial and error' is an efficient and simple method to define reasonable values for the parameters of an ANN model. The ANN model includes three layers: (i) an input layer, which consists of 12 neurons (one for each input variable of Figure 9 and the coefficient of restitution), (ii) a hidden layer of 15 neurons, and (iii) an

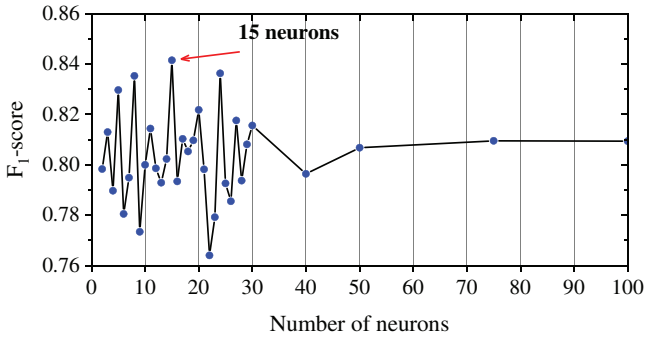

FIGURE 10 Optimal number of neurons in the hidden layer for the classification problem.

Actual \ Predicted	Class 0	Class 1	
Class 0	True Negative **529**	False Positive **9**	Specificity **98.3 %**
Class 1	False Negative **20**	True Positive **76**	Recall **79.2 %**
	Negative Predictive Value **96.4 %**	Precision **89.4 %**	Accuracy **95.5 %**
	F_1-score **0.840**	AUC **0.979**	PRAUC **0.908**

FIGURE 11 Confusion matrix and related metrics of the classifier based on the input variables of Figure 9 and the coefficient of restitution.

output layer of a single neuron, which represents the binary outcome of the classification problem, that is, safe rocking or overturning. Figure 10 illustrates the optimal number of neurons in the hidden layer, based on which the F_1-score becomes maximum. Furthermore, this study considers one hidden layer.[86,87] The connection weights and biases in the hidden layer are adjusted by the Levenberg–Marquart (LM) backpropagation algorithm, while the activation function is assumed to follow a Softmax function. In the output layer, a linear activation function is utilised to scale the output of the model to the desired values. Finally, the maximum epoch is 1000 based on 'trial and error', which is sufficient for this study.

The input dataset used for the classification problem consists of 2710 safe rocking and 484 overturning samples (Figure 4). 80% of all samples (i.e., 2560) serves as the training-validation dataset, and the remaining 20% (i.e., 634) is considered for the test process. Figure 11 evaluates the generated ANN classification model on the test dataset through a confusion matrix using the metrics of Section 6.1.2. The diagonal entries of the confusion matrix denote the classes that are correctly classified, and the off-diagonal entries represent the misclassified classes. Specifically, the model correctly classifies 529 out of 538 (actual) safe rocking cases (i.e., 98.3%) and 76 out of 96 (actual) overturning cases (i.e., 79.2%). The classification model shows a prediction accuracy of 95.5 % (i.e., 605 out of 634 cases are correctly classified). In addition, Figure 11 shows that even though Recall indicates an ordinary level of the model's performance (i.e., 79.2%), AUC, PRAUC, and F_1-score show an excellent prediction ability of the classifier, 97.9%, 90.8%, and 84%, respectively. Recall that probabilistic metrics, such as AUC and PRAUC, represent more reliable metrics for classifiers and, thus, illustrate more efficiently the reliability of the trained classification model in predicting the binary response (safe rocking or overturning) of a rocking block when subjected to recorded earthquakes.

To identify the significant predictors that contribute the most to the binary outcome of the classification problem, this study adopts the improved stepwise method, based on which the input parameters are ranked according to their influence on the output.[84,86] To this aim, the AUC, PRAUC and F_1-score metrics serve as objective functions. The lower

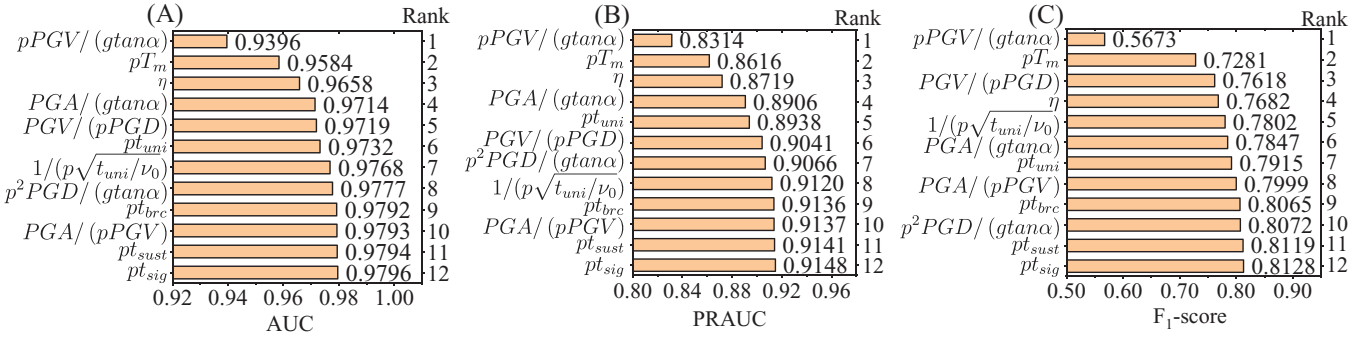

FIGURE 12 Significant predictors that contribute the most to the classification of rocking response (safe rocking or overturning) based on the (A) AUC, (B) PRAUC, and (C) F_1-score.

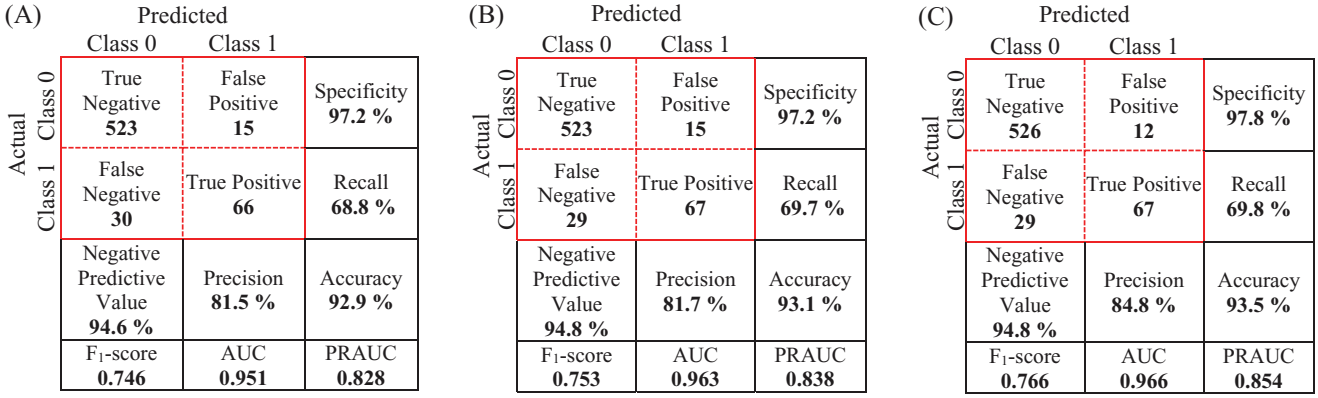

FIGURE 13 Confusion matrix and related evaluation metrics of the classifier based on the two most significant predictors of Figure 12 (i.e., $pPGV/(gtana)$ and pT_m).

the AUC, PRAUC, and F_1-score, the more effective the parameter. Figure 12 ranks the contribution of each input variable to the output of the classification problem. Figure 12 shows that the distinction between safe rocking and overturning is mainly dictated by the velocity characteristics of the seismic signal, herein *PGV*. Furthermore, Figure 12 shows that frequency characteristics, such as T_m, are also critical to the outcome of the classification problem. In addition, both velocity (*PGV*) and frequency (T_m) characteristics show consistency regardless of the adopted objective function, that is, AUC (Figure 12A), PRAUC (Figure 12B), and F_1-score (Figure 12C). Interestingly, the coefficient of restitution η and the peak ground acceleration *PGA* are also important to the classification problem. This is based on the realisation that the coefficient of restitution η controls the amount of energy dissipated at each impact, and the intensity-based *PGA* dominates the response of stockier blocks since those blocks require higher levels of ground acceleration to initiate rocking, which, soon after, cause overturning.

Based on the outcomes of Figure 12, Figure 13 evaluates a classification model that is trained using the two most important and consistent input variables, that is, $pPGV/(gtana)$ and pT_m, via a confusion matrix. As expected, Figure 13 shows that the classification model is less accurate (on average 93%) compared to the model of Figure 11, where all parameters are considered (95.5%). However, the difference is marginal. Thus, the classification model based on the two most important parameters of Figure 12 is accurate and reliable enough considering the uncertainties that characterise an ANN model (e.g., number of hidden layers and neurons per hidden layer, training function, number of samples, etc.). In general, Figures 12 and 13 reveal that a vector-valued IM consists of $pPGV/(gtana)$ and pT_m shows high contribution to the binary outcome of the classification problem. Thus, it could be adopted as an optimal vector-valued $IM_{OT} = [pPGV/(gtana),\ pT_m]$ to predict the probability of overturning of rocking structures when subjected to recorded earthquakes. However, this topic merits further investigation, which is beyond the scope of the present study.

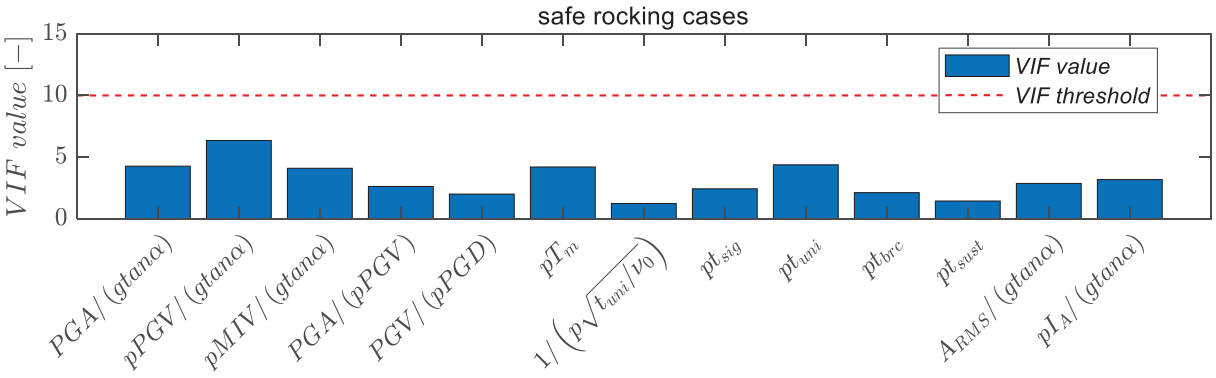

FIGURE 14 VIF analysis on the examined intensity measures of Table 1 for the regression problem after removing linear dependency.

8 | PREDICTION OF ROCKING AMPLITUDE WITH THE USE OF ANN AND THE ROLE OF THE COEFFICIENT OF RESTITUTION

This section focuses on the safe rocking cases, where the adopted structures survive the applied ground motions and return to their initial (rest) position. The aim of this section is twofold. Firstly, it generates an ANN model to predict the rocking amplitude. Secondly, it reveals the contribution of each input variable to the rocking amplitude with special emphasis on the coefficient of restitution.

8.1 | Dimensionality reduction for the prediction of rocking amplitude

Similarly, the identification and elimination of collinearity/multicollinearity is a crucial step prior to the training of the ANN model. This section conducts correlation and *VIF* analyses to detect and minimise the effect of collinearity/multicollinearity on the prediction. For brevity, those results are omitted. Figure 14 presents the adopted (linearly) independent variables to serve as input for training the ANN model. Again, as an already independent parameter, the coefficient of restitution is not considered in the correlation and *VIF* analyses. However, it is still considered in the training process of the ANN model.

8.2 | An ANN model to predict rocking amplitude and identify the most significant predictors

The ANN model for the regression problem consists of three parts: (i) an input layer, (ii) a number of hidden layers and (iii) an output layer. A 'trial and error' procedure is utilised to identify the most optimal number of hidden layers and neurons in each hidden layer. Figure 15A shows that a single hidden layer is adequate to train the ANN model efficiently.[88,89] In

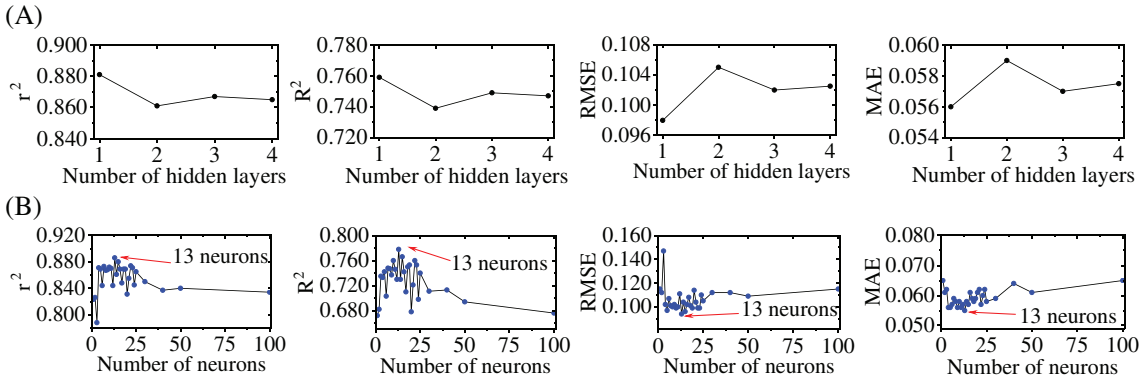

FIGURE 15 Optimal number of (A) hidden layers and (B) neurons in each hidden layer for the regression problem.

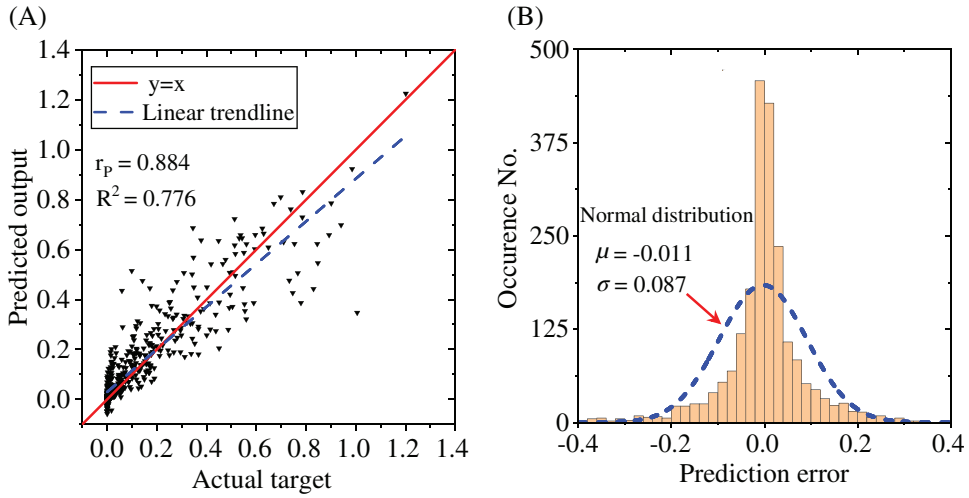

FIGURE 16 (A) Actual targets versus predicted outputs, and (B) distribution of the prediction error.

addition, Figure 15B shows that the hidden layer with 13 neurons in which the tangent hyperbolic activation function transforms the input signal into the output provides the optimal solution. The input layer consists of 14 neurons (i.e., one for each input variable of Figure 14 and the coefficient of restitution). A linear activation function is used in the output layer. Again, a maximum epoch of 1000 is sufficient for this study.

Figure 16A plots the actual target with respect to the corresponding predicted output. In addition, Figure 16A presents the linear trendline and emphasises the difference between the current prediction (dashed blue line) and the perfect one (solid red line). The predicted outputs present a Pearson correlation coefficient as high as $r_P = 0.884$ and a coefficient of determination $R^2 = 0.776$, illustrating the sufficient predictive ability of the ANN model. Observe in Figure 16A that for close to zero actual targets, the ANN model predicts some negative outputs, which have no physical meaning. Nevertheless, the ANN model accurately predicts the peak safe rocking response, considering that both targets and outputs are on a natural scale where a certain level of scatter is expected. Finally, Figure 16B highlights the distribution of prediction error values based on the test dataset. It shows that the mean value of the prediction error is close to zero ($\mu = -0.011$) with a standard deviation of $\sigma = 0.087$. Based on the observed characteristics of the residuals, the error conforms to a normal distribution. This normality of errors implies that the model is unbiased, and the errors are random, indicating that the ANN model is a reliable and valid predictor of the rocking amplitude.

To reveal the contribution of each input variable (i.e., the IMs of Figure 14 and the coefficient of restitution η) to the rocking amplitude, this section adopts the trained ANN model to rank them from the most to the least important through the Garson, Olden, Stepwise, and Profile methods. Figure 17 shows that each method provides a slightly different ranking. This is due to the different processes each approach follows to quantify the contribution of each input variable (Section 6.2.1). Nevertheless, the different methods show an adequate level of consistency. Specifically, Figure 17 unveils that rocking amplitude is governed by a combination of the frequency-based pT_m, the duration-based pt_{uni}, and the intensity-based $pMIV/(gtana)$ as well as $pPGV/(gtana)$ regardless of the adopted approach. Note that the frequency of appearance of the individual impulses of exceedance $|\Delta\dot{u}_g|$ captured by $1/(p\sqrt{t_{uni}/v_0})$, the intensity-based $PGA/(gtana)$, and the frequency-based $PGA/(pPGV)$ play a considerable, yet not dominant, role. On the contrary, the contribution of the coefficient of restitution η on rocking amplitude is inconsistent. In general, η has a marginal effect on the rocking amplitude. Interestingly, Figure 17 unveils the consistent significance of a novel IM for the rocking literature, the intensity-based $pMIV/(gtana)$, which represents the maximum impulse of exceedance $|\Delta\dot{u}_g|$, that is, the maximum enclosed area between the accelerogram and the rocking initiation threshold (Figure 2). Further corroboration of those results comes from a sensitivity analysis. The analysis shows that solely the contribution of the top input variables (denoted with shaded bars in Figure 17) remains consistent and thus reliable. Finally, note that the Olden method also reveals the sign of contribution, implying that when all input variables are considered, for instance, the frequency-based $PGV/(pPGD)$, due to its interaction with the remaining variables, does not simultaneously increase with the rocking amplitude.

Table 2 compares the predictive ability of the ANN model when it considers all input variables (of Figure 14 and the coefficient of restitution) versus solely the most significant ones from Figure 17 using, respectively, the Garson, Olden, Stepwise, and Profile methods. As expected, the performance of the ANN model deteriorates when fewer input variables

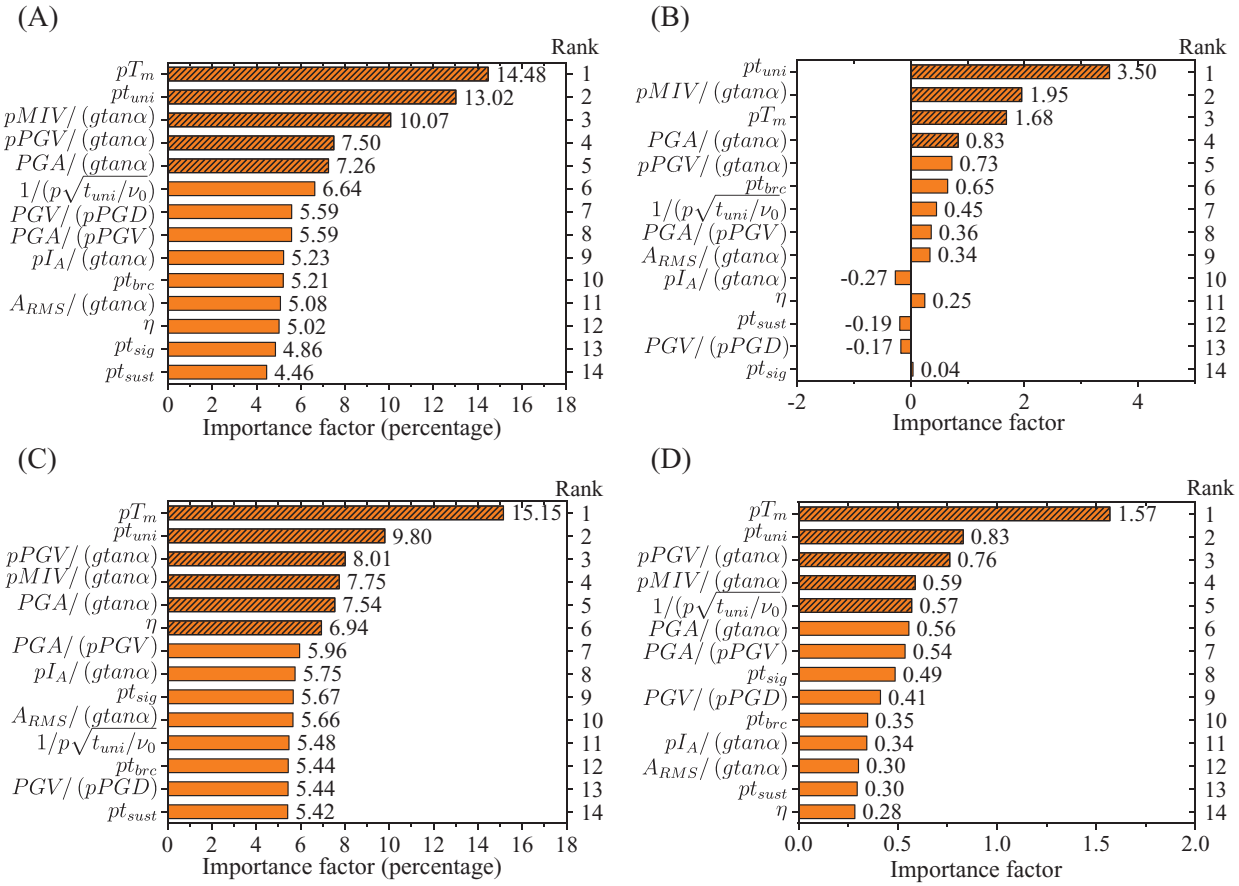

FIGURE 17 Significant predictors that contribute the most to the rocking amplitude based on the (A) Garson, (B) Olden, (C) Stepwise, and (D) Profile methods.

are considered. Specifically, when the four most significant predictors of Figure 17 are considered, the prediction accuracy of the rocking amplitude is characterised by $R^2 = 0.726$, a marginal decrease from $R^2 = 0.776$ when all input variables are considered. Thus, the four most important input variables can be potentially adopted to reduce the dimensionality and simplify the rocking problem. By further decreasing the number of input variables, the prediction accuracy of the ANN model remains sufficiently high. Table 2 shows that the difference a single parameter can make does not considerably affect the prediction accuracy. However, this is not the case when solely the most significant parameter is considered. Table 2 reveals that the ANN model fails to efficiently predict the rocking amplitude when trained with only the most significant predictor of Figure 17. An exception is the uniform duration (pt_{uni}), based on which the ANN model shows a moderate predictive ability ($R^2 = 0.45$). This highlights the importance of the uniform duration as an efficient scalar IM,[26] especially if one considers that the input variables are on a natural (and not a logarithmic) scale, where a certain level of scatter is expected.

From a practical/engineering viewpoint, using ANN might be challenging as it contains the definition of many parameters, for example, number of hidden layers and neurons per hidden layer, training function, number of samples, among others. To facilitate a rapid estimation of the rocking amplitude, a simple model expresses a relation between seismic demand, represented by the absolute peak safe rocking response y, and a set of explanatory parameters, represented by the most significant predictors $\mathbf{X} = \{x_i, x_j, \ldots x_n\}$ (Figure 17).

$$y = \mathbf{g}^{\mathbf{T}} \cdot f(\mathbf{X}) + \varepsilon \tag{11}$$

where $\mathbf{g}^{\mathbf{T}} = \{a, b, c, d, \ldots\}^{\mathbf{T}}$ is the vector of the unknown constant (regression) coefficients, while ε refers to the error term. The function $f(\bullet)$ determines the type of regression (i.e., linear or nonlinear).[90] When the error ε follows a normal distribution with zero mean, the product $\mathbf{g}^T \cdot f(\mathbf{X})$ represents the mean value of the predicted response $\bar{y}$.[91]

The most well-known and widely used seismic demand model is a single-parameter regression model that estimates the seismic demand as a function of a ground motion IM.[92] However, the classic regression technique may not accurately

TABLE 2 Predictive ability of the ANN model using either all input variables or the most significant ones based on the Garson, Olden, Stepwise, and Profile methods.

Metric	All input variables	Four most significant parameters (Figure 17)				Three most significant parameters (Figure 17)			
		Garson	Olden	Stepwise	Profile	Garson	Olden	Stepwise	Profile
r_P	0.884	0.853	0.836	0.853	0.853	0.834	0.834	0.803	0.803
R^2	0.776	0.726	0.696	0.726	0.726	0.694	0.694	0.642	0.642
RMSE	0.095	0.106	0.111	0.106	0.106	0.111	0.111	0.118	0.118
sMAPE	0.452	0.456	0.45	0.456	0.456	0.482	0.482	0.455	0.455
MAE	0.057	0.062	0.062	0.062	0.062	0.061	0.061	0.065	0.065
Metric	**Two most significant parameters (Figure 17)**				**Most significant parameter (Figure 17)**				
	Garson	**Olden**	**Stepwise**	**Profile**	**Garson**	**Olden**	**Stepwise**	**Profile**	
r_P	0.815	0.761	0.815	0.815	0.418	0.73	0.418	0.418	
R^2	0.662	0.574	0.662	0.662	0.172	0.453	0.172	0.172	
RMSE	0.116	0.13	0.116	0.116	0.181	0.136	0.181	0.181	
sMAPE	0.472	0.482	0.472	0.472	0.546	0.48	0.546	0.546	
MAE	0.068	0.076	0.068	0.068	0.122	0.081	0.122	0.122	

capture a more complex nonlinear relation between input and output variables.[90,93] Thus, alternative formats have been investigated to enhance the prediction ability of the regression model. Pan et al.[94] found nonlinear and multivariate regression models more efficient than conventional linear regression models in predicting the response of multi-span highway bridges. Soleimani and Liu[34] compared variations of a polynomial nonlinear regression model with the classic linear regression model and revealed the superiority of the former in predicting the response of bridge components. Ghosh et al.[95] and Seo and Park[96] adopted a 2nd-order polynomial model to estimate the fragility of bridges. Finally, Sichani et al.[97] also adopted a 2nd-order polynomial model to predict the maximum rocking angle of dry casks under seismic loads. Consequently, this section investigates the ability of 1st- and 2nd-order polynomial regression models to predict the rocking amplitude of rigid rocking blocks when subjected to earthquake ground motions.

$$\bar{y} = a + \sum_{i=1}^{n} b_i x_i + \sum_{i=1}^{n} c_i x_i^2 + \sum_{i=1}^{n} \sum_{j=2, j>i}^{n} d_{ij} x_i x_j \tag{12}$$

where n corresponds to the number of explanatory parameters represented by the most influential input variables that contribute to the rocking amplitude.

The best combination of the most influential input variables is investigated using the stepwise regression approach.[98] Stepwise regression is an iterative method that involves adding variables (forward selection), removing variables (backward elimination), or a combination of both (bidirectional elimination) from a nonlinear regression model based on their contribution to the model's fit. This study employs the stepwise method with a bidirectional elimination. Considering the four most influential and most consistent input variables of Figure 17 (i.e., $IM_{SR} = [pT_m,\ pt_{uni},\ pMIV/(gtana),\ pPGV/(gtana)]$), Equation (13) provides an estimation of the seismic demand (after conducting nonlinear regression analysis on the test data characterised by $R^2 = 0.58$) using a 1st-order polynomial model:

$$\frac{|\bar{\theta}|_{\max}}{\alpha} = -0.032 + 0.114\left(\frac{pPGV}{g\tan\alpha}\right) + 0.359\left(\frac{pMIV}{g\tan\alpha}\right) + 0.005\,(pt_{uni}) \tag{13}$$

and Equation (14) provides an estimation of the seismic demand (with $R^2 = 0.7$) using a 2nd-order polynomial model:

$$\begin{aligned}\frac{|\bar{\theta}|_{\max}}{\alpha} = {} & -0.047 + 0.312\left(\frac{pPGV}{g\tan\alpha}\right) + 0.201\left(\frac{pMIV}{g\tan\alpha}\right) + 0.003\,(pT_m) - 0.013\,(pt_{uni}) \\ & -0.076\left(\frac{pPGV}{g\tan\alpha}\right)\left(\frac{pMIV}{g\tan\alpha}\right) - 0.143\left(\frac{pPGV}{g\tan\alpha}\right)(pT_m) + 0.023\left(\frac{pPGV}{g\tan\alpha}\right)(pt_{uni}) \\ & +0.139\left(\frac{pMIV}{g\tan\alpha}\right)(pT_m) - 0.032\left(\frac{pMIV}{g\tan\alpha}\right)(pt_{uni}) + 0.044\,(pT_m)(pt_{uni})\end{aligned} \tag{14}$$

The higher the order of the polynomial, the higher the prediction accuracy at the cost, though, of a more complex model.

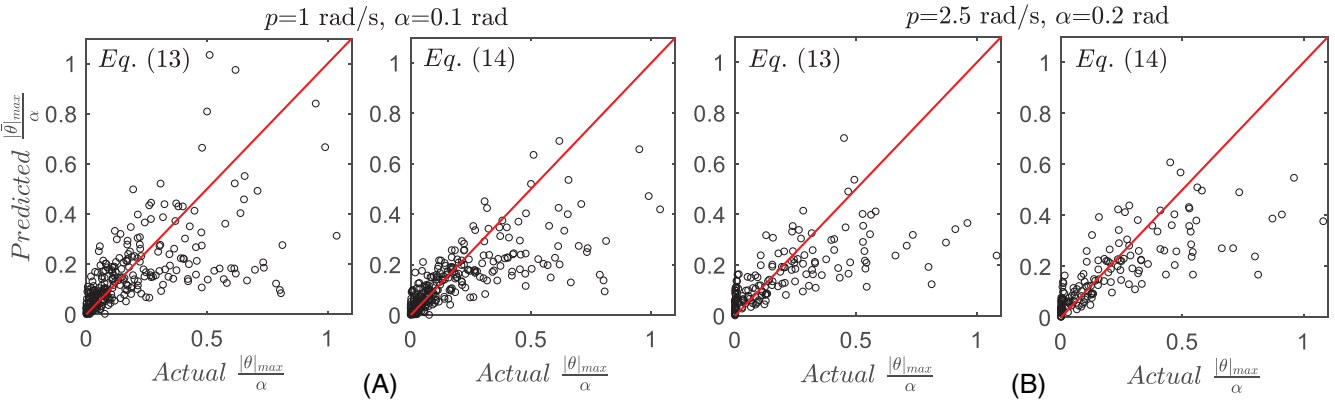

FIGURE 18 Actual vs predicted (via the seismic demand models of Equations (13) and (14)) rocking amplitude of a block with (A) $p = 1$ rad/s, $\alpha = 0.1$ rad, (B) $p = 2.5$ rad/s, $\alpha = 0.2$ rad, and coefficient of restitution $\eta = 0.92$.

Figure 18 illustrates the accuracy of the proposed seismic demand models of Equations (13) and (14) in predicting the rocking amplitude of two different blocks with (i) $p = 1$ rad/s, $\alpha = 0.1$ rad (Figure 18A) and (ii) $p = 2.5$ rad/s, $\alpha = 0.2$ rad (Figure 18B) when the coefficient of restitution is $\eta = 0.92$. The adopted structural configurations can represent different sizes and slenderness of real-life rocking structures. The structure of Figure 18A has a height of 14.6 m and a base of 1.47 m (aspect ratio of 10), representing, for example, a rocking bridge column or a masonry wall/façade. The structure of Figure 18B has a height of 2.3 m and a base of 0.47 m (aspect ratio of 4.9), representing, for example, a lab or household equipment or a museum object. Importantly, for the purposes of Figure 18, a suite of unseen (new) strong ground motions is utilised.[99] The selected recorded earthquakes are characterised by moment magnitude $M_w > 6$ and rapture distance $R_{rup} < 25$ km. In total, 375 nonlinear dynamic analyses are conducted, which lead to 305 safe rocking cases for the rocking structure of Figure 18A, and 166 safe rocking cases for the stockier rocking structure of Figure 18B. The actual rocking amplitude is plotted in the abscissa of Figure 18, while the ordinate represents the predicted amplitudes via Equations (13) and (14). Figure 18 illustrates the acceptable prediction accuracy of the proposed seismic demand models for both structural configurations. In addition, it shows that both demand models underestimate the prediction of higher rocking amplitudes ($|\theta|_{max}/\alpha > 0.5$), while they accurately predict lower rocking amplitudes ($|\theta|_{max}/\alpha < 0.5$). Nevertheless, the proposed seismic demand models show adequate prediction accuracy, which implies that they can serve as alternative and easy-to-use demand estimates of various rocking blocks when subjected to recorded earthquakes. This outcome becomes even more important if one considers that Equations (13) and (14) are formulated on a natural (and not a logarithmic) scale and tested under new (unseen) recorded earthquakes.

9 | CONCLUSIONS

This paper investigates the ability of ANN to treat the rocking problem. Specifically, it focuses on predicting the response of rigid rocking blocks of different sizes and slenderness when subjected to recorded earthquakes. Two response modes are considered: the block either overturns or safely returns to its initial (rest) position after the end of the ground excitation. In this context, an ANN classifier is trained to accurately categorise rocking response into safe rocking or overturning based on the structural parameters, ground motion characteristics, and the coefficient of restitution. It further investigates the role of each input variable in the overturning mode, revealing the substantial contribution of velocity and frequency characteristics of the ground excitation, while the role of the coefficient of restitution is also important to the outcome of the classification problem. In addition, an ANN predictor is trained to predict the safe rocking amplitude and identify the most critical input variables. The results unveil that a combined consideration of duration, frequency, and intensity characteristics of the ground motion consistently shows a substantial contribution to the rocking amplitude. Among those, this paper identifies a novel ground motion characteristic for the rocking literature, that is, the maximum incremental velocity that shows a substantial correlation with the response, while the role of the coefficient of restitution becomes less important to the outcome of the regression problem. Finally, even though the formulated ANN models accurately predict the seismic response of rocking blocks, encompassing both overturning and safe rocking scenarios, there is an inherent complexity stemming from the various parameters that can influence the predictive ability of an ANN model,

for instance, number of hidden layers and neurons per hidden layer, training function, number of samples, among others. Acknowledging the associated challenges the use of ANN may bring to engineering practitioners, this study proposes closed-form expressions combining the most influential input variables for a rapid, yet adequately accurate, estimation of the rocking amplitude.

ACKNOWLEDGEMENTS

This study has been partly funded by the STAND4HERITAGE project (new STANDards FOR seismic assessment of built cultural HERITAGE) that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant No. 833123) as an Advanced Grant. This work was also partly financed by FCT/MCTES through national funds (PIDDAC) under the R&D Unit Institute for Sustainability and Innovation in Structural Engineering (ISISE), under reference UIDB/04029/2020 (doi.org/10.54499/UIDB/04029/2020), and under the Associate Laboratory Advanced Production and Intelligent Systems ARISE under reference LA/P/0112/2020. The opinions and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the sponsoring organisations.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Seyed Amir Banimahd https://orcid.org/0000-0003-3235-4455
Anastasios I. Giouvanidis https://orcid.org/0000-0002-5850-3095
Shaghayegh Karimzadeh https://orcid.org/0000-0003-3753-1676
Paulo B. Lourenço https://orcid.org/0000-0001-8459-0199

REFERENCES

1. Housner GW. The behavior of inverted pendulum structures during earthquakes. *Bull Seismol Soc Am*. 1963;53(2):403-417.
2. Makris N. The role of the rotational inertia on the seismic resistance of free-standing rocking columns and articulated frames. *Bull Seismol Soc Am*. 2014;104(5):2226-2239.
3. Fragiadakis M, Diamantopoulos S. Fragility and risk assessment of freestanding building contents. *Earthq Eng Struct Dyn*. 2020;49(10):1028-1048.
4. Kazantzi AK, Lachanas CG, Vamvatsikos D. Seismic response distribution expressions for on-ground rigid rocking blocks under ordinary ground motions. *Earthq Eng Struct Dyn*. 2021;50(12):3311-3331.
5. Funari MF, Mehrotra A, Lourenço PB. A tool for the rapid seismic assessment of historic masonry structures based on limit analysis optimisation and rocking dynamics. *Appl Sciences*. 2021;11(3):942.
6. Vlachakis G, Giouvanidis AI, Mehrotra A, Lourenço PB. Numerical block-based simulation of rocking structures using a novel universal viscous damping model. *J Eng Mech*. 2021;147(11):04021089.
7. Vlachakis G, Colombo C, Giouvanidis AI, Savalle N, Lourenço PB. Experimental characterisation of dry-joint masonry structures: interface stiffness and interface damping. *Constr Build Mater*. 2023;392:130880.
8. Bachmann JA, Vassiliou MF, Stojadinović B. Dynamics of rocking podium structures. *Earthq Eng Struct Dyn*. 2017;46(14):2499-2517.
9. Dimitrakopoulos EG, Giouvanidis AI. Seismic response analysis of the planar rocking frame. *J Eng Mech*. 2015;141(7):04015003.
10. Giouvanidis AI, Dimitrakopoulos EG. Seismic performance of rocking frames with flag-shaped hysteretic behavior. *J Eng Mech*. 2017;143(5):04017008.
11. Giouvanidis AI, Dong Y. Seismic loss and resilience assessment of single-column rocking bridges. *Bull Earthq Eng*. 2020;18(9):4481-4513.
12. Manzo RN, Vassiliou MF. Cyclic tests of a precast restrained rocking system for sustainable and resilient seismic design of bridges. *Eng Struct*. 2022;252:113620.
13. Manzo RN, Vassiliou MF, Mouzakis H, Badogiannis E. Shaking table tests of a resilient bridge system with precast reinforced concrete columns equipped with springs. *Earthq Eng Struct Dyn*. 2022;51(1):213-239.
14. Giouvanidis AI, Dimitrakopoulos EG, Lourenço PB. Chattering: an overlooked peculiarity of rocking motion. *Nonlin Dyn*. 2022;109(2):459-477.
15. Giouvanidis AI, Dimitrakopoulos EG. Nonsmooth dynamic analysis of sticking impacts in rocking structures. *Bull Earthq Eng*. 2017;15:2273-2304.
16. Yim CS, Chopra AK, Penzien J. Rocking response of rigid blocks to earthquakes. *Earthq Eng Struct Dyn*. 1980;8(6):565-587.
17. Ishiyama Y. Motions of rigid bodies and criteria for overturning by earthquake excitations. *Earthq Eng Struct Dyn*. 1982;10(5):635-650.
18. Psycharis IN, Fragiadakis M, Stefanou I. Seismic reliability assessment of classical columns subjected to near-fault ground motions. *Earthq Eng Struct Dyn*. 2013;42(14):2061-2079.

19. Sieber M, Vassiliou MF, Anastasopoulos I. Intensity measures, fragility analysis and dimensionality reduction of rocking under far-field ground motions. *Earthq Eng Struct Dyn*. 2022;51(15):3639-3657.
20. Dimitrakopoulos EG, Paraskeva TS. Dimensionless fragility curves for rocking response to near-fault excitations. *Earthq Eng Struct Dyn*. 2015;44(12):2015-2033.
21. Petrone C, Di Sarno L, Magliulo G, Cosenza E. Numerical modelling and fragility assessment of typical freestanding building contents. *Bull Earthq Eng*. 2017;15:1609-1633.
22. Lachanas CG, Vamvatsikos D, Dimitrakopoulos EG. Intensity measures as interfacing variables versus response proxies: the case of rigid rocking blocks. *Earthq Eng Struct Dyn*. 2023;52(6):1722-1739.
23. Liu H, Huang Y, Liu X. An intensity measure for the rocking fragility analysis of rigid blocks subjected to floor motions. *Sustainability*. 2023;15(3):2418.
24. Pappas A, Sextos A, Da Porto F, Modena C. Efficiency of alternative intensity measures for the seismic assessment of monolithic free-standing columns. *Bull Earthq Eng*. 2017;15:1635-1659.
25. Kavvadias IE, Vasiliadis LK, Elenas A. Seismic response parametric study of ancient rocking columns. *Int J Archit Heritage*. 2017;11(6):791-804.
26. Giouvanidis AI, Dimitrakopoulos EG. Rocking amplification and strong-motion duration. *Earthq Eng Struct Dyn*. 2018;47(10):2094-2116.
27. Karimzadeh S, Funari MF, Szabó S, Hussaini SS, Rezaeian S, Lourenço PB. Stochastic simulation of earthquake ground motions for the seismic assessment of monumental masonry structures: source-based vs site-based approaches. *Earthq Eng Struct Dyn*. 2024;53(1):303-330.
28. Xie Y, Sichani EM, Padgett JE, DesRoches R. The promise of implementing machine learning in earthquake engineering: a state-of-the-art review. *Earthq Spectra*. 2020;36(4):1769-1801.
29. Xie Y, Zhang J, DesRoches R, Padgett JE. Seismic fragilities of single-column highway bridges with rocking column-footing. *Earthq Eng Struct Dyn*. 2019;48(7):843-864.
30. Hu S, Wang W, Alam MS. Probabilistic nonlinear displacement ratio prediction of self-centering energy-absorbing dual rocking core system under near-fault ground motions using machine learning. *J Earthq Eng*. 2023;27(3):488-519.
31. Gajan S. Data-driven modeling of peak rotation and tipping-over stability of rocking shallow foundations using machine learning algorithms. *Geotechnics*. 2022;2(3):781-801.
32. Achmet Z, Diamantopoulos S, Fragiadakis M. Rapid seismic response prediction of rocking blocks using machine learning. *Bull Earthq Eng*. 2023:1-19. doi:10.1007/s10518-023-01680-4
33. Mignan A, Broccardo M. Neural network applications in earthquake prediction (1994–2019): meta-analytic and statistical insights on their limitations. *Seismol Res Letters*. 2020;91(4):2330-2342.
34. Soleimani F, Liu X. Artificial neural network application in predicting probabilistic seismic demands of bridge components. *Earthq Eng Struct Dyn*. 2022;51(3):612-629.
35. Mohammadi A, Karimzadeh S, Banimahd SA, Ozsarac V, Lourenço PB. The potential of region-specific machine-learning-based ground motion models: application to Turkey. *Soil Dyn Earthq Eng*. 2023;172:108008.
36. Mohammadi A, Karimzadeh S, Yaghmaei-Sabegh S, Ranjbari M, Lourenço PB. Utilising artificial neural networks for assessing seismic demands of buckling restrained braces due to pulse-like motions. *Buildings*. 2023;13(10):2542.
37. Karimzadeh S, Mohammadi A, Hussaini SMS, Caicedo D, Askan A, Lourenço PB. ANN-based ground motion model for Turkey using stochastic simulation of earthquakes. *Geophys J Int*. 2023;236(1):413-429.
38. Mangalathu S, Jeon JS. Machine learning–based failure mode recognition of circular reinforced concrete bridge columns: comparative study. *J Struct Eng*. 2019;145(10):04019104.
39. Tran VL, Kim SE. Efficiency of three advanced data-driven models for predicting axial compression capacity of CFDST columns. *Thin-Walled Struct*. 2020;152:106744.
40. Solhmirzaei R, Salehi H, Kodur V, Naser MZ. Machine learning framework for predicting failure mode and shear capacity of ultra high performance concrete beams. *Eng Struct*. 2020;224:111221.
41. Tang Q, Dang J, Cui Y, Wang X, Jia J. Machine learning-based fast seismic risk assessment of building structures. *J Earthq Eng*. 2022;26(15):8041-8062.
42. ElGawady MA, Ma Q, Butterworth J, Ingham J. Effects of interface material on the performance of free rocking blocks. *Earthq Eng Struct Dyn*. 2011;40:375-392.
43. Aslam M, Scalise DT, Godden WG. Earthquake rocking response of rigid bodies. *J Struct Division*. 1980;106(2):377-392.
44. Kramer SL. *Geotechnical Earthquake Engineering*. Prentice-Hall International; 1996.
45. Rathje EM, Abrahamson NA, Bray JD. Simplified frequency content estimates of earthquake ground motions. *J Geotech Geoenviron Eng*. 1998;124(2):150-159.
46. Dimitrakopoulos EG, Kappos AJ, Makris N. Dimensional analysis of yielding and pounding structures for records without distinct pulses. *Soil Dyn Earthq Eng*. 2009;29(7):1170-1180.
47. Arias A. A measure of earthquake intensity. In: Hansen RJ, ed. *Seismic Design for Nuclear Power Plants*. MIT Press, Cambridge, MA; 1970:438-483.
48. Fajfar P, Vidic T, Fischinger M. A measure of earthquake motion capacity to damage medium-period structures. *Soil Dyn Earthq Eng*. 1990;9(5):236-242.
49. Park YJ, Ang AHS, Wen YK. Seismic damage analysis of reinforced concrete buildings. *J Struct Eng*. 1985;111(4):740-757.
50. Nau JM, Hall WJ. *An Evaluation of Scaling Methods for Earthquake Response Spectra*. University of Illinois at Urbana-Champaign; 1982.

51. Riddell R, Garcia JE. Hysteretic energy spectrum and damage control. *Earthq Eng Struct Dyn*. 2001;30(12):1791-1816.
52. Trifunac MD, Brady AG. A study on the duration of strong earthquake ground motion. *Bull Seismol Soc Am*. 1975;65(3):581-626.
53. Bolt BA. Duration of strong ground motion. Proceedings of the Fifth World Conference on Earthquake Engineering, Vol. 1, Rome; 1973;1304-1313.
54. Zhou Y, Katayama T. Effects of magnitude, epicentral distance and site conditions on the duration of strong ground motion. *Seisan Kenkyu*. 1985;37(12):10-13.
55. Bommer JJ, Martinez-Pereira A. The effective duration of earthquake strong motion. *J Earthq Eng*. 1999;3(2):127-172.
56. Bertero VV, Mahin SA, Herrera RA. Aseismic design implications of near-fault San Fernando earthquake records. *Earthq Eng Struct Dyn*. 1978;6(1):31-42.
57. Sorrentino L, AlShawa O, Decanini LD. The relevance of energy damping in unreinforced masonry rocking mechanisms. Experimental and analytic investigations. *Bull Earthq Eng*. 2011;9:1617-1642.
58. Galvez F, Sorrentino L, Dizhur D, Ingham JM. Seismic rocking simulation of unreinforced masonry parapets and façades using the discrete element method. *Earthq Eng Struct Dyn*. 2022;51(8):1840-1856.
59. Ancheta TD, Darragh RB, Stewart JP, et al. NGA-West2 database. *Earthq Spectra*. 2014;30(3):989-1005.
60. Lachanas CG, Vamvatsikos D, Vassiliou MF. The influence of the vertical component of ground motion on the probabilistic treatment of the rocking response of free-standing blocks. *Earthq Eng Struct Dyn*. 2022;51(8):1874-1894.
61. Kallioras S, Graziotti F, Penna A, Magenes G. Effects of vertical ground motions on the dynamic response of URM structures: comparative shake-table tests. *Earthq Eng Struct Dyn*. 2022;51(2):347-368.
62. Vamvatsikos D, Cornell CA. Incremental dynamic analysis. *Earthq Eng Struct Dyn*. 2002;31(3):491-514.
63. Shome N, Cornell CA. *Probabilistic Seismic Demand Analysis of Non-Linear Structures*. Report No. RMS-35, RMS Program, Stanford University, Palo Alto, California; 1999.
64. Bazzurro P, Luco N. Damage potential of near-source ground motion records. 8th U.S. National Conference on Earthquake Engineering, April 18–22, San Francisco, California; 2006.
65. Hair J, Black W, Babin B, Anderson R. *Multivariate Data Analysis*. Prentice Hall; 2010.
66. Belsley DA, Kuh E, Welsch RE. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons; 2005.
67. Mason CH, Perreault WD. Collinearity, power, and interpretation of multiple regression analysis. *J Mark Res*. 1991;28(3):268-280.
68. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. *Applied Linear Statistical Models*. McGraw-Hill; 1996.
69. Liu KH, Zheng JK, Pacheco-Torgal F, Zhao XY. Innovative modeling framework of chloride resistance of recycled aggregate concrete using ensemble-machine-learning methods. *Constr Build Mater*. 2022;337:127613.
70. O'Brien RM. A caution regarding rules of thumb for variance inflation factors. *Qual Quant*. 2007;41(5):673-690.
71. Haykin S. *Kalman Filtering and Neural Networks*. John Wiley & Sons; 2004.
72. Auer P, Burgsteiner H, Maass W. A learning rule for very simple universal approximators consisting of a single layer of perceptrons. *Neural Netw*. 2008;21(5):786-795.
73. Adamowski J, Karapataki C. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. *J Hydrol Eng*. 2010;15(10):729-743.
74. Bishop CM. *Pattern Recognition and Machine Learning*. Springer; 2006.
75. Levenberg K. A method for the solution of certain non-linear problems in least squares. *Q Appl Math*. 1944;2(2):164-168.
76. Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math*. 1963;11(2):431-441.
77. Ying X. An overview of overfitting and its solutions. *J Phys Conf Ser*. 2019;1168(2):022022. IOP Publishing.
78. Prechelt L. *Early Stopping-But When?* Springer; 2012.
79. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media; 2009.
80. Marcot BG, Hanea AM. What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Comput Stat*. 2020;36(3):2009-2031.
81. Brownlee J. *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning*. Machine Learning Mastery; 2020.
82. Garson DG. Interpreting neural network connection weights. *Artif Intell Expert*. 1991;6:47-51.
83. Olden JD, Jackson DA. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecol Modell*. 2002;154(1-2):135-150.
84. Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol Modell*. 2003;160(3):249-264.
85. Lek S, Belaud A, Baran P, Dimopoulos I, Delacoste M. Role of some environmental variables in trout abundance models using neural networks. *Aquat Living Resour*. 1996;9(1):23-29.
86. Morfidis K, Kostinakis K. Seismic parameters' combinations for the optimum prediction of the damage state of R/C buildings using neural networks. *Adv Eng Softw*. 2017;106:1-16.
87. DeLautour OR, Omenzetter P. Prediction of seismic-induced structural damage using artificial neural networks. *Eng Struct*. 2009;31(2):600-606.
88. Nikbin IM, Rahimi S, Allahyari H. A new empirical formula for prediction of fracture energy of concrete based on the artificial neural network. *Eng Fract Mech*. 2017;186:466-482.

89. Arslan MH. An evaluation of effective design parameters on earthquake performance of RC buildings using neural networks. *Eng Struct*. 2010;32(7):1888-1898.
90. Huang HH, Hsiao CK, Huang SY. Nonlinear regression analysis. In: *International Encyclopedia of Education*. Elsevier; 2010:339-346.
91. Dobson AJ, Barnett AG. *An Introduction to Generalized Linear Models*. CRC press; 2018.
92. Cornell CA, Jalayer F, Hamburger RO, Foutch DA. Probabilistic basis for 2000 SAC Federal Emergency Management Agency steel moment frame guidelines. *J Struct Eng*. 2002;128(4):526-533.
93. Bates DM, Watts DG. *Nonlinear Regression Analysis and Its Applications*. Wiley; 1988.
94. Pan Y, Agrawal AK, Ghosn M. Seismic fragility of continuous steel highway bridges in New York State. *J Bridge Eng*. 2007;12(6):689-699.
95. Ghosh J, Padgett JE, Dueñas-Osorio L. Surrogate modeling and failure surface visualization for efficient seismic vulnerability assessment of highway bridges. *Probabilistic Eng Mech*. 2013;34:189-199.
96. Seo J, Park H. Probabilistic seismic restoration cost estimation for transportation infrastructure portfolios with an emphasis on curved steel I-girder bridges. *Struct Safety*. 2017;65:27-34.
97. Sichani ME, Padgett JE, Bisadi V. Probabilistic seismic analysis of concrete dry cask structures. *Struct Saf*. 2018;73:87-98.
98. Efroymson MA. *Multiple Regression Analysis, Mathematical Methods for Digital Computers*. Wiley; 1960.
99. Giardini D, Woessner J, Danciu L, et al. Seismic hazard harmonization in Europe (SHARE): online data resource. *Swiss Seismological Service*, ETH Zurich, Switzerland; 2013:10. doi:10.12686/SED-00000001-SHARE

How to cite this article: Banimahd SA, Giouvanidis AI, Karimzadeh S, Lourenço PB. A multi-level approach to predict the seismic response of rigid rocking structures using artificial neural networks. *Earthquake Engng Struct Dyn*. 2024;1-24. https://doi.org/10.1002/eqe.4110