



Article

Structural Damage Prediction of a Reinforced Concrete Frame Under Single and Multiple Seismic Events Using Machine Learning Algorithms

Petros C. Lazaridis ^{1,*} , Ioannis E. Kavvadias ^{1,*} , Konstantinos Demertzis ^{1,2,*} , Lazaros Iliadis ¹ and Lazaros K. Vasiliadis¹

¹ Department of Civil Engineering, Democritus University of Thrace, Campus of Kimmeria, 67100 Xanthi, Greece; liliadis@civil.duth.gr (L.I.), lvasilia@civil.duth.gr (L.K.V.)

² School of Science & Technology, Informatics Studies, Hellenic Open University, Greece)

* Correspondence: plazarid@civil.duth.gr (P.C.L.); ikavvadi@civil.duth.gr (I.E.K.); kdemertz@fmenr.duth.gr (K.D.)

Abstract: Advanced machine learning algorithms, have the potential to be successfully applied to many areas of system modelling. In the present study the capability of ten machine learning algorithms in predicting the structural damage of an 8-storey reinforced concrete frame building subjected to single and successive ground motions is examined. From this point of view, the initial damage state of the structural system, as well as 16 well known ground motion intensity measures are adopted as the features of the machine-learning algorithms that aim to predict the structural damage after each seismic event. The structural analyses are performed considering both real and artificial mainshock–aftershock sequences, while the structural damage is expressed in terms of two overall damage indices. The comparative study results in the most efficient damage index, as well as the most promising machine learning algorithm in predicting the structural response of a reinforced concrete building under single or multiple seismic events. Finally, the configured methodology deployed in a user-friendly web-application.

Keywords: Seismic Sequence; Machine Learning Algorithms; Repeated Earthquakes, Structural Damage Prediction, Intensity Measures, Damage Accumulation, Machine Learning, Artificial Neural Network

1. Introduction

During earthquake events, it is common to observe aftershocks following a mainshock. Moderate-to-strong aftershocks may lead to additional structural damage and even collapse of buildings that sustained damage from the mainshock. Thus, the seismic performance of structural systems subjected to successive ground motions receives increasing attention the last years. The recent disaster held on March 2021 in Tyrnavos-Elassona region, Thessaly of Greece due to a pair of compatible magnitude ($M_w=6.3$, $M_w=6.1$) [1] shallow earthquakes with more than 1800 damaged or non-serviceable buildings, emerge the necessity of predicting the damage potential caused by mainshock–aftershock sequences in order to assess the seismic risk. It should be noted that the final, accumulated, damage includes the initial damage caused by major earthquake and the incremental damage caused by the following seismic sequence. The effect of successive seismic events on the structural performance has been thoroughly examined by many researchers [2–6]. Specifically, Amadio et al. [7] studied the influence of repeated shocks on the response of nonlinear Single Degree of Freedom (SDOF) systems using different hysteretic models. Hatzigeorgiou and Beskos [8] conducted an exhausting parametric study on SDOF systems and proposed an empirical relation to calculate inelastic displacement ratio under repeated earthquakes. Hatzigeorgiou and Liolios [9] examined the nonlinear behaviour of Reinforced Concrete (RC) frames subjected to multiple shocks considering a set of eight frames that varied both

at height regularity and dimensioning practice. Hatzivassiliou and Hatzigeorgiou [10] studied the accumulation of damage and ductility demands due to seismic sequence on three dimensional RC structures. Hosseinpour and Abdelnaby [11] studied the impact of different aspects, as earthquake direction, aftershock polarity and the influence of the vertical component, on the nonlinear response of RC frames under successive earthquakes. Also, more recently, Kavvadias et al. [12] and Zhou et al. [13] investigated the correlation between aftershock related Intensity Measures (IMs) and final structural damage indices. Additionally, multiple researchers [14–17] have been evaluating the fragility of buildings and infrastructures against seismic sequences, in the past.

In recent years advanced Machine Learning Algorithms (MLAs), such as Artificial Neural Networks (ANNs), have been successfully applied to many areas of system modelling. Their success is based on the thorough processing of data that captures the behaviour of a system. By detecting patterns in the collected data, valuable information can be extracted and predictions can be made that automate the decision-making process. That fact makes Machine Learning (ML) an advanced tool in modern engineering modelling. From this point of view, the utilize of MLAs in earthquake engineering raises year by year, examining mainly the capability of such models in predicting the seismic structural damage [18–20]. Among others, De Latour and Omenzetter [21] investigated the efficiency of ANNs on the prediction of seismic damage on numerous RC frames, while Alvanitopoulos et al. [22] also examined regular RC structures and, by incorporating fuzzy layers in ANN configuration (architecture). Subsequently, Morfidis and Kostinakis [23] used feature selection methods in a dataset of 3 dimensional RC buildings to found the more damage correlated set of seismic IMs. More recently, the same authors [24] examined the effectiveness of ANNs on the damage prediction of non-regular at height structures. Applications of Recurrent Neural Networks (RNNs) on Earthquake Engineering are presented by González et al. [25] and Mangalathu and Burton [26]. Furthermore, Zhang et al. [27] developed a Long-Short Term Memory (LSTM) network to predict structural response. For the same purpose, Convolutional Neural Networks (CNNs) have been applied by Li et al. [28] and Oh et al. [29]. Additionally, Thaler et al. [30] proposed a combination of Monte Carlo simulation and ANNs to predict the post seismic structural statistics of an elasto-plastic frame structure. The application of different fuzzy and crisp ML techniques in localization and predicting the amount of damage to an RC frame under individual earthquakes has been evaluated by Vrochidou et al. [31]. The common characteristic of the above studies is that the initial structural damage state of the structure is omitted. However, Lazaridis et al. [32] used an Ensemble Neural Network to predict the structural damage after a sequence of two seismic shocks employed as input features, both damage after the first earthquake and the IMs of the second one.

In the present study the reliability of MLAs in predicting the seismic structural damage of an 8-storey RC frame structure subjected to both single and successive seismic events (consisting of double seismic shocks) is examined. Due to the fact that the effect of each seismic excitation on the structural response is examined individually, to manipulate the data in total, the initial structural damage is taken into account even if the structure is intact i.e., in case of single seismic event (mainshock). The initial damage, as well as the ground motion intensity, which is expressed in terms 16 well-known IMs, are considered as the features of the ML problem, while the post-earthquake damage as the target. By this, the ML model could be applied even in case of multiple aftershock events given the characteristics of the complete seismic activity.

2. Primitive Data

2.1. Ground Motion Records

For the purpose of this study both artificial and natural seismic sequences are considered. By this, a sufficient set of data is ensured. Randomized seismic sequences are synthesized, using a suite of 318 individual natural acceleration records, to generate artificial seismic sequences accelerograms taking into account the differences of the ground

motion features [33]. The descriptive statistics of the aforementioned excitation suite are listed in Table A1. Specifically, every record of aforementioned suite is combined randomly in pairs with another six records from the same suite. As a result, 1908 pairs of first and second shock are constructed. These seismic sequences and the corresponding structural responses are used as the major part of the overall dataset for the specific ML problem. As a minor part of the overall dataset, 111 natural pairs of sequential shocks records are considered. The assumed natural sequences are occurred from 1972 to 2020, while the time gap between the occurrence of the successive shocks is smaller than fifteen months. It has to be mentioned that each mainshock-aftershock record is obtained by the same station. As a result, the natural set is consisted of 41 real seismic sequences recorded by 63 stations. Both sequential and individual records are selected from ESM [34] and PEER NGA West [35] databases. The natural seismic sequences are listed in Table A2. In order to construct the accelerograms composed of two successive seismic records, an intermediate zero ceasing time gap of 20 seconds is added to eliminate the overlap between the building oscillations. It should be noted that Non Linear Time History Analysis (NLTHAs) are performed not only using the seismic sequences but also using the first shock of each sequence, as the scope of this study is to examine the seismic structural response not only under seismic sequences but also under single ground motion records.

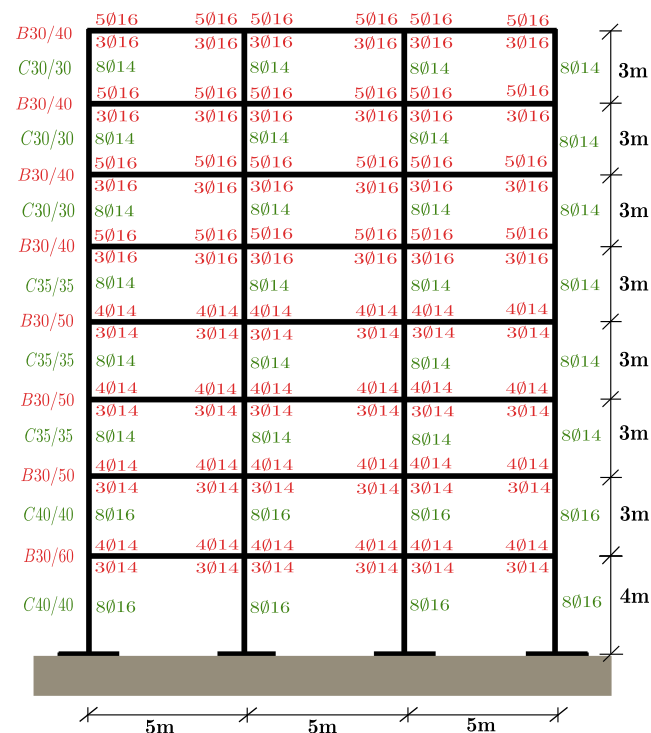


Figure 1. Reinforced Concrete Frame designed by Hatzigeorgiou and Liolios [9].

2.2. Reinforced Concrete Structure

Existing buildings designed and constructed without earthquake provisions comprise the majority of structures both in Greece and worldwide. That fact raises particular concern about their response to a potential earthquake. In this view, an 8-storey planar regular RC frame (Figure 1) designed only for gravity loads by Hatzigeorgiou and Liolios [9] is examined in the present study. The finite element simulation of the frame held in IDARC 2D [36], using spread plasticity concept and three-parameter Park hysteretic model [37]. Every floor considered to have only one horizontal degree of freedom to take into account the huge plane stiffness of RC slab as a rigid diaphragm. Sparsely placed stirrups with poor anchor details are assumed in order to be in accordance with obsolete design codes. Thus, a nonlinear deformation-stress model for concrete without confinement is adopted. As a

result, the concrete with mean compressive strength equal to 28 MPa is modeled by a curve defined by the initial modulus of elasticity ($E_0 = 31.42 \text{ GPa}$), the strain at the maximum stress ($\epsilon_{c0} = 2\%$), the ultimate strain in compression ($\epsilon_{cu} = 3.5\%$), stress at tension cracking $\sigma_t = 0.0022 \text{ GPa}$, and slope of the post-peak falling branch ($E_{fb} = -6.2 \text{ GPa}$). Furthermore, for steel grade S500s a bilinear curve with hardening employed. The with yield and ultimate strengths equal to 550 MPa and 660 MPa respectively and the corresponding strains equal to 2.75% and 45%, according to Eurocode-2 [38] provisions. The initial elastic fundamental period of the structure is equal to 1.27 seconds. The generation of IDARC 2D input files and the post-processing of the results are performed through GNU Octave [39,40] code.

3. Features, Targets and Dataset Generation

3.1. Ground Motion IMs

The basic parameters adopted in order to perform a seismic structural damage prediction analysis are the characteristics of the ground motion. By this, the identification of the seismic parameters that affect the dynamic response is of outmost importance. For this purpose, a set of 16 ground motion IMs is calculated. Amplitude parameters such as the maximum absolute values of ground accelerations ($a_g(t)$), velocities ($v_g(t)$), and displacements ($d_g(t)$) signals, which referenced as PGA, PGV, and PGD [41], respectively are examined. Additionally, the Arias Intensity (I_A) [42] and the Cumulative Absolute Velocity (CAV) [43], which are calculated by integral of the accelerogram time history are considered.

An inherent feature of signals is the frequency content which varies dynamically over the time in case of ground motion records. However, it can be quantified using the equivalent frequency PGA/PGV [41] as if it was a sinusoid signal. Another quantity that is related to the frequency content of a ground motion is the Potential Destructiveness Measure after Araya and Saragoni (I_{AS}) [44], determined by the zero crossings number of the acceleration signal (u_0) per unit of time.

Various definitions have been given in the past for the strong motion duration of a seismic excitation, in order to identify the time interval of the signal in which the vast amount of its total intensity is released. In this work, the strong motion durations defined by Trifunac and Brady (SMD_{TB}) [45], and by Reinoso, Ordaz and Guerrero (SMD_{ROG}) [46] are assumed. Both of these are based on the time evolution of Arias Intensity according to Husid Diagram [47]. Also, the bracketed duration as described by Bolt (SMD_{Bolt}) [48], which is defined by the first and last exceedance of the 5 percent of g , is employed.

Combining the above parameters results in more complex measures such as Power P_{90} [41], a_{rms} [41], Characteristic Intensity (I_c) [41], the potential damage measure according to Fajfar, Vidic and Fischinger (I_{VFF}) [49] and the IM after Riddell and Garcia (I_{RG}) [50].

It has to be mentioned that seismic parameters that depend on the fundamental structural period, such as individual spectral values were not calculated. These parameters could not be used due to the elongation of the elastic period during the first seismic event. Instead, the Housner Intensity [51] (SI_H) which accumulate pseudo-spectral velocities (PSV) to a constant range of possible eigen periods and demonstrates high correlation with the structural damage [23,52,53], is employed. All of the mathematical expressions of the examined IMs are summarised in Table 1. The elastic spectra are defined using OpenSeismoMatlab [54], while values of the IMs are computed through Python [55] code.

3.2. Damage Indicators

For the ML modeling of the present study, the structural damage is assumed both as an input feature, to take into account the initial damage due to the former seismic shock, as well as a target one in order to describe the damage accumulation after the examined ground motion. The structural response is assessed in terms of two overall seismic damage indices. Namely, the overall damage index after Park and Ang ($DI_{C,PA}$) [36] and the damage index after DiPasquale and Çakmak (DI_{DC}) [56].

Table 1. Mathematical Expressions of IMs.

Num	Name	Expression	Ref.	Num	Name	Expression	Ref.
1	PGA	$\max a_g(t) $	[41]	9	SMD_{ROG}	$t(H_d = 97.5\%) - t(H_d = 2.5\%)$	[46]
2	PGV	$\max v_g(t) $	[41]	10	SMD_{Bolt}	$t_{last}^{a_g > 0.05g} - t_{1st}^{a_g > 0.05g}$	[48]
3	PGD	$\max d_g(t) $	[41]	11	P_{90}	$\frac{I_A(H_d=95\%) - I_A(H_d=5\%)}{SMD_{TB}}$	[41]
4	I_A	$\frac{\pi}{2g} \int_0^{t_{end}} a_g^2(t) dt$	[42]	12	a_{rms}	$\sqrt{\frac{1}{SMD_{TB}} \int_{t_{5\%}}^{t_{95\%}} a_g(t)^2 dt}$	[41]
5	CAV	$\int_0^{t_{end}} a_g(t) dt$	[41]	13	I_c	$a_{rms}^{1.5} \cdot SMD_{TB}^{0.5}$	[41]
6	PGA/PGV	$\frac{PGA}{PGV}$	[41]	14	I_{FVF}	$PGV \cdot SMD_{TB}^{0.25}$	[49]
7	I_{AS}	$\frac{I_A}{u_o}$	[44]	15	I_{RG}	$PGD \cdot SMD_{TB}^{\frac{1}{3}}$	[50]
8	SMD_{TB}	$t(H_d = 95\%) - t(H_d = 5\%)$	[45]	16	SI_H	$\int_{0.1}^{2.5} PSV(T, \xi = 0.05) dT$	[51]

The originally introduced damage index after Park and Ang (DI_{PA}) [57] results from summation of the maximum flexural responses and the hysteretic energy consumption of the plastic hinges and is calculated by Equation 1 modified by Kunnath et al. [58] (Equation 2). The overall damage index ($DI_{G,PA}$) [36] is calculated as a weighted average of the sub-factors, weighted by the percentages of the total energy consumed by each member of the construction, according to Equation 3. The value of $DI_{G,PA}$ as close to zero as possible implies a complete damage-free structural system with an elastic response. However, a structure is characterized as near to collapse when $DI_{G,PA}$ takes values over the unit.

$$DI_{PA} = \frac{\delta_m}{\delta_u} + \frac{\beta}{Q_y \delta_u} \int dE \quad (1)$$

$$DI_{PA,component} = \frac{\theta_m - \theta_r}{\theta_u - \theta_r} + \frac{\beta}{\theta_u M_y} E_h \quad (2)$$

$$DI_{G,PA} = \frac{\sum E_i DI_{PA,component}}{\sum E_i} \quad (3)$$

where δ_m is the maximum element displacement response, δ_u is the ultimate element displacement, β is the model constant parameter for strength deterioration proposed by Park et al. [59], $\int dE$ is the cumulative hysteretic energy consumed by the element during its response, Q_y is the yield strength of the element, θ_m is the maximum element rotation during the time history response, θ_u is the ultimate capacity of the element and θ_r is the recoverable element rotation during unloading.

During high-intensity seismic events, it is known that the cross-sections in plastic hinges areas of a building can be severely cracked or even present steel yielding, resulting in structural stiffness degradation. Therefore, increase of the building's flexibility, and as such its fundamental period is expected [60]. The DI_{DC} is based on the above mentioned increase of the fundamental period and is calculated according to Equation 4.

$$DI_{DC} = 1 - \frac{T_{0_{initial}}}{T_{0_{equivalent}}} \quad (4)$$

where $T_{0_{initial}}$ is the fundamental period before the starting of analysis and $T_{0_{equivalent}}$ is the fundamental period at the end of the analysis.

3.3. Datasets Configuration

The scope of this study is to examine the capability of MLAs in predicting the structural damage of an RC frame under single or multiple ground motion records. To achieve this, the intensity of each seismic event and the corresponding response is treated individually, taking into account the initial structural damage just before the certain oscillation. In case of sequential seismic events the damage occurred by the first shock is considered as the

initial damage of the structure subjected to the aftershock. In order to provide a universal model that can predict the structural damage, regardless the pre-earthquake state of the building, the initial damage is taken into account even if the structure is intact i.e., in case of single seismic events. In such case the value of the damage indices is set as zero. It has to be mentioned that from the initially 1908 artificially generated sequences finally 1528 of them are considered. The rest sequences are omitted either due to convergence problems of the NLTHAs, either due to the absence of structural damage under the first shock. Thus, the dataset in total, comprises 1528 artificial and 111 natural seismic sequences. Additionally, 429 single seismic events are considered. As such, 2068 NLTHAs are performed.

4. Exploratory Data Analysis (EDA)

For the most complete and effective decision-making, statistical analysis of the examined data is required, in order to capture the technical characteristics of the problem. This exploratory analysis includes a set of numerical and graphical methods, which allow us to obtain an initial consideration about the features of the data that will be used in the ML Models. The purpose of the aforementioned analysis is the practical (non-scientific) interpretation of the data by unveiling the main characteristics of the data format, as well as their origin. This technique is a necessary step before the application of statistical inference methods, in order to thoroughly check the suitability of the data, the formulation of the adopted hypotheses and the selection of the appropriate method.

In particular, based on the problem analyzed in this paper, problematic values can be identified, i.e. values that are cut off from the main corpus and can be characterized as outliers or even incorrect, and appropriately treated them. Moreover, the normality of the data population can be checked. This is particularly important as many of the implemented methods require normality of the data.

The Table 2 lists the most important statistics of the ground motions' IMs. Mean (μ) estimates the average value in the population, for symmetric or nearly symmetric distributions, also σ estimates the standard deviation in the population. When the standard deviation is elevated we know that there are values of the variable sufficiently far from the mean. In a Normal Distribution, 95% of the values of the variable are within the limits $\mu \pm 2\sigma$. Moreover the minimum (min) and the maximum (max) values of each variable indicate the wide range of the seismic parameters.

Table 2. Descriptive statistics for the IMs of the overall dataset.

	PGA	PGV	PGD	I _A	CAV	PGA PGV	SMD							I _{FVF}	I _{RG}	SI _H
							I _{AS}	TB	ROG	Bolt	P ₉₀	a _{rms}	I _c			
	$\frac{cm}{s^2}$	$\frac{cm}{s}$	cm	$\frac{cm}{s}$	$\frac{cm}{s}$	s^{-1}	$\frac{cm}{s}$	s	s	s	$\frac{cm}{s^2}$	$\frac{cm}{s^2}$	$\frac{cm^{1.5}}{s^{2.5}}$	$cm \cdot s^{-0.75}$	$cm \cdot s^{\frac{1}{3}}$	cm
μ	299.6	29.1	49.4	131.7	738.7	13.0	3.4	13.5	16.9	11.6	15.8	76.4	2267.1	52.4	135.1	90.4
σ	244.1	24.2	120.6	186.2	527.0	8.6	5.0	10.0	11.3	10.0	25.9	63.5	2430.7	42.3	383.3	73.3
min	7.4	0.8	0.0	0.2	28.0	1.7	0.0	0.5	0.7	0.0	0.0	1.8	12.5	1.4	0.1	1.5
max	1465.2	148.2	1314.2	1332.4	3354.8	75.5	41.5	49.5	56.9	58.6	170.7	326.2	13323.4	243.0	4625.5	457.7

The statistical representation of the data is shown in Figures 2 and 3, where clear information is provided about the centre of the data, the symmetry, the skewness, the type of any asymmetry and the outliers. Information on the distortion and curvature of the distribution is also sought. Distortion refers to any deviation over the normal distribution. If the curve shifts to the left or right, it is said to be skewed. Skewness can be quantified as a representation of the degree to which a given distribution differs from a normal distribution. A normal distribution is non-skewed, while for example a lognormal distribution, exhibits right skewness. Distributions can exhibit right (positive) skewness or left (negative) skewness in varying degrees. The skewness, is the degree of asymmetry that observed in a probability distribution. A distribution with positive asymmetry possess, a shifted to lower values median. Obviously, the opposite is occurred in case of negative asymmetry.

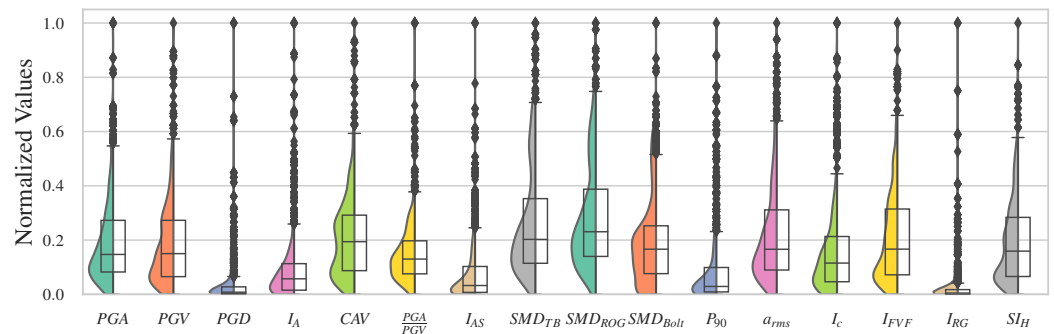


Figure 2. Violin and box plots of the IMs.

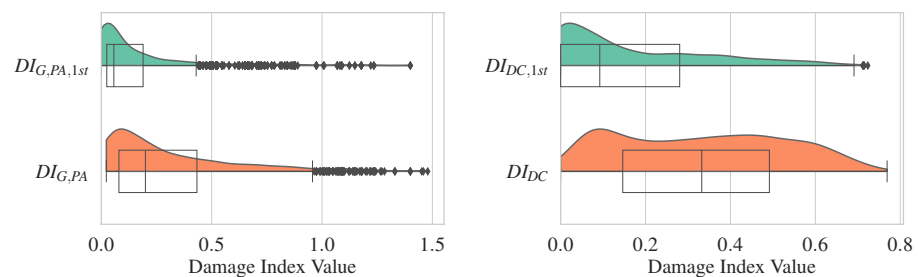


Figure 3. Violin and box plots of the damage indices.

In Figure 2 the distributions of all the examined IMs, with their values normalized in $[0,1]$, are presented. Moreover, in Figure 3 the distribution of the structural damage after the single ($DI_{G,PA,1st}$, $DI_{DC,1st}$) and the successive ($DI_{G,PA}$, DI_{DC}) seismic events are depicted comparatively. Under seismic sequences damage accumulation can be observed, as the distributions of both damage indices are shifted to higher values compared to those that correspond to the damage after the first shocks. It has been mentioned that the Figures 2 and 3, although that offer meaningful information, constitutes a tool of data exploratory analysis without leading to definitive conclusions.

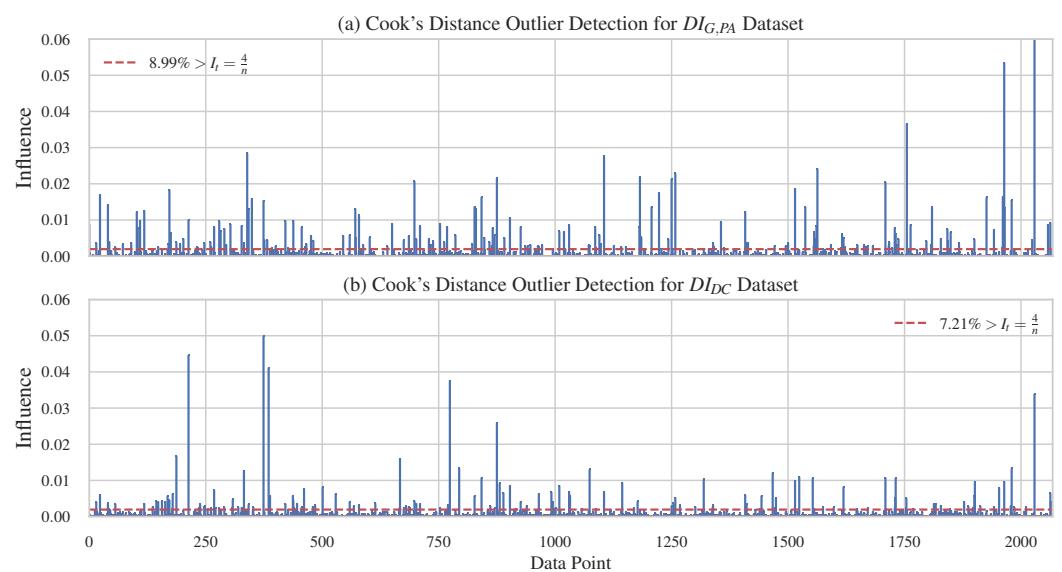


Figure 4. Cook's Distance of each data point for (a) $DI_{G,PA}$ dataset and (b) DI_{DC} dataset.

Values are characterised as extremes or outliers, are merely "suspect" values, i.e. values which may be incorrect or unusual. The number of points clarified as outliers depends on the sample size and the shape of the distribution. To identify the outliers, Cook's

distance [61,62] values is calculated according to Equation 5 and illustrated in Figure 4(a) and Figure 4(b) for $DI_{G,PA}$ and DI_{DC} datasets respectively. In the same figure the threshold which is equal to $I_t = \frac{4}{n}$ (n: number of observations) is depicted. The percentage of the potential outliers, with influence that exceeds the above threshold, is equal to 8.99% and 7.21% for $DI_{G,PA}$ and DI_{DC} datasets respectively.

$$D_i = \sum_{i=1}^n \frac{(\hat{y}_j - \hat{y}_{j(i)})^2}{p \cdot MSE} \quad (5)$$

where \hat{y}_j is the j^{th} predicted value, $\hat{y}_{j(i)}$ is the j^{th} predicted value, where the fit does not include observation i , MSE is the mean squared error, p is the number of coefficients in the regression model.

Subsequently, a correlation analysis is carried out in order to determine the degree of linear correlation between each pair of involved variables X and Y with variances σ_X , σ_Y respectively, and covariance $\sigma_{XY} = COV(X, Y) = E(X, Y) - E(X)E(Y)$. The results of Pearson coefficient [63] calculation, according to Equation 6, are shown in Figure 5.

$$\rho_{X,Y} = \frac{COV(X, Y)}{\sigma_X \sigma_Y} \quad (6)$$

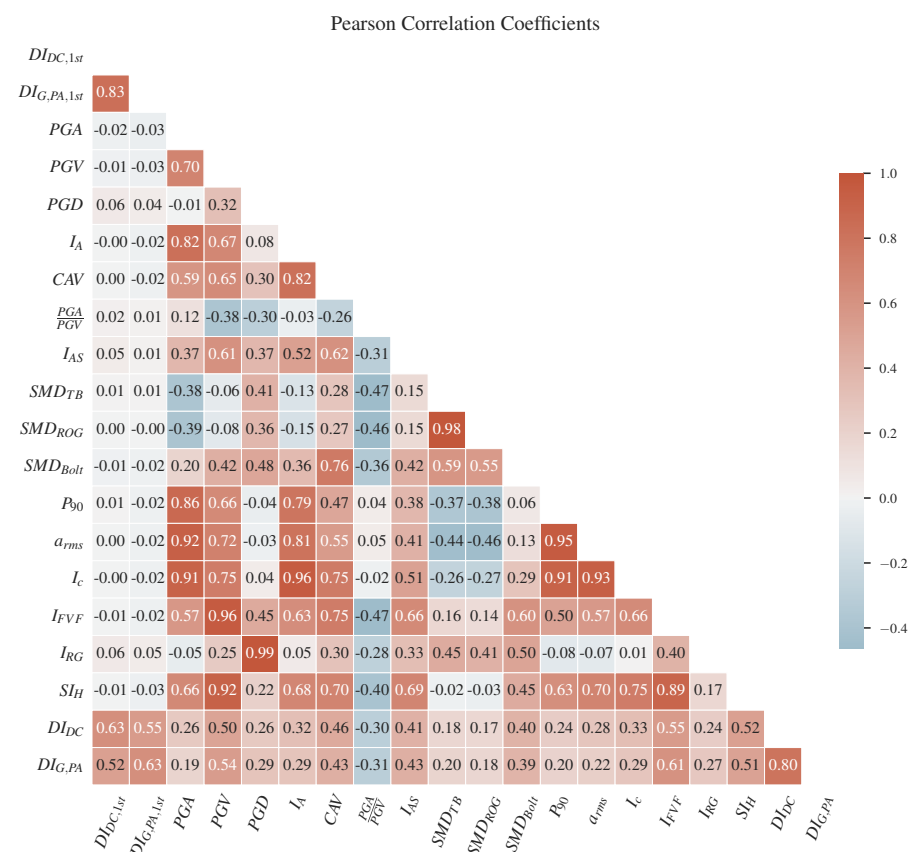


Figure 5. Heatmap of Pearson Correlation Coefficient for every pair of the examined variables including input features and targets.

Because there is a proven inability of the Pearson method to detect nonlinear correlations such as sinusoidal wave, quadratic curve, etc., the Predictive Power Score (PPS) [64] technique is also used to summarize the predictive relationships between the available data, explaining how variable A informs variable B more than variable B informs variable A. Technically, the score is a measurement in the interval [0, 1] of the success of a model in predicting a target variable with the help of an out-of-sample predictor variable. From this

method, hidden patterns of the data can be identified and as such facilitated the selection of appropriate prediction variables. The results of the PPS calculation are shown in the Figure 6.

After this analytical investigation of the data of the considered research, the analytical statistical hypothesis testing, for testing hypotheses related to the distribution of X (with its unknown parameters and its shape), hypothesis and independence testing related to the comparison of the unknown parameters of the problem variables, it was shown that this dataset is suitable for the correct application of ML methods. It is particularly important to understand the logic, meaning and limits of application of the data in question, so that this knowledge will allow us to correctly interpret the results and make correct conclusions with an awareness of the magnitude of the uncertainty in them. The logic question is directly related to the ongoing research question of whether the data used in the application being developed are appropriate and actually model the problem.

In conclusion, the above investigation allows us to interpret whether a MLA can extract a confident value associated with the option available to it. Similarly, whether it can abstain from trusting the choice when a particular output is too low. Finally, it is possible to explore algorithms that can be more effectively integrated into larger tasks, in a way that partially or completely avoids the problem of error propagation.

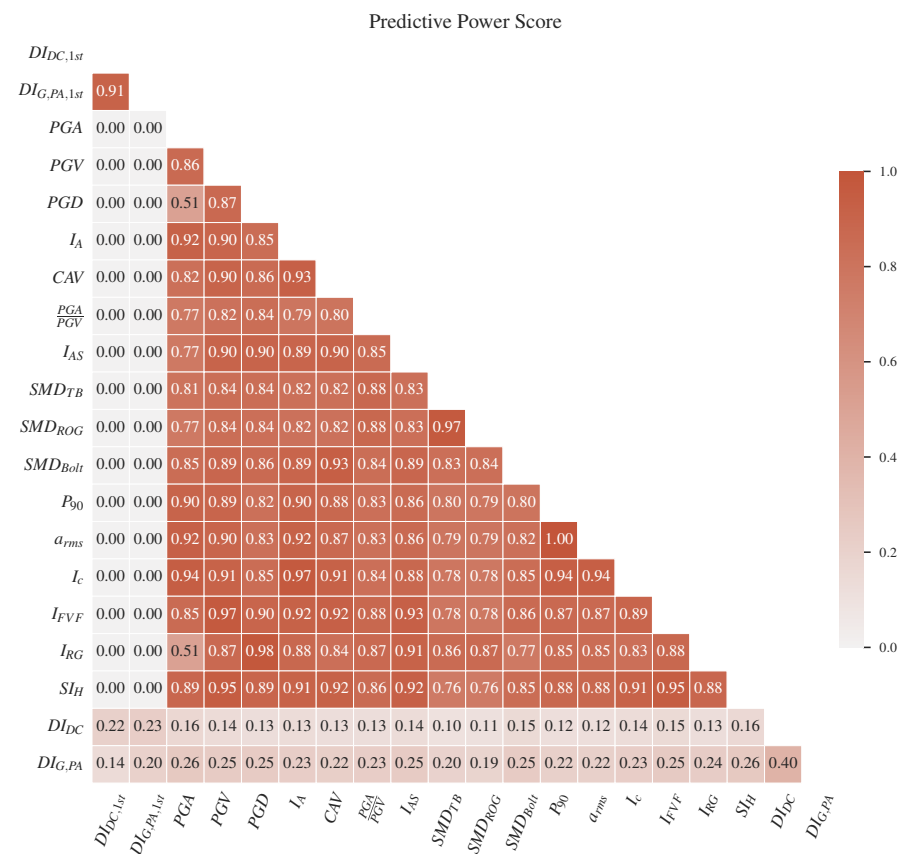


Figure 6. Heatmap of Predictive Power Score (PPS) for every pair of the examined variables including input features and targets.

5. Results

5.1. Comparative Performance Analysis of the Examined MLAs

The selection of the proper MLA, in modelling the seismic demand prediction of a RC frame under single and multiple ground motion records, is of utmost importance. This selection has to be made taking into account the particularities of the current data, the EDA and restrictions of the examined algorithms. In order to result into the most efficient algorithm, a thorough comparatively investigation among 10 different MLAs is performed.

Their performance is assessed by conducting sensitivity and accuracy analysis regarding the estimated errors obtained by the provided data.

The ten examined MLAs are: the Adaboost Regressor (ABR) [65], the Bayesian Ridge (BR) [66], the Decision Tree Regressor (DTR) [67], the Extra Trees Regressor (ETR) [68], the Gradient Boosting Regressor (GBR) [69], the K-Nearest Neighbors (KNN) [70,71], the Light Gradient Boosting Machine (LGBM) [72], the Linear Regressor (LR) [73], the Multi-Layer Feed-Forward Neural Network (MLNN) [74], the Random Forest Regressor (RFR) [75]. In this sense, an extensive and detailed comparison of 10 different MLAs on the two provided datasets was carried out. The MLAs implemented using the Scikit-learn [76], and LightGBM [72] Python packages as well as evaluated with Yellowbrick library [77,78]. It should be said that the following metrics [73] for comparison and cost analysis of the correct regression errors were taken into account and are listed below:

1. Mean Absolute Error (MAE) is a measure of errors between the estimated and the observed values, and it is given by the following expression:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (7)$$

where \hat{y}_i are the predicted value, y_i the real one of the i^{th} observation and n is the total number of observations

2. Mean Square Error (MSE):

$$RMSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (8)$$

3. Root Mean Squared Error (RMSE) calculates the average error between the estimated values and the observed values :

$$MAE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (9)$$

4. Coefficient of Determination R^2 expresses the variation in the dependent variable that is predictable from the independent variables:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

where \bar{y} is the average of the observed values

5. Root Mean Squared Log Error (RMLSE) is an extension of MSE that is used mainly when the predicted values display high deviation:

$$RMLSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(\hat{y}_i + 1) - \log(y_i + 1))^2} \quad (11)$$

6. Mean Absolute Percentage Error (MAPE) calculates the accuracy, as a ratio, and is defined by the following formulation:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (12)$$

In Figure 7 the metrics of each MLA for both datasets are presented. It can be seen that considering $DI_{G,PA}$, higher performance of the ML modelling is obtained. Specifically, the algorithm with the best prediction capability is the Extra Trees Regressor. This method is based on Decision Trees and randomize Decision Trees for random sub-samples in order to minimize the over-fitting. In particular, given a data sample $\mathbf{X} = \mathbf{x}_1, \dots, \mathbf{x}_n$ and the respected values $\mathbf{y} = y_1, \dots, y_n$ is chosen repeatedly a random sample without substitution

from the learning ensemble in order to estimate the target values using decision trees. Extra Trees algorithm performs like Random Forest, as multiple trees are generated and the nodes are separated using randomly chosen subsets of features. However, there are two main differences: the sampling is carried out without replacement, which means that there is no bootstrap, and the nodes are separated randomly among a random subset of features that are chosen for every node. The aforementioned randomness is based on the random separations of the total sample. Thus low variance is achieved. Another important feature is that the predictions are calculated by multiple decision trees and as such there is high prediction accuracy to new data. Moreover, the algorithm reduces the risk of over-fitting due to the randomness that is introduced in the model. Based on the above characteristics the model can be efficiently generalized.

Considering the DI_{DC} , the algorithm with the higher prediction capacity is the Gradient Boosting Regressor. The boosted trees algorithms are a combination of Boosting and Decision Trees. Boosting is a meta-algorithm for reducing bias in supervised learning. In the case of boosting [79], predictive regressors are used in order to develop weighted trees. The features of the regression trees and the boosting algorithms are combined to produce boosting trees. The Gradient Boosting method produces a prediction model comprised by a set of weak prediction, usually decision trees, models. It builds gradually the model and generalizes it by optimizing a loss function. In other words, at each iteration a new weak regressor is trained and extended the previous ones in order to increase the accuracy of the model.

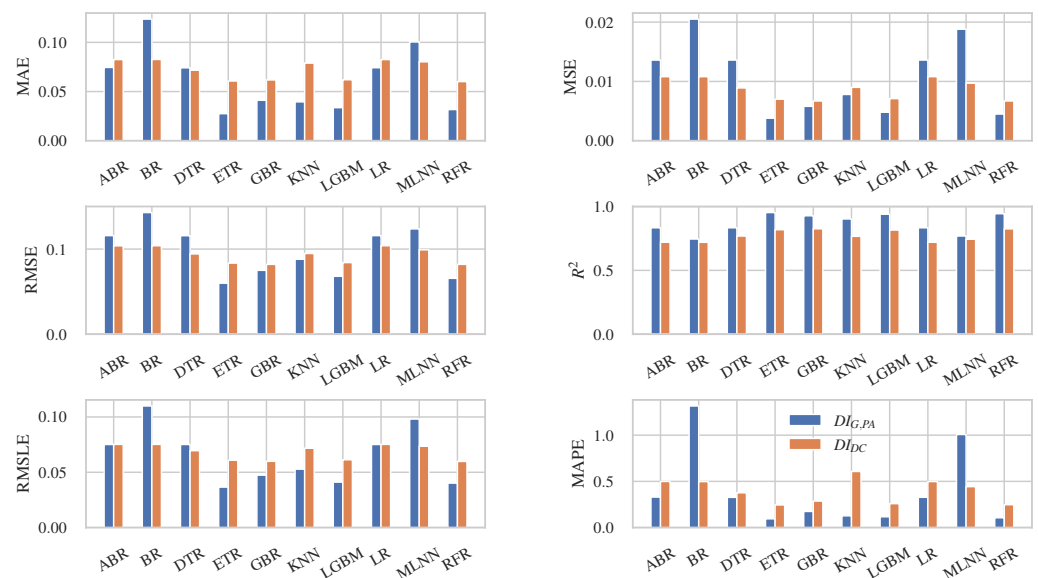


Figure 7. Performance metrics of the examined MLAs.

5.2. Evaluation of the MLAs with the Higher Prediction Ability

A major control procedure of the adjustment and the response of the algorithms is based on the learning curve, which depicts the learning performance as a function of the gained experience on the scale of time. It is a widely used tool that assesses the training and the validation data after each update of the measured error performance. Via this method, problems such as under- or over- fitting of the model, and the adequacy of the training or the validation data, could be emerged. In Figure 8 the learning curves of the distinctive MLA for each dataset are depicted. Specifically, Figure 8(a) illustrates the learning curves of the Extra Trees Regressor algorithm, when the structural damage is assessed in terms of $DI_{G,PA}$, while Figure 8(b) presents the learning curves of the Gradient Boosting Regressor considering structural damage in terms of DI_{DC} .

It is obvious that the training curve of Figure 8(a) improves over time without any trends of over-fitting. The prediction ability of the model is highlighted due to the high

performance since the starting point of the procedure, with score over 0.87. An increasing trend is depicted with a quite tight confidence interval, a fact that reflects the quality of the model. Moreover, there is no loss of the training and as such the distance between training and the validation curve reduces in relation to the experience. This distance is referred as “generalization gap” and defines the quality of the model. It is obvious that smaller gap between the two curves implies higher accuracy of the model.

From training and the validation curves of Figure 8(b) it could be easily observed that the algorithm fits quite well. Specifically, there are no trends of under- or over-fitting. The sufficient fit is noticed due to that fact that the training score is higher than the cross validation one, while the generalization gap reduces in relation to the experience, and tends to a constant value. Moreover, based on the aforementioned curves the quality of the considered data is assessed. Particularly, the total data set is representative to gain a solution, as the training data provide adequate information to train the problem, in relation to the data that is used to validate it. It can be remarked that sufficient samples are provided in order to lead to generalization. The training curve seems to be improving, without tends of over-fitting. Over-fitting could be identified by small alteration similarly presented in both training and cross validation curve. Moreover, there are no cases of validation loss lower than the training loss, a fact that indicates that the model can predict easier values of the validation data set compared to the training one.

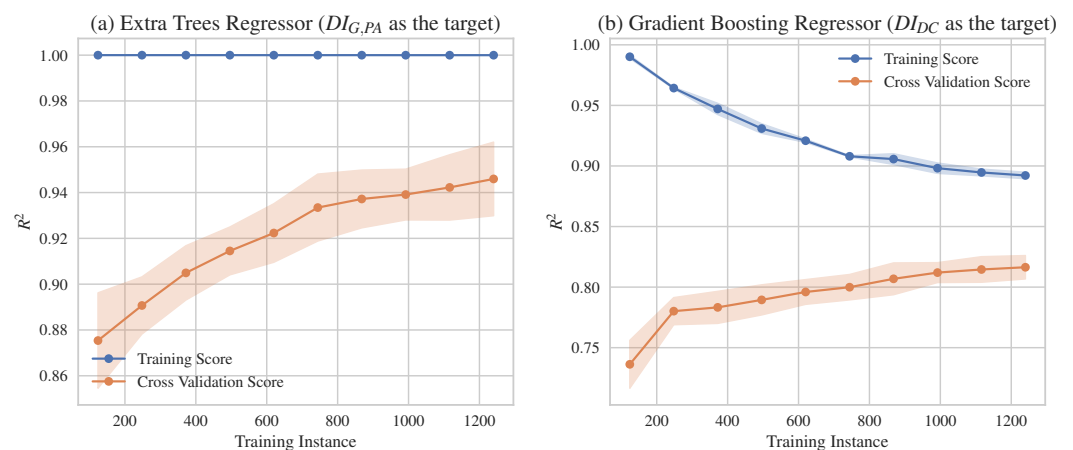


Figure 8. Learning Curves of the MLAs with the best predictive capacity for (a) the prediction of $DI_{G,PA}$ (b) the prediction of DI_{DC} .

In order to thoroughly assess the ML methods, the K-Fold cross validation methodology [76] is performed. During this method the total data set is divided in K subsets. Each randomly defined subset consists of different observations. One subset is used as the cross validation one, while the K-1 others are merged and used as the training set. This process is performed K times, using different set as the validation one and the K-1 rests of them as the training set. The performance of the MLA is evaluated for each case and on average. By this, the performance of the method in relation to the prediction error are determined. Specifically, the statistical properties, the bias and the variance of the regression prediction error are recorded and analyzed using a 10-fold cross validation procedure. A decomposition of the variability of the 10-fold cross validation sample, taking into account its variability sources. In Figure 9 the error metrics are depicted comparatively for the qualified MLA of each data set. In general, Extra Trees Regressor algorithm exhibit higher performance than the Gradient Boosting Regressor algorithm considering $DI_{G,PA}$ and DI_{DC} structural damage index, respectively. Considering $DI_{G,PA}$ it is observed a generally smooth shift of the error that expresses the probability of a given number of events occurring over a fixed period of time, taking into account the observations that occur at a known average constant rate and are independent of their appearance. There is only one case of significant error fluctuation which describes repetitive non-periodic

alterations that could not be predicted by this algorithm. On the other hand, regarding the DI_{DC} it is observed that there is higher dispersion of the error, which normalizes in the later folds. This fact translates the randomness of the samples into some folds, which are independent of the time period of their occurrence. The normalization and stabilization of the error after the initial fluctuations describes some possible repetitive non-periodic changes which are satisfactorily predicted by the specific algorithm. By this fact the model can be characterized by its constant generalization ability, and as such correct output values can be calculated for inputs that are new and different from those with which it is trained.

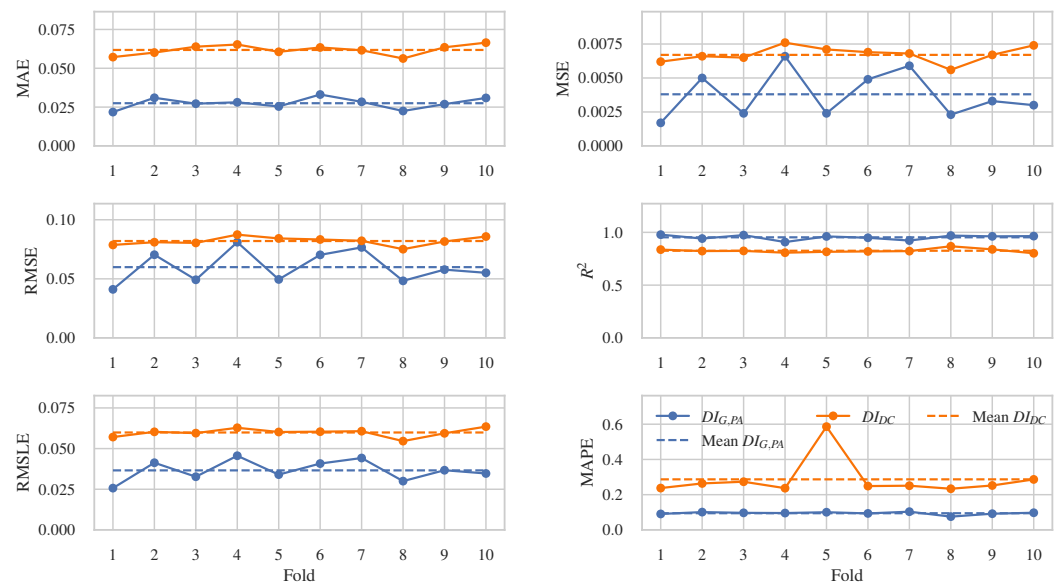


Figure 9. Evolution of performance metrics during 10-fold cross validation.

Finally in order to further analyze the introduced errors and determine their influence on the prediction ability of the examined models, the residual plots are presented in Figure 10. These scatter plots show the vertical deviations with respect to the regression line. These deviations, referred as residuals, are obtained by subtracting the observed responses from the predicted ones. In Figure 10(a) the residuals of the Extra Tree Regressor algorithm is predicted. The vertical deviations in relation to the regression line are quite limited both in train ($R^2 = 1.0$) and test ($R^2 = 0.950$). The residuals which are obviously very limited and their dispersion is minimal can be considered as cases of small population samples that do not follow a statistically central trend. Thus, these values are not related with the position of the center of the distribution and their mean value does not approach the actual one. As a result, the random error increases as the sample size increases too. The model holds high percentages of accuracy as the aforementioned samples are few enough, while the level of error is independent of the observation occurrence. In conclusion, the model understands the structure of residuals and manages to reduce the generalization error, while the predictive ability is exponentially increasing, without the requirement of special interventions in the hyperparameters of the model. Additionally, Figure 10(b) presents the residuals of the Gradient Boosting Regressor algorithm for training ($R^2 = 0.893$) and validation ($R^2 = 0.833$). The predicted response is calculated by the Gradient Boosting Regressor, since all the unknown parameters of the model have been calculated from the NLTHA results data. Careful examination of the residuals allows us to determine whether the adopted model is appropriate and the assumptions are reasonable. In our case the residuals can be considered as variables that compose general errors independently distributed with an average of 0.0. That fact implies that the model mistakenly predicts the response in a random way, i.e. the model predicts values higher or lower than the real ones with equal probability. In addition, the error is independent of the time or the magnitude of the observations, or even of the adjustment factors involved in making the

prediction. In conclusion, the residuals from these assumptions mean that the errors contain a structure that is not taken into account in the model, due to the inability to limit the error by generalizing the way of parameterizing the variation of its predictive capability. The identification of this structure, in theory, could lead to an enhanced model by adding representative terms. However, this consideration will lead to a model that will accumulate significant bias that would not lead to generalized solutions.

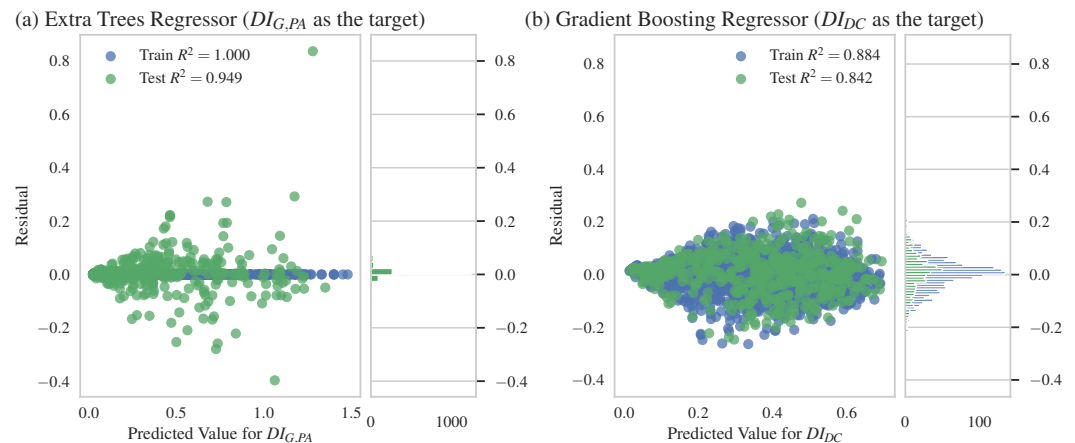


Figure 10. Residuals of the MLAs with the best predictive capacity for (a) the prediction of $DI_{G,PA}$ (b) the prediction of DI_{DC} .

To this end, the structural damage prediction of an RC frame under single and multiple ground motion records is more efficient to be modeled adopting the $DI_{G,PA}$. Due to the nature of the examined problem, it is evident that $DI_{G,PA}$ can assess both the initially occurred damage as well as the damage accumulation due to the successive ground motion records.

The comparative thorough analysis emerges the high performance of the Extra Trees Regressor. The ability of this algorithm can be explained due to its parametric nature, where it summarizes the data with a constant size set of parameters, regardless the number of the training instances. This fact leads to a learning system that achieves noteworthy results in relation to the competing systems. Another important observation is that the method produces accurate results without repetitive problems of indefinite causes, because all the intermediate partitions in the examined data set are handled very efficiently. In addition, one of the main advantages gained from the results is the high reliability resulting from the R^2 values, combined with the very low error rate. That fact arise as a result of receiving data without bootstrap, which allows the maintenance of more relevant data for the forthcoming predictions. Similarly, in the case of small population samples that do not follow the statistically central trend, the algorithm managed to achieve low variance, so that the sample data is close to the projections of the target function. This observation parallels the sensitivity of the model's correlative hyperparameters related with the data, which offers better predictability and stability as the overall behavior of the model is less noisy, while the overall risk of a particularly poor solution that may arise by undersampling is reduced. The above consideration is also supported by the dispersion of the expected error, which is concentrated close to the average error value, a fact that rigidly states the reliability of the system and its generalizability.

5.3. Web-Application Development

In this end, the authors decide to utilize the described methodology in a user-friendly web-application (Appendix B) that incorporate the trained models in order to deliver the results of this study via an interactive tool. After uploading an acceleration record file in PEER or ESM format, the IMs and the response spectra are calculated. The final structural damage is calculated considering as input features the IMs and the initial damage. The

initial damage, that express the damage state before the current seismic excitation, can be set through a slide bar. By adjusting the initial damage, the final damage is recalculated in real-time. Thus, several scenarios can be reproduced considering either an undamaged structure subjected to a seismic excitation, that could be considered as a main shock, or an already damaged structure, by a potential previous shock, subjected to an aftershock. The application is deployed using Streamlit [80] Python framework.

6. Discussion and Conclusions

In the present study, we proposed a ML approach of the structural damage prediction after a single or multiple seismic events on an RC frame in terms of $DI_{G,PA}$ and DI_{DC} damage indices using input features the IMs of the second shock and the established damage after the first one. The ability of ten MLAs to model the problem of structural damage prediction of a RC frame under single or multiple ground motions was thoroughly investigated. For this purpose multiple error metrics was adopted in order to assess the predictive capacity of the examined MLAs. Then it took place thorough comparison between the now known MLAs, moreover the generalization of the most efficient algorithms evaluated. The investigation relied entirely on evaluative methods of sensitivity analysis, variability and error analysis.

This algorithm facilitates the learning of specialized functions for extracting useful representations in complex learning dependencies, utilizes the use Random Decision Trees to learn without causing uncertainty issues, moreover overfitting is avoided, while utilizes significantly reducing computing training costs and time, producing improved training stability, high generalization performance and remarkable determination accuracy. In addition, the algorithm leads to much better predictive results, and high generalization ability, with reduced bias and variance. Therefore, a robust forecasting model capable of respond to the highly complex problem, as this of structural damage prediction, is deduced. It should also be emphasized that this methodology deals with great accuracy the noisy scattered residuals points.

Conclusively, as an outcome of this research the authors developed a user-friendly web-application that incorporates the results of this study. Future improvements should focus on the expanding dataset to incorporate different force resisting mechanisms and structural features simultaneously with IMs as input of sequential deep learning models to predict the final seismic damages as regression or classification problem.

Author Contributions: Conceptualization, P.C.L., I.E.K. and K.D.; methodology, P.C.L., I.E.K. and K.D.; software, P.C.L., I.E.K. and K.D.; validation, P.C.L., I.E.K. and K.D.; investigation, P.C.L., I.E.K. and K.D.; data curation, P.C.L.; writing—original draft preparation, P.C.L., I.E.K. and K.D.; writing—review and editing, P.C.L., I.E.K. and K.D.; visualization, P.C.L.; supervision, L.I., L.K.V.; project administration, P.C.L., I.E.K.

Funding: This research received no external funding

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

Acknowledgments: The first author needs to thank his parents for their significant support during his studies.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ML	Machine Learning
MLA	Machine Learning Algorithm
SDOF	Single Degree Of Freedom
RC	Reinforced Concrete
IM	Intensity Measure
ANN	Artificial Neural Network
LSTM	Long Short Term Memory
CNN	Convolutional Neural Network
NLTHA	Non Linear Time History Analysis
PGA	Peak Ground Acceleration
PGV	Peak Ground Velocity
PGD	Peak Ground Displacement
I_A	Arias Intensity
CAV	Cumulative Absolute Velocity
I_{AS}	Seismic Intensity after Araya and Saragoni
SMD_{TB}	Strong Motion Duration after Trifunac and Brady
SMD_{ROG}	Strong Motion Duration after Reinoso, Ordaz and Guerrero
SMD_{Bolt}	Strong Motion Duration after Bolt
a_{rms}	Root Mean Squared of Ground Acceleration Signal
I_c	Characteristic Intensity
I_{FVF}	Potential Damage Measure after Fajfar, Vidic and Fischinger
I_{RG}	Intensity Measure after Riddel and Garcia
PSV	Pseudo-Spectrum Velocities
H_d	Husid Diagram
SI_H	Spectral Intensity after Housner
$DI_{G,PA,1st}$	The overall Park and Ang damage index after the first seismic shock (input feature)
$DI_{G,PA}$	The overall Park and Ang damage index after the second seismic shock (target)
$DI_{DC,1st}$	DiPasquale and Çakmak damage index after the first seismic shock (input feature)
DI_{DC}	DiPasquale and Çakmak damage index after the second seismic shock (target)
EDA	Exploratory Data Analysis
PPS	Predictive Power Score
ABR	AdaBoost Regressor
BR	Bayesian Ridge
DTR	Decision Tree Regressor
ETR	Extra Trees Regressor
GBR	Gradient Boosting Regressor
KNN	K Nearest Neighbors Regressor
LGBM	Light Gradient Boosting Machine
LR	Linear Regressor
MLNN	Multi-Layer Feed-Forward Neural Network
RFR	Random Forest Regressor

Appendix A

Table A1. Descriptive statistics for the IMs of the 318 individual records.

	PGA	PGV	PGD	I_A	CAV	$\frac{PGA}{PGV}$	SMD						I_c	I_{FVF}	I_{RG}	SI_H
							I_{AS}	TB	ROG	Bolt	P_{90}	a_{rms}				
	$\frac{cm}{s^2}$	$\frac{cm}{s}$	cm	$\frac{cm}{s}$	$\frac{cm}{s}$	s^{-1}	$\frac{cm}{s}$	s	s	s	$\frac{cm}{s^2}$	$\frac{cm}{s^2}$	$\frac{cm^{1.5}}{s^{2.5}}$	$cm \cdot s^{-0.75}$	$cm \cdot s^{\frac{1}{3}}$	cm
μ	302.5	29.5	52.0	134.1	754.4	12.9	3.5	13.8	17.2	12.0	16.0	76.8	2301.9	53.4	143.2	91.8
σ	247.5	24.4	124.7	186.4	516.5	8.5	5.0	10.0	11.3	10.0	26.2	63.9	2455.5	42.1	398.7	72.8
min	32.7	1.2	0.1	0.7	28.0	1.7	0.0	0.6	1.9	0.0	0.1	7.4	55.3	1.4	0.1	2.1
max	1465.2	132.8	1314.2	1332.4	3119.3	75.5	41.5	49.5	56.9	58.6	150.1	306.4	13323.4	201.8	4625.5	387.1

Table A2. Seismic Metadata for Natural Sequences.

Region	1st Shock		2nd Shock		Station Code / Name	Component	PGA _{1st} (g)	PGA _{2nd} (g)	
	Date	M	Date	M					
Ancona Friuli	1972-06-14	4.2	1972-06-21	4.0	ANP	N-S	0.220	0.410	
	1976-09-11	5.8	1976-09-15	6.1	BUI	N-S	0.233	0.110	
Montenegro	1979-04-15	6.9	1979-04-15	5.8	PETO	E-W	0.108	0.093	
						E-W	0.299	0.644	
	1979-05-24	6.2	BAR	N-S	0.304	0.089			
				E-W	0.371	0.201			
	HRZ	N-S	0.215	0.066					
		E-W	0.254	0.076					
	ULO	N-S	0.282	0.033					
		E-W	0.236	0.030					
	Imperial Valley	1979-10-15	6.5	1979-10-15	5.0	Holtville Post Office	315	0.221	0.254
	Mammoth Lakes	1980-05-25	6.1	1980-05-25	5.7	Convict Creek	90	0.419	0.371
Irpinia	1980-11-23	6.9	1980-11-24	5.0	BGI	N-S	0.129	0.031	
						E-W	0.189	0.033	
	STR	N-S	0.224	0.018					
		E-W	0.320	0.032					
Gulf of Corinth	1981-02-24	6.6	1981-02-25	6.3	KORA	Trans	0.296	0.121	
Coalinga	1983-07-22	5.8	1983-07-25	5.2	Elm (Old CHP)	Logn	0.240	0.121	
						90	0.519	0.677	
Kalamata	1986-09-13	5.9	1986-09-15	4.8	KAL1	0	0.341	0.481	
						Trans	0.269	0.140	
	KALA	Logn	0.232	0.237					
		Trans	0.296	0.152					
Spitak	1988-12-07	6.7	1988-12-07	5.9	GUK	Logn	0.216	0.334	
						N-S	0.181	0.144	
						E-W	0.182	0.099	
Georgia	1989-01-08	4.0	1989-01-08	4.1	NAB	E-W	0.206	0.217	
	1991-05-03	5.6	1991-05-03	5.2	SAMB	N-S	0.354	0.208	
						E-W	0.504	0.122	
Erzican	1992-03-13	6.6	1992-03-15	5.9	AI178 ERC MET	N-S	0.411	0.032	
						E-W	0.487	0.039	
Ilia Northridge	1993-03-26	4.7	1993-03-26	4.9	PYR1	Logn	0.109	0.100	
	1994-01-17	6.7	1994-01-17	5.9	Moorpark - Fire Station	90	0.193	0.139	
						180	0.291	0.184	
	1994-01-17	6.7	1994-01-17	5.2	Pacoima Kagel Canyon	360	0.432	0.053	
						228	0.874	0.529	
	1994-03-20	6.7	1994-03-20	5.3	Rinaldi Receiving Station	270	0.752	0.102	
						90	0.605	0.181	
Umbria Marche	1997-09-26	5.7	1997-09-26	6.0	CLF	N-S	0.276	0.197	
						E-W	0.256	0.227	
Kalamata	1997-10-13	6.5	1997-11-18	6.4	KRN1	N-S	0.395	0.502	
						Trans	0.119	0.071	
						Logn	0.118	0.092	
Bovec	1998-04-12	5.7	1998-08-31	4.3	FAGG	N-S	0.024	0.023	
						E-W	0.023	0.026	
Azores Islands	1998-07-09	6.2	1998-07-11	4.7	HOR	N-S	0.405	0.082	
						E-W	0.369	0.092	
Izmit	1999-08-17	7.6	1999-11-12	7.3	ARC	N-S	0.210	0.007	
						E-W	0.132	0.007	
	ATK	N-S	0.102	0.016					
		E-W	0.167	0.016					
	DHM	N-S	0.090	0.017					
		E-W	0.084	0.017					
	FAT	N-S	0.181	0.034					
		E-W	0.161	0.024					
	KMP	N-S	0.102	0.014					
		E-W	0.127	0.017					
	ZYT	N-S	0.119	0.021					
		E-W	0.109	0.029					
	Athens	1999-09-07	5.9	1999-09-07	4.3	SPLB	Trans	0.324	0.059
							Logn	0.341	0.071
Chi-Chi	1999-09-20	7.6	1999-09-20	6.2	TCU071	N-S	0.651	0.382	
						E-W	0.528	0.193	
	TCU129	N-S	0.624	0.398					
		E-W	1.005	0.947					
	1999-09-25	6.3	TCU078	N-S	0.307	0.387			
				E-W	0.447	0.266			
	TCU079	N-S	0.424	0.626					
E-W		0.592	0.776						
Duzce	1999-11-12	7.3	1999-11-12	4.7	AI010 BOL	E-W	0.820	0.060	

Continued on next page

Region	1st Shock		2nd Shock		Station Code / Name	Component	PGA _{1st} (g)	PGA _{2nd} (g)	
	Date	M	Date	M					
Bingöl	2003-05-01	6.3	2003-05-01	3.5	AI049 BNG	N-S	0.519	0.147	
L Aquila	2009-04-06	6.1	2009-04-07	5.5	AQK	E-W	0.291	0.068	
						N-S	0.353	0.081	
						E-W	0.330	0.090	
						AQV	N-S	0.545	0.146
						E-W	0.657	0.129	
Darfield	2010-09-03	7.0	2009-04-09	5.4	AVZ	N-S	0.069	0.021	
			2011-02-21	6.2	AQA	N-S	0.442	0.057	
			Botanical Gardens	S01W	0.190	0.452			
				N89W	0.155	0.552			
				S80E	0.251	0.349			
			Cashmere High School	N26W	0.194	0.384			
				N64E	0.233	0.478			
Christchurch Hospital	N01W	0.209	0.346						
Emilia	2012-05-20	6.1	2012-05-29	6.0	MRN	N-S	0.263	0.294	
	2012-06-03	5.1	2012-06-12	4.9	T0827	E-W	0.262	0.222	
Central Italy	2016-08-24	6.0	2016-08-24	5.4	AQK	E-W	0.263	0.234	
			2016-08-26	4.8	AMT	N-S	0.375	0.336	
			E-W	0.867	0.325				
	2016-10-26	5.4	2016-10-26	5.9	CMI	N-S	0.341	0.308	
						E-W	0.720	0.651	
						E-W	0.556	0.537	
	2016-10-30	6.5	2016-10-30	6.5	CIT	N-S	0.052	0.213	
						E-W	0.092	0.325	
	2016-10-26	5.9	2016-10-30	6.5	CLO	N-S	0.193	0.582	
						E-W	0.183	0.427	
						CNE	N-S	0.380	0.294
						MMO	N-S	0.168	0.188
						E-W	0.170	0.189	
2016-10-30	6.5	2016-10-31	4.2	T1213	E-W	0.215	0.311		
					N-S	0.867	0.185		
2017-01-18	5.5	2017-01-18	5.4	PCB	E-W	0.794	0.212		
					N-S	0.586	0.561		
Dodecanese Islands	2019-08-08	4.8	2020-10-30	7.0	GMLD	E-W	0.408	0.388	
						N-S	0.450	0.899	
						E-W	0.673	0.763	

Appendix B

https://share.streamlit.io/plazarid/ml_rc_frame/main/ML_stream_app.py

References

- Papadopoulos, G.A.; Agalos, A.; Karavias, A.; Triantafyllou, I.; Parcharidis, I.; Lekkas, E. Seismic and Geodetic Imaging (DInSAR) Investigation of the March 2021 Strong Earthquake Sequence in Thessaly, Central Greece. *Geosciences* **2021**, *11*. doi:10.3390/geosciences11080311.
- Goda, K.; Taylor, C.A. Effects of aftershocks on peak ductility demand due to strong ground motion records from shallow crustal earthquakes. *Earthquake Engineering & Structural Dynamics* **2012**, *41*, 2311–2330. doi:10.1002/eqe.2188.
- Iervolino, I.; Giorgio, M.; Chioccarelli, E. Closed-form aftershock reliability of damage-cumulating elastic-perfectly-plastic systems. *Earthquake engineering & structural dynamics* **2014**, *43*, 613–625. doi:10.1002/eqe.2363.
- Yu, X.H.; Li, S.; Lu, D.G.; Tao, J. Collapse capacity of inelastic single-degree-of-freedom systems subjected to mainshock-aftershock earthquake sequences. *Journal of Earthquake Engineering* **2020**, *24*, 803–826. doi:10.1080/13632469.2018.1453417.
- Ghosh, J.; Padgett, J.E.; Sánchez Silva, M. Seismic damage accumulation in highway bridges in earthquake-prone regions. *Earthquake Spectra* **2015**, *31*, 115–135. doi:10.1193/120812EQS347M.
- Ji, D.; Wen, W.; Zhai, C.; Katsanos, E.I. Maximum inelastic displacement of mainshock-damaged structures under succeeding aftershock. *Soil Dynamics and Earthquake Engineering* **2020**, *136*, 106248. doi:10.1016/j.soildyn.2020.106248.
- Amadio, C.; Fragiacommo, M.; Rajgelj, S. The effects of repeated earthquake ground motions on the non-linear response of SDOF systems. *Earthquake Engineering & Structural Dynamics* **2003**, *32*, 291–308. doi:10.1002/eqe.225.
- Hatzigeorgiou, G.D.; Beskos, D.E. Inelastic displacement ratios for SDOF structures subjected to repeated earthquakes. *Engineering Structures* **2009**, *31*, 2744–2755. doi:10.1016/j.engstruct.2009.07.002.
- Hatzigeorgiou, G.D.; Liolios, A.A. Nonlinear behaviour of RC frames under repeated strong ground motions. *Soil Dynamics and Earthquake Engineering* **2010**, *30*, 1010–1025. doi:10.1016/j.soildyn.2010.04.013.
- Hatzivassiliou, M.; Hatzigeorgiou, G.D. Seismic sequence effects on three-dimensional reinforced concrete buildings. *Soil Dynamics and Earthquake Engineering* **2015**, *72*, 77–88. doi:10.1016/j.soildyn.2015.02.005.

11. Hosseinpour, F.; Abdelnaby, A. Effect of different aspects of multiple earthquakes on the nonlinear behavior of RC structures. *Soil Dynamics and Earthquake Engineering* **2017**, *92*, 706–725. doi:10.1016/j.soildyn.2016.11.006.
12. Kavvadias, I.E.; Rovithis, P.Z.; Vasiliadis, L.K.; Elenas, A. Effect of the aftershock intensity characteristics on the seismic response of RC frame buildings. In Proceedings of the Proceedings of the 16th European Conference on Earthquake Engineering, Thessaloniki, Greece, 2018, pp. 18–21.
13. Zhou, Z.; Yu, X.; Lu, D. Identifying Optimal Intensity Measures for Predicting Damage Potential of Mainshock–Aftershock Sequences. *Applied Sciences* **2020**, *10*, 6795. doi:10.3390/app10196795.
14. Yu, X.; Zhou, Z.; Du, W.; Lu, D. Development of fragility surfaces for reinforced concrete buildings under mainshock-aftershock sequences. *Earthquake Engineering & Structural Dynamics* **2021**. doi:10.1002/eqe.3542.
15. Jeon, J.S.; DesRoches, R.; Lowes, L.N.; Brilakis, I. Framework of aftershock fragility assessment—case studies: older California reinforced concrete building frames. *Earthquake Engineering & Structural Dynamics* **2015**, *44*, 2617–2636. doi:10.1002/eqe.2599.
16. Hosseinpour, F.; Abdelnaby, A. Fragility curves for RC frames under multiple earthquakes. *Soil Dynamics and Earthquake Engineering* **2017**, *98*, 222–234. doi:10.1016/j.soildyn.2017.04.013.
17. Abdelnaby, A.E. Fragility curves for RC frames subjected to Tohoku mainshock-aftershocks sequences. *Journal of Earthquake Engineering* **2018**, *22*, 902–920. doi:10.1080/13632469.2016.1264328.
18. Sun, H.; Burton, H.V.; Huang, H. Machine Learning Applications for Building Structural Design and Performance Assessment: State-of-the-Art Review. *Journal of Building Engineering* **2020**, p. 101816. doi:10.1016/j.job.2020.101816.
19. Xie, Y.; Ebad Sichani, M.; Padgett, J.E.; DesRoches, R. The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. *Earthquake Spectra* **2020**, p. 8755293020919419. doi:10.1177/8755293020919419.
20. Harirchian, E.; Hosseini, S.E.A.; Jadhav, K.; Kumari, V.; Rasulzade, S.; Işık, E.; Wasif, M.; Lahmer, T. A Review on Application of Soft Computing Techniques for the Rapid Visual Safety Evaluation and Damage Classification of Existing Buildings. *Journal of Building Engineering* **2021**, p. 102536. doi:10.1016/j.job.2021.102536.
21. De Lautour, O.R.; Omenzetter, P. Prediction of seismic-induced structural damage using artificial neural networks. *Engineering Structures* **2009**, *31*, 600–606. doi:10.1016/j.engstruct.2008.11.010.
22. Alvanitopoulos, P.; Andreadis, I.; Elenas, A. Neuro-fuzzy techniques for the classification of earthquake damages in buildings. *Measurement* **2010**, *43*, 797–809. doi:10.1016/j.measurement.2010.02.011.
23. Morfidis, K.; Kostinakis, K. Seismic parameters' combinations for the optimum prediction of the damage state of R/C buildings using neural networks. *Advances in Engineering Software* **2017**, *106*, 1–16. doi:10.1016/j.advengsoft.2017.01.001.
24. Kostinakis, K.; Morfidis, K. Application of Artificial Neural Networks for the Assessment of the Seismic Damage of Buildings with Irregular Infills' Distribution. In *Seismic Behaviour and Design of Irregular and Complex Civil Structures III*; Springer, 2020; pp. 291–306. doi:10.1007/978-3-030-33532-8_23.
25. González, J.; Yu, W.; Telesca, L. Earthquake Magnitude Prediction Using Recurrent Neural Networks. In Proceedings of the Multidisciplinary Digital Publishing Institute Proceedings, 2019, Vol. 24, p. 22.
26. Mangalathu, S.; Burton, H.V. Deep learning-based classification of earthquake-impacted buildings using textual damage descriptions. *International Journal of Disaster Risk Reduction* **2019**, *36*, 101111. doi:10.1016/j.ijdr.2019.101111.
27. Zhang, R.; Chen, Z.; Chen, S.; Zheng, J.; Büyükoztürk, O.; Sun, H. Deep long short-term memory networks for nonlinear structural seismic response prediction. *Computers & Structures* **2019**, *220*, 55–68. doi:10.1016/j.compstruc.2019.05.006.
28. Li, J.; He, Z.; Zhao, X. A data-driven building's seismic response estimation method using a deep convolutional neural network. *IEEE Access* **2021**, p. 1. doi:10.1109/ACCESS.2021.3065837.
29. Oh, B.K.; Park, Y.; Park, H.S. Seismic response prediction method for building structures using convolutional neural network. *Structural Control and Health Monitoring* **2020**, *27*, e2519. doi:10.1002/stc.2519.
30. Thaler, D.; Stoffel, M.; Markert, B.; Bamer, F. Machine-learning-enhanced tail end prediction of structural response statistics in earthquake engineering. *Earthquake Engineering & Structural Dynamics* **2021**. doi:10.1002/eqe.3432.
31. Vrochidou, E.; Bizergianidou, V.; Andreadis, I.; Elenas, A. Assessment and Localization of Structural Damage in r/c Structures through Intelligent Seismic Signal Processing. *Applied Artificial Intelligence* **2021**, pp. 1–26. doi:10.1080/08839514.2021.1935589.
32. Lazaridis, P.C.; Kavvadias, I.E.; Demertzis, K.; Iliadis, L.; Papaleonidas, A.; Vasiliadis, L.K.; Elenas, A. Structural Damage Prediction Under Seismic Sequence Using Neural Networks. In Proceedings of the 8th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, 2021. doi:10.7712/120121.8750.18752.
33. Li, Y.; Song, R.; Van De Lindt, J.W. Collapse fragility of steel structures subjected to earthquake mainshock-aftershock sequences. *Journal of Structural Engineering* **2014**, *140*, 04014095. doi:10.1061/(ASCE)ST.1943-541X.0001019.
34. Luzi, L.; Lanzano, G.; Felicetta, C.; D'Amico, M.; Russo, E.; Sgobba, S.; Pacor, F.; ORFEUS Working Group 5. Engineering Strong Motion Database (ESM) (Version 2.0). *Istituto Nazionale di Geofisica e Vulcanologia (INGV), Rome, Italy* **2020**. doi:10.13127/ESM.2.
35. Ancheta, T.D.; Darragh, R.B.; Stewart, J.P.; Seyhan, E.; Silva, W.J.; Chiou, B.S.; Wooddell, K.E.; Graves, R.W.; Kottke, A.R.; Boore, D.M.; et al. Peer NGA-West2 database. Technical report, Pacific Earthquake Engineering Research Center Berkeley, CA, 2013.
36. Valles, R.; Reinhorn, A.M.; Kunnath, S.K.; Li, C.; Madan, A. IDARC2D version 4.0: A computer program for the inelastic damage analysis of buildings. Technical report, US National Center for Earthquake Engineering Research (NCEER), 1996.
37. Park, Y.J.; Reinhorn, A.M.; Kunnath, S.K. *IDARC: Inelastic Damage Analysis of Reinforced Concrete Frame–Shear–Wall Structures*; National Center for Earthquake Engineering Research Buffalo, NY, 1987.

38. CEN. EN 1992-1-1 Eurocode 2: Design of concrete structures - Part 1-1: General rules and rules for buildings. Brussels: European Committee for Standardization 2005.
39. Eaton, J.W. GNU Octave and reproducible research. *Journal of Process Control* 2012, 22, 1433–1438. doi:10.1016/j.jprocont.2012.04.006.
40. Eaton, J.W.; Bateman, D.; Hauberg, S.; Wehbring, R. *GNU Octave version 6.1.0 manual: a high-level interactive language for numerical computations*, 2020.
41. Kramer, S.L. *Geotechnical earthquake engineering*; Pearson Education India, 1996.
42. Arias, A. A measure of earthquake intensity. Seismic Design for Nuclear Power Plants. *Massachusetts Institute of Technology* 1970.
43. EPRI, A. Criterion for determining exceedance of the operating basis earthquake. Rapport NP-5930 2848-16. *Electric Power Research Institute, USA* 1988.
44. Araya, R. Earthquake accelerogram destructiveness potential factor. In Proceedings of the Proc. 8th World Conference on Earthquake Engineering, 1985. 7, 1985, Vol. 11, pp. 835–843.
45. Trifunac, M.D.; Brady, A.G. A study on the duration of strong earthquake ground motion. *Bulletin of the Seismological Society of America* 1975, 65, 581–626.
46. Reinoso, E.; Ordaz, M.; Guerrero, R. Influence of strong ground-motion duration in seismic design of structures. In Proceedings of the Proceedings, 12th World Conference on Earthquake Engineering, 2000, Vol. 1151.
47. Husid, R. Características de terremotos. Análisis general. *Revista IDIEM* 1969, 8, ág–21.
48. Bolt, B.A. Duration of strong ground motion. In Proceedings of the 5th World Conference on Earthquake Engineering, 1973, Vol. 292, pp. 25–29.
49. Fajfar, P.; Vidic, T.; Fischinger, M. A measure of earthquake motion capacity to damage medium-period structures. *Soil Dynamics and Earthquake Engineering* 1990, 9, 236–242. doi:10.1016/S0267-7261(05)80002-8.
50. Riddell, R.; Garcia, J.E. Hysteretic energy spectrum and damage control. *Earthquake Engineering & Structural Dynamics* 2001, 30, 1791–1816. doi:10.1002/eqe.93.
51. Housner, G.W. Spectrum intensities of strong-motion earthquakes 1952.
52. Masi, A.; Vona, M.; Mucciarelli, M. Selection of Natural and Synthetic Accelerograms for Seismic Vulnerability Studies on Reinforced Concrete Frames. *Journal of Structural Engineering* 2011, 137, 367–378. doi:10.1061/(ASCE)ST.1943-541X.0000209.
53. Lazaridis, P.C.; Kavvadias, I.E.; Vasiliadis, L.K. Correlation between Seismic Parameters and Damage Indices of Reinforced Concrete Structures. In Proceedings of the 4th Panhellenic Conference on Earthquake Engineering and Engineering Seismology, 2019.
54. Papazafeiropoulos, G.; Plevris, V. OpenSeismoMatlab: A new open-source software for strong ground motion data processing. *Heliyon* 2018, 4, e00784. doi:10.1016/j.heliyon.2018.e00784.
55. Rossum, G. Python reference manual, 1995.
56. DiPasquale, E.; Çakmak, A. Detection of seismic structural damage using parameter-based global damage indices. *Probabilistic Engineering Mechanics* 1990, 5, 60–65. doi:10.1016/0266-8920(90)90008-8.
57. Park, Y.J.; Ang, A.H.S. Mechanistic seismic damage model for reinforced concrete. *Journal of Structural Engineering* 1985, 111, 722–739. doi:10.1061/(ASCE)0733-9445(1985)111:4(722).
58. Kunnath, S.K.; Reinhorn, A.M.; Lobo, R. IDARC Version 3.0: A program for the inelastic damage analysis of reinforced concrete structures. Technical report, US National Center for Earthquake Engineering Research (NCEER), 1992.
59. Park, Y.J.; Ang, A.H.; Wen, Y.K. Damage-limiting aseismic design of buildings. *Earthquake Spectra* 1987, 3, 1–26. doi:10.1193/1.1585416.
60. Katsanos, E.; Sextos, A. Inelastic spectra to predict period elongation of structures under earthquake loading. *Earthquake Engineering & Structural Dynamics* 2015, 44, 1765–1782. doi:10.1002/eqe.2554.
61. Cook, R.D. Detection of influential observation in linear regression. *Technometrics* 1977, pp. 15–18. doi:10.1080/00401706.1977.1048949.
62. Cook, R.D. Influential observations in linear regression. *Journal of the American Statistical Association* 1979, pp. 169–174. doi:10.1080/01621459.1979.10481634.
63. Gibbons, J.D.; Chakraborti, S. Nonparametric Statistical Inference 2010. doi:10.1201/9781439896129.
64. Wetschoreck, F.; Krabel, T.; Krishnamurthy, S. 8080labs/ppscore: zenodo release, 2020. doi:10.5281/zenodo.4091345.
65. Drucker, H. Improving regressors using boosting techniques. In Proceedings of the ICML. Citeseer, 1997, Vol. 97, pp. 107–115.
66. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *Journal of machine learning research* 2001, 1, 211–244.
67. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and regression trees*; Routledge, 2017. doi:10.1201/9781315139470.
68. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Machine learning* 2006, 63, 3–42. doi:10.1007/s10994-006-6226-1.
69. Friedman, J.H. Greedy function approximation: a gradient boosting machine. *Annals of statistics* 2001, pp. 1189–1232.
70. Fix, E.; Hodges, J.L. Discriminatory analysis. Nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique* 1989, 57, 238–247. doi:10.2307/1403797.
71. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 1992, 46, 175–185. doi:10.1080/00031305.1992.10475879.
72. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the Advances in Neural Information Processing Systems; Guyon, I.; Luxburg, U.V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R., Eds. Curran Associates, Inc., 2017, Vol. 30.
73. Glantz, S.A.; Slinker, B.K. *Primer of applied regression & analysis of variance, ed*; McGraw-Hill, Inc., New York, 2001.

-
74. Minsky, M.; Papert, S.A. *Perceptrons: An introduction to computational geometry*; MIT Press, 2017.
 75. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32. doi:10.1023/A:1010933404324.
 76. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
 77. Bengfort, B.; Bilbro, R. Yellowbrick: Visualizing the Scikit-Learn Model Selection Process. *The Journal of Open Source Software* **2019**, *4*. doi:10.21105/joss.01075.
 78. Bengfort, B.; Bilbro, R.; Johnson, P.; Billet, P.; Roman, P.; Deziel, P.; McIntyre, K.; Gray, L.; Ojeda, A.; Schmierer, E.; et al. Yellowbrick v1.3, 2021. doi:10.5281/zenodo.4525724.
 79. Friedman, J.H. Stochastic gradient boosting. *Computational statistics & data analysis* **2002**, *38*, 367–378. doi:10.1016/S0167-9473(01)00065-2.
 80. Teixeira, T.; et al. Streamlit. 0.69. 0. *Github*. URL: <https://github.com/streamlit/streamlit> **2020**.