A machine learning-based analysis for predicting fragility curve parameters of buildings

Dabiri, Hamed; Faramarzi, Asaad; Dall'Asta, Andrea; Tondi, Emanuele ; Micozzi, Fabio

DOI:
10.1016/j.jobe.2022.105367

Document Version
Peer reviewed version

Citation for published version (Harvard):
Dabiri, H, Faramarzi, A, Dall'Asta, A, Tondi, E & Micozzi, F 2022, 'A machine learning-based analysis for predicting fragility curve parameters of buildings', *Journal of Building Engineering*, vol. 62, 105367. https://doi.org/10.1016/j.jobe.2022.105367

Fragility curves of buildings; a critical review and a machine learning-based study

Abstract

Fragility curves are one of the substantial means required for seismic risk assessment of buildings in the framework of performance-based earthquake engineering (PBEE). Driving fragility curves, however, needs an extensive analytical analysis which makes it time-consuming and sometimes inaccurate due to errors. In this study, hence, machine learning (ML)-based models are proposed for predicting fragility parameters of structures namely dispersion, β, and median, μ. Firstly, a critical review on the analytical models proposed for deriving fragility curves is provided as well as ML-based models developed so far. Then, to achieve the research objective, a comprehensive database including 238 datasets from peer-reviewed international publications is collected. It is then divided into training (85%) and testing (15%) sub-datasets for the purpose of training the models and assessing the results, respectively. The most effective parameters on the target outputs are defined as input variables including construction material, building plan area, building height, damage state, buildings' period and soil classification. β and μ were estimated utilizing various ML-based techniques namely nonlinear regression, decision tree, random forest, KNN and ANN. The actual values and the values predicted by the proposed models are compared. Moreover, the models' accuracy is assessed through performance metrics and Taylor diagram. The results demonstrated the high ability of the models for learning the relationship between inputs and outputs. According to the accuracy assessment, Decision tree and nonlinear regression were the most accurate models for predicting β and μ, respectively. A sensitivity study was also conducted by changing input variables and estimation equations are provided accordingly.

Keywords: fragility curve, machine learning, Regression, Decision Tree, Random Forest, KNN, ANN, parametric study.

1. Introduction

1.1.background

Among various natural hazards, earthquakes cause a notable loss portion of life and likelihood and thus are recognised as the greatest threat to human beings [1, 2]. As a result, risk assessment is undoubtedly of high importance which could help humankinds in the sense of life and economics [3]. In this context, Performance-based Earthquake Engineering (PBEE) and potential failure mode analysis (PFMA) have been developed over the past years [4, 5]. PBEE is known as a process with quantitative measures as the output which reflect the response of structures under seismic loads [6]. One of the main issues in seismic risk analysis is various sources of uncertainties (e.g., randomness in ground motion, structural modelling related uncertainties, damage state, etc.) which is addressed by considering fragility curves in the PBEE framework [4, 7].

1.2. Fragility curves definition

Various definitions have been provided for clarifying the concept of fragility curves by researchers [8-13]. Simply stated, fragility is described as the probability of reaching or exceeding a specified damage measure (DM) under a given earthquake Intensity Measure (IM) for a structure [3, 8, 14]. In the viewpoint of mathematic, fragility curves could be expressed as Eq. 1 [765-main-good].

$$F_R = P\left(\theta_{max} \geq \frac{y}{IM} = x\right) \qquad (1)$$

where θ_{max} is the maximum Engineering Damage Parameter (EDP), and y and x are the values of θ_{max} and IM, respectively. The fragility curves could be drawn using lognormal distribution functions which are defined by two parameters: median (μ) and lognormal standard deviation (or dispersion, β). These parameters, therefore, are known as fragility curve's parameters and could be estimated by the maximum likelihood method [4, 15, 16]. As a result, the fragility curves could be expressed as Eq. 2:

$$F_R(IM) = \Phi\left[\frac{\ln(IM/\mu)}{\beta}\right] \qquad (2)$$

where $\Phi(.)$ is the standardized normal distribution function, μ is the logarithmic mean and β is the logarithmic standard deviation defining the lognormal distribution [12, 15].

As could be figured out from the above-mentioned explanations, the chosen IM could affect fragility curve of a building significantly. A macroseismic IM does not lead to accurate fragility curves because (i) it can cause interdependency between the vulnerability and the IM itself because it is obtained from observation of earthquake consequences on buildings, and (ii) it is a subjective parameter which means that it is influenced by the sensitivity and judgment of the surveyor [12, 14]. Hence, parameters namely peak ground acceleration (PGA), peak ground velocity (PGV) and peak ground displacement (PGD) have been applied instead of macro seismic IMs [12]. Among these parameters, PGA is more common than others because of its simplicity for deriving fragility curves of complex buildings [17]. It should be also stated that different parameters could be used as EDP such as interstory drift ratio (IDR), ultimate rotation, etc. The most usual EDP, however, is IDR since it could reflect the structures' damage state perfectly [14].

1.3. Review on the methods for deriving fragility curves

fragility curves derivation is mainly based on incremental dynamic analysis (IDA). The output of IDA is a curve which reflects the relationship between the defined IM and EDP. IDA could be performed through five main steps: (a) modelling the structure by a Finite Element (FE) analysis software (e.g., OpenSees) in order to simulate nonlinear behaviour of the structure, (b) collecting a series of seismic records, (c) defining the IM (e.g., PGA) and EDP (e.g., drift ratio), (d) scaling seismic records in order to reach the collapse state, and (e) repeating step (d) for different earthquake records to reach an appropriate number of data for each damage state which reflects the building's response accurately [8, 10, 18].

Distribution of the obtained curves through the above-mentioned steps, intersected by a vertical line passing from each damage state will be a lognormal distribution with median (μ) and dispersion (β) which are used for deriving the fragility curve of the corresponding damage state. Fig. 1 displays the fragility curve derivation schematically.

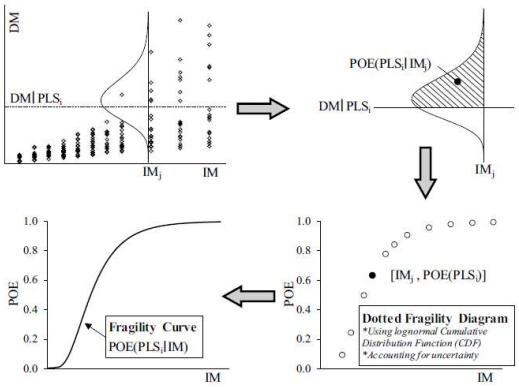


***Figure 1**. Schematically illustration of driving fragility curves [11]*

The method shown in Fig. 1 is recognized as the conventional method for generating fragility curves. Several novel approaches have been developed and presented in the last few decades as well. The fundamental concepts of these methods are summarized in Table 1.

***Table 1**. A summary of the methods developed for deriving fragility curves.*

reference	***aim***	***Methodology and outcomes***
Park et al. [8]	*Evaluating seismic fragility of low-rise unreinforced masonry buildings*	*A simplified spring model was presented for simulating the highly nonlinear dynamic behaviour of URM buildings.*
Rota et al. [9]	*Developing a methodology for driving fragility curves of masonry buildings*	*Mechanical properties were considered as random variables with a reasonable value. The input variables were extracted through Monte Carlo method from the distributions. The probability distribution of the damage states was defined by pushover analysis. Novelty: unlike other studies, results of nonlinear stochastic analyses of a prototype building is used while simplified models of buildings and approximate analysis are generally considered by researchers.*

Cardone et al. [10]	*Driving Fragility curves of RC buildings through a hybrid approach*	*In HAZUS, fragility curves are developed using inelastic static analysis, in this study, however, fragility parameters are obtained by a hybrid method. Median values are predicted by comprehensive loss assessment analysis and dispersion values are evaluated according to the results of accurate inelastic dynamic analysis.*
Alwaeli et al. [11]	*Proposing a new methodology for driving fragility curves with less computational efforts and time*	*A novel record selection criterion and a fragility curve tolerance factor are provided for developing rigorous (refined) and less-demanding fragility relations for RC high-rise buildings.*
Dona et al. [19]	*Developing a mechanics-based fragility model for Italian residential URM buildings*	*A database including 500 building information was used to proposed a fragility model which was based on the classification of the buildings in terms of age and story numbers. The verification was made by simulating 2009 L'Aquila earthquake, proved the acceptable reliability of the model.*
Sandoli et al. [14]	*Proposing a hybrid method for driving fragility curves of masonry buildings*	*The innovative method is a combination of: (i) an expert judgment by classifying buildings in the viewpoint of construction age, structural typology, seismic behaviour and damage of buildings caused by the earthquakes, and (ii) numerical analysis results.*
Beilic et al. [20]	*Developing out-of-plane fragility functions using probabilistic analysis and Monte Carlo simulation*	*The study novelty was that, unlike other studies which consider demand, this study used the uncertainties in the capacity namely aspect ratio, level of In-Plane damage, position of the infill walls and type of masonry (solid or hollow units).*
Cardone et al. [21]	*Generating collapse fragility curves for base-isolated RC buildings*	*Fragility curves of RC building with either low- or high-seismic resistance systems retrofitted by different isolation systems are driven.*

The main aim of the recently developed models is to ease the process of deriving fragility curves by (i) proposing building simulation assumptions [8, 9], (ii) proposing simpler methods by combining simpler techniques [10, 14], (iii) using the assets of building classification in terms of age, structural typology, story numbers and seismic behaviour [14, 19], (iv) presenting

novel earthquake record selection [11] and (v) considering the uncertainties of capacity rather than those of demand [20]. Fragility curves of buildings with novel resisting systems (e.g., seismic isolators) have also been presented and discussed by researchers [21].

1.4. Review on the application of ML-based techniques for generating fragility curves

Due to complicated and time-consuming analysis required for IDA and fragility assessment, there has been an increasing interest in the implementation of quicker and simpler models for deriving fragility curves. Artificial Intelligence (AI) and Machine Learning (ML) have been increasingly applied by researchers in various fields [22-24] as well as structural and earthquake engineering [25-28].

During the last years, a few attempts have been made to apply ML-based techniques in the process of emerging fragility curves. A summary of the relevant studies is given in Table 2. The main objective of the studies listed in Table 2 is to boost the process of generating fragility curves by either making the process quicker [4] or reducing the uncertainty degrees of different parameters [15, 29]. As an example, maximum story drift was estimated through ML-based methods in the models proposed by Kiani et al [4] and Hwang et al. [7]. Jia and Wu [29]have also developed a novel model for predicting a dimensionless parameter which was used for obtaining failure probability. The most remarkable conclusion of the studies is that the results of the prediction models developed based on ML or Neural Network (NN) are in line with the results of the conventional method with an acceptable level of accuracy. More specifically, their evaluations have clarified that tree-based approaches (e.g., DT or RF) led to more accurate outcomes compared to other ML-based approaches (e.g., ANN) [7, 29].

The notable issue with the proposed models, on the other hand, is their limitations including (i) being capable only for a specific building, and (ii) excluding the effective parameters (e.g., construction material, soil type, building's location, etc.) in the prediction models. Accordingly, further studies are definitely required to provide a more generalized model for emerging fragility curves.

Table 2. *Literature review on the application of ML-based models for obtaining fragility curves.*

reference	*aim*	*methodology*	*conclusions*	*Predicted parameters (output)*	*Limitations*
Mitropoulou and Papadrakakis [15]	*Developing fragility curves based on neural network IDA predictions using Harmony search Optimization algorithm.*	*(a) Analysing an 8-story regular plan and a 5-story irregular plan RC buildings in OpenSees, (b) considering Arias Intensity (I_A), Characteristic intensity (I_C) and Cumulative Absolute Velocity (CAV) IMs reflecting respectively the amplitude, the frequency content and the duration of a strong ground motion were considered as inputs while spectral acceleration in different DSs defined as outputs (4 nodes), (c) considering SA and maximum inter-story drift ratio (MIDR) as IM and EDP, respectively, (d) predicting fundamental period spectral acceleration (seismic demand) by NN, (e) developing four limit state fragility curves, (f) examining the computational cost of the neurocomputing scheme.*	*The obtained results with NN were in line with the results of the conventional method.*	*Period spectral acceleration was predicted.*	*(i) The results were limited to two buildings, (ii) Arias Intensity, Characteristic intensity and Cumulative Absolute Velocity IMs were considered as inputs while other parameters were not taken into account.*
Kiani et al. [4]	*(a) Application of Machine learning methods for deriving fragility curves, (b) evaluating influence of input uncertainties (i, e., GM variability), (C) investigating the influence of training samples' size on the results of classification techniques.*	*(a) Modelling an 8-story steel special moment resisting frame with the period of 2.3s located in Los Angeles, US in OpenSees, (b) considering spectral velocity as IM, (c) considering as EDP in two groups (MIDR>0.03 rad and MIDR<0.03 rad), (e) predicting structure response in terms of Displacement Spectrum Intensity Ratio (DSIR)*	*(i) RF, SVM and DT were the most accurate methods, (ii) lasso regression and Naïve Bayes were not affected by training sample size while QDA was the most sensitive method.*	*DSIR is predicted.*	*(i) The results were limited to an 8-story steel frame, (ii) only one limit state (MIDR=0.03) was considered), (iii) other input parameters including structural systems, damage measures, thresholds of structural responses, and sites were not considered.*
Jia and Wu [29]	*Driving fragility curves of RC frame-shear wall structures using ensemble Neural network.*	*(a) Analysing a 4-story shear wall RC structure with a basic acceleration of 0.2 g located on Type II soil in China using SAP2000, (b) Considering MDIR and peak floor acceleration (PFA) as EDP and PGA as IM, (c) Defining four common damage states, (d) Defining 26 parameters reflecting GM and structure properties as inputs and an index (L) as output.*	*The proposed ensemble ANN model predicted more accurate values in comparison to back propagation (BP), cascade BP and Alman ANN.*	*a dimensionless parameter (L) was predicted which is used for calculating failure probability.*	*(i) The results were limited to a 4-story shear wall RC structure, (ii) parameters namely material, period, soil type, and location were not considered in the model.*
Hwang et al. [7]	*(a) Predicting seismic response and structural collapse of RC frames, (b) identifying different input variables on RC frames' seismic collapse.*	*(a) Analysing a 4-story and an 8-story special moment resisting frame located on stiff soil in California using OpenSees, (b) Considering Sa and maximum story drift as IM and EDP, respectively, (c) using modelling-related parameters and ground motion intensity measures as input for predicting MSD in the first prediction model, (d) using MSD in addition to all the input models in the first part, as input for predicting survival-failure vector.*	*(i) Influence of structural modelling uncertainties on seismic demand could be safely neglected at minor to moderate damage levels for low- to mid-rise RC frame buildings, (ii) extreme gradient boosting algorithm and tree-based techniques (dt and RF) led to more accurate results compared to other methods.*	*Maximum story drift and collapse status were predicted.*	*(i) The results were limited to 4- and 8- story RC frame buildings, (ii) parameters namely material, period, soil type, and location were not considered in the model.*

2. Significance, novelty and methodology of the study

Obtaining fragility curves is an inevitable key step in seismic risk assessment in the performance-based earthquake engineering which accounts for the uncertainties of risk due to seismic events. Deriving fragility curves, however, is generally time-consuming due to a huge amount of analytical analysis performed for IDA. Moreover, it needs powerful operating systems, particularly in the case of studying tall and complex buildings.

Although a few innovative models have been developed so far (Table 2) [4, 7, 15, 29], they suffer from shortcomings which were mentioned in the previous section. More importantly, the models do not estimate fragility curves directly. Otherwise noted, they could be used for predicting neither IM or EDM. This means that the time-consuming analysis still needs to be performed even if the prediction models are applied.

The main objective of this research, therefore, is to propose a prediction model for deriving fragility curves using regression- and ML-based techniques. The remarkable novelty of this paper is proposing models which output fragility curves' parameters (β and μ) directly. In other words, the time-consuming IDA is eliminated and consequently the fragility curves could be obtained quickly by defining the inputs. Furthermore, all the parameters which are proved to have a strong effect on fragility curves are considered. More importantly, the proposed models are not limited to a specific building and could be utilized for any reinforced concrete (RC), steel or masonry buildings.

Overall, the benefits of the proposed prediction models are: (i) predicting μ and β directly, (ii) eliminating time-consuming IDA, (iii) considering all the effective parameters and (iv) being generalized and therefore applicable for a huge number of buildings.

To this aim, a comprehensive database is gathered and various ML-based methods including nonlinear regression (NLR), Decision Tree (DT), Random Forest (RF), K-nearest Neighbours (KNN) and Artificial Neural Network (ANN) are used to develop models for deriving fragility curves. The accuracy of the models is assessed through Taylor diagram and performance metrics and the most accurate model is introduced. Eventually, a parametric study is conducted to figure out the effect of input variables on the output parameters and prediction equations are provided accordingly.

3. Data collection

A database including 238 results of the fragility assessment of buildings was collected from peer-reviewed international publications [1-3, 8, 9, 12, 17, 18, 30-36]. It is worth mentioning that the reliability of the collected database was improved by removing outliers and incomplete datasets. It should be also explained that during the data collection, special attention was given to consider the parameters which affect fragility parameters significantly based on the literature results. As an example, Gaudio et al. [12] have proved that buildings' height has a significant influence on its fragility while influence of its construction age could be neglected.

Accordingly, the parameters considered as inputs for the prediction models are: construction materials (i. e., RC, steel and masonry), buildings' plan area (m^2), buildings' height (m), lateral resisting system (i. e., shear wall, bearing masonry wall, bracing system or moment resisting frame-MRF), buildings' location, damage state, buildings' period, soil classification (e. g., rock). The output, on the other hand, are the fragility parameters: median (μ) and dispersion

(β). One of the parameters which could reveal if the input variables are defined appropriately is Pearson correlation coefficient and is defined as the ratio of *x*, *y* covariance, *cov (x, y)*, to the production of their standard deviation ($\sigma_x\sigma_y$) as given in Eq. 3 [37, 38]:

$$\rho_{x,y} = \frac{cov\ (X,Y)}{\sigma_x \sigma_y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \qquad (3)$$

Pearson coefficient of two parameters is a value in the range of (0,1). $\rho \approx 1$ reflects the high influence of the parameters on each other while $\rho \approx 0$ stands for the linear dependency of the parameters. It should be explained that Pearson correlation exhibits linear relationship between parameters, which means that $\rho \approx 0$ does not necessarily represent complete indecency of variables and they might have nonlinear dependency [39, 40]. Pearson correlation coefficients of the input variables, and β and μ are reported in Table 3 and Table 4, respectively.

Table 3. *Pearson correlation coefficients between input variables and β.*

	Material	*Plan area (m^2)*	*Height*	*Resistant system*	*Location*	*State*	*Code*	*Period (s)*	*Soil type*	*Dispersion*
Material	*1.000*									
Plan area (m^2)	*0.230*	*1.000*								
Height (m)	*0.420*	*0.555*	*1.000*							
Resistant system	*-0.723*	*0.032*	*-0.137*	*1.000*						
Location	*-0.491*	*-0.576*	*-0.340*	*0.335*	*1.000*					
State	*0.210*	*-0.364*	*-0.143*	*0.035*	*-0.015*	*1.000*				
Code	*-0.369*	*-0.349*	*-0.267*	*0.445*	*0.870*	*0.070*	*1.000*			
Period (s)	*0.302*	*0.307*	*0.801*	*0.009*	*-0.036*	*-0.122*	*-0.007*	*1.000*		
Soil type	*-0.169*	*-0.381*	*-0.135*	*-0.112*	*0.557*	*0.059*	*0.276*	*-0.082*	*1.000*	
Dispersion	***0.181***	***0.379***	***0.171***	***0.015***	***0.076***	***-0.067***	***0.235***	***0.105***	***0.047***	***1.000***

Table 4. *Pearson correlation coefficients between input variables and μ.*

	Material	*Plan area (m^2)*	*Height*	*Resistant system*	*Location*	*State*	*Code*	*Period (s)*	*Soil type*	*Median*
Material	*1.000*									
Plan area (m^2)	*0.230*	*1.000*								
Height (m)	*0.420*	*0.555*	*1.000*							
Resistant system	*-0.723*	*0.032*	*-0.137*	*1.000*						
Location	*-0.491*	*-0.576*	*-0.340*	*0.335*	*1.000*					
State	*0.210*	*-0.364*	*-0.143*	*0.035*	*-0.015*	*1.000*				
Code	*-0.369*	*-0.349*	*-0.267*	*0.445*	*0.870*	*0.070*	*1.000*			
Period (s)	*0.302*	*0.307*	*0.801*	*0.009*	*-0.036*	*-0.122*	*-0.007*	*1.000*		
Soil type	*-0.169*	*-0.381*	*-0.135*	*-0.112*	*0.557*	*0.059*	*0.276*	*-0.082*	*1.000*	
Median	***-0.184***	***0.024***	***-0.006***	***0.064***	***0.384***	***0.045***	***0.229***	***-0.039***	***0.609***	***1.000***

According to Tables 3 and 4, it could be claimed that plan area and building location have the highest linear influence on β and μ, respectively. Resistant system and height, on the other hand, exhibited the lowest linear dependency respectively on β and μ.

Distribution of the input variable and output parameter(s) influence the applicability and generalization of prediction models; the broader range of input values is, the more general the model will be. On that account, it is tried to include the datasets which lead to a high range of values for each input. Distribution of the input parameters and outputs is demonstrated in Fig. 2. The statistical characteristics including minimum, maximum, median, variance and standard deviation of the data are also given in Table 5.

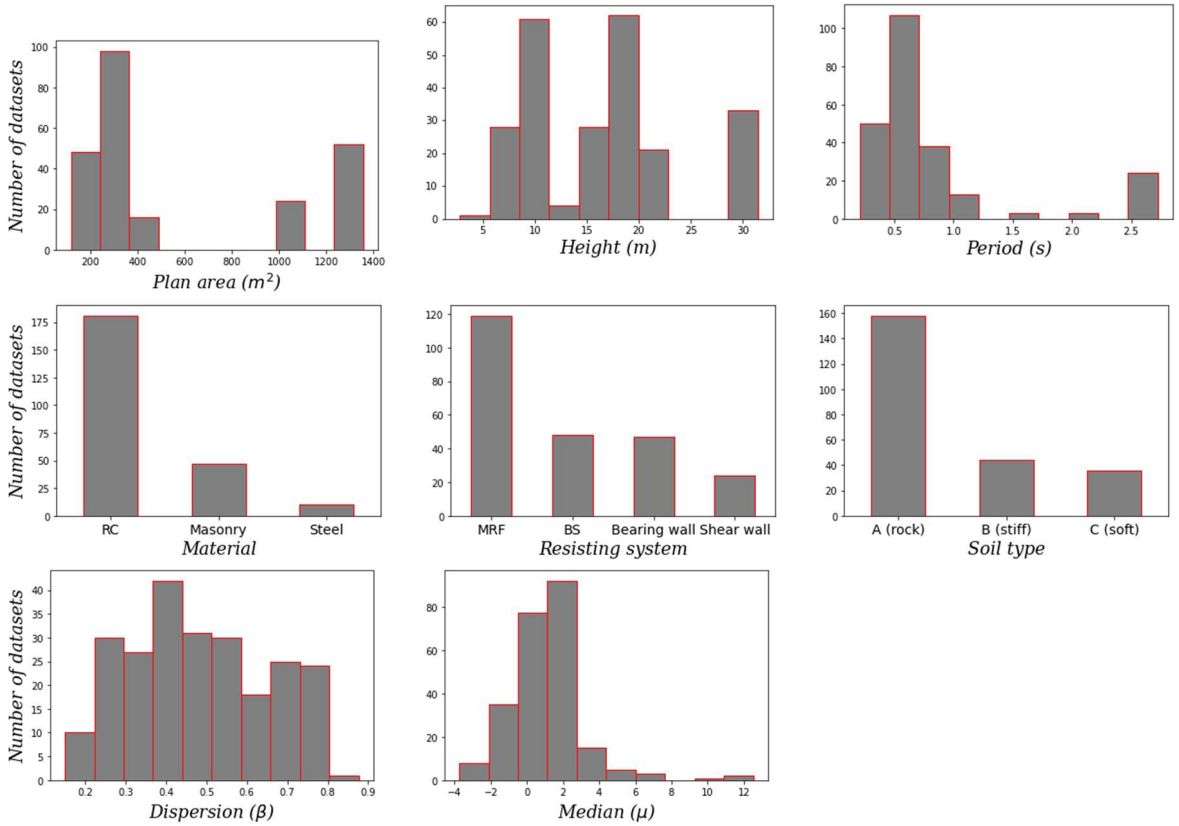


Figure 2. *Distribution of the input and output parameters.*

Table 5. *Statistical properties of the quantitative input and output parameters.*

	Minimum	*Maximum*	*Mean*	*Median*	*Variance*	*Standard deviation*
Plan area (m^2)	*53.29*	*1296.00*	*539.84*	*274.40*	*220402.62*	*469.47*
Height (m)	*1.35*	*30.00*	*15.31*	*14.50*	*53.57*	*7.32*
Period (s)	*0.08*	*2.60*	*0.72*	*0.48*	*0.49*	*0.70*
Dispersion (β)	*0.11*	*0.84*	*0.45*	*0.42*	*0.03*	*0.17*
Median (μ)	*-4.56*	*11.70*	*0.39*	*0.30*	*4.16*	*2.04*

The database was divided into two sub-databases namely training (85%) and testing (15%). The former is used for training the models the relationship between the inputs and outputs and the latter is used for assessing the accuracy of the predicted values. Otherwise mentioned, the testing datasets are not used for training purpose while they are utilized for validity of the prediction models.

4. Prediction models

The prediction models are developed in two parts in this study. In the first part models are presented to predict dispersion while median is estimated in the second part. Various techniques including NLR, DT, RF, KNN and ANN were used for proposing prediction models. An extensive definition of the methods could be found in the literature; therefore, they are introduced here briefly for the sake of shortness.

4.1. Nonlinear regression (NLR)

Simply stated, a regression-based model is a model which fits an equation to a set of data. The equation refers to a line which processes the least difference between the mean and each data. The obtained regression equation is known as linear and nonlinear when the fit line is a straight line and a curve, respectively. In the sense of mathematic, a regression model could be expressed as Eq. 4. When f is linear in θ, y will be linear while when f is nonlinear in θ, y will be a nonlinear regression model [25, 26, 41].

$$y = f(x_i : \theta) + \varepsilon \qquad (4)$$

where f is the function showing the relationship between inputs and outputs, x_i are the inputs, θ are the parameters and ε is a random variable error with mean=0 and standard deviation=σ. NLR generally yields more accurate predicted values than linear regression since it can fit a much wider range of curves (nonlinear relationship) between its variables [25, 42, 43]. Moreover, NLR finds the most reliable fit by minimizing the sum of squares of the distance between the actual and model prediction values (RSS), as given in Eq. 5 [44]:

$$RSS\ (P_1, \dots, P_m) = \sum_{i=1}^{n} e_i^2, \qquad e_i = z_i - g(t_i, P_1, \dots, P_m) \qquad (5)$$

where z_i is the prediction values, g are the actual values and P_i are the parameters.

4.2. Decision Tree (DT)

A decision tree is a method which uses a tree-shape graph for moving a dataset sample to the most accurate target output based on its characteristics. DT could be used for solving both classification (finite set of values) and regression (continues values) problems [45, 46]. The process of predicting a value is schematically depicted in Fig. 3. The most assets of DT technique are: (i) being simple for understanding and interpreting, (ii) being independent of the nonlinear relation between parameters, and (iii) being flexible to be adopted to new scenarios [47, 48].

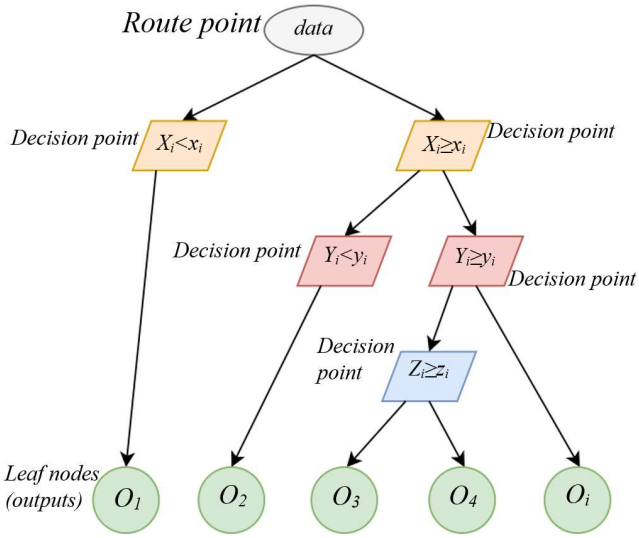


Figure 3. *Schematic illustration of a DT prediction model.*

4.3. Random Forest (RF)

In the simplest concept, RF consists of many decision trees and target output is predicted by considering either the average of the DTs' predicted values or the most voted value [9t,10t]. To explain more precisely, RF is basically the combination of Bagging and Random selection of features by creating various decision trees. The most notable point about RF is that selecting training dataset for each tree is done through Bootstrap sampling and the feature which is chosen as the decision node is a random subset of the main dataset [49-51]. The above-mentioned explanations are simply demonstrated in Fig. 4.

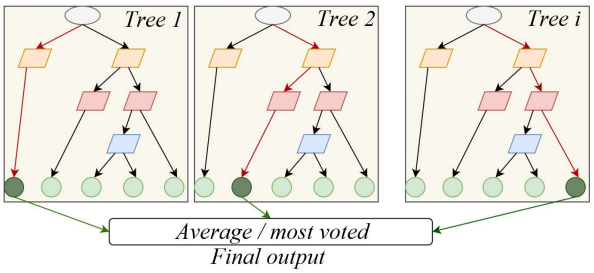
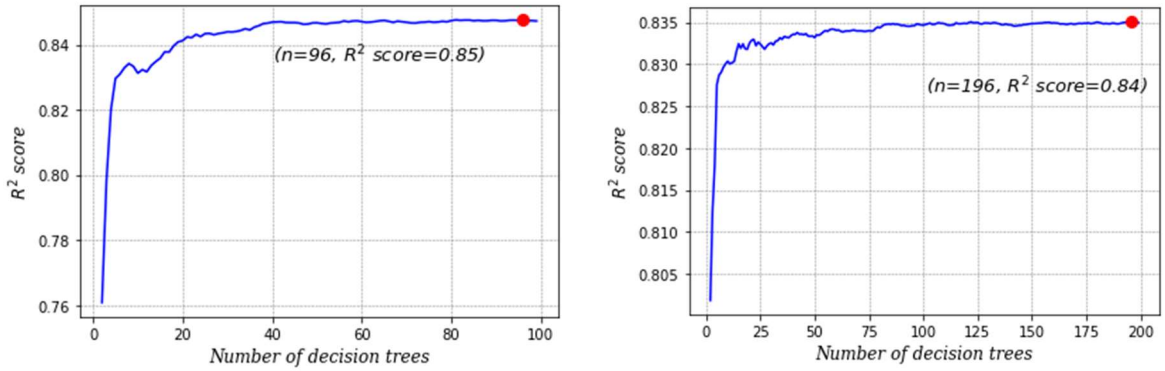


Figure 4. *A simplified illustration of a RF model.*

In order to find out the most efficient number of trees in our RF model, R^2-score of various RFs with different numbers of trees was obtained (Fig. 5). As could be observed in Fig. 5, 96 and 196 trees led to the highest accuracy of the RF models developed for predicting β and μ, respectively.

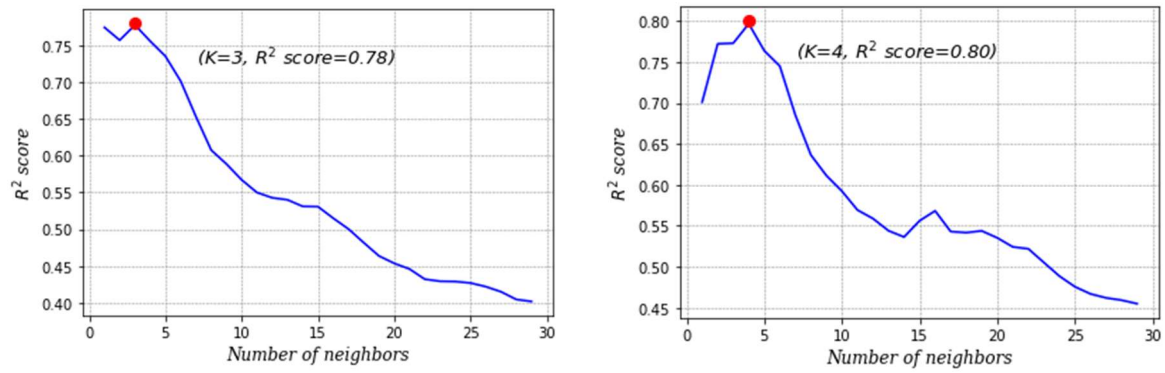


(a) (b)

Figure 5. *The most efficient number of trees in the RF models developed for predicting (a) β and (b) μ.*

4.4. K-Nearest Neighbours

K-nearest Neighbours (KNN) is recognized as a non-parametric prediction method which means that the prediction process is not influenced by the relationship between input and output parameters [52]. Therefore, KNN is known as the simplest classification approach by data scientists. In this method, the data are plotted in a multi-dimensional space where the axis are the data's features. When a new data is added to the space according to its characteristics, the average of its "K" nearest neighbours is the target output [53, 54]. As a result, the most effective parameter which could enhance the accuracy of predicted values is the number of nearest neighbours defined as "K". Although some researchers have claimed that square root of the total number of datasets could be the most accurate K [53], it is generally obtained by trial-and-error process. The best K values obtained for predicting β and μ are 3 and 4, respectively, as depicted in Fig. 6.



(a) (b)

Figure 6. *Finding the most accurate K for the KNN models for predicting (a) β and (b) μ.*

4.5. Artificial Neural Network

ANN is basically a combination of computation and mathematics which is inspired from the human brain by simulating the performance of the brain's nervous system [54, 55]. The architecture of an ANN model consists of (i) input layers: the number of the nodes in this layer is equal to the number of model inputs, (ii) hidden layer(s) which might be considered to enhance the model accuracy, and (iii) output layers: like input layer, the number of nodes is determined based on the number of target outputs. The nodes are connected to each other by weights which are updated in each irritation in order to reach an acceptable estimated output [55, 56]. Other optional components which might be considered in an ANN model are: (iv)

bias values which could be defined for hidden and output layers and (v) activation functions which might be added to the weights in order to allow the ANN to account for nonlinear behaviour in the training dataset. Otherwise noted, an Ann model without an activation function can perform linearly with unreliable accuracy [56, 57]. An ANN model could be formulized as below:

$$O_j = f\sum(w_{ij}I_i + b) \qquad (5)$$

Where O_j is the model output, w_{ij} is the associated weight which is updated in each epoch, I_i is input data and b is bias [25].

Based on the above-mentioned definitions, number of hidden layers, number of nodes in each hidden layer and type of the activation function are the main factors which affect the performance of an ANN model significantly. These parameters are typically obtained through a trial-and-error process. The architecture of the ANN model developed for presenting β and μ is displayed in Fig. 7. It should be noted that rectified linear activation function or ReLu function was defined as the activation function in the models.

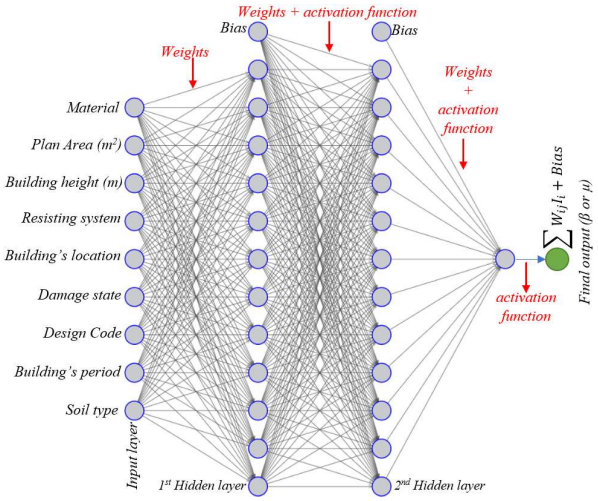


***Figure** 7. The architecture of the ANN models developed in this study for predicting (a) β and (b) μ.*

5. Results

The prediction models were developed by adjusting their characteristics using training sub-databases. Then, the target outputs of both training and testing datasets were predicted by the models. The correlation between the predicted and actual values is shown in Fig. 8 and Fig. 9 respectively for β and μ.

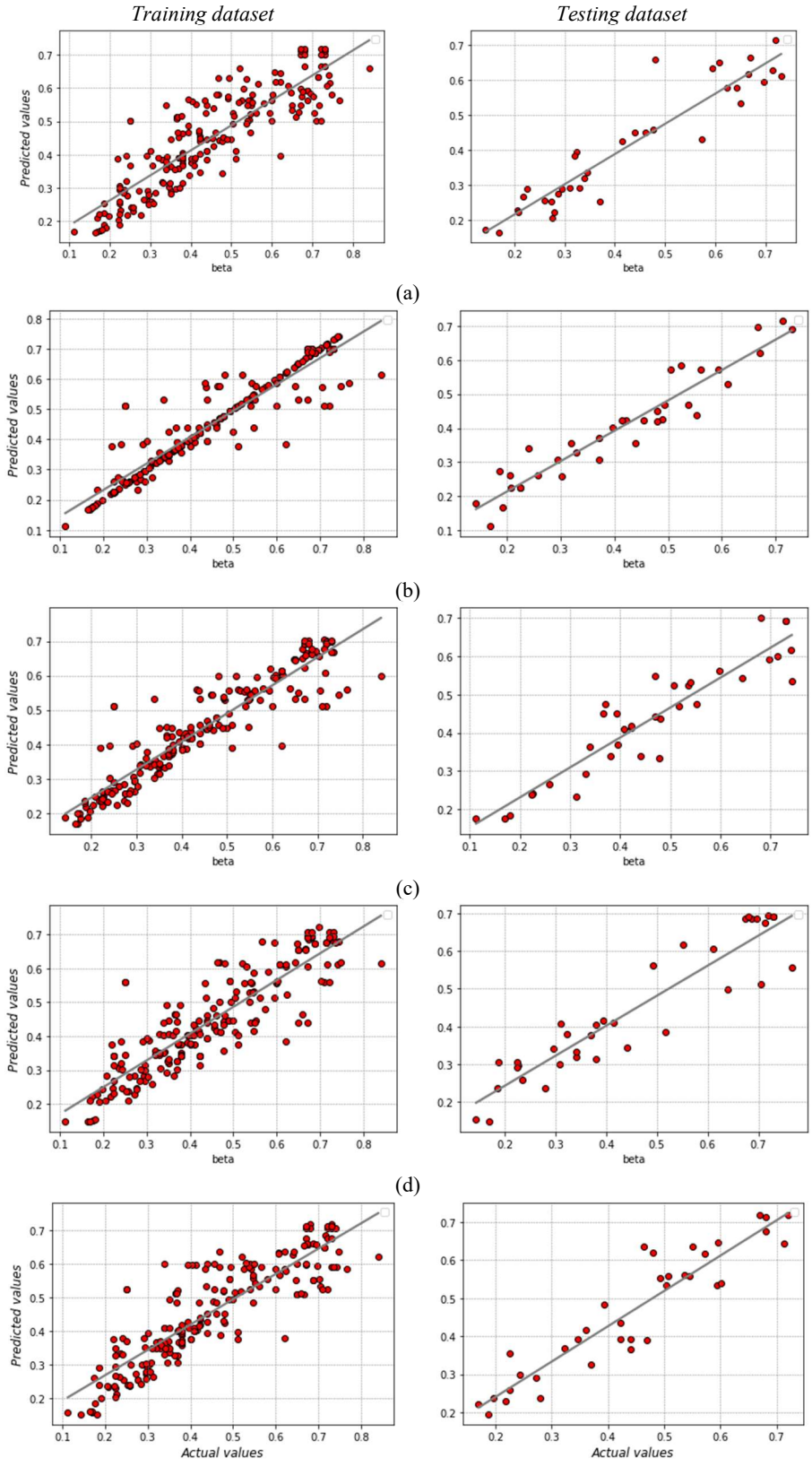


Figure 8. *Correlation between actual and predicted values of β.*

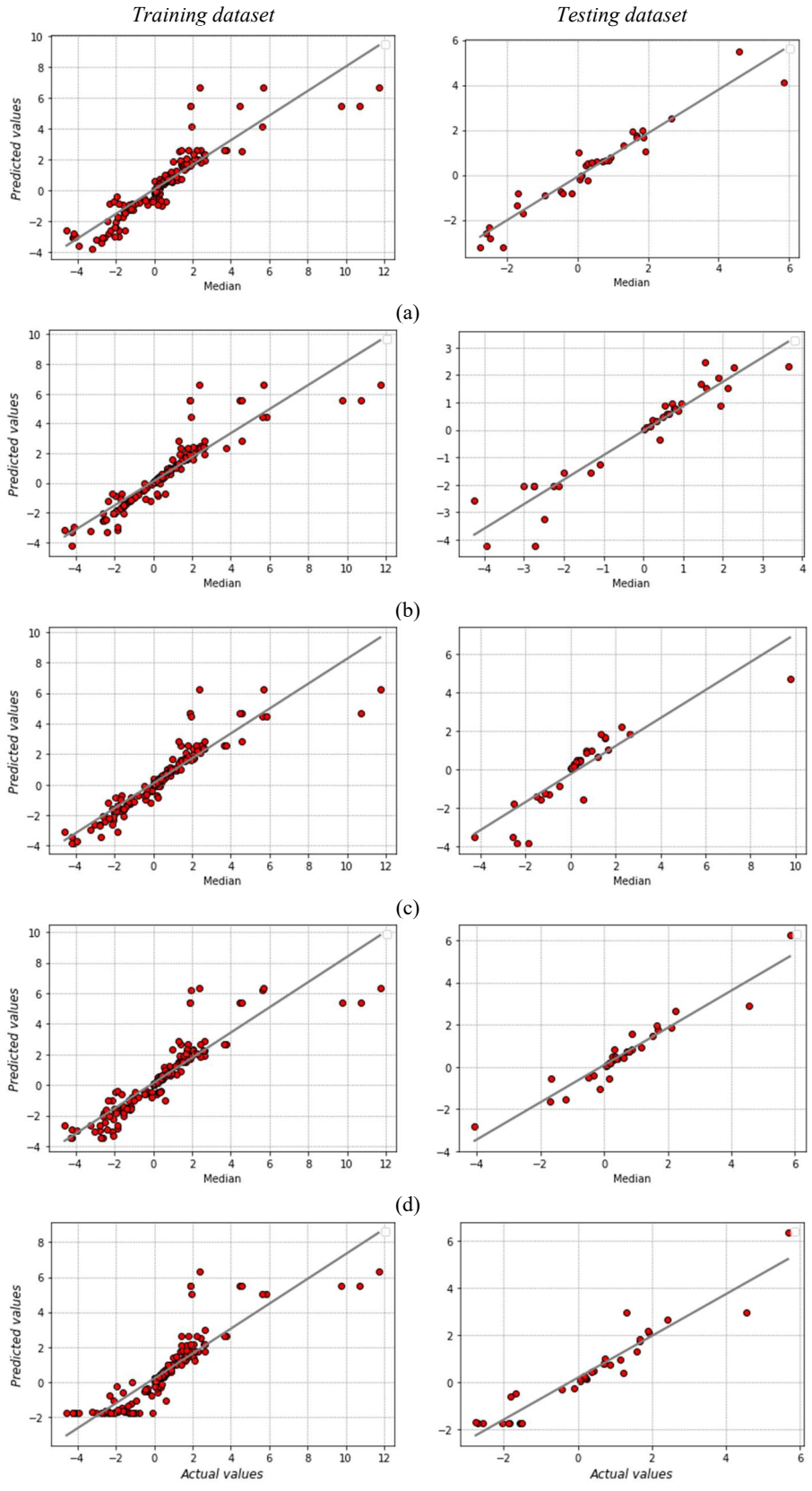


Figure 9. *Correlation between actual and predicted values of μ.*

In order to figure out the accuracy of the models more precisely, each predicted value of β and μ is compared with its corresponding actual value as displayed in Fig. 10 and Fig. 11, respectively.

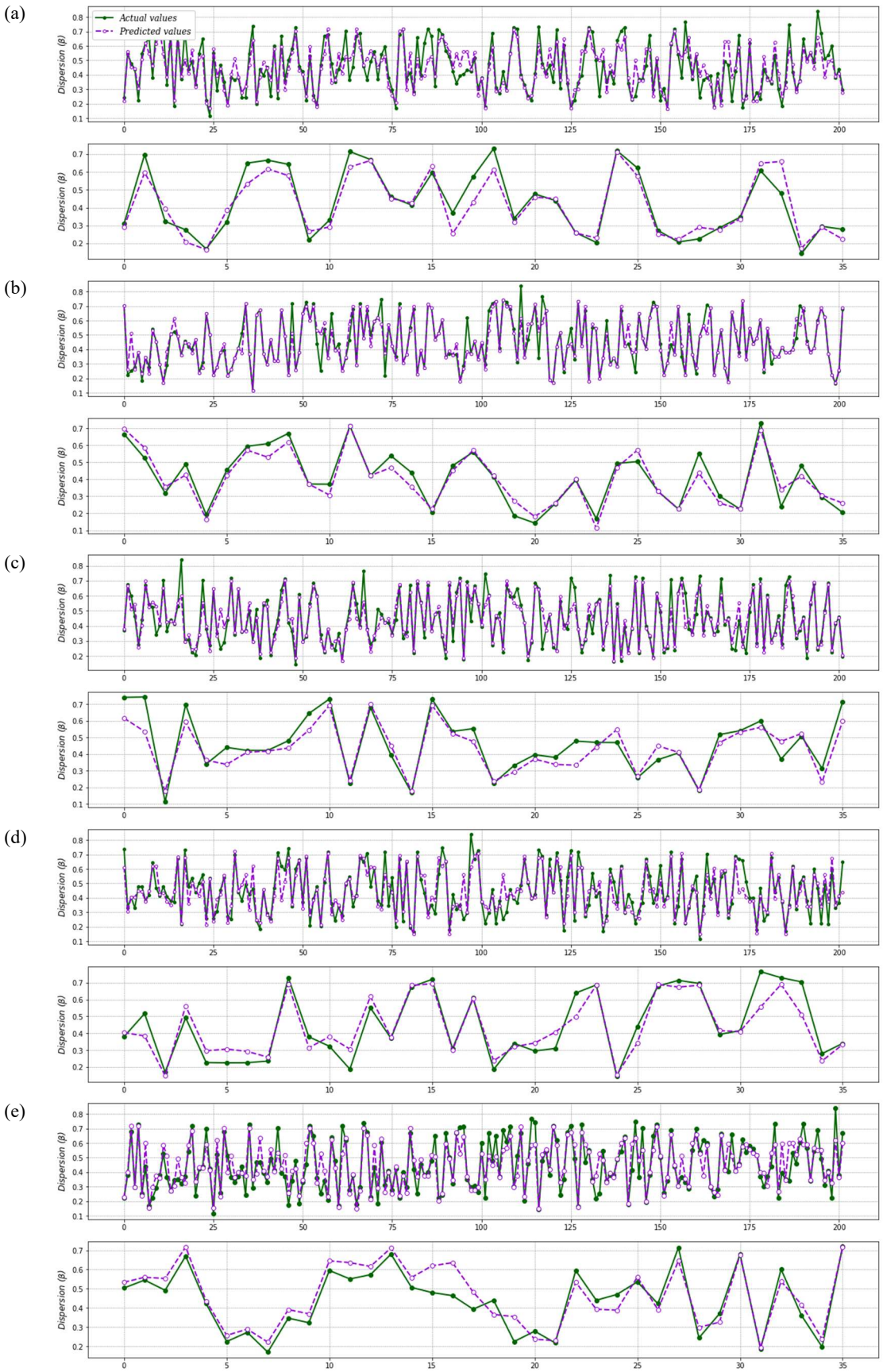


Figure 10. *Comparing the predicted and actual values of β.*

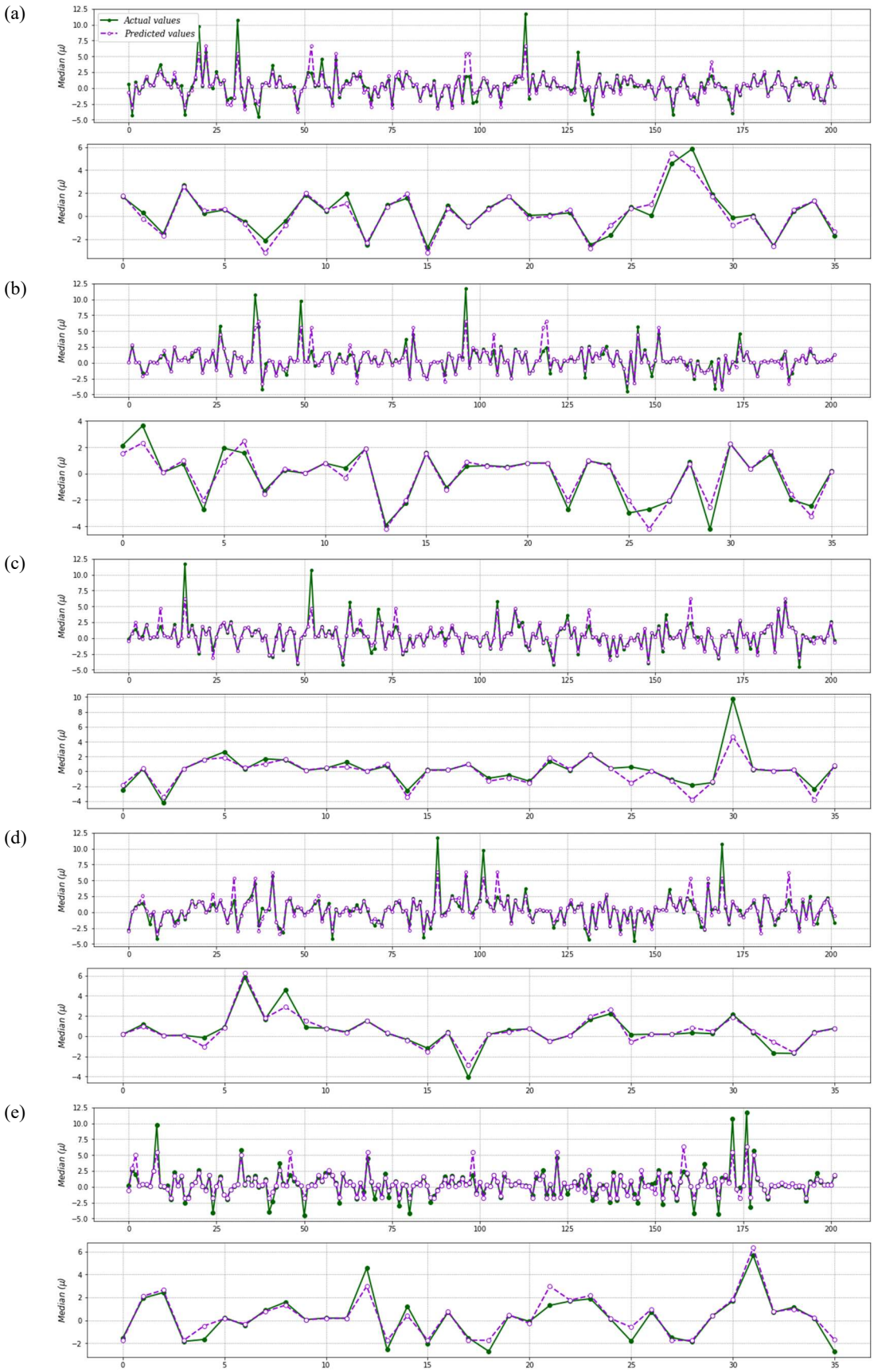


Figure 11. *Comparing the predicted and actual values of μ.*

At the first glance on figures 7-10, it could be claimed that almost all the methods exhibited a high ability to learn the relationship between input variables and two target outputs. More

specifically, DT and NLR showed good correlation between actual and predicted values of β and μ in Figures 7(b) and 8(a). However, the models need to be assessed more precisely as presented in the next section.

6. Accuracy assessment

The accuracy of the prediction models proposed in this study is assessed through common performance metrics and Taylor diagram.

6.1. Performance metrics

The performance metrics which are commonly used for assessing accuracy of a model are Root of Mean Square Error (RMSE), Mean Absolut Error (MAE), Mean Absolute Percentage Error (MAPE) and R2-score. These parameters are calculated using Eq. (6-9).

$$RMSE = \left(\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2\right)^{0.5} \quad (6)$$

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

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \quad (8)$$

$$R^2 = 1 - \frac{\sum_i(\hat{y}_i - y_i)^2}{\sum_i(y_i - \bar{y}_i)^2} \quad (9)$$

where y is the actual output, $\hat{y}$ is the predicted output, n is the number of data records and $\bar{y}$ is the mean of the dataset. Table 6 compares the performance metrics of all the models for estimating β and μ quantitatively.

***Table 6**. Performance metrics of the developed models.*

	Dispersion prediction models				*Median prediction models*			
	R^2	*RMSE*	*MAE*	*MAPE*	*R^2*	*RMSE*	*MAE*	*MAPE*
Nonlinear	*0.87*	*0.06*	*0.05*	*11.83*	***0.92***	***0.51***	***0.35***	***157.93***
DT	***0.91***	***0.05***	***0.04***	***11.27***	*0.91*	*0.60*	*0.38*	*26.78*
RF	*0.83*	*0.07*	*0.05*	*12.23*	*0.76*	*1.06*	*0.50*	*42.49*
KNN	*0.86*	*0.07*	*0.05*	*60.07*	*0.91*	*0.48*	*0.27*	*340.78*
ANN	*0.84*	*0.06*	*0.05*	*13.49*	*0.90*	*0.59*	*0.38*	*31.81*

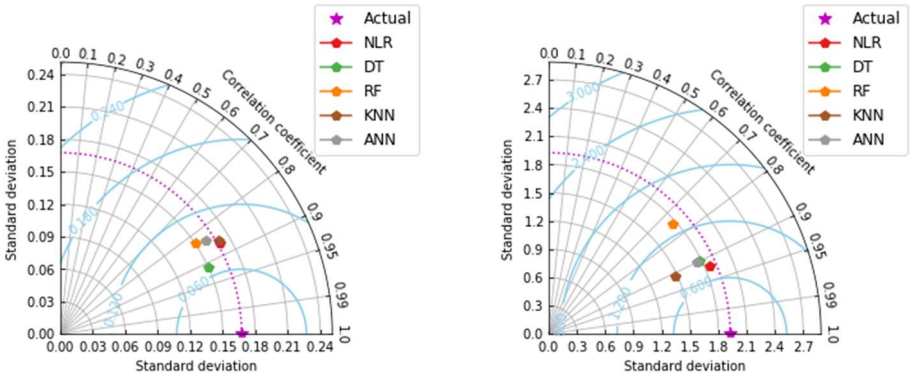
Regarding the values reported for the β-prediction models, it could be observed that the DT model processes the highest R^2-score (91%) and hence is introduced as the most accurate model. The RF model, on the other hand, demonstrated the lowest reliability with R^2=67%.

Considering the models developed for predicting μ, it could be stated that the NLR, DT and KNN models exhibited similar accuracy (R^2≈91%). ANN and RF, then, showed lower accuracy compared to other models.

It is worth mentioning that almost all the proposed models have an acceptable level of ability to learn the relationship between the inputs and outputs which reflect the high reliability of the ML-based methods for predicting fragility curves of buildings.

6.2. Taylor diagram

In order to compare the accuracy of the models more easily, they are compared in Taylor diagram shown in Fig. 12. It should be explained that Taylor diagram consists of three main parts: (i) horizontal and vertical axis which reflects standard deviation, (ii) circular curves centred at actual values which reflect RSME, and (iii) radial line which reflects R^2-score. In other words, each model is plotted in a Taylor diagram by its standard deviation, RMSE and R^2-score. The closest model to the actual value is known as the most accurate model [25, 58, 59].



(a) (b)

Figure 12*. Using Taylor diagram for comparing the models developed for predicting (a) β and (b) μ.*

Considering Diagrams illustrated in Fig. 11, and the above-mentioned explanations, the models DT and NLR are the closest model to the star point (actual values) and thus are introduced in as the most accurate models for predicting respectively β and μ, as concluded in the previous section.

Overall, it could be clarified that a ML-based model combined of DT and NLR could be used for deriving fragility curves with a high level of accuracy.

7. Sensitivity analysis

In this section of the paper, the influence of buildings properties namely construction material (i.e., RC, steel and masonry), plan area, building height and period, and soil type on the fragility parameters is assessed. A two-story (6.20m height) masonry building located on soil type C (according to EC8 [60] soil classification) in L'Aquila, Italy is selected from the literature [] and considered as the reference dataset. Then, the above-mentioned characteristics are varied to generate a series of 26 datasets, as reported in Table 10. The first row of Table 10 represents the reference building which is used for assessing the influence of parameters' variation on both β and μ. The material of the second and third building is changed to reinforced concrete and steel, respectively. Then, the plan area of the reference building is varied in the range of 100-1000 m^2 by the step of 100. The building height is also increased from 3m to 30 m by the step of 3. It should be noted that building period is influenced by its height and therefore the variation of either β or μ by changing building's period is almost the same as that of changing building's height. However, to figure out the effect of building period on the fragility parameters, it is varied according to height variation. Eventually, three soil classifications

including A (rock), B (stiff) and C (soft) are taken into account. Fragility parameters, β and μ, of the generated database are estimated by the prediction model developed using the Decision Tree model. It is worth explaining that the DT model is used for prediction because it showed a high accuracy (R^2=*0.91*) for predicting both β and μ. The predicted values and their differences compared to the reference case is provided in Table 7.

***Table** 7. Variation of β and μ by changing building's properties and soil type.*

	Material	*Area (m2)*	*Height (m)*	*Period*	*Soil type*	β	$(\beta_i-\beta_0)/\beta_0$	μ	$(\mu_i-\mu_0)/\mu_0$
1	*Masonry*	*220.00*	*6.20*	*0.12*	*C (soft)*	*0.38*	*0.00*	*1.34*	*0.00*
2	*RC*	*220.00*	*6.20*	*0.12*	*C (soft)*	*0.74*	*0.95*	*6.59*	*3.92*
3	*Steel*	*220.00*	*6.20*	*0.12*	*C (soft)*	*0.72*	*0.88*	*6.59*	*3.92*
4	*Masonry*	*100.00*	*6.20*	*0.12*	*C (soft)*	*0.62*	*0.63*	*1.90*	*0.42*
5	*Masonry*	*200.00*	*6.20*	*0.12*	*C (soft)*	*0.36*	*-0.07*	*1.90*	*0.42*
6	*Masonry*	*300.00*	*6.20*	*0.12*	*C (soft)*	*0.37*	*-0.03*	*1.68*	*0.25*
7	*Masonry*	*400.00*	*6.20*	*0.12*	*C (soft)*	*0.37*	*-0.03*	*1.68*	*0.25*
8	*Masonry*	*500.00*	*6.20*	*0.12*	*C (soft)*	*0.37*	*-0.03*	*1.68*	*0.25*
9	*Masonry*	*600.00*	*6.20*	*0.12*	*C (soft)*	*0.37*	*-0.03*	*1.68*	*0.25*
10	*Masonry*	*700.00*	*6.20*	*0.12*	*C (soft)*	*0.37*	*-0.03*	*1.68*	*0.25*
11	*Masonry*	*800.00*	*6.20*	*0.12*	*C (soft)*	*0.37*	*-0.03*	*1.68*	*0.25*
12	*Masonry*	*900.00*	*6.20*	*0.12*	*C (soft)*	*0.37*	*-0.03*	*1.68*	*0.25*
13	*Masonry*	*1000.00*	*6.20*	*0.12*	*C (soft)*	*0.37*	*-0.03*	*1.68*	*0.25*
14	*Masonry*	*220.00*	*3.00*	*0.13*	*C (soft)*	*0.36*	*-0.07*	*1.90*	*0.42*
15	*Masonry*	*220.00*	*6.00*	*0.21*	*C (soft)*	*0.42*	*0.11*	*2.27*	*0.69*
16	*Masonry*	*220.00*	*9.00*	*0.29*	*C (soft)*	*0.24*	*-0.38*	*2.27*	*0.69*
17	*Masonry*	*220.00*	*12.00*	*0.35*	*C (soft)*	*0.18*	*-0.54*	*2.27*	*0.69*
18	*Masonry*	*220.00*	*15.00*	*0.42*	*C (soft)*	*0.18*	*-0.54*	*2.27*	*0.69*
19	*Masonry*	*220.00*	*18.00*	*0.48*	*C (soft)*	*0.18*	*-0.54*	*2.27*	*0.69*
20	*Masonry*	*220.00*	*21.00*	*0.54*	*C (soft)*	*0.18*	*-0.54*	*2.27*	*0.69*
21	*Masonry*	*220.00*	*24.00*	*0.60*	*C (soft)*	*0.18*	*-0.54*	*2.27*	*0.69*
22	*Masonry*	*220.00*	*27.00*	*0.65*	*C (soft)*	*0.18*	*-0.54*	*2.27*	*0.69*
23	*Masonry*	*220.00*	*30.00*	*0.71*	*C (soft)*	*0.18*	*-0.54*	*2.27*	*0.69*
24	*Masonry*	*220.00*	*6.20*	*0.12*	*A (rock)*	*0.45*	*0.18*	*2.43*	*0.81*
25	*Masonry*	*220.00*	*6.20*	*0.12*	*B (stiff)*	*0.45*	*0.18*	*2.43*	*0.81*
26	*Masonry*	*220.00*	*6.20*	*0.12*	*C (soft)*	*0.36*	*-0.07*	*1.90*	*0.42*

7.1. Influence of soil type and building properties on β

Figure 13 displays the variation of β when soil type or building parameters change. In terms of construction material (Fig. 13a), it could be claimed that RC and steel structures with similar response, exhibited considerably higher dispersion than that of masonry buildings. As far as building geometry is concerned, it could be realized that by increasing both plan area and height of a building, dispersion alters notably. Fig. 13b, reveals that dispersion drops when building area is increased and then, it remains almost constant for larger areas (in this case study 200 m^2 with the height to area ratio of 3.1%). Roughly the same variation could be figured out that for building's height variation according to Fig. 13c. More clearly, when the height increases from 3 (low-rise) to 6 (mid-rise), dispersion increases by 27.61%. Then, a sharp decrease is

observed for higher buildings with 6-12 m (high-rise) height. Dispersion of tall buildings (more than 12 m height in this study), however, remains approximately unchanged. Fig. 13d shows that the relationship between building's period and dispersion is the same as that of building's height, since period is directly affected by the building's height as explained previously. Using the dispersion values predicted by the developed DT model and the quantitative building properties namely plan area (A, m^2), height (H, m) and period (T, s), fitting curves are drawn as depicted by black dotted lines in Fig.13 b-d. Accordingly, the following equations are suggested for obtaining dispersion:

$$\beta = 0.05083 \times 10^{-7} A^2 - 0.0007A + 0.5845 \qquad (10)$$

$$\beta = 6 \times 10^{-4} H^2 - 0.0283H + 0.4708 \qquad (11)$$

$$\beta = 1.268T^2 - 1.425T + 0.5615 \qquad (12)$$

Fig. 13e illustrates how soil type could affect dispersion. Based on this figure, it could be claimed that dispersion of the buildings located on soft soil (type C) possess lower dispersion compared to those located on either rock (type A) or stiff soil (type B). The same results have been reported in similar studies evaluating the effect of soil type on fragility parameters of buildings [17, 34].

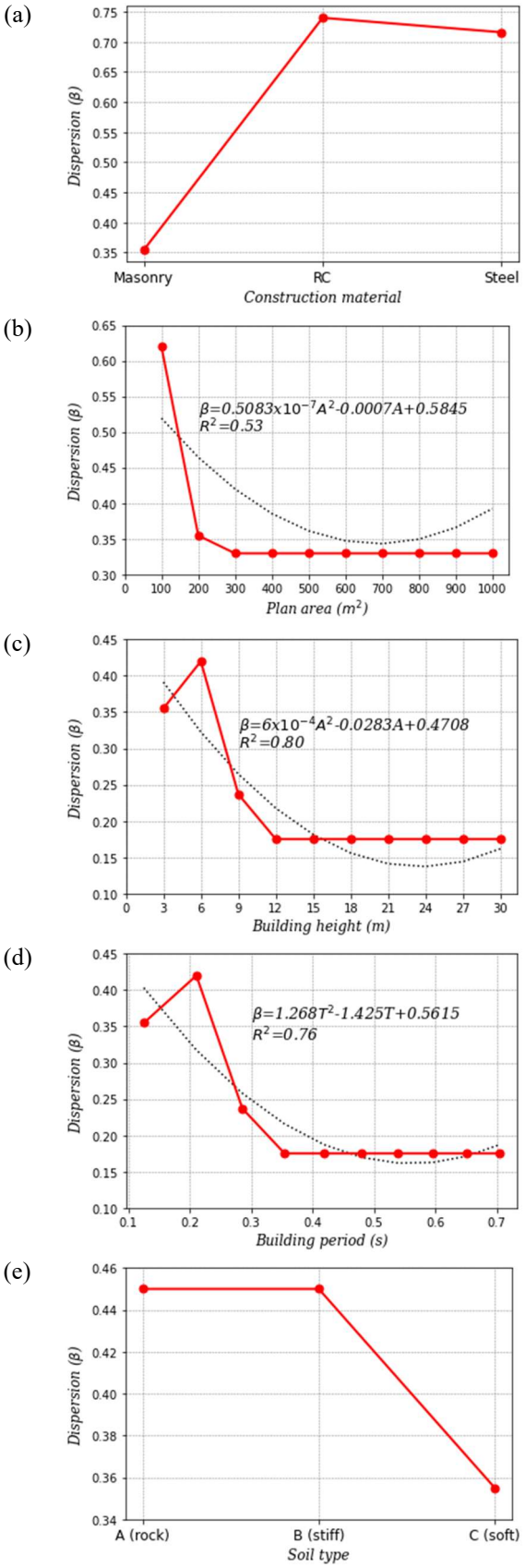


***Figure 13**. Influence of building properties and soil type on dispersion (β).*

7.2. In fluence of soil type and building properties on μ

The effect of building properties and soil type on median is demonstrated in Fig. 14. Just like β, median values of RC and steel structures are higher than that of the masonry building (Fig. 14a). Furthermore, by increasing the building's plan area, median value reduces notably first. Then, it remains almost the same for larger areas (300 m^2 in this study), as displayed in Fig. 14b. The median variation regarding soil type is also the same as that of dispersion; buildings on soil types A and B have higher μ in comparison to those on soil type C as depicted in Fig. 14e. By increasing the building's height or period, unlike dispersion variation, median values jump first (building height=6 m and height/area=2.73% in this study). Then after, it remains roughly constant for higher buildings representing mid- and high-rise buildings. The prediction equations suggested based on the estimated values and corresponding fitting curves are given in Eq. (13-15):

$$\mu = 9.1667 \times 10^{-7} A^2 - 0.0011A + 2 \qquad (13)$$

$$\mu = -9 \times 10^{-4} H^2 + 0.0376H + 1.937 \qquad (14)$$

$$\mu = -2.1533T^2 + 2.1531T + 1.775 \qquad (15)$$

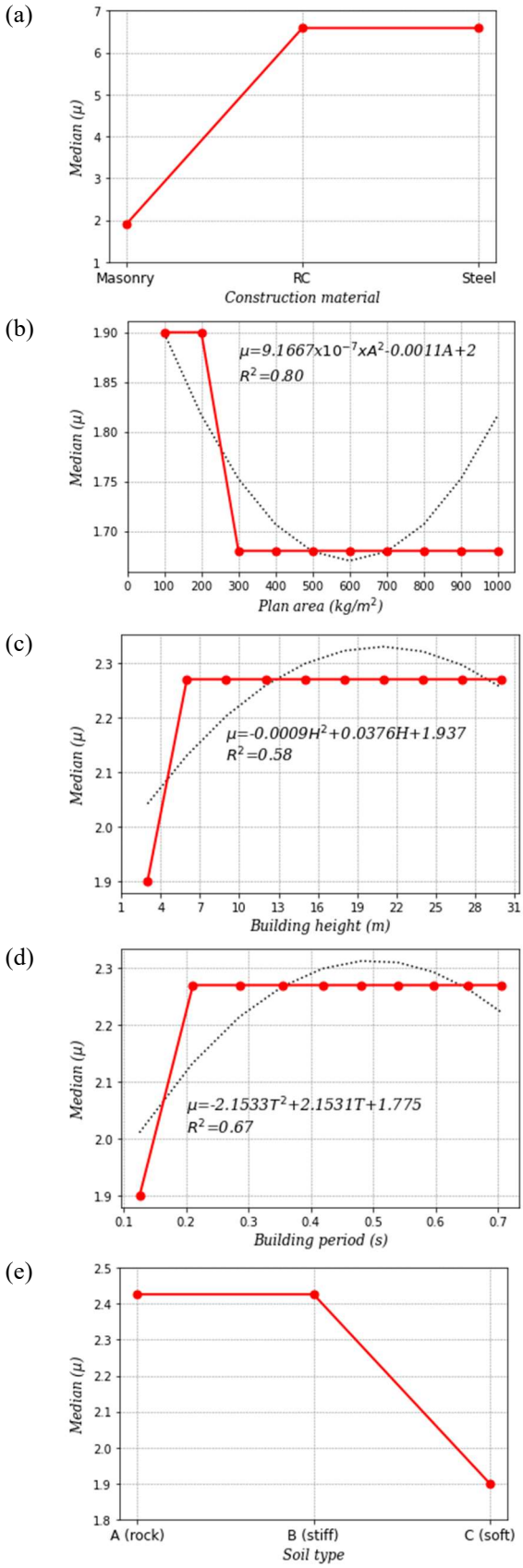


Figure 14. *Influence of building properties and soil type on median (μ).*

8. Summary and conclusion

Fragility curves are one of the crucial means which should be obtained for risk assessment of buildings in the PEER framework. The process of deriving fragility curves, however, is time-consuming and complicated. These issues might increase inaccuracy of the fragility curves. Therefore, proposing a quick and error-free alternative approach for generating fragility curves has become one of the main researchers' concerns in the field of structural and earthquake engineering. In this study, an attempt has been made to develop ML-based models for predicting fragility parameters of buildings and hence deriving fragility curves. To this end, totally 238 datasets are collected from peer-reviewed publications. They were then divided into training (85%) and testing (15%) sub-databases. Parameters which are proven to have the highest effect on fragility parameters are considered as the inputs for predicting dispersion (β) and median (μ) of fragility curves. Various ML-based prediction models namely NLR, DT, RF, KNN and ANN were developed. The predicted values were compared with the actual values and the accuracy of the models were assessed through common performance metrics and Taylor diagram. Eventually, a parametric study conducted and equations were presented for calculating fragility parameters. The main conclusions are:

- Almost all the ML-based techniques showed a high ability to learn the relationship between input and output parameters. They could be therefore considered as a quick and accurate model for estimating fragility curves instead of time-consuming and inaccurate analytical analysis.
- All the developed models for predicting β exhibited high reliability with R^2-score≥0.83. The DT model with R^2-score=0.91, however, was the most accurate model.
- Among the models proposed for estimating μ, NLR processed the highest R^2-score compared to other models and hence is introduced as the most accurate model.
- Based on the study results, a hybrid ML-based model with the combination of DT and NLR is recommended for deriving fragility curves of buildings.
- It is also worth mentioning that, unlike other studies carried out for developing models for obtaining fragility curves, the models of the present study have the following benefits: (i) generating fragility curves directly by estimating β and μ, (ii) considering significant parameters which influence fragility curves notably, and (iii) being generalized which means that they could be applied for RC, steel and masonry buildings with different heights and plan areas located on different soil types.
- The parametric study on a case study illustrated that, (i) β and μ of RC or steel buildings were higher than that of the masonry building, (ii) β and μ of the buildings located on either rock or stiff soil were higher than those of the buildings located on soft soil, (iii) by increasing building plan area, β and μ reduced first (up to height/area ratio=3.1 and 2.73%, respectively) and then remained almost constant, and (iv) increasing building height led to initial reduction and increase respectively in β and μ, while fragility curves of tall buildings remained unchanged.

References

1. Frankie, T.M., B. Gencturk, and A.S. Elnashai, *Simulation-based fragility relationships for unreinforced masonry buildings.* Journal of Structural Engineering, 2013. **139**(3): p. 400-410.
2. Özel, A.E. and E.M. Güneyisi, *Effects of eccentric steel bracing systems on seismic fragility curves of mid-rise R/C buildings: A case study.* Structural Safety, 2011. **33**(1): p. 82-95.
3. Prieto, J.A., et al., *Development of structural debris flow fragility curves (debris flow buildings resistance) using momentum flux rate as a hazard parameter.* Engineering Geology, 2018. **239**: p. 144-157.
4. Kiani, J., C. Camp, and S. Pezeshk, *On the application of machine learning techniques to derive seismic fragility curves.* Computers & Structures, 2019. **218**: p. 108-122.
5. Saouma, V.E. and M.A. Hariri-Ardebili, *Performance Based Earthquake Engineering*, in *Aging, Shaking, and Cracking of Infrastructures*. 2021, Springer. p. 517-528.
6. Lange, D., S. Devaney, and A. Usmani, *An application of the PEER performance based earthquake engineering framework to structures in fire.* Engineering Structures, 2014. **66**: p. 100-115.
7. Hwang, S.-H., et al., *Machine learning-based approaches for seismic demand and collapse of ductile reinforced concrete building frames.* Journal of Building Engineering, 2021. **34**: p. 101905.
8. Park, J., et al., *Seismic fragility analysis of low-rise unreinforced masonry structures.* Engineering Structures, 2009. **31**(1): p. 125-137.
9. Rota, M., A. Penna, and G. Magenes, *A methodology for deriving analytical fragility curves for masonry buildings based on stochastic nonlinear analyses.* Engineering Structures, 2010. **32**(5): p. 1312-1323.
10. Cardone, D., M. Rossino, and G. Gesualdi, *Estimating fragility curves of pre-70 RC frame buildings considering different performance limit states.* Soil Dynamics and Earthquake Engineering, 2018. **115**: p. 868-881.
11. Alwaeli, W., et al., *Rigorous versus less-demanding fragility relations for RC high-rise buildings.* Bulletin of Earthquake Engineering, 2020. **18**(13): p. 5885-5918.
12. Del Gaudio, C., et al., *Empirical fragility curves for masonry buildings after the 2009 L'Aquila, Italy, earthquake.* Bulletin of earthquake engineering, 2019. **17**(11): p. 6301-6330.
13. Dall'Asta, A., et al., *Influence of time-dependent seismic hazard on structural design.* Bulletin of Earthquake Engineering, 2021. **19**(6): p. 2505-2529.
14. Sandoli, A., et al., *Fragility curves for Italian URM buildings based on a hybrid method.* Bulletin of Earthquake Engineering, 2021. **19**(12): p. 4979-5013.
15. Mitropoulou, C.C. and M. Papadrakakis, *Developing fragility curves based on neural network IDA predictions.* Engineering Structures, 2011. **33**(12): p. 3409-3421.
16. Shinozuka, M., et al., *Statistical analysis of fragility curves.* Journal of engineering mechanics, 2000. **126**(12): p. 1224-1231.
17. Pejovic, J. and S. Jankovic, *Seismic fragility assessment for reinforced concrete high-rise buildings in Southern Euro-Mediterranean zone.* Bulletin of Earthquake Engineering, 2016. **14**(1): p. 185-212.
18. Aljawhari, K., et al., *Effects of ground-motion sequences on fragility and vulnerability of case-study reinforced concrete frames.* Bulletin of Earthquake Engineering, 2021. **19**(15): p. 6329-6359.
19. Donà, M., et al., *Mechanics-based fragility curves for Italian residential URM buildings.* Bulletin of Earthquake Engineering, 2021. **19**(8): p. 3099-3127.

20. Beilic, D., et al., *Seismic fragility curves of single storey RC precast structures by comparing different Italian codes.* Earthq. Struct, 2017. **12**(3): p. 359-374.
21. Cardone, D., G. Perrone, and V. Piesco, *Developing collapse fragility curves for base-isolated buildings.* Earthquake Engineering & Structural Dynamics, 2019. **48**(1): p. 78-102.
22. Faramarzi, A., A.M. Alani, and A.A. Javadi, *An EPR-based self-learning approach to material modelling.* Computers & Structures, 2014. **137**: p. 63-71.
23. Alani, A.M. and A. Faramarzi, *An evolutionary approach to modelling concrete degradation due to sulphuric acid attack.* Applied Soft Computing, 2014. **24**: p. 985-993.
24. Faramarzi, A., A.A. Javadi, and A. Ahangar-Asr, *Numerical implementation of EPR-based material models in finite element analysis.* Computers & Structures, 2013. **118**: p. 100-108.
25. Dabiri, H., K. Rahimzadeh, and A. Kheyroddin. *A comparison of machine learning- and regression-based models for predicting ductility ratio of RC beam-column joints*. in *Structures*. 2022. Elsevier.
26. Dabiri, H., et al., *Compressive strength of concrete with recycled aggregate; a machine learning-based evaluation.* Cleaner Materials, 2022. **3**: p. 100044.
27. Sun, H., H.V. Burton, and H. Huang, *Machine learning applications for building structural design and performance assessment: state-of-the-art review.* Journal of Building Engineering, 2021. **33**: p. 101816.
28. Salehi, H. and R. Burgueño, *Emerging artificial intelligence methods in structural engineering.* Engineering structures, 2018. **171**: p. 170-189.
29. Jia, D.-W. and Z.-Y. Wu, *Seismic fragility analysis of RC frame-shear wall structure under multidimensional performance limit state based on ensemble neural network.* Engineering Structures, 2021. **246**: p. 112975.
30. Kirçil, M.S. and Z. Polat, *Fragility analysis of mid-rise R/C frame buildings.* Engineering Structures, 2006. **28**(9): p. 1335-1345.
31. Saruddin, S.N.A. and F.M. Nazri, *Fragility curves for low-and mid-rise buildings in Malaysia.* Procedia Engineering, 2015. **125**: p. 873-878.
32. Zucconi, M., R. Ferlito, and L. Sorrentino, *Simplified survey form of unreinforced masonry buildings calibrated on data from the 2009 L'Aquila earthquake.* Bulletin of Earthquake Engineering, 2018. **16**(7): p. 2877-2911.
33. Karafagka, S., S. Fotopoulou, and D. Pitilakis, *Fragility curves of non-ductile RC frame buildings on saturated soils including liquefaction effects and soil–structure interaction.* Bulletin of Earthquake Engineering, 2021. **19**(15): p. 6443-6468.
34. Suzuki, A. and I. Iervolino, *Seismic fragility of code-conforming Italian buildings based on SDoF approximation.* Journal of Earthquake Engineering, 2021. **25**(14): p. 2873-2907.
35. Abo-El-Ezz, A., M.-J. Nollet, and M. Nastev, *Seismic fragility assessment of low-rise stone masonry buildings.* Earthquake Engineering and Engineering Vibration, 2013. **12**(1): p. 87-97.
36. Kumar, P. and A. Samanta. *Seismic fragility assessment of existing reinforced concrete buildings in Patna, India*. in *Structures*. 2020. Elsevier.
37. Nettleton, D., *Commercial data mining: processing, analysis and modeling for predictive analytics projects*. 2014: Elsevier.
38. Berman, J.J., *Principles and practice of big data: preparing, sharing, and analyzing complex information*. 2018: Academic Press.
39. Profillidis, V. and G. Botzoris, *Chapter 5—statistical methods for transport demand modeling.* Modeling of Transport Demand; Profillidis, VA, Botzoris, GN, Eds, 2019.

40. Berman, J.J., *Data simplification: taming information with open source tools*. 2016: Morgan Kaufmann.
41. Huang, H.-H., et al., *Nonlinear regression analysis.* International encyclopedia of education, 2010: p. 339-346.
42. Wood, D.A. and J. Cai, *Sustainable Geoscience for Natural Gas SubSurface Systems*. 2021: Elsevier.
43. Hasanuzzaman, M. and N. Abd Rahim, *Energy for sustainable development: demand, supply, conversion and management*. 2019: Academic Press.
44. Sanft, R. and A. Walter, *Exploring mathematical modeling in biology through case studies and experimental activities*. 2020: Academic Press.
45. Liu, S., et al., *Computational and statistical methods for analysing big data with applications*. 2015: Academic Press.
46. Bellini, T., *IFRS 9 and CECL Credit Risk Modelling and Validation: A Practical Guide with Examples Worked in R and SAS*. 2019: Academic Press.
47. Shobha, G. and S. Rangaswamy, *Chapter 8-Machine Learning Handbook of Statistics*. 2018, Elsevier.
48. Kotu, V. and B. Deshpande, *Data science: concepts and practice*. 2018: Morgan Kaufmann.
49. Paul, S. and D. Bhatia, *Smart Healthcare for Disease Diagnosis and Prevention*. 2020: Academic Press.
50. Mao, W. and F. Wang, *New advances in intelligence and security informatics*. 2012: Academic Press.
51. Williams, B., et al., *Data-Driven Model Development for Cardiomyocyte Production Experimental Failure Prediction*, in *Computer Aided Chemical Engineering*. 2020, Elsevier. p. 1639-1644.
52. Chanal, D., et al., *Online Diagnosis of PEM Fuel Cell by Fuzzy C-Means Clustering.* 2021.
53. Subasi, A., et al., *Human activity recognition using machine learning methods in a smart healthcare environment*, in *Innovation in health informatics*. 2020, Elsevier. p. 123-144.
54. Richman, J.S., *Multivariate neighborhood sample entropy: a method for data reduction and prediction of complex data*, in *Methods in enzymology*. 2011, Elsevier. p. 397-408.
55. Malekian, A. and N. Chitsaz, *Concepts, procedures, and applications of artificial neural network models in streamflow forecasting*, in *Advances in Streamflow Forecasting*. 2021, Elsevier. p. 115-147.
56. Sadiq, R., M.J. Rodriguez, and H.R. Mian, *Empirical models to predict disinfection by-products (DBPs) in drinking water: an updated review.* 2019.
57. Walczak, S. and N. Cerpa, *Heuristic principles for the design of artificial neural networks.* Information and software technology, 1999. **41**(2): p. 107-117.
58. Wadoux, A.M.-C., D.J. Walvoort, and D.J. Brus, *An integrated approach for the evaluation of quantitative soil maps through Taylor and solar diagrams.* Geoderma, 2022. **405**: p. 115332.
59. Shariati, M., et al., *A novel hybrid extreme learning machine–grey wolf optimizer (ELM-GWO) model to predict compressive strength of concrete with partial replacements for cement.* Engineering with Computers, 2020: p. 1-23.
60. *Eurocode 8: Design of structures for earthquake resistance - Part 1: General rules, seismic actions and rules for buildings, European Standard EN 1998-1:2004, 2004.*