Results in Engineering 21 (2024) 101750

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Utilizing advanced machine learning approaches to assess the seismic fragility of non-engineered masonry structures

Ehsan Harirchian [a,*], Seyed Ehsan Aghakouchaki Hosseini [b], Viviana Novelli [c], Tom Lahmer [a], Shahla Rasulzade [d]

[a] Institute of Structural Mechanics (ISM), Bauhaus-Universität Weimar, 99423 Weimar, Germany
[b] Department of Built Environment Engineering, Faculty of Design and Creative Technologies, School of Future, Environments, Auckland University of Technology (AUT), 1010 Auckland, New Zealand
[c] School of Engineering, Cardiff University, Queen's Buildings, The Parade, Cardiff CF24 3AA, UK
[d] Leibniz-Institute for Research and Information in Education, Educational Computer Science (DIPF), 60323 Frankfurt Am Main, Germany

ARTICLE INFO

Keywords:
Machine learning
Seismic vulnerability
Fragility curves
Masonry structures

ABSTRACT

Seismic fragility assessment provides a substantial tool for assessing the seismic resilience of these buildings. However, using traditional numerical methods to derive fragility curves poses significant challenges. These methods often overlook the diverse range of buildings found in different regions, as they rely on standardized assumptions and parameters. Consequently, they may not accurately capture the seismic response of various building types. Alternatively, extensive data collection becomes essential to address this knowledge gap by understanding local construction techniques and identifying the relevant parameters. This data is crucial for developing reliable analytical approaches that can accurately derive fragility curves. To overcome these challenges, this research employs four Machine Learning (ML) techniques, namely Support Vector Regression (SVR), Stochastic Gradient Descent (SGD), Random Forest (RF), and Linear Regression (LR), to derive fragility curves for probability of collapse in terms of Peak Ground Acceleration (PGA). To achieve the research objective, a comprehensive input/output dataset consisting of on-site data collected from 646 masonry walls in Malawi is used. Adopted ML models are trained and tested using the entire dataset and then again using only the most highly correlated features. The study includes a comparative analysis of the efficiency and accuracy of each ML approach and the influence of the data used in the analyses. Random Forest (RF) technique emerges as the most efficient ML approach for deriving fragility curves for the surveyed dataset in terms of achieved lowest values for evaluation metrics of the ML methods. This technique scored the lowest Mean Absolute Percentage Error (MAPE) of 16.8 %, and the lowest Root Mean Square Error (RMSE) of 0.0547. These results highlight the potential of ML techniques, particularly RF, in derivation of fragility curves with proper levels of accuracy.

1. Introduction

Seismic activity poses a significant threat to the stability of buildings worldwide, particularly in developing countries where the predominant building stock consists mostly of non-engineered structures. These structures are often constructed without proper engineering design, lack quality control measures, and utilize locally sourced materials and traditional construction techniques that do not meet seismic-resistant standards. Furthermore, these techniques are often modified and adapted to suit local resources, further compromising their resilience to seismic events [1–3]. In East Africa, for instance, non-engineered structures are prevalently built with clay bricks and poor mortar, which is highly vulnerable to seismic events due to its high likelihood of failing out of plane. As a result, these buildings pose significant risks to human lives, property, and essential infrastructure [4,5]. Notable earthquakes in East Africa such as the devastating Karonga Earthquake struck Malawi on December 19, 2009, measuring 6.2 in magnitude, and resulted in widespread damage and loss of life in Karonga. The 2004 Lake Victoria Earthquake, measuring 6.8 in magnitude near the Tanzania-Uganda border caused considerable damage and casualties in

* Corresponding author.
E-mail addresses: ehsan.harirchian@uni-weimar.com (E. Harirchian), ehsan.hosseini@aut.ac.nz (S.E. Aghakouchaki Hosseini), NovelliV@cardiff.ac.uk (V. Novelli), tom.lahmer@uni-weimar.de (T. Lahmer), s.rasulzade@dipf.de (S. Rasulzade).

https://doi.org/10.1016/j.rineng.2024.101750
Received 15 November 2023; Received in revised form 19 December 2023; Accepted 2 January 2024
Available online 5 January 2024

Bukoba (Tanzania) and Kampala (Uganda). Additionally, a magnitude 6.4 earthquake hit Arusha in northern Tanzania on July 14, 2001, causing significant damage to buildings and infrastructure. The Great Rift Valley experienced a powerful earthquake on February 20, 1994, measuring 6.7 in magnitude near the Rwanda-Democratic Republic of Congo border, resulting in widespread destruction and loss of life [6–8]. These events highlight the urgent need for enhanced seismic resilience and disaster preparedness measures.

In East Africa, as well as other seismic prone countries across the world, the assessment of risks plays a crucial role in safeguarding human lives and economies [9–11]. Various approaches to disaster risk reduction (DRR) have been developed and refined over the years [11–13]. DRR encompasses a wide range of strategies, policies, and measures aimed at mitigating the impacts and vulnerabilities associated with both natural and human-induced disasters. Within the realm of DRR analysis, one area of significant concern is the multitude of uncertainties, including uncertainties in structural modelling and variations in damage states. To effectively address these uncertainties, one needs incorporating fragility curves into the DRR framework, which enables a more comprehensive understanding of risks while facilitating targeted efforts to reduce them. Probabilistic and analytical approaches such as Bayesian Inference, Performance-Based Design (PBD), Capacity Spectrum Method (CSM), Incremental Dynamic Analysis (IDA), and Response Spectrum Method (RSM) have been widely used to derive fragility curves, providing valuable insights into the structural behaviour under seismic loading.

The primary objective of this study is to use the four chosen ML-based techniques to derive fragility curves based on the calculated PGA values corresponding to collapse state of surveyed buildings and perform a comparative analysis to evaluate the efficacy of each approach and the influence of the data employed. This eliminates the need for time-consuming numerical modelling to assess building capacity under seismic loads, enabling the rapid generation of fragility curves by defining the inputs. Furthermore, the study identifies the key parameters that significantly influence fragility curves by establishing correlations between the input and output of the available dataset. Consequently, these ML-based techniques can generate fragility curves applicable to a wide range of non-engineered constructions with similar characteristics. To accomplish the research objective, a comprehensive database is compiled and four ML-based methods, such as Support Vector Regression (SVR), Stochastic Gradient Descent (SGD), Random Forest (RF), and Linear Regression (LR), are employed to develop fragility curve models. The accuracy of these models is evaluated using performance metrics, ultimately identifying the most precise model. Additionally, a parametric study is conducted to analyse the influence of input variables on output parameters, enabling the formulation of prediction equations. The four ML estimators for the purpose of data analysis in this research have been employed using the Scikit-learn package in Python language.

2. Fragility functions

Fragility curves play a crucial role in quantifying the vulnerability of structures to seismic events and assessing their associated risks. Researchers have offered several definitions to clarify the concept of fragility curves [14]. In simple terms, fragility refers to the probability of a structure reaching or surpassing a specified Damage Measure (DM) given a particular earthquake Intensity Measure (IM). An example of a DM could be the extent of structural deformation, such as the displacement or drift experienced by a building during an earthquake. It quantifies the level of damage that a structure may undergo based on specific criteria. On the other hand, an IM represents a parameter that characterizes the intensity or severity of an earthquake. It can be a physical quantity, such as Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV), or Spectral Acceleration (SA). These measures reflect the ground motion experienced during an earthquake and are used as inputs to estimate the probability of damage to structures. From a mathematical perspective, fragility curves can be represented as the probability of damage as follows,

$$P[DM|IM] = \varphi \frac{\ln(IM)\text{-}\ln(\eta)}{\beta} \tag{1}$$

Where the terms can be defined as follows.

- φ: The standard normal distribution function, which calculates the probability of a value occurring in a standard normal distribution.
- η and β: equivalent median and logarithmic standard deviation (i.e., dispersion) of the fragility curves in relation to IM

In Eq. (1), the standard normal distribution function (φ) is applied to obtain the conditional probability of the damage state.

2.1. Vulnerability assessment methods

Over the past five decades, significant progress has been achieved in the development of a wide array of methodologies for deriving fragility curves, aimed at assessing the vulnerability of existing building stock [15–19]. Broadly classified into two main categories, namely empirical and analytical methods, seismic vulnerability analysis, particularly for masonry structures, has traditionally relied on empirical methods. Empirical techniques involve the statistical processing of data, using real damage observations from past earthquakes to calibrate and estimate the vulnerability of distinct classes of buildings. This includes methods that classify buildings based on their damage propensity and inspection and rating methods that assign scores to significant vulnerability components [20–23]. Empirical methodologies, commonly utilized in seismic vulnerability assessment, are subject to intrinsic constraints that influence their adaptability and credibility [24–26]. The dependence on historical data presents challenges in areas characterized by limited seismic activity or incomplete records. The inherent site-specific nature of fragility curves derived through empirical means complicates their extrapolation to various geographical contexts and diverse building typologies. Assumptions of stationarity, a fundamental premise in these methods, may encounter inadequacies in dynamic seismic environments, thereby potentially compromising predictive accuracy. Additionally, the sensitivity to data quality and the intricacy of integrating nonlinear effects serve as additional impediments to the efficacy of empirical approaches in capturing the complexities inherent in seismic vulnerability assessment. Empirical methodologies, commonly utilized in seismic vulnerability assessment, are subject to intrinsic constraints that influence their adaptability and credibility. The Performance-Based Design (PBD) approach stands out as a prime example of the shift towards analytical methods in seismic vulnerability assessment. By leveraging probabilistic methods, PBD provides a sophisticated framework for evaluating the performance of structures under seismic events. Fragility curves are derived by considering the uncertainties associated with ground motions, structural properties, and component response. PBD incorporates statistical methods, such as Monte Carlo simulations, to quantify the probabilities of exceeding predefined performance levels. This approach allows for the integration of advanced structural modelling techniques, enabling more accurate and detailed fragility curve estimations [27–29]. In a parallel vein to the PBD approach, Bayesian inference emerges as another analytical method that addresses the limitations of relying solely on historical data. While PBD utilizes probabilistic methods and advanced structural modelling to assess structural performance, Bayesian inference takes a statistical approach to continually refine fragility curves based on observed data. It integrates prior knowledge with the likelihood of observed data, estimating the posterior distribution of fragility parameters. This approach, akin to PBD, acknowledges and quantifies uncertainties but uniquely does so by iteratively refining fragility curves as more data becomes available. Bayesian methods offer a flexible framework, allowing the incorporation of diverse information sources, including expert judgment and experimental data. By bridging the gap between historical observations and evolving data, Bayesian inference ensures a dynamic and adaptive approach to seismic vulnerability assessment, reinforcing the analytical shift towards more comprehensive and refined methodologies [30–32]. Expanding on the analytical landscape, the Capacity Spectrum Method (CSM) serves as another noteworthy approach in seismic vulnerability assessment. In a manner similar to PBD and Bayesian inference, CSM offers a comprehensive analytical framework. CSM is particularly effective for multi-degree-of-freedom (MDOF) structures, where it involves the development of capacity curves depicting the relationship between structural capacity and the demand imposed by seismic ground motions. Fragility curves are then established by comparing these capacity curves with demand curves, considering uncertainties in ground motion intensity, structural properties, and response. This method not only captures structural behaviour under seismic loading but also contributes to a probabilistic assessment of performance. By integrating advanced modelling techniques, CSM aligns with the analytical shift, emphasizing a thorough understanding of structural dynamics and embracing uncertainties for more accurate and detailed fragility curve estimations [33,34]. Conversely, the Response Spectrum Method (RSM) presents a distinct yet widely embraced analytical approach for the derivation of fragility curves. Relying on response spectra, RSM furnishes a graphical representation of the maximum structural response across various periods of vibration, facilitating the estimation of fragility curves. Particularly advantageous when confronted with limited data or necessitating swift fragility estimations, RSM's simplicity is evident. However, this simplicity comes with a caveat, potentially leading to an oversimplification of structural behaviour and compromising the accuracy of fragility curve estimations [38,39]. Advancing further within the analytical spectrum, Incremental Dynamic Analysis (IDA) emerges as a paramount technique, distinguished by its prowess in capturing the nonlinear behaviour inherent in structures subjected to seismic loading. In this sophisticated approach, the structure undergoes exposure to a series of ground motion records, progressively intensifying in magnitude. Fragility curves materialize through a meticulous examination of structural responses at each intensity level, enabling the quantification of the probability of surpassing predefined damage states. IDA stands unparalleled in its ability to account for the complete spectrum of structural behaviour, offering a nuanced and comprehensive characterization of fragility curves. The meticulous nature of IDA, capturing nonlinearities with precision, positions it as an exemplary method, particularly lauded for its ability to provide detailed insights into structural vulnerability under seismic loading conditions [35–37]. The reviewed approaches above present their own strengths and limitations, and the choice of the method should be based on the specific requirements of the analysis, including available data, and desired level of accuracy.

2.2. Fitting methods for fragility curves

Various methods are employed for fitting fragility curves, each presenting unique strengths and considerations tailored to different analytical needs. Logistic Regression, a classical statistical approach, stands out for its simplicity and interpretability. By modelling the probability of surpassing a damage state based on predictor variables such as ground motion intensity, Logistic Regression provides a transparent representation of the relationship between variables. Its ease of implementation and comprehensibility make it a valuable choice in scenarios where a straightforward interpretation of the fragility curve is essential [40,41]. On the other hand, Maximum Likelihood Estimation (MLE) offers versatility by estimating model parameters through the maximization of the likelihood of observed data. Particularly advantageous for handling larger datasets, MLE ensures statistically robust parameter estimates, contributing to the reliability of the resulting fragility curves [42,43]. Non-linear Regression becomes crucial in situations where the interplay between variables exhibits nonlinearity, allowing for a more precise representation of intricate fragility patterns that might be oversimplified by linear models. It accommodates complexities inherent in the relationship between ground motion intensity and structural response, contributing to a more accurate portrayal of vulnerability [44,45]. While the discussion on ML techniques is reserved for the next paragraph, it is important to acknowledge that these modern approaches provide a powerful alternative. Capable of capturing intricate patterns and nonlinearities in the data, ML techniques enhance the predictive capacity of fragility curves, albeit with considerations related to model complexity and interpretability. The choice of fitting method should be a deliberate decision, considering the specific characteristics of the dataset, the inherent vulnerabilities of the structures under consideration, and the ultimate goals of the fragility analysis. Aligning the selected approach with these considerations ensures a tailored and effective application of fragility curve fitting methods.

2.3. Machine leaning for derivation of fragility curves

ML techniques have emerged as a promising approach for effectively deriving fragility curves and assessing the vulnerability of buildings to seismic events. Several studies have explored the application of ML algorithms in rapid seismic assessment and damage prediction of buildings [28–31,46]. Damage prediction for historic structures has been investigated using Support Vector Regression (SVR), Naive Bayes classifiers with Gaussian distribution, and artificial neural networks (ANN) [47,48]. Techniques such as LR and K-nearest neighbors (KNN) have been utilized for predicting damage in masonry walls [34], and a hybrid ANN-Genetic Algorithm has been employed for the vulnerability assessment of existing buildings [49].

The advent of image analysis, web-based, and smartphone applications has facilitated the adoption of cutting-edge methods for vulnerability analysis [35]. For instance, techniques such as ML-EHSAPP [50], EZRVS [51], VULMA [52], IRAEHSA [53], or RASDA [54] have been developed as robust methods for vulnerability analysis of existing buildings and operate as efficient ML-based applications. The ability of ML methods to handle complex and diverse data from various sources makes them highly suitable for accurately and reliably generating fragility curves for building stocks. This is particularly valuable in regions with limited resources and expertise, where ML can significantly enhance seismic risk mitigation strategies [37,38]. Moreover, some methods are developed to derive seismic vulnerability of existing buildings by using mechanical models derived from data obtained by the transfer learning technique [55] or user-reported data and modern Internet of Things (IoT) [56].

3. Methodology

To implement the selected ML techniques for derivation of fragility curves for building collapse in this study, the following steps have been performed.

1. Selection of input data: This is the initial step in creating an ML model. The independent variables or properties serve as the input data. The ML model effectively captures the relationship between the input data characteristics and the corresponding outcomes, demonstrating how the properties of the input data can impact the results.
2. Pre-processing of dataset: Numerical, categorical, and ordinal data types are the three main forms of statistical data. However, machine learning models are primarily designed to handle numerical data. To have the model operate properly, categorical and ordinal features of data should be transformed into numerical features. This conversion allows the model to effectively process and analyse the data, enabling it to make accurate predictions and derive meaningful insights.

3. Parameter selection: Multi-parametric issues can be solved with ML algorithms. The selection criteria for parameters depends on the field of research. In the context of this research and considering its objectives, specific structural parameters that have a significant impact on the building collapse in terms of PGA were meticulously selected based on the data obtained during the field survey. These carefully chosen parameters play a crucial role in understanding and assessing the behaviour and performance of buildings when subjected to seismic conditions.
4. Splitting dataset: The dataset is split into training and testing subsets. The training subset is used to build the predictive models, while the testing subset is used to evaluate their performance. Each observation in the dataset includes predictors (independent variables) and corresponding class labels. The predictors represent consistent and available independent variables. In this research, the focus is on the PGA thresholds for building collapse. These thresholds are known in the training set, but undefined in the testing set, as they are the target values to be predicted by the trained models.
5. Model selection: This is the method of selecting a model for the training set.
6. Model performance: The purpose of this step is to assess the model's performance.
7. Model utilization: Unknown circumstances can also be analysed using an appropriate ML model. In such cases, many factors influence the model performance including the quantity and quality of input data, feature selection, and the frequency of outliers.

Fig. 1 depicts a schematic of the workflow for the applied ML techniques. During the learning process, the ML models acquire expertise by employing pre-defined algorithms, primarily for regression tasks [57]. In supervised learning, there are several methods available for multi-class categorization [58]. In this study, the selection of the most suitable regression algorithm involved the initial training and testing of the ML techniques using all features in the input datasets. Subsequently, their performance was evaluated and compared. Furthermore, the correlation between input variables and output was captured, and the top 10 most relevant features were identified. The ML models were then re-trained, re-tested, and re-evaluated using only these selected features.

3.1. Dataset

The dataset used in this study is obtained from the data collected and processed over a three-year period as part of the PREPARE project.[1] The primary objective of the PREPARE project is to improve the resilience and preparedness of East African countries in the face of seismic events. The initial focus of the project was on Malawi, which is considered representative of the local construction practices and characteristics within the region. Kloukinas et al. [59] have emphasized the limitations of relying exclusively on international databases such as WHE-PAGER [60] which provides fragility curves based on data that may not accurately reflect the specific conditions and characteristics of buildings in the country being examined. Depending solely on these international databases may result in the formulation of fragility curves that do not truly capture the attributes of buildings in the given area. In light of these limitations, Novelli et al. [61] carried out a comprehensive building survey in the central and southern regions of Malawi with the objective of understanding local construction techniques and collecting geometric and structural parameters to develop numerical modelling to derive fragility curves. The on-site inspection involved a total of 323 houses, resulting in the examination of 646 walls. Two walls were inspected for each building. The surveyed houses were in Salima, Blantyre, Lifidzi, and Golomoti, covering a range of areas within the regions.

The survey form comprises a single page dedicated to collecting data per wall. On-site data collection involves direct observations supplemented by additional detailed information derived from measurements and sketches. The efficiency of this process is highlighted by the possibility to swiftly complete the form, requiring a maximum of 10 min. Notably, the form has been tailored to meet the unique requirements of efficiently gathering data for houses in Malawi. Originating from existing templates [62], the form underwent modifications to suit the specific context. The data collection process entailed manual recording, utilizing a printed version of the survey form. Following this initial phase, a thorough electronic verification and reporting process was undertaken to establish the comprehensive database. This database encompasses the input parameters discussed in Section 3.1.1, providing a detailed and accurate representation of the collected data.

3.1.1. Input parameters

The data collected on-site, which forms the initial dataset for building ML techniques to derive fragility curves, consists of 21 features gathered for each individual wall inspected per building. These features offer valuable insights into the structural characteristics and attributes of the walls. The following is a summary of the features included in the dataset.

1. **Structural fundamental period (sec)**: This feature represents the fundamental period of vibration of the structure, which is an important parameter in seismic analysis and design.
2. **RIGHT Connection** and **3. LEFT Connection (good** or **bad)**: These features describe the type of connection on the left and right sides of the wall. Connection quality can have a significant impact on the overall structural integrity.
4. **Total area opening (m^2)**: This feature quantifies the total area of openings such as windows and doors present in the wall. Openings can affect the structural behaviour and vulnerability of the building.
5. **Wall thickness (mm)**: This feature affects the overall stability and load-bearing capacity of the wall.
6. **Wall Length (m)** and **7. Wall Height (m)**: These features indicate the overall length and height of the wall, providing information about the size and extent of the building's external wall.
8. **Masonry type (unfired** and **fired** bricks), **9. brick height (mm), 10. brick length (mm)**, and **11. brick staggering (mm)**: These features include the type of masonry and dimensions of the masonry blocks used in the construction, such as height, length, and the overlapping arrangement of bricks. These dimensions contribute to the structural behaviour and strength of the wall.
12. **Mortar type (concrete** or **mud)**: This feature specifies the type of mortar used in the masonry construction. The mortar properties play a crucial role in the overall structural performance.
13. **Length of the wall perpendicular to the inspected wall (m)**: This feature represents the length of the wall perpendicular to the inspected wall, providing additional geometric information about the building.
14. **Number of internal walls perpendicular to the inspected wall (number)**, **15. Number of internal walls parallel to the inspected wall (number),** and **16. Number of internal walls perpendicular to the back wall and parallel to the inspected wall (number)**: These features provide an insight into the building's overall layout and structural configuration.
17. **Roof type (thatched roof** and **metallic sheet)**: This feature describes the type of roof used in the building.
18. **Roof Orientation** (**Parallel** or **Orthogonal**): This feature indicates the orientation of the roof respect to inspected wall.
19. **Gable (yes** or **no)**: Presence of gable. This feature denotes whether the wall has a gable, which is a triangular portion of a wall between the edges of intersecting roof pitches.

[1] https://www.bristol.ac.uk/engineering/research/international-development/natural-disasters/prepare-africa/.

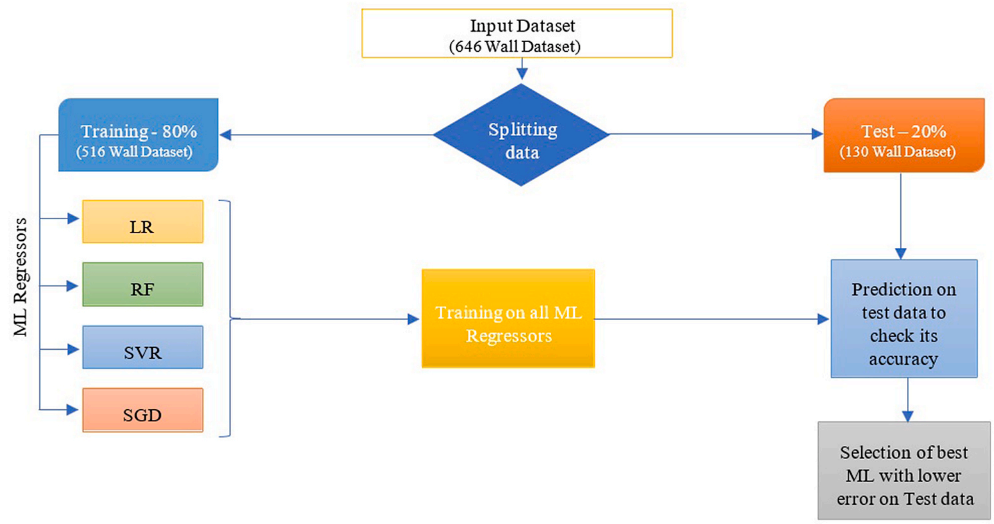


Fig. 1. General workflow of using supervised ML regressors in this study.

20. **Gable height (m):** Gable height. This feature represents the height of the gable, providing additional information about its structural characteristics.
21. **Spandrel height (m):** This feature describes the height of the spandrel, which is the space between the top of a window in one story and the sill of the window directly above it.

These features collectively provide a comprehensive set of parameters for analysing and understanding the structural attributes of the walls and the buildings they belong to. They offer valuable insights into the geometric, material, and connection properties that influence the seismic behaviour and vulnerability of the structures. The set of data was collected to derive fragility curves following the methodology detailed in Ngoma et al. [63].

Histograms shown in Fig. 2 (a) to (u) show the overall variation and distribution of all input parameters for the entire dataset of surveyed buildings in Malawi. The *y*-axis indicates the number of counts (buildings) in each histogram, while the *x*-axis represents the input parameter (inspected feature) for the surveyed buildings.

3.1.2. *Output parameter as PGA*

In addition to the building survey, an experimental campaign was carried out to characterize the mechanical properties of the local construction materials [46,64]. This campaign involved testing samples of materials used in the construction of the surveyed houses. By analysing the mechanical properties of these materials, the researchers could gain a better understanding of their strength and other relevant factors that contribute to the overall performance and vulnerability of the structures under seismic loads. The data obtained from the on-site inspections and the experimental characterization of material properties were used to Static Push-Over (SPO) curves for all 646 walls, building upon the previous work of D'Ayala [65].

The singular SPO curve (Fig. 3 (a)) serves as an illustration of the structural response exhibited by each examined wall at different levels of damage, ranging from minor structural damage to complete collapse. Following this, the SPO curve for each of the 646 walls underwent individual transformation into 646 Incremental Dynamic Analysis (IDA) curves (Fig. 3 (b)) using the SPo2IDA tool [66]. The resultant IDA curves were then utilized to formulate 646 seismic fragility curves. This process resulted in a unique IDA curve for each wall.

Subsequently, the resulting IDA curves were employed to formulate 646 seismic fragility curves. These curves provide a quantitative assessment of the probability of a building reaching or exceeding a specific damage state, given a particular level of PGA. For details on deriving fragility curves for different damage limit states, please refer to the paper by Novelli et al. [61]. Here, we focus on simplifying and presenting the methodology solely for deriving fragility curves for an individual inspected wall at the collapse state. To achieve this, Eq. (1) was reconfigured into Eq. (2). In this equation, DM is defined as Collapse (C), and IM represents the PGA for the single wall at collapse, where 'n' denotes the number of analysed walls:

$$\mathrm{P}[\mathrm{C}|\mathrm{PGA}(\mathrm{g})] = \frac{1}{\mathrm{n}}\sum_{1=1}^{\mathrm{n}} \mathrm{P_i}[\mathrm{C}|\mathrm{PGA}(\mathrm{g})] = \times\frac{1}{\mathrm{n}}\sum_{1=1}^{\mathrm{n}} \varphi \frac{\ln\left(\mathrm{PGA}(\mathrm{g})_{\mathrm{i}}\right)\text{-}\ln(\eta)}{\beta} \tag{2}$$

The final fragility curves (Fig. 3 (c)) are obtained as the mean fragility across the curves derived for the individually analysed walls. This suggests that the individual fragility curves at collapse represent the mean of the 646 fragility curves obtained for the walls under inspection.

3.1.3. *Data pre-processing*

Data pre-processing is a significant step in the data analysis as the quality of a dataset directly influences the efficiency of the ML algorithm due to its learning process. In this research, the data pre-processing is performed by standardization via StandardScaler which standardize features by removing the mean and scaling them to unit variance. The standard score (z) of a sample x is calculated as:

$$z = \frac{x - u}{s} \tag{3}$$

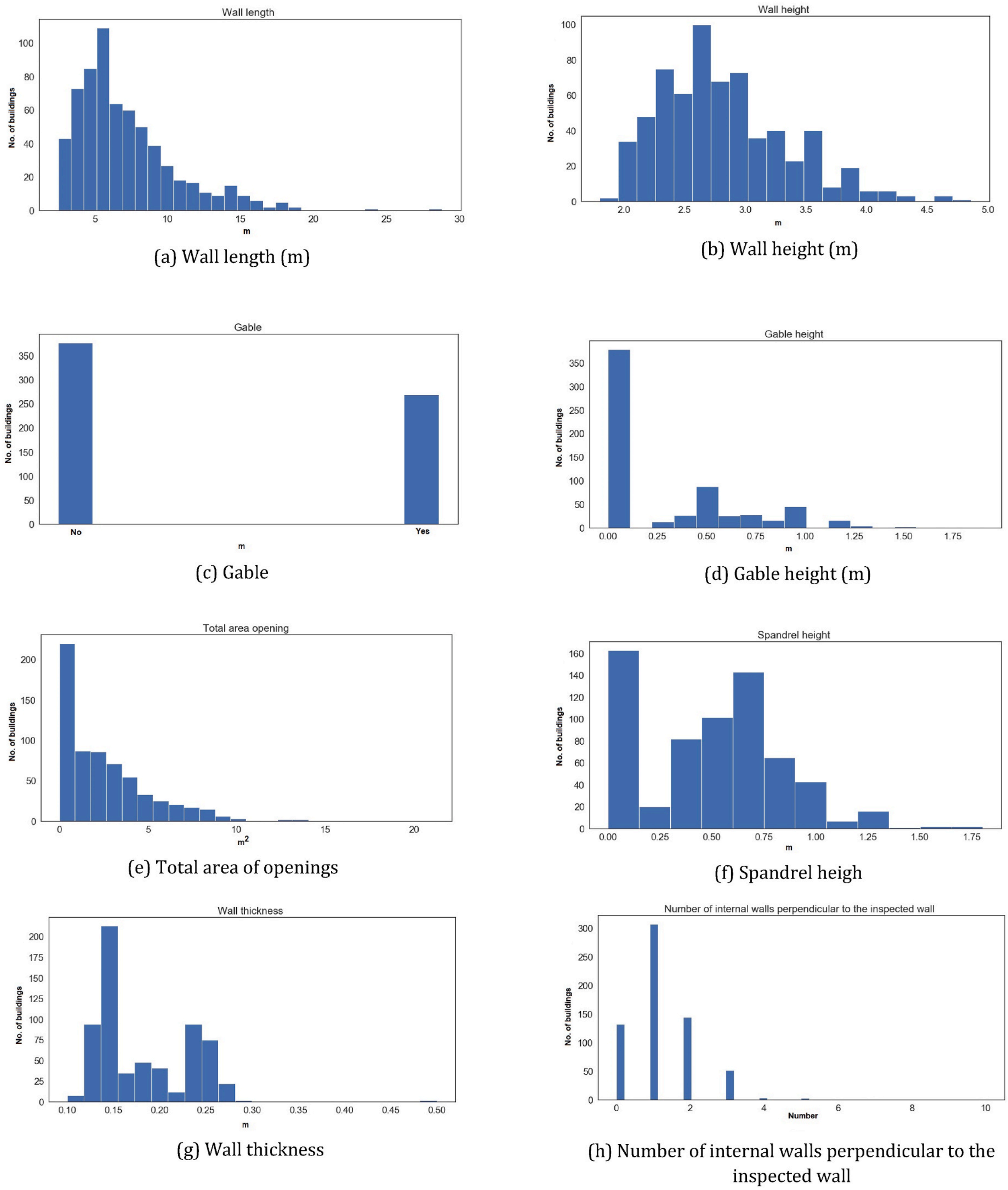
where u is the mean of the training samples, and s is the standard deviation of the training samples. Centring and scaling happen independently on each feature by computing the relevant statistics on the samples in the training set.

3.1.4. *Splitting of dataset*

The dataset has been divided into two subsets, namely training and test sets. It is common practice to allocate 80 % of the data for training purposes and reserve 20 % for the test subset. Following this standard, the same division was applied in this research, resulting in 516 data points for training and 130 data points for testing. The training set comprises data with known outputs, allowing the model to learn and adjust its parameters. On the other hand, the test subset is used to evaluate the performance of the model's predictions.

4. Model implementation

Four ML estimators have been applied using Scikit-learn package and



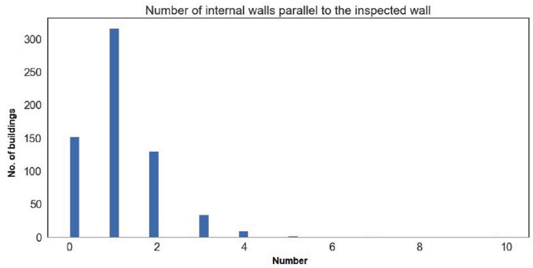
(a) Wall length (m)
(b) Wall height (m)
(c) Gable
(d) Gable height (m)
(e) Total area of openings
(f) Spandrel heigh
(g) Wall thickness
(h) Number of internal walls perpendicular to the inspected wall

Fig. 2. Histograms representing the distribution of data collected from inspected buildings (a) Wall length, (b) Wall height, (c) Gable, (d) Gable height, (e) Total area of openings, (f) Spandrel height, (g) Wall thickness, (h) Number of internal walls perpendicular to the inspected wall, (i) Number of internal walls parallel to the inspected wall, (j) Number of internal walls perpendicular to the back wall and parallel to the inspected wall, (k) Length of the wall perpendicular to the inspected wall, (m) Roof type, (l) Roof orientation, (n) Masonry type, (o) Mortar type, (p) Brick length, (q) Brick height, (r) Brick staggering, (s) Left connection, (t) Right connection, and (u) Fundamental period of the structure, for the entire data of Malawi.

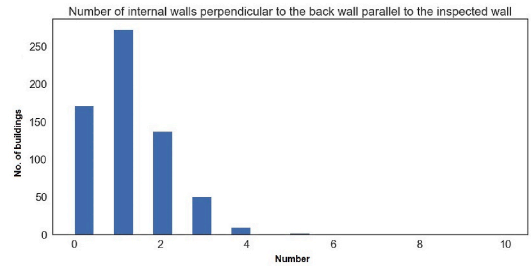
Python language. The fundamental libraries required for analyses include Scikit-learn, Pandas, NumPy, and Matplotlib. Following sections describe the selected ML estimators in brief.

4.1. Support Vector Regression (SVR)

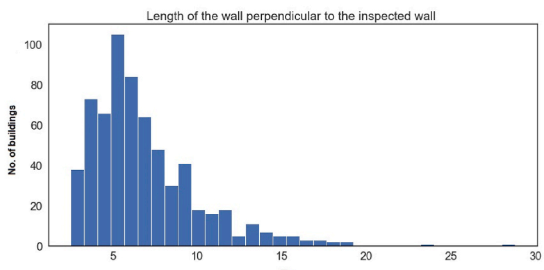
SVR is a powerful and versatile ML technique used for regression analysis [67]. It works by finding a hyperplane that best fits the data points in a higher-dimensional space. The primary goal of SVR is to find a function that maps input data to continuous output values while minimizing the error between the predicted and actual values. SVR is useful for handling nonlinear and complex relationships between the input and output variables by using a kernel trick to map the data into a higher-dimensional feature space. This technique is especially useful when dealing with datasets that have high noise levels and a small number of samples.



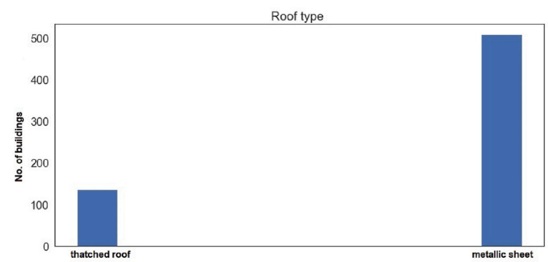
(i) Number of internal walls parallel to the inspected wall



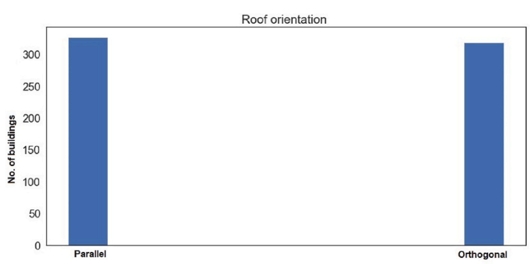
(j) Number of internal walls perpendicular to the back wall and parallel to the inspected wall



(k) Length of the wall perpendicular to the inspected wall

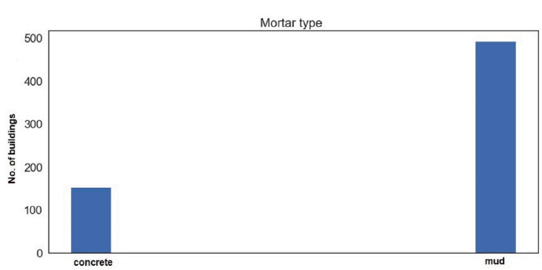


(l) Roof type

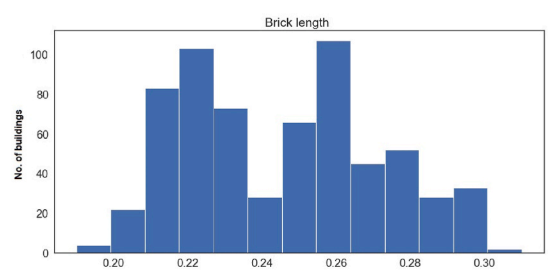


(m) Roof orientation

(n) Masonry type



(o) Mortar type



(p) Brick's length

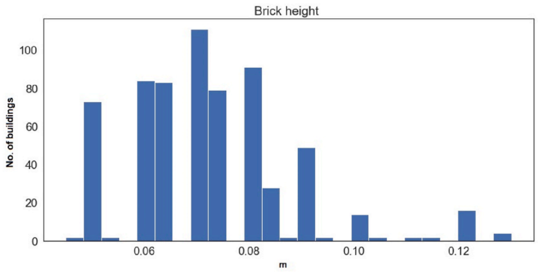
Fig. 2. (*continued*).

4.2. *Linear Regression (LR)*

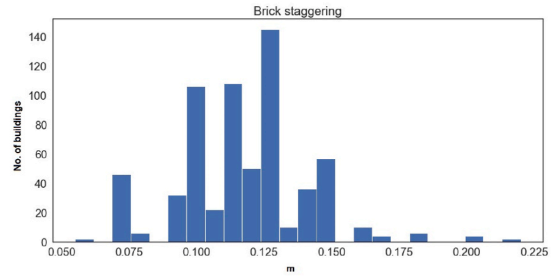
LR is a fundamental and widely used ML technique that aims to establish a relationship between a dependent variable and one or more independent variables. It involves fitting a straight line or hyperplane to a dataset to predict the values of the dependent variable based on the values of the independent variables [68]. LR models can be used for both regression and classification tasks, while their simplicity and interpretability make them popular for exploratory analysis and a baseline model for more complex algorithms.

4.3. *Random Forest regressor (RF regressor)*

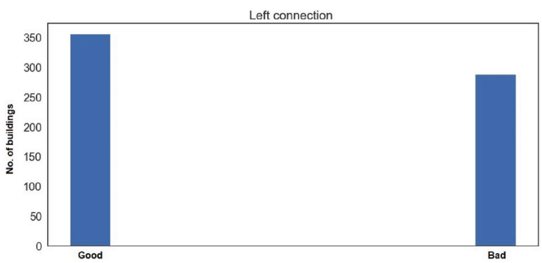
It is a variant of the RF algorithm that combines the predictions of multiple decision trees to make more accurate predictions. In RF regression, a set of decision trees are created using randomly selected subsets of the training data and features [69]. Then, the output of each tree is averaged to obtain the final prediction. This approach helps to reduce overfitting and improve the model's accuracy and generalization performance. RF regression can handle both linear and nonlinear relationships between the input and output variables and is suitable for



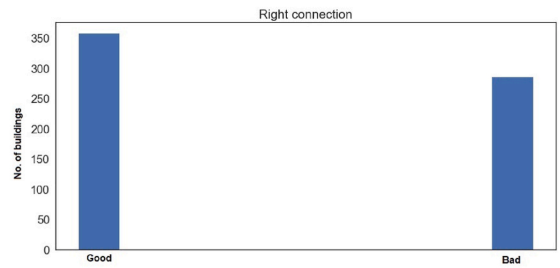
(q) Brick's height



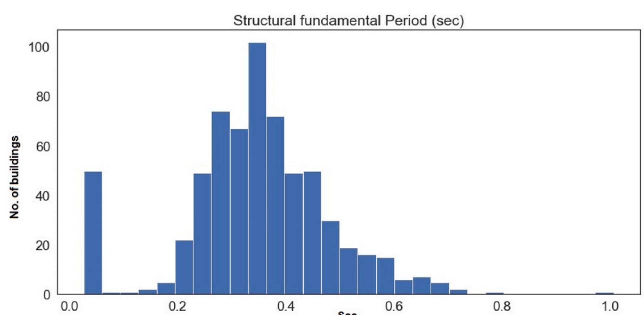
(r) Brick staggering



(s) Left connection



(t) Right connection



(u) Fundamental period of the structure

Fig. 2. (*continued*).

datasets with large numbers of features. The interpretability and scalability of the algorithm make it a popular choice for practical regression problems.

4.4. Stochastic Gradient Descent (SGD)

SGD regressor is a machine learning technique used for regression analyses that optimize the cost function by minimizing the error between predicted and actual values. It is a variant of the gradient descent algorithm that uses a stochastic approach to update the model parameters by computing the gradient of the cost function on small, randomly selected subsets of the training data [70]. This approach helps to improve the convergence rate and reduce the computational cost compared to batch gradient descent. The SGD regressor is particularly useful for large-scale datasets with a high number of features, where the computational cost of the traditional batch gradient descent is prohibitive. The algorithm is also robust to noisy data and can handle both linear and nonlinear relationships between the input and output variables.

4.5. Evaluate regression model

Mean-Squared Error (MSE), Root-Mean-Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) are metrics used to evaluate an ML regression model. These metrics illustrate how accurate the predictions of ML models are, and what is the amount of deviation from the actual values.

MSE calculates the squared difference between the actual and predicted values as presented in Eq. (4),

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y})^2 \tag{4}$$

where, n is the total number of samples, y and $\widehat{y}$ are the actual values and predicted output values, respectively.

RMSE can be any non-negative value, including values greater than 1. RMSE as presented in Eq. (5), is a measure of the difference between predicted and actual values and is typically used to evaluate the performance of a regression model. The lower the RMSE, the better the model is at making accurate predictions. However, the value of RMSE depends on the units of measurement of the dependent variable and is not restricted to a range of 0–1.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y})^2} \tag{5}$$

MAPE is a measure of the prediction accuracy of a forecasting

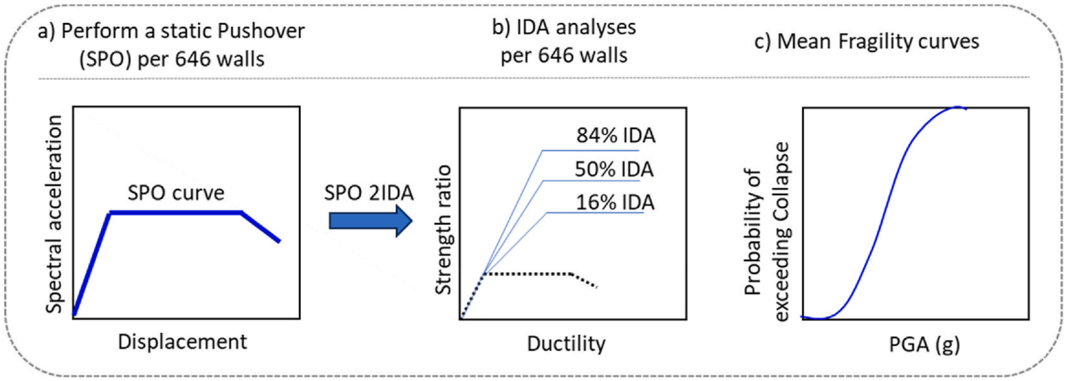


Fig. 3. a) SPO curve b) IDA curve at 16, 50, 84 % and c) mean fragility curve.

method in statistics. It usually expresses the accuracy as a ratio defined by Eq. (6).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}}{y_i}\right| \tag{6}$$

where, y and $\hat{y}$ are the actual values and predicted output values, respectively. Their difference is divided by the actual value y. The absolute value of this ratio is summed for every predicted point in time and divided by the number of fitted points n.

5. Results and discussion

5.1. Prediction by ML

The four selected ML estimators have been utilized on the complete dataset, which includes all the buildings inspected on-site in Salima, Lifidzi, Golomoti, and Blantyre in Malawi. In Figs. 4–7, the actual and predicted PGA(g) values of the walls at the collapse state are presented for the test data using SVR, LR, RF, and SGD regression models, respectively. These figures demonstrate that the ML estimators employed in this study have produced highly promising results, with the predicted values closely aligning with the actual values. The use of sophisticated algorithms and advanced statistical techniques has enabled the development of accurate predictive models that consider a wide range of complex variables and interconnected factors.

5.2. Compare evaluations

A comparative study is conducted on the performance results obtained from the four implemented ML regressors over the used dataset with all 21 features. Table 1 shows the metrics of each implemented ML method on the dataset. The MAPE value of 16.8 % by RF means that the average difference between the forecasted and actual values is as low as possible. Therefore, RF performs best among other ML estimators with the lowest MAPE and RMSE. Therefore, RF has been selected as this study's most proper ML method.

5.3. Feature correlation

Understanding features' importance and correlation is crucial when building an ML model. It helps to find the best set of features that allows one to build optimized models of studied phenomena. Correlation is a measure of the linear relationship between variables. Fig. 8 illustrates the Pearson correlation of inputs and output of the dataset. Fig. 9 shows that the features of period and right connection are the most important features for the derived correlation. In Fig. 8, the y-axis represents the F-values that were estimated from the correlation values as presented in Fig. 8. Table 2 shows the ranking of the correlation in terms of their feature score based on the correlation values shown in Figs. 8 and 9. Based on these scores the most important features can be fed into the RF technique as the input to derive the output. This would be described in detail in the next section.

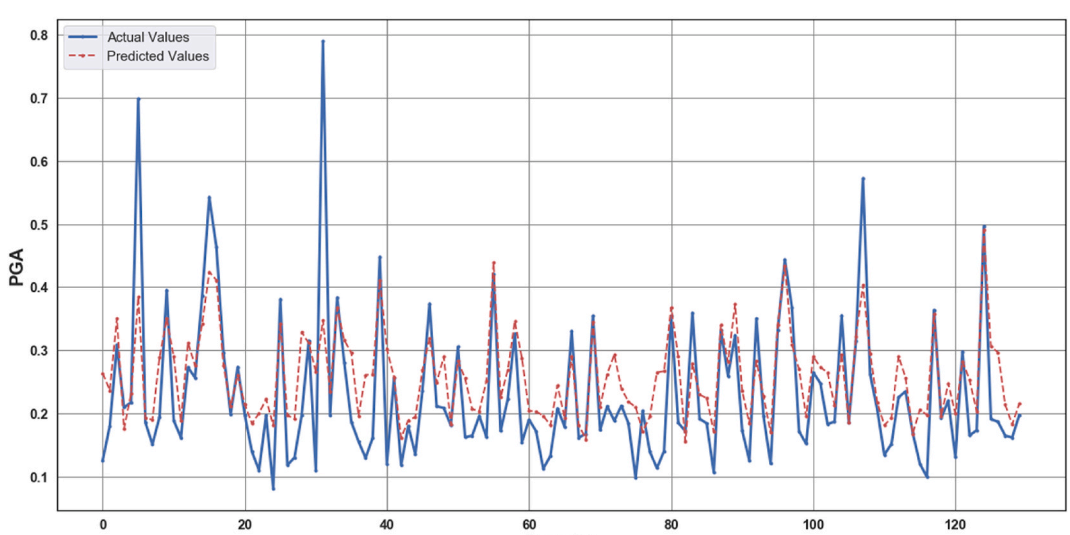


Fig. 4. Actual and predicted values by SVR.

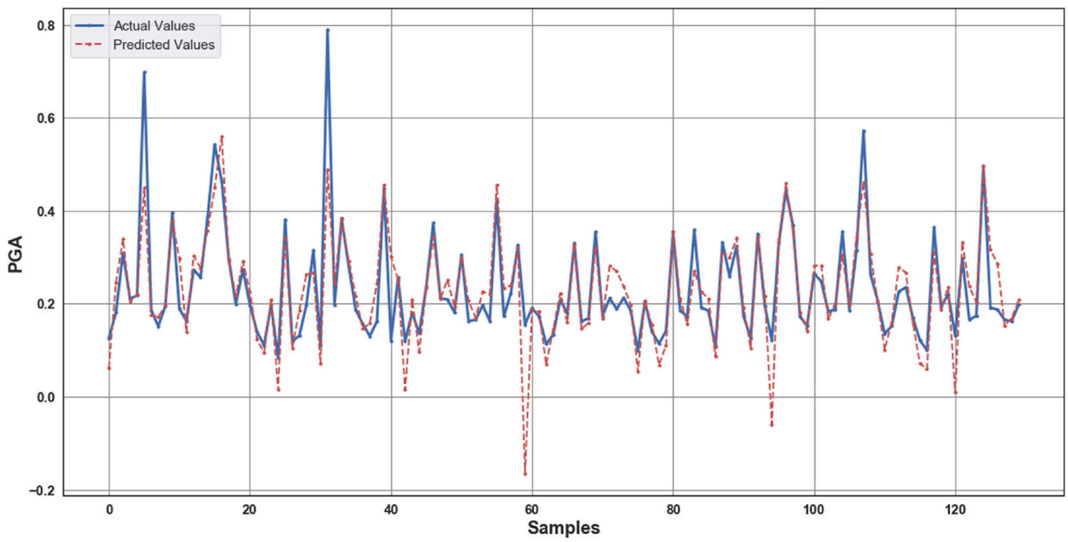


Fig. 5. Actual and predicted values by LR.

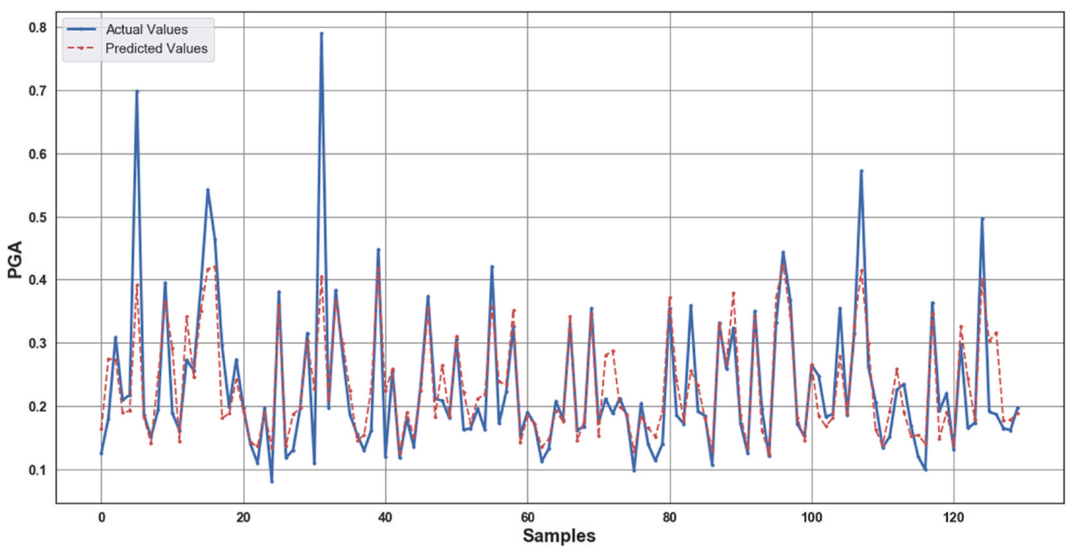


Fig. 6. Actual and predicted values by RF.

5.4. *Random Forest for reduced features*

Here, RF has been used to train and test the data by considering the best features and running the model by removing the less correlated features. The accuracy was significantly improved by reducing the number of less correlated features used in the RF model. Specifically, a marked reduction in MAPE and other key measures of model performance was observed when the number of features were decreased from 21 to 10. In each step, the reduced feature was the least correlated feature such as spandrel height, brick length, roof orientation, gable, mortar type, gable height, brick staggering, length of the wall perpendicular to the inspected wall, brick height, number of internal walls parallel to the inspected wall, and wall height, respectively. In fact, the top 10 most correlated features (ranks 1 to10) in Table 2 were fed into the RF to derive the output.

This indicates that many initial features added little value to the model and that focusing on the most important and highly correlated variables can achieve more accurate and reliable predictions. These findings underscore the importance of carefully selecting and refining input features when building ML models and suggest that a targeted data-driven approach can yield significant improvements in model performance and prediction accuracy.

Table 3 presents the result of implementing RF by considering the best features based on the correlation of features. In each stage, one of the lowest correlated features was removed, and the model used the rest of features for analysis until it only used the top 10 most correlated features including: 1-Structural fundamental period, 2-Right connection, 3-Left connection, 4-Wall thickness, 5-Total area opening, 6-Wall length, 7-Number of internal walls perpendicular to the inspected wall, 8-Masonry type, 9-Number of internal walls perpendicular to the back wall and parallel to the inspected wall, and 10-Roof type.

Fig. 10 illustrates the MAPE values by RF. This figure shows the good performance of the RF method for output evaluation using the selected reduced features.

Fig. 11 shows the derived fragility curves using actual data set that has been illustrated comparatively with those obtained using four ML techniques in this research based on all data set and finally the one acquired via RF and only reduced features. Small amount of error value (MAPE) represented in Table 3 can be visually seen in this figure which shows accuracy and success of the applied ML techniques for deriving

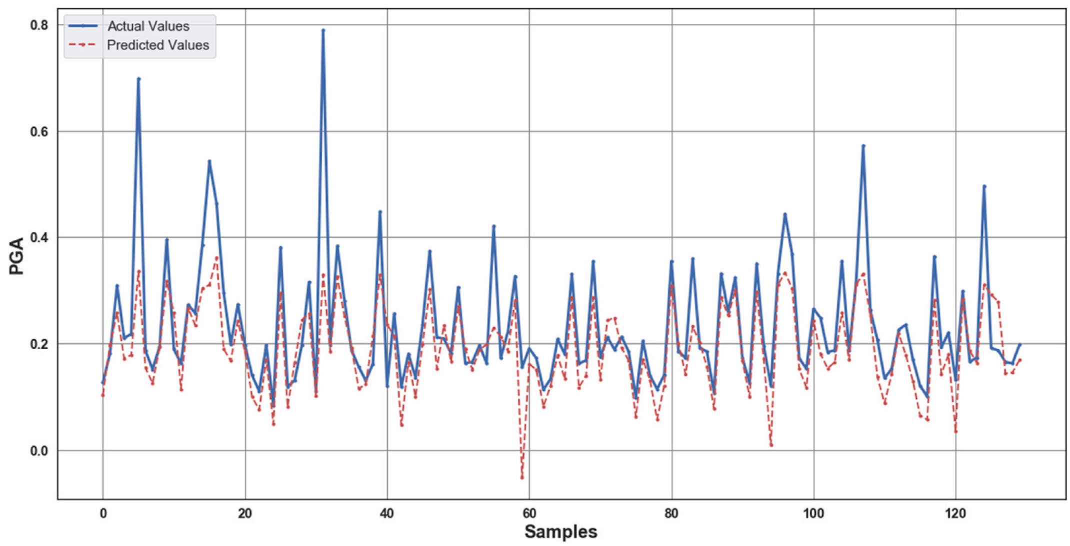


Fig. 7. Actual and predicted values by SGD.

Table 1
A comparison between different metrics for each ML method by considering all features.

Metrics	ML method			
	LR	SVR	RF	SGD
MSE	0.004	0.006	0.003	0.006
RMSE	0.064	0.0798	0.0547	0.0816
MAPE (%)	20.373	30.327	16.807	22.458

fragility curves to estimate building collapse probability in terms of PGA.

6. Conclusion

Conducting research on the seismic vulnerability and deriving fragility curves for non-engineered masonry buildings in developing countries is of the paramount importance to ensure the safety of people and infrastructure in earthquake-prone areas. These buildings are particularly susceptible to significant damage when subjected to future

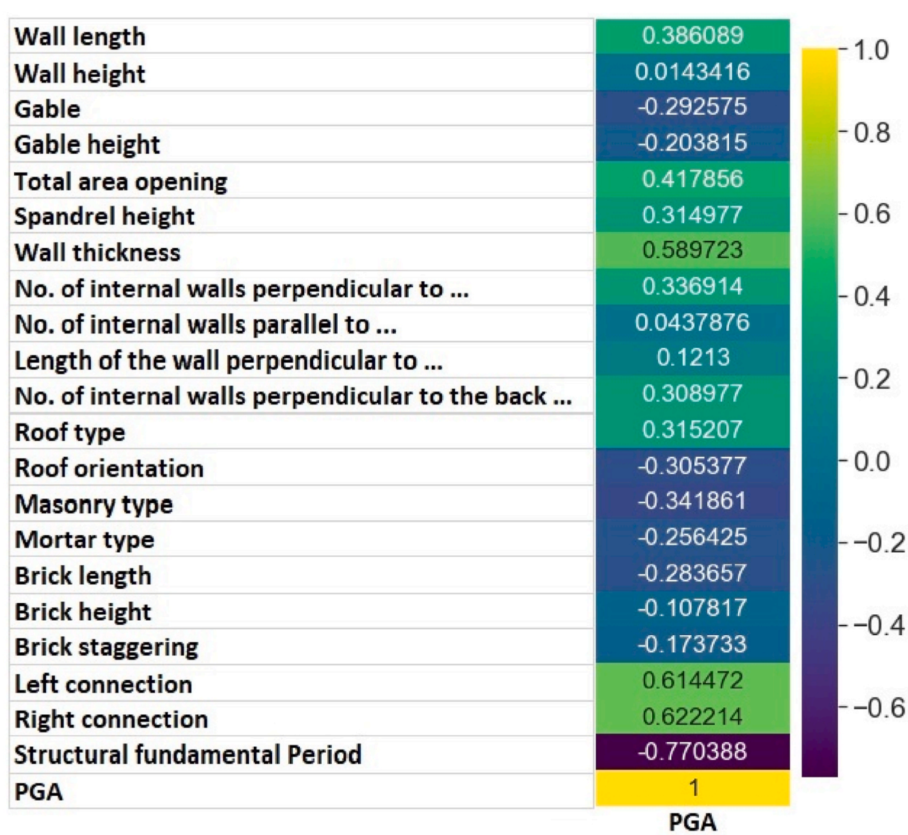


Fig. 8. Correlation between inputs and output.

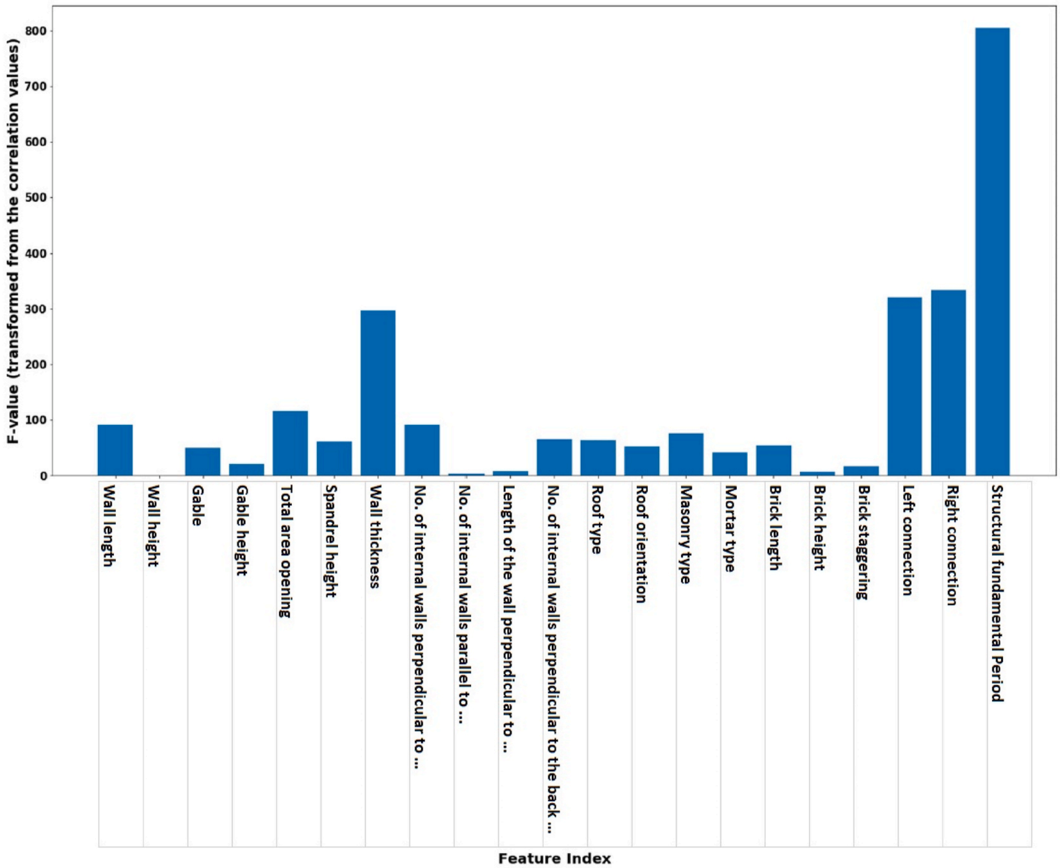


Fig. 9. Feature index.

Table 2
Sorting most correlated features to the output and the feature score.

Rank	Feature	Score
1	Structural fundamental period	804.64
2	Right connection	333.01
3	Left connection	321.13
4	Wall thickness	297.43
5	Total area opening	116.01
6	Wall length	91.82
7	No. of internal walls perpendicular to the inspected wall	91.34
8	Masonry type	75.78
9	No. of internal walls perpendicular to the back wall parallel to the inspected wall	65.29
10	Roof type	64.59
11	Spandrel height	61.98
12	Brick length	54.22
13	Roof orientation	52.09
14	Gable	49.39
15	Mortar type	41.91
16	Gable height	21.49
17	Brick staggering	17.63
18	Length of the wall perpendicular to the inspected wall	8.77
19	Brick height	7.08
20	No. of internal walls parallel to the inspected wall	2.8
21	Wall height	0.36

seismic ground excitations, highlighting the need for in-depth investigations into their dynamic responses considering the complexities of masonry materials. The devastating toll and losses caused by past earthquakes have spurred decision-makers to seek scientific collaborations and international research projects to develop comprehensive mitigation plans and retrofitting techniques.

The advancements in data science and artificial intelligence, with data driven and ML techniques at their core, offer powerful tools for

Table 3
A comparison between different metrics for RF by reducing features to top 10.

Metrics	Number of Features										
	20	19	18	17	16	15	14	13	12	11	10
MSE	0.0038	0.0028	0.0029	0.0022	0.0024	0.0028	0.0028	0.003	0.0026	0.0029	0.003
RMSE	0.0617	0.0536	0.0546	0.0477	0.049	0.0537	0.0537	0.0557	0.0514	0.0538	0.054
MAPE (%)	17.674	14.988	15.106	14.101	13.054	13.526	13.764	14.238	13.109	13.205	14.451

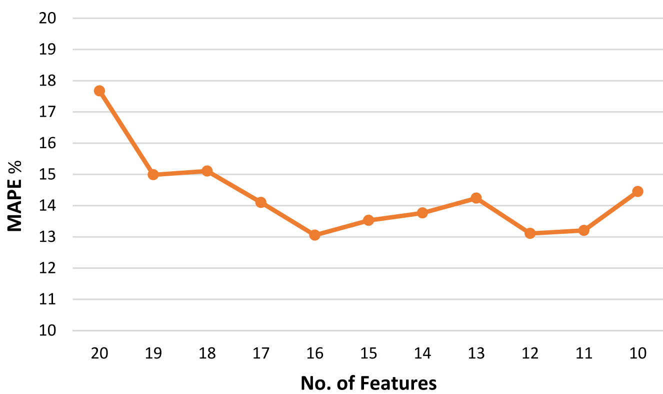


Fig. 10. MAPE values by RF for reduced features.

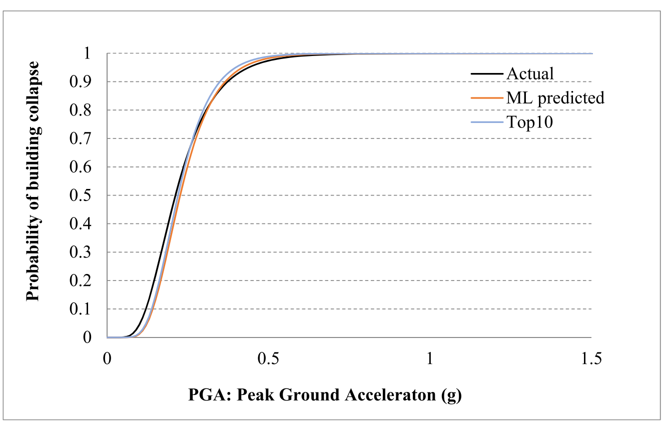


Fig. 11. Fragility curves obtained using actual data set compared with that of predicted via LR, SVR, RF, and SGD techniques based on all features, and RF using reduced features to top 10.

tackling complex problems. These include classification, regression, and image processing, among others. In this study, four different ML estimators, namely SVR, LR, RF, and SGD, were applied to a dataset obtained through on-site visual inspections of 646 masonry walls in Malawi. The goal was to evaluate the probability of collapse for these buildings by generating fragility curves that provide valuable insights into their seismic vulnerability and risk. By determining the key parameters that have the greatest impact on the derived output parameter, the process of data collection was streamlined towards analyses using the selected ML approaches in this study for the purpose of fragility curves derivation and collapse probability estimation of the surveyed buildings in the targeted area.

Among the ML techniques evaluated, the RF approach demonstrated the best performance. Furthermore, by reducing the number of features from 21 to 10, the model's performance improved from 17.67 % to 14.45 % in terms of the MAPE value. The most important and the most correlated parameters were determined as structural fundamental period, right connection, left connection, wall thickness, total area of openings, wall length, number of internal walls perpendicular to the inspected wall, masonry type, number of internal walls perpendicular to the back wall and parallel to the inspected wall, and the roof type.

Therefore, based on the analyses and results conduced in this research, it is concluded that ML techniques can be employed for the purpose of seismic fragility assessment of non-engineered buildings. Obtained results proved a high level of accuracy for these techniques, especially for the RF method which followed the fragility curve acquired from actual data very closely as shown in Fig. 10. Enlarging dataset in terms of number of buildings and seismic events, exploring other ML methods, and incorporating additional and/or different features for the surveyed non-engineered buildings for a targeted region could enhance the prediction performance which could be explored in further research.

CRediT authorship contribution statement

Ehsan Harirchian: Conceptualization, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Seyed Ehsan Aghakouchaki Hosseini:** Investigation, Writing – original draft, Writing – review & editing. **Viviana Novelli:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. **Tom Lahmer:** Funding acquisition, Project administration, Supervision. **Shahla Rasulzade:** Writing – original draft, Writing – review & editing.

Declaration of competing interest

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

Acknowledgments

We acknowledge the support of the German Research Foundation (DFG) and the Bauhaus-Universität Weimar within the Open-Access Publishing Programme.

Data availability

Data will be made available on request.

Abbreviations

ANN	Artificial Neural Network
CSM	Capacity Spectrum Method
DRR	Disaster Risk Reduction
IDA	Incremental Dynamic Analysis
LR	Linear Regression
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MSE	Mean Squared Error
PBD	Performance-Based Design
PGA	Peak Ground Acceleration
PGV	Peak Ground Velocity
RF	Random Forest
RMSE	Root Mean Square Error
SA	Spectral Acceleration
SGD	Stochastic Gradient Descent
SVR	Support Vector Regression

References

[1] S. Galano, A. Calabrese, State of the art of seismic Protection Technologies for non-engineered buildings (N-EBs) in developing regions of the world, J. Earthq. Eng. (2023) 1–27.

[2] D. Losanno, N. Ravichandran, F. Parisi, Seismic fragility models for base-isolated unreinforced masonry buildings with fibre-reinforced elastomeric isolators, Earthq. Eng. Struct. Dynam. 52 (2) (2023) 308–334.

[3] Y. Zhang, J.F. Fung, K.J. Johnson, S. Sattar, Review of seismic risk mitigation policies in earthquake-prone countries: lessons for earthquake resilience in the United States, J. Earthq. Eng. 26 (12) (2022) 6208–6235.

[4] N. Paul, V. Silva, D. Amo-Oduro, Development of a uniform exposure model for the African continent for use in disaster risk assessment, Int. J. Disaster Risk Reduc. 71 (2022), 102823-102823.

[5] K. Goda, et al., Scenario-based earthquake risk assessment for central-southern Malawi: the case of the Bilila-Mtakataka Fault, Int. J. Disaster Risk Reduc. 67 (2022), 102655-102655.

[6] J.K. Mulwa, F. Kimata, S. Suzuki, Z.N. Kuria, The seismicity in Kenya (East Africa) for the period 1906–2010: a review, J. Afr. Earth Sci. 89 (2014) 72–78.

[7] J. Biggs, E. Nissen, T. Craig, J. Jackson, D. Robinson, Breaking up the hanging wall of a rift-border fault: the 2009 Karonga earthquakes, Malawi, Geophys. Res. Lett. 37 (11) (2010).

[8] A.S. Macheyeki, D. Delvaux, M. De Batist, A. Mruma, Fault kinematics and tectonic stress in the seismically active Manyara–dodoma rift segment in Central Tanzania–implications for the East African rift, J. Afr. Earth Sci. 51 (4) (2008) 163–188.

[9] K. Vasileiou, J. Barnett, D.S. Fraser, Integrating local and scientific knowledge in disaster risk reduction: a systematic review of motivations, processes, and outcomes, Int. J. Disaster Risk Reduc. (2022) 103255.

[10] M. Kuglitsch, et al., Artificial intelligence for disaster risk reduction: opportunities, challenges, and prospects, Bulletin nº 71 (1) (2022).

[11] F. Freddi, et al., Innovations in earthquake risk reduction for resilience: recent advances and challenges, Int. J. Disaster Risk Reduc. 60 (2021) 102267.

[12] R. Monteiro, et al., Development of fragility curves for large-scale seismic risk of RC structures, in: Proceedings of the 17th World Conference on Earthquake Engineering, JPN, 2020.

[13] C. Galasso, et al., Risk-based, pro-poor urban design and planning for tomorrow's Cities, Int. J. Disaster Risk Reduc. 58 (2021) 102158.

[14] P. Gehl, D.M. Seyedi, J. Douglas, Vector-valued fragility functions for seismic risk evaluation, Bull. Earthq. Eng. 11 (2) (2013) 365–384.

[15] S. Ruggieri, et al., An archetype-based automated procedure to derive global-local seismic fragility of masonry building aggregates: meta-forma-xl, Int. J. Disaster Risk Reduc. 95 (2023) 103903.

[16] S. Ruggieri, F. Porco, G. Uva, D. Vamvatsikos, Two frugal options to assess class fragility and seismic safety for low-rise reinforced concrete school buildings in Southern Italy, Bull. Earthq. Eng. 19 (2021) 1415–1439.

[17] M.M. Kassem, F.M. Nazri, E.N. Farsangi, The seismic vulnerability assessment methodologies: a state-of-the-art review, Ain Shams Eng. J. 11 (4) (2020) 849–864.

[18] D. D'Ayala, V. Novelli, Seismic Vulnerability Assessment: Masonry Structures, Springer Berlin Heidelberg, 2014.

[19] G.M. Calvi, R. Pinho, G. Magenes, J.J. Bommer, L.F. Restrepo-Vélez, H. Crowley, Development of seismic vulnerability assessment methodologies over the past 30 years, ISET J. Earthq. Technol. 43 (3) (2006) 75–104.

[20] C. Del Gaudio, A. Rosti, A. Penna, P. Ricci, M. Rota, G.M. Verderame, Empirically based approaches for the derivation of fragility curves of Italian RC building typologies, Procedia Struct. Integr. 44 (2023) 259–266.

[21] H. Azizi-Bondarabadi, N. Mendes, P.B. Lourenço, N.H. Sadeghi, Empirical seismic vulnerability analysis for masonry buildings based on school buildings survey in Iran, Bull. Earthq. Eng. 14 (2016) 3195–3229.

[22] T. Rossetto, I. Ioannou, D. Grant, T. Maqsood, Guidelines for the Empirical Vulnerability Assessment, 2014.

[23] F. Sabetta, A. Goretti, A. Lucantoni, Empirical fragility curves from damage surveys and estimated strong ground motion, in: Proceedings of the 11th European Conference on Earthquake Engineering, 1998, pp. 1–11. Paris, France.

[24] A. Du, X. Wang, Y. Xie, Y. Dong, Regional Seismic Risk and Resilience Assessment: Methodological Development, Applicability, and Future Research Needs–An Earthquake Engineering Perspective, Reliability Engineering & System Safety, 2023 109104.

[25] V. Novelli, et al., Seismic mitigation framework for non-engineered masonry buildings in developing countries: application to Malawi in the East African rift, Resilient Structures and Infrastructure (2019) 195–223.

[26] A. Preciado, A. Ramírez-Gaytán, R. Salido-Ruiz, J.L. Caro-Becerra, R. Lujan-Godinez, Earthquake Risk Assessment Methods of Unreinforced Masonry Structures: Hazard and Vulnerability, 2015.

[27] S. Tesfamariam, Performance-based design of tall timber buildings under earthquake and wind multi-hazard loads: past, present, and future, Frontiers in Built Environment 8 (2022) 848698.

[28] C. Dai, S. Hu, W. Wang, Performance-based design of steel frames with self-centering modular panel, J. Build. Eng. 57 (2022) 104841.

[29] O. Idels, O. Lavan, Performance based formal optimized seismic design of steel moment resisting frames, Comput. Struct. 235 (2020) 106269.

[30] J. Guo, P. Zhang, J. Wang, S. Li, Z. Guan, A novel framework for seismic fragility analysis with the combination of Box-Cox transformation and Bayesian inference, Eng. Struct. 277 (2023) 115436.

[31] J. Alam, D. Kim, B. Choi, Seismic risk assessment of intake tower in Korea using updated fragility by Bayesian inference, Struct. Eng. Mech. 69 (3) (2019) 317–326.

[32] M.S. Azad, D.-D. Nguyen, B. Thusa, T.-H. Lee, Seismic Fragility Assessments of APR-1400 Containment Building Using Bayesian Inference Framework, 2022.

[33] A. Nettis, R. Gentile, D. Raffaele, G. Uva, C. Galasso, Cloud Capacity Spectrum Method: accounting for record-to-record variability in fragility analysis using nonlinear static procedures, Soil Dynam. Earthq. Eng. 150 (2021) 106829.

[34] M.B. Baylon, M.E.P. Sevilla, M.D. Cutora, R.M.S. Villa, P.M.P. Reynes, J.M. V. Montemayor, Development of fragility curves for seismic vulnerability assessment: the case of Philippine General Hospital spine building, International Research Journal of Science, Technology, Education, & Management (IRJSTEM) 2 (4) (2022).

[35] M. Miari, R. Jankowski, Incremental dynamic analysis and fragility assessment of buildings with different structural arrangements experiencing earthquake-induced structural pounding, in: Computational Science–ICCS 2022: 22nd International Conference, Springer, London, UK, 2022, pp. 117–124. June 21–23, 2022, Proceedings, Part II.

[36] K. Mahmoodi, A. Noorzad, A. Mahboubi, M. Alembagheri, Seismic performance assessment of a cemented material dam using incremental dynamic analysis, in: Structures, vol. 29, Elsevier, 2021, pp. 1187–1198.

[37] A. Moayedifar, H.R. Nejati, K. Goshtasbi, M. Khosrotash, Seismic fragility and risk assessment of an unsupported tunnel using incremental dynamic analysis (IDA), Earthquakes and Structures 16 (6) (2019) 705.

[38] M. Zhao, Z. Gao, P. Wang, X. Du, Response spectrum method for seismic analysis of monopile offshore wind turbine, Soil Dynam. Earthq. Eng. 136 (2020) 106212.

[39] S. Xiao, M. Wang, Response spectrum method for building structures with general nonviscous damping models, in: Structures, vol. 40, Elsevier, 2022, pp. 571–580.

[40] W. Yang, X. Zou, M. Wang, P. Liu, A multinomial logistic regression model-based seismic risk assessment method for museum exhibition halls, J. Build. Eng. 69 (2023) 106312.

[41] A. Suzuki, I. Iervolino, Intensity measure conversion of fragility curves, Earthq. Eng. Struct. Dynam. 49 (6) (2020) 607–629.

[42] I. Iervolino, Estimation uncertainty for some common seismic fragility curve fitting methods, Soil Dynam. Earthq. Eng. 152 (2022) 107068.

[43] C.-T. Dang, T.-P. Le, P. Ray, A novel method based on maximum likelihood estimation for the construction of seismic fragility curves using numerical simulations, Compt. Rendus Mec. 345 (10) (2017) 678–689.

[44] V. Manfredi, A. Masi, G. Nicodemo, A. Digrisolo, Seismic fragility curves for the Italian RC residential buildings based on non-linear dynamic analyses, Bull. Earthq. Eng. 21 (4) (2023) 2173–2214.

[45] A. Stocchi, C. Giry, S. Capdevielle, I. Zentner, E. Nayman, F. Ragueneau, A simplified non-linear modelling strategy to generate fragility curves for old masonry buildings, Comput. Struct. 254 (2021) 106579.

[46] E. Voyagaki, et al., Masonry Panel Testing in Malawi, 2020.

[47] E. Nazarian, T. Taylor, T. Weifeng, F. Ansari, Machine-learning-based approach for post event assessment of damage in a turn-of-the-century building structure, Journal of Civil Structural Health Monitoring 8 (2) (2018) 237–251.

[48] P.C. Lazaridis, I.E. Kavvadias, K. Demertzis, L. Iliadis, L.K. Vasiliadis, Structural damage prediction of a reinforced concrete frame under single and multiple seismic events using machine learning algorithms, Appl. Sci. 12 (8) (2022), 3845-3845.

[49] M.A. Bülbül, E. Harirchian, M.F. Işık, S.E. Aghakouchaki Hosseini, E. Işık, A hybrid ANN-GA model for an automated rapid vulnerability assessment of existing RC buildings, Appl. Sci. 12 (10) (2022) 5138.

[50] E. Harirchian, K. Jadhav, V. Kumari, T. Lahmer, ML-EHSAPP: a prototype for machine learning-based earthquake hazard safety assessment of structures by using a smartphone app, European Journal of Environmental and Civil Engineering (2021) 1–21.

[51] D. Saadati, A. Moghadam, EZRVS: an AI-based web application to significantly enhance seismic rapid visual screening of buildings, J. Earthq. Eng. (2023) 1–18.

[52] A. Cardellicchio, S. Ruggieri, V. Leggieri, G. Uva, A machine learning framework to estimate a simple seismic vulnerability index from a photograph: the VULMA project, Procedia Struct. Integr. 44 (2023) 1956–1963.

[53] E. Harirchian, T. Lahmer, Improved rapid assessment of earthquake hazard safety of structures via artificial neural networks, in: IOP Conference Series: Materials Science and Engineering, vol. 897, IOP Publishing, 2020 012014, 1.

[54] K. Morfidis, S. Stefanidou, O. Markogiannaki, A rapid seismic damage assessment (RASDA) tool for RC buildings based on an artificial intelligence algorithm, Appl. Sci. 13 (8) (2023) 5100.

[55] S. Ruggieri, A. Cardellicchio, G. Uva, Using transfer learning technique to define seismic vulnerability of existing buildings through mechanical models, Procedia Struct. Integr. 44 (2023) 1964–1971.

[56] G. Uva, V. Sangiorgio, S. Ruggieri, F. Fatiguso, Structural vulnerability assessment of masonry churches supported by user-reported data and modern Internet of Things (IoT), Measurement 131 (2019) 183–192.

[57] T.M. Mitchell, Machine Learning, vol. 45, McGraw-hill, New York, 1997, pp. 870–877.

[58] R. Caruana, A. Niculescu-Mizil, An Empirical Comparison of Supervised Learning Algorithms, 2006, pp. 161–168.

[59] P. Kloukinas, V. Novelli, I. Kafodya, I. Ngoma, J. Macdonald, K. Goda, A building classification scheme of housing stock in Malawi for earthquake risk assessment, J. Hous. Built Environ. 35 (2) (2020) 507–537.

[60] K. Jaiswal, D.J. Wald, Creating a Global Building Inventory for Earthquake Loss Assessment and Risk Management, US Geological Survey Denver, CO, 2008.

[61] V. Novelli, et al., Fragility curves for non-engineered masonry buildings in developing countries derived from real data based on structural surveys and laboratory tests, Soft Comput. 25 (2021) 6113–6138.

[62] V. Novelli, D. D'Ayala, N. Makhloufi, D. Benouar, A. Zekagh, A procedure for the identification of the seismic vulnerability at territorial scale. Application to the Casbah of Algiers, Bull. Earthq. Eng. 13 (2015) 177–202.
[63] I. Ngoma, I. Kafodya, P. Kloukinas, V. Novelli, J. Macdonald, K. Goda, Building classification and seismic vulnerability of current housing construction in Malawi, Malawi J. Sci. Technol. 11 (1) (2019) 57–72.
[64] P. Kloukinas, I. Kafodya, I. Ngoma, V. Novelli, J. Macdonald, K. Goda, Strength of Materials and Masonry Structures in Malawi, 2019.
[65] D.F. D'ayala, Force and displacement based vulnerability assessment for traditional buildings, Bull. Earthq. Eng. 3 (2005) 235–265.
[66] D. Vamvatsikos, C. Allin Cornell, Direct estimation of the seismic demand and capacity of oscillators with multi-linear static pushovers through IDA, Earthq. Eng. Struct. Dynam. 35 (9) (2006) 1097–1117.
[67] A. Géron, Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems O'Reilly Media, 2017, pp. 54–56. Sebastopol, CA.
[68] X. Su, X. Yan, C.L. Tsai, Linear regression, Wiley Interdisciplinary Reviews: Comput. Stat. 4 (3) (2012) 275–294.
[69] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[70] S.-i. Amari, Backpropagation and stochastic gradient descent method, Neurocomputing 5 (4–5) (1993) 185–196.