International Journal of Disaster Risk Reduction 90 (2023) 103655

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

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

Towards disaster risk mitigation on large-scale school intervention programs

Rafael Fernández [a,b], Juan Francisco Correal [a], Dina D'Ayala [b], Andrés L. Medaglia [c,*]

[a] Centro de Investigación en Materiales y Obras Civiles (CIMOC), Department of Civil and Environmental Engineering, Universidad de los Andes, Bogotá, Colombia
[b] Department of Civil Engineering and Geomatics, University College London, London, United Kingdom
[c] Centro para la Optimización y Probabilidad Aplicada (COPA), Department of Industrial Engineering, Universidad de los Andes, Bogotá, Colombia

ARTICLE INFO

Keywords:
Disaster mitigation
Disaster preparedness
School infrastructure
Clustering analysis
Unsupervised machine learning
Machine learning
Education

ABSTRACT

Education infrastructure is one of the main barriers on school quality in Low- and Middle-Income Countries (L&MICs), since it is insufficient and unevenly distributed. Improving the school infrastructure is needed to provide a high-quality education environment. Although research on how to improve the infrastructure is available, there is still a lack of a consistent and systematic approach to develop large-scale interventions at the national or regional level. To fill this gap, we propose a data-driven methodology with the purpose of developing a prioritization of interventions to carry out a seismic disaster risk reduction program. The method starts by identifying groups of similar buildings using clustering analysis, starting with a seismic taxonomy as descriptor (i.e., model input). Then, domain experts analyze the suggested clusters to design scalable interventions for the representative building of each cluster. The proposed data-driven methodology requires experts' criteria in each step to validate the results and make them applicable, but significantly reduces the bias by automating the decision-making process. We use as case study the Dominican Republic public school infrastructure and present the results of the application of the proposed method. The method presented herein is extensible to other infrastructure portfolios, as well as to other types of hazards.

1. Introduction

Several barriers affect the education quality in Low- and Middle-Income Countries (L&MICs). The Non-Governmental Organization *Educate a Child* considers school infrastructure as one of nine barriers for high quality education [1]. Likewise, the World Bank argues that quality education should be achieved through five pillars, being one of them, school infrastructure [2]. The critical role that infrastructure plays in education quality has become evident with the COVID-19 pandemic [2]. This crisis has also shown that there is a need for strong institutions in countries to address school infrastructure challenges at a large scale, in a well-structured and systematic way [2]. One of the greatest challenges related to school infrastructure is the safety against natural hazards, and in particular against earthquakes. This criticality is due to the vulnerability of the occupants and the fact that the effect of earthquakes cannot be mitigated through early warning. Therefore, the focus is on how to improve seismic safety in school infrastructure.

Multilateral agencies have been promoting large-scale school infrastructure safety improvement programs at different levels. For

* Corresponding author.
E-mail addresses: ri.fernandez1110@uniandes.edu.co (R. Fernández), jcorreal@uniandes.edu.co (J.F. Correal), d.dayala@ucl.ac.uk (D. D'Ayala), amedagli@uniandes.edu.co (A.L. Medaglia).

https://doi.org/10.1016/j.ijdrr.2023.103655
Received 16 December 2022; Received in revised form 21 March 2023; Accepted 21 March 2023
Available online 23 March 2023

example, the United Nations Educational, Scientific, and Cultural Organization (UNESCO), developed the VISUS methodology for characterizing safety conditions of school facilities, which has been successfully applied in different countries in Latin America and Asia [3]. Similarly, school infrastructure safety is considered in the Comprehensive School Safety program and the Worldwide Initiative for Safe Schools developed by the United Nations Office for Disaster Risk Reduction (UNDRR), in collaboration with the Global Alliance for Disaster Risk Reduction & Resilience in the Education Sector (GADRRRES) [4,5]. These initiatives provide evidence that it is possible to develop a global framework to support activities related to safe learning facilities and school disaster risk management.

Development banks are also actively working on improving school infrastructure, playing a vital role in technical support, training, capacity building, and program financing. In particular, the World Bank set up the Global Program for Safer Schools (GPSS), an initiative that leverages large-scale investment to enhance safety and resilience of school infrastructure at risk from natural hazards. Through the GPSS, the World Bank has helped governments improve the quality of learning environments for children in Central and East Asia, Sub-Saharan Africa, and the Americas [6]. In the Latin American and the Caribbean (LAC) region, the Inter-American Development Bank (IDB) conducted a study named "*Learning in twenty-first-century schools*" [7] which established architectural and functional characteristics to ensure a quality learning environment for public schools. These programs have helped improve the school infrastructure in L&MICs, bridging the disparities of education infrastructure in the region, while considering the scarce resources and reduced intervention budgets.

The literature also provides a body of knowledge on school infrastructure improvement, particularly focused on disaster mitigation. Naja & Bytiyeh [8] advocate for school buildings retrofitting as opposed to buildings replacement. The authors show cost-benefit analysis of different retrofitting strategies. However, the authors do not present a methodology for the prioritization or optimization of investment in the implementation of such program. Similarly, Grimaz & Malisan [9] present the VISUS methodology and some case studies. The authors indicate the benefits of elaborating and implementing plans based on visual inspections and the implementation of pre-codified algorithms to replicate expert reasoning. This novel approach helps unveil the current conditions and the safety gap of a given school portfolio, but it does not provide indications on how to choose and plan the interventions. Large scale interventions are the focus of Grant et al. [10] who developed a method to select schools suitable for retrofitting by comparing the age of the building with the seismic provisions of the building code of the time and the current seismic demand in Italy. This method proposes a novel way of selecting vulnerable buildings, but it does not propose a way to develop interventions that could be applicable to a large set of buildings. Chi et al. [11] proposed a machine learning algorithm for the classification of school buildings based on structural information of the public-school infrastructure in Taiwan. This method also fails to identify the type of interventions to be adopted for each facility.

Muñoz et al. [12] analyze the success, innovation, and challenges of school safety and disaster education (disaster risk as part of the curriculum) in several countries in Latin America and the Caribbean. This analysis shows how efforts are addressed differently by each country, despite most of them are in line with the Comprehensive School Safety Framework and Sendai Framework. In the case of Colombia, the retrofitting of school buildings became essential, since the implementation of the national seismic resistant building code (NSR-10). This document reclassified school buildings to "essential" facilities, therefore requiring higher performance for the same level of hazard. In this context, the city of Bogotá invested more than US$ 400 million to upgrade, retrofit, and build new school infrastructure; other cities in Colombia are replicating this effort [12]. Conversely, more systematic intervention strategies have been developed for other type of infrastructures. For instance, for the maintenance of a bridge network, Jaafaru & Agbelie [13] developed a comprehensive bridge maintenance framework using machine learning, multi-criteria decision analysis, and evolutionary optimization. Similar frameworks have been developed for resilient infrastructure systems such as power [14] and residential buildings involving transportation and drainage systems [15].

Most of the initiatives from multilateral agencies, development banks, and academia usually lack of a systematic approach to identify gaps and propose solutions in relation to school infrastructure. A notable exception is the Global Library of School Infrastructure (GLOSI) taxonomical system, developed under the Global Program for Safer Schools (GPSS) [16,17]. This taxonomy was designed to characterize different school buildings from their construction attributes to identify their expected seismic response. This response is characterized by the identification of attributes of 12 parameters, including the structural system, number of stories, seismic design level, irregularities, and structural health condition, among others. However, in the case of large-scale seismic retrofitting programs, it is not possible to determine particular interventions for each building. Therefore, there is a clear need to identify groups of similar buildings, to develop generic interventions that can be implemented at a national and regional level. Each group should be represented by a typical blueprint building, referred herein as representative building [18]. Representative buildings have been used to develop retrofitting strategies in different countries, such as Italy [19] and Peru [20]. However, in these cases the identification of the homogeneous groups was based on information and expert criteria, rather than using a systematic approach. Identifying the set of representative buildings is a critical step since this allows the design of target interventions that can be rolled out to large number of buildings with similar response. Nonetheless, the correlation between a taxonomical classification and the identification of groups of buildings for which a given strengthening strategy is suitable, is not a trivial step, and one that has not been investigated so far.

To fill this gap, we propose a data-driven methodology to prioritize interventions in a seismic disaster risk reduction program. The method is based on a clustering algorithm that find groups of similar buildings using the GLOSI seismic taxonomy for a given portfolio of school buildings. The clustering analysis works on datasets with no prior information on their relationships, which makes it highly appropriate to identify the main typologies and group the exposed assets based on the categorical information from the taxonomy string definition. After the clustering analysis phase, domain experts analyze the final clusters and design scalable interventions for each cluster's representative building (characterized by its representative taxonomy string). We illustrate the method to support a seismic risk reduction program with data from the public-school building portfolio of the Dominican Republic. The information on school infrastructure was gathered in a project funded by The World Bank [21]. We show that it is possible to identify clusters, analyze, and develop a large-scale intervention strategy based on seismic retrofitting alternatives on representative buildings. Finally, we discuss how it is possible to extend this methodology to other countries contingent on the available information. In addition, the methodology is applicable to other type of infrastructure assets, such as hospitals or bridges, subject to other natural hazards, like floods, hurricanes, or landslides.

2. Theoretical background: clustering

Clustering is the process of identifying groups of elements belonging to a larger set, where the elements within each group share a number of attributes, while they are considerably different from elements in other groups. To measure this similarity, a *distance* between elements should be calculated. Several distance metrics are available in the literature and the choice of the most appropriate should be based on the type and characteristics of the data set [22]. This grouping process, however, is done without prior knowledge of the distribution within the set, therefore it is often classified as an unsupervised machine learning algorithm. Nonetheless, expert judgement is often part of the process as it helps interpreting and justifying the correct number of clusters in which the data should be divided.

Clustering in disaster risk assessment has been difficult to apply due to the variability in construction characteristics, the large number of infrastructure elements, and the uncertainty in the structural behavior of buildings against natural hazards, among other factors [23]. Nevertheless, several authors have proposed clustering techniques as a way to approach disaster management challenges. Aleskerov et al. [24] developed a decision support system for disaster management in Turkey to predict damages and losses in seismic scenarios based on clusters. Likewise, Prasad et al. [25] identified clusters based on the socioeconomic level of the occupants in the city of Dehradun in India to develop a risk assessment. The authors conclude that people with low income are exposed to higher seismic risk due to the high vulnerability of their dwellings. Gunasekera et al. [26] developed a global exposure model based on clustering analysis using satellite images (with 1 km^2 resolution) to support the generation of country disaster risk profiles. The IDB have used clustering techniques in projects in Ecuador related to earthquake [27] and landslide [28] hazards, using the risk assessment results as the clustering metric. This approach has also been used to develop risk assessments for other types of hazards such as flooding [29] or coastal erosion [30].

Clustering algorithms are based on different approaches, e.g. *K*-means, *K*-modes, Mean Shift, Expectation-Maximization using Mixture Models, and Hierarchical clustering, among others [31]. The choice for the most suitable approach depends on the characteristics of the data and the purpose of the study. For instance, the *K*-means algorithm is one of the most used due to its simplicity in the implementation [32]. For this algorithm, the first assumption is the number of possible clusters, namely *K*, and the selection of *K* elements as centroid of each cluster. Then, the distance to each one of these centroids is calculated for each element, to assign the "closest" one to each, finding with this a first distribution of clusters. The next step is to recalculate the centroid of the cluster, assigning it based on the mean characteristics of the cluster. This process is repeated iteratively until convergence to find the final distribution of clusters.

However, the implementation of the *K*-means requires to have objects described by continuous numerical parameters, such as distances or income. One distinctive aspect of the present study is that the objects (e.g., buildings) are classified based on categorical data, such as the building material or the structural system. For this reason, we need clustering algorithms designed to work with categorical and binary datasets. The *K*-modes algorithm, adapted from the *K*-means algorithm [33], is a possible choice. In this algorithm the iterative process to find the clusters is similar as the one in the *K*-means algorithm with the modification of using the mode of the cluster to recalculate the centroid rather than using the mean. However, this algorithm is sensitive to the initialization step, so different implementations could lead to different results, even on the same data [34].

As an alternative to the *K*-modes, the Hierarchical Clustering Algorithm (HCA) also works with different types of data and uses different distance metrics, yet it is not sensitive to the initialization step. Two possible approaches can be followed: agglomerative and divisive. The first assumes as starting point as many clusters as the number of elements in the dataset and proceeds with successive cycles of aggregation based on a distance metric until all data are in one cluster. On the other hand, the divisive approach starts with all elements in one cluster and segregates sets based on a distance metric until each element belongs to one cluster. This algorithm has the advantage that there is no need to select a *K* number prior to the implementation, the user can select it after the implementation since all the possible distributions of clusters are calculated. One of the earlier developments of HCA by Day & Eldesbrunner [35], presented a number of dissimilarity measures such as the Euclidean, Manhattan, Chebychev, and Minkowsky metrics. The literature shows reliable results of HCA when compared to other algorithms and in applications with different distance metrics [36–38]. One of the main advantages of HCA is how the dissimilarity metric can be adapted to any type of data. Despite its flexibility, HCA has one characteristic that may be considered a drawback in some applications. As no assumptions are required for the number of clusters, the algorithm presents the results as a dendrogram and leaves the decision to the user. This characteristic can be positive in many contexts, but with large datasets the algorithm may become less sensitive, data-order dependent, and inefficient, particularly for categorical variables [39,40].

Besides the *K*-modes and the HCA, a third clustering algorithm that handles categorical data is the Bernoulli Mixture Model (BMM). The mixture models are used to find the underlying probability distributions of the data. In the BMM case, the distribution is known to follow a Bernoulli distribution, and therefore the objective of the method is to find its parameters. This algorithm was developed initially to work with data in a binary representation, making the algorithm particularly suitable for text categorization [41] and binary images [42]. The most common algorithm for the implementation of the BMM is the Expectation-Maximization (EM) algorithm

[43]. Due to its efficiency and flexibility, the BMM is chosen as the clustering algorithm in this study, and further details on its application to the problem of clustering buildings is provided in section 3.2.

3. Proposed methodology

The main objective of the proposed methodology is to assign a schematic retrofitting option to each building in a school buildings portfolio and prioritize them based on their risk reduction and cost. Although the focus is on seismic retrofitting as the disaster risk reduction strategy, our methodology also applies to other types of hazards. The first step is to determine the taxonomy coding –in this case the GLOSI taxonomy string– for all the school buildings in the portfolio, which could be an expensive and time-consuming task [18]. If the information is not readily available for all buildings in the portfolio, one option is to use allocation algorithms or proxy models to fill the missing data [18], at the expense of reducing reliability in the retrofitting strategy assignment. This dataset will be the main input of the model. The next step is to apply the clustering algorithm, namely the BMM, on the buildings data set. The conditions under which the BMM clustering is applied should be such that the resulting clusters are meaningful from a structural point of view. Once the clusters are formed, the most common taxonomy string in each cluster is selected and a *representative building* is assigned to each cluster. In the next step, a structural model is produced for each representative building. This structural model is used to determine the collapse mechanisms, assign the vulnerability function, and identify a suitable retrofitting strategy. Then, the vulnerability function is computed again for the retrofitted model and the risk reduction is computed as the difference between the risk before and after the strengthening, given the same earthquake hazard profile. With this information, along with the cost of the implementation, it is possible to design a regional risk reduction strategy to prioritize the interventions, that will be the main output of the method. Fig. 1 summarizes the steps of the proposed methodology.

As the focus of this paper is the design and application of the clustering algorithm, the GLOSI taxonomy is illustrated in section 3.1 as a necessary step to prepare the clustering algorithm in section 3.2. Although the scope of the study is neither an exhaustive review of structural retrofitting strategies, nor the detailed development of a specific retrofitting strategy for the case study, for the sake of completeness, section 3.3 includes a review of the relevant international guidelines and regulations for structural seismic retrofitting.

3.1. GLOSI seismic taxonomy

The GLOSI taxonomy was developed within the Global Program for Safer Schools (GPSS) framework [16,18,17]. Table 1 presents the summary of the GLOSI parameters and their attributes. A total of 12 taxonomy parameters are divided into two categories: primary and secondary. The primary category includes the main structural system, the height range, and the seismic design level, which are sufficient to determine basic deficiencies. The second category include parameters that affect specific failure modes. For example, a building with structural irregularities or a flexible diaphragm will be more prone to localized failures than a building without irregularities or a stiff diaphragm. The first step of the proposed methodology is to identify the taxonomical string of each building, i.e., to choose the specific attribute representing that building for each of the parameters. For example, a reinforced concrete building with moment resisting frame and short column, two stories, low design level, rigid diaphragm, no irregularities, short span, regular column, rigid foundation, no pounding risk, no retrofitting, in good condition, and without vulnerable non-structural elements, is coded by the taxonomy string RC3/MR/LD/RD/NI/SS/RO/RF/NP/OS/GC/NN (see Fig. 2).

3.2. Clustering algorithm

The first step before applying the clustering algorithm is data preparation. As each building is coded by a taxonomy string (see

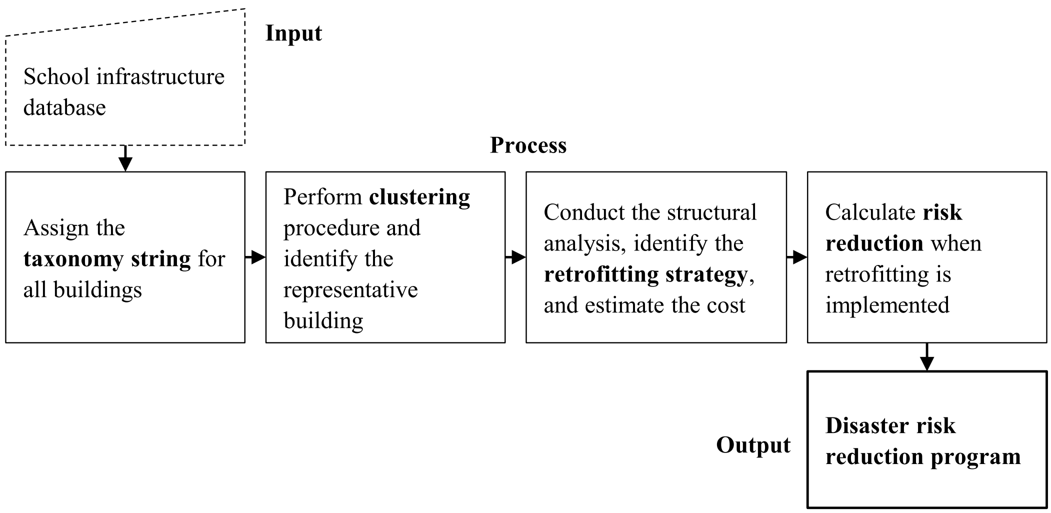

Fig. 1. Proposed methodology.

Table 1
GLOSI taxonomy parameters and attributes.

No.	Parameter	Attributes	
1	Main structural system	For load bearing masonry (LBM) buildings: **URM1 - URM7 -** Unreinforced masonry **A -** Adobe **CM -** Confined masonry with a rectangular block in cement mortar **RM -** Reinforced masonry with a rectangular block in cement mortar	For reinforced concrete (RC) buildings: **RC1 -** Bare framed construction **RC2 -** Infilled framed construction **RC3 -** RC framed construction with short columns **RC4 -** Dual or combined framed construction **RC5 -** Non-engineered framed construction
2	Height range	**LR -** Low-rise **MR -** Mid-rise **HR -** High-rise	
3	Seismic design level	**PD -** Poor design **LD -** Low design **MD -** Medium design **HD -** High design	
4	Diaphragm type	**FD:** Flexible diaphragm **RD:** Rigid diaphragm	
5	Structural irregularity	**NI -** No irregularities **HI -** Horizontal **VI -** Vertical **HV -** Both horizontal and vertical	
6	Wall panel length/Span length	LBM: **SP -** Short panel **LP -** Long panel	RC: **SS -** Short span **LS -** Long span
7	Wall Openings/Pier type	LBM: **SO -** Small openings **LO -** Large openings	RC: **RO -** Regular column **SW -** Slender-weak column
8	Foundation type	**FF -** Flexible foundation **RF -** Rigid foundation	
9	Seismic pounding risk	**PR -** Pounding risk **NP -** No pounding	
10	Effective seismic retrofitting	**OS -** Original structure **RS -** Retrofitted structure	
11	Structural health condition	**PC -** Poor condition **GC -** Good condition	
12	Non-structural components	**VN -** Vulnerable non-structural components **NN -** Non-vulnerable non-structural components	

Fig. 2. RC3/MR/LD/RD/NI/SS/RO/RF/NP/OS/GC/NN building.

§3.1), we convert them into binary vectors using one-hot encoding [44]. In this process, the taxonomy string (12-attribute combination) is translated into a binary vector of 40 positions representing all the 40 possible attributes in the GLOSI taxonomy string. For each position of the vector, a 1 specifies that the attribute is present in the taxonomy string; and a 0 represents absence. This encoding is represented in Fig. 3.

Once buildings are encoded into binary vectors, then the clustering algorithm follows. In this methodology we propose the Bernoulli Mixture Model (BMM) implemented with the Expectation-Maximization (EM) algorithm, henceforth labeled EM-BMM [45]. The EM algorithm is an iterative process of the maximum likelihood estimation with latent variables, demonstrated to be effective for the BMM [46]. The BMM is particularly appropriate on binary vectors, since there are only two possible attributes in each position. The EM-BMM starts with random Bernoulli distribution parameters for each possible cluster and starts a loop of an expectation step (E) and a maximization step (M). In the first step, the algorithm finds the likelihood of each element to follow the previously defined Bernoulli distribution and assigns the element to a cluster. Then, the second step computes a new distribution based on the new clusters formed in the previous step. These two steps iterate until the clustering process converges. For further details of the EM-BMM methodology the reader is referred to Refs. [45,46].

To implement the above-mentioned clustering algorithm, we need to specify the number of clusters in which we want to divide the data, namely, K. To find this number, an iterative process starts specifying a small number for K such as 2. Then, the clustering algorithm runs and the decision process analyzes if the cluster distribution is logical, interpretable, and acceptable for the domain experts. If this is not the case, then the number of clusters should be increased ($K = K + 1$) and the iterative process repeats. In general, a good practice is to stop at the smallest number of clusters K that generate an acceptable distribution of buildings within the clusters. The acceptability of the distribution requires input from technical structural experts, who evaluate if the typologies included in the same cluster can be intervened with the same strategy. For instance, if two buildings with different structural systems, such as Adobe (A) and Unreinforced masonry (URM), have the same structural deficiencies: poor design level, a flexible diaphragm, and a poor structural health condition, they could be grouped in the same cluster due to the similarity of their taxonomical strings, and the intervention for this cluster could be the replacement for new buildings. Conversely, a cluster including two buildings with CM and an RM structural systems, with similar secondary parameters (low-design level, rigid diaphragm, no pounding risk, bad structural health condition, and non-vulnerable non-structural elements), could be included in the same cluster. In this group, the intervention could be designed to target the main deficiency of the structural system that could be improved by the implementation of jacketing or splint and bandage, but there will be no need to install a ring beam (see next section). The maximum number of clusters to be considered, K_{max}, is the number of unique taxonomy strings in the dataset. This case is certainly not desirable, but possible if the set of buildings is highly heterogenous. Fig. 4 presents a schematic overview of this process.

3.3. Retrofitting principles

Once we have assembled the building clusters, it is necessary to design a retrofitting strategy for each representative building, applicable to all members of the clusters. Presenting all the possible retrofitting strategies is beyond the scope of this paper, but it is important to highlight that retrofitting options should be designed based on national or international guidelines and regulations, such as ASCE 41–17 [47], FEMA E−74 [48], FEMA 308 [49], EUROCODE 8 [50] or the International Building Code [51], among others, to ensure the quality and the applicability of the strategy. The identification of possible mitigation and intervention options must account for the expected sequence of damages and failures for each representative building. Intervention options shall consider technical, economic, social, and political feasibility. Possible interventions may include the replacement of the entire building (due to important deficiencies that retrofitting would not correct or improve), roof intervention, foundation retrofitting, structural retrofitting, non-structural components intervention, or any other type of intervention depending on the construction type. Table 2 presents a compendium of different deficiencies and structural retrofitting strategies limited to masonry and reinforced concrete buildings. These interventions have been successfully applied in school buildings in Peru [52,53], Indonesia [54], Cyprus [55], Lebanon [8], Nepal [56], Italy [57], Portugal, and Spain [58], among others. The costing of these retrofitting measures should include materials, workmanship, indirect costs, and demolition costs, when needed. The total cost of the retrofit shall be normalized to the replacement cost of the building.

The intervention choice should be made by structural engineers with knowledge of the construction types and after a detailed analysis of the buildings. However, a preliminary selection of possible techniques could be made using the taxonomy string. For instance, an LBM building with poor design level (parameter 3) could be improved with the implementation of jacketing or splint and bandage (depending on characteristics of the building). This same building with a flexible roof (parameter 4) could also be improved with the implementation of a horizontal band beam. If the building has a poor structural health condition (parameter 11), the retrofitting strategy may also include injection technique as part of the retrofitting strategy. Some interventions could be implemented at

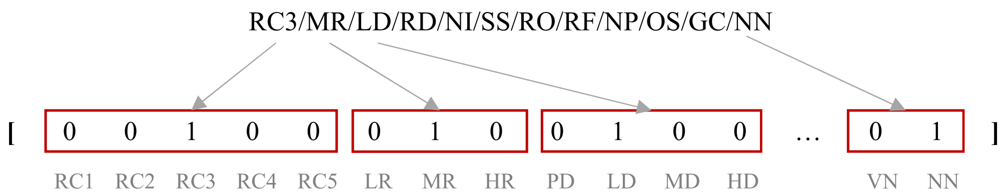

Fig. 3. Illustrative example of the one-hot encoding process using the GLOSI taxonomy.

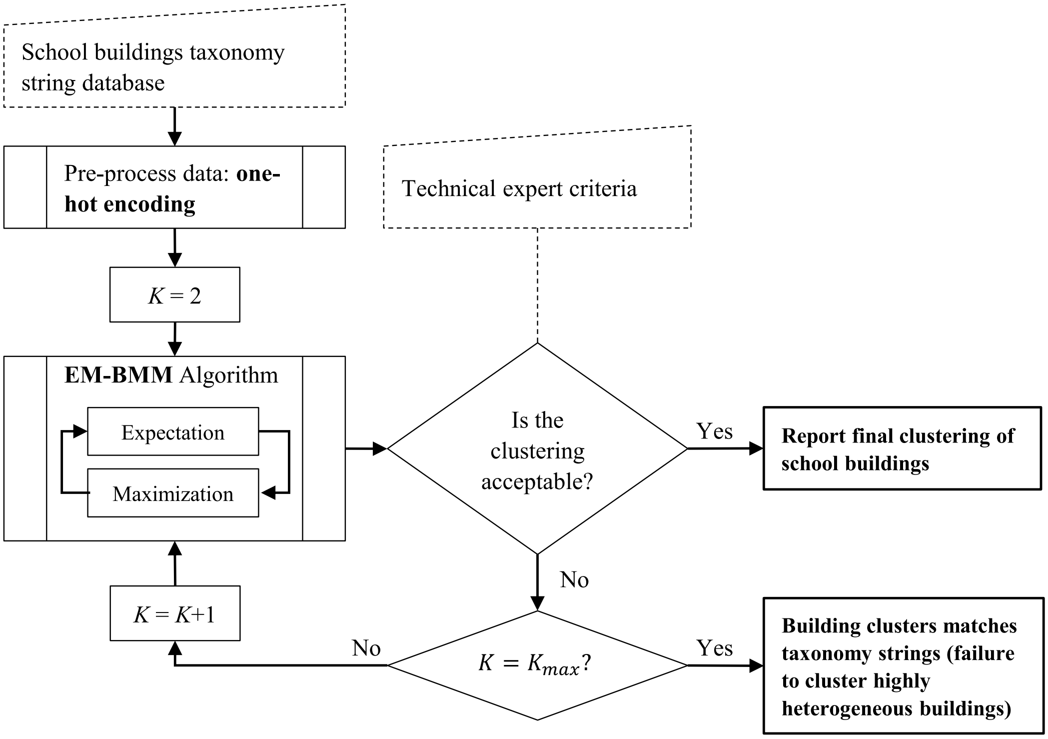

Fig. 4. EM-BMM clustering algorithm.

the same time in a building, such as the ones presented above, or an infill isolation and steel braced frames for RC buildings. In these cases, each intervention will target different deficiencies at the same time. This list of interventions is presented as a reference, it is not exhaustive, and additional interventions could be considered depending on each specific case. Note that the strengthening methods are those reported in the literature, yet not independently verified by the authors. Once a given retrofitting intervention method is chosen, it is still necessary to customize the design of these interventions for a real school building in a later stage.

4. Case study

We illustrate the proposed methodology by applying it to the Dominican Republic public-school infrastructure. Details of the school portfolio are provided in section 4.1, followed by the taxonomic string assignment in section 4.2, clustering in section 4.3, choice of the retrofitting strategies in section 4.4., comparison of alternative risk reduction programs in section 4.5 and finally, implications for decision makers in section 4.6.

4.1. School infrastructure database

This portfolio includes 6,087 school facilities with 18,820 buildings which host approximately 1.5 million students (see Fig. 5). We use data gathered in a study funded by the World Bank and conducted jointly by Universidad de Los Andes and University College London in 2020 [21]. The GLOSI taxonomy string assignment to each school was underpinned by field surveys, online surveys to the school directors, satellite images, secondary data from previous surveys, and existing information available through the Ministry of Education database. For un-surveyed buildings, we used secondary proxy data to classify them using an allocation algorithm [21].

4.2. Descriptive analysis of the seismic taxonomic strings

Table 3 presents the distribution of the taxonomy strings in the portfolio. To account for all typologies existing in the Dominican Republic, we introduce a new mixed structural system, namely JRM, consisting of reinforced masonry walls and a central RC frame in the long direction of the building, being the fifth most common taxonomy in the country (the J is included since this typology is locally known as Johnson). The most common school building typology is a reinforced concrete moment resistant frames with short column (RC3), mid-rise, with a low-design level (see Table 1); the second most common are the RC3, low-rise, low-design buildings; and the third most common are the reinforced masonry, low-rise with low-design level buildings. In total, there are 17 unique taxonomy strings in this portfolio of school buildings.

Since the main objective of this methodology is to undertake large-scale school intervention programs, it is necessary to understand the similarities and differences between these 17 seismic taxonomic strings and, based on this, to group them into clusters of buildings

Table 2
Generic retrofitting strategies.

Material	Retrofitting strategy	Description	Structural deficiencies	References
Load Bearing Masonry	Horizontal Band Beam	The horizontal beam installed over all the walls (including over the gables). Usually made of RC.	Limited capacity due to independent behavior of individual walls; Out of plane failure of walls; Corner separation; Excessive deformation. Roof structure failure	[59,60]
	Column Tie	Vertical columns installed usually in the middle or in the corners of the walls. Usually made of RC.	Out of plane failure of walls; Excessive deformation; Poor connection between cross walls.	[61,62]
	Injection Technique	Cement based grout or epoxy resins used as the injection material.	Limited tensile strength, cohesion, and friction due to poor quality or deteriorated materials; Localized failure of walls; Existing cracks; Poorly connected wythes in multi-wythes walls	[49,63–66]
	Jacketing	Local type of intervention technique usually applied on both sides (inner and outer) of the walls.	Limited shear and flexural capacity of walls due to poor quality of materials; Localized failure of walls; Poor connection between cross walls	[56,67]
	Splint and Bandage	Local intervention similar to jacketing but applied in the form of vertical column (splint) and/or horizontal beam (bandage) usually applied on both sides (inner and outer) of the walls.	Limited shear and flexural capacity of walls due to poor quality of materials; Localized failure of walls; Poor connection between cross walls	[56,67–69]
	Roof Strengthening	Intervention on horizontal structure applied to poorly built roofs lacking in plane stiffness.	Timber or Steel roof structure with rafters and purlins only (no diagonal bracing); Poorly connected joints; Limited in-plane strength/ stiffness of roof structure	[68]
Reinforced Concrete	Infills Isolation	Isolation of non-structural masonry walls from the structure and posterior retrofitting to eliminate the out of plane failure.	Soft story; Captive column.	[47,48]
	Reinforced Concrete Walls	RC walls in existing bays to provide stiffness to the building and eliminate the short column (or weak story) collapse mechanism.	Excessive building flexibility; Soft story; Captive column; Low horizontal capacity and resistance.	[47,48,52]
	Steel Braced Frames	Steel braced frames in bays to provide stiffness to the building and eliminate the short column (or weak story) collapse mechanism.	Excessive building flexibility; Soft story; Captive column; Low horizontal capacity and resistance.	[47,48,53, 70]
	Reinforced Concrete Buttress	RC buttress in the exterior to provide stiffness to the building and eliminate the short column (or weak story) collapse mechanism.	Excessive building flexibility; Soft story; Captive column; Low horizontal capacity and resistance.	[47,48]

with similar characteristics and deficiencies. Even though the first three taxonomy strings account for around 70% of the population of buildings, with this process we aim to understand if these could be included in the same cluster or if they should be separated and how the remaining set of taxonomic groups could be distributed. Correspondingly, the number of clusters (K) to be found using the methodology (see §3.2) should be a value between 2 and 17.

4.3. Clustering

Table 4 presents the clusters' distribution obtained with the EM-BMM algorithm for different values of K. We do not present the results for lower numbers than five since their distribution is too coarse and lacks interpretability. The algorithm finds a relatively stable number of clusters for the first two iterations ($K = 5, 6$), with the largest cluster having more than 50% of the buildings. This large cluster, containing more than half of the buildings, includes the following two typologies: RC3/MR/LD/RD/NI/SS/SW/RF/NP/OS/PC/VN and RC3/LR/LD/RD/NI/SS/RO/RF/NP/OS/PC/VN. These two taxonomy strings include the RC3 structural system that generates short column failure. However, generally this effect is more critical in buildings with two or more stories [71]. Also, the strings differ in the story height and in the fact that the first string include a weak column effect (SW), opposite to the second (RO). Considering this, one can state that these two strings have different structural behavior and therefore different retrofitting strategies. The buildings with the first taxonomy string could be retrofitted with the isolation of infill walls and a stiffening intervention, such as the steel braced frames. Conversely, the buildings with the second taxonomy string will only need the isolation of the infill walls, since they are one story, and the RC frames are stiff enough (this applies to this case study considering the local construction characteristics, but is not applicable to all RC3 low rise [21]). Considering the composition of the cluster and the types of retrofitting, this large cluster is not viable. With seven or more clusters ($K \geq 7$), the largest cluster splits into smaller ones. Note also that the first three clusters remain stable after $K = 7$. This occurs since these three clusters are each comprised by only one taxonomy string, which is consistent with the first three taxonomy strings presented in Table 3, while greater values of K lead to the fragmentation of the smaller clusters (where the dissimilarity between the taxonomies inside is larger). The next step is to analyze the clusters generated for $K = 7$ to understand if their buildings composition allows for the same strengthening provisions.

Table 5 presents the clusters' composition obtained from the EM-BMM algorithm for $K = 7$. Each line contains the cluster distribution as a percentage of the total buildings of each taxonomy string (showing only the first three attributes). The most common eight typologies are included explicitly in the table while the others are included in one column as "Other". The final column indicates the percentage of buildings included in each cluster.

The first two clusters are consistent from a structural point of view since JRM-LR-LD and RM-LR-LD buildings are included in

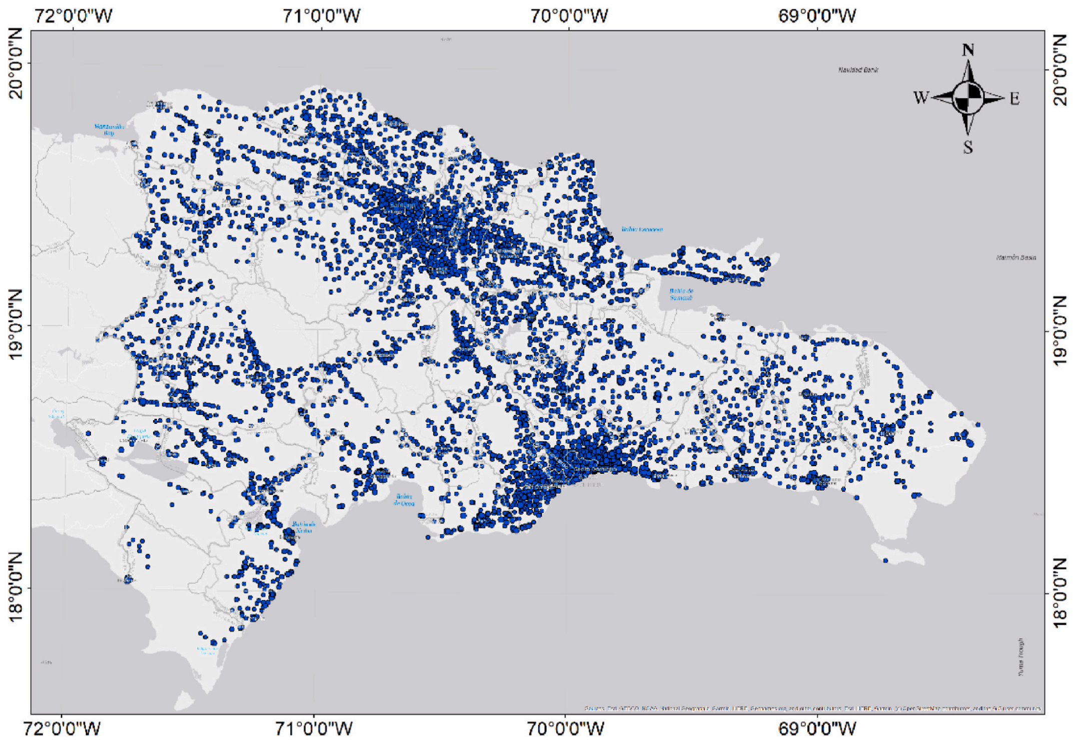

Fig. 5. Geographical distribution of school facilities in the Dominican Republic.

Table 3
Main typologies in the Dominican Republic.

Seismic taxonomy strings	Number of buildings	Participation in portfolio (%)
RC3/MR/LD/RD/NI/SS/SW/RF/NP/OS/PC/VN	4,653	25.87%
RC3/LR/LD/RD/NI/SS/RO/RF/NP/OS/PC/VN	4,597	25.56%
RM/LR/LD/FD/NI/SP/LO/RF/NP/OS/GC/NN	3,736	20.77%
RC1/MR/HD/RD/NI/SS/RO/RF/NP/OS/GC/NN	1,516	8.43%
JRM/LR/LD/RD/NI/LP/SO/FF/NP/OS/PC/VN	1,102	6.13%
RC1/LR/HD/RD/NI/SS/RO/RF/NP/OS/GC/NN	803	4.46%
OT1/LR/HD/RD/NI/SS/RO/RF/NP/OS/GC/NN	697	3.88%
TF/LR/PD/FD/NI/SP/LO/FF/NP/OS/PC/VN	281	1.56%
OT2/MR/PD/FD/NI/SP/LO/FF/NP/OS/PC/VN	262	1.46%
RM/LR/PD/FD/NI/SP/LO/FF/NP/OS/PC/VN	248	1.38%
OT1/LR/LD/RD/NI/SS/SW/RF/NP/OS/PC/VN	30	0.17%
OT3/LR/PD/RD/NI/SS/RO/RF/NP/OS/PC/VN	29	0.16%
OT2/LR/PD/FD/NI/SP/LO/FF/NP/OS/PC/VN	18	0.10%
OT3/MR/PD/RD/NI/SS/RO/RF/NP/OS/PC/VN	6	0.03%
OT1/LR/PD/RD/NI/SS/SW/RF/NP/OS/PC/VN	4	0.02%
OT2/LR/LD/FD/NI/SP/LO/FF/NP/OS/PC/VN	2	0.01%
OT1/MR/LD/RD/NI/SS/SW/RF/NP/OS/PC/VN	1	0.01%

different clusters with no other strings. Cluster ID # 3 includes RM-LR-PD and TF-LR-PD buildings. The entire strings in both cases include a flexible diaphragm (FD), flexible foundation (FF), poor condition (PC) and vulnerable non-structural elements (VN), which make them similar enough to be included in the same cluster. Considering this, both are highly vulnerable buildings, and therefore should be replaced with new ones, so it is an appropriate and valid cluster. Similarly, Cluster ID # 4 includes RC1-LR-HD and RC1-MR-HD buildings, which correspond to newer buildings. These buildings are both high design (HD), have a rigid diaphragm (RD), rigid foundation (RF) and have a good structural health condition (GC). Also, these buildings were built after 2012 and therefore already comply with the current building codes, so should not be intervened. Cluster ID # 5 includes the RC3-LR-LD while Cluster ID # 6 the RC3-MR-LD. As stated before, the separation of these buildings is critical since the retrofitting for the RC3-LR-LD should be less extensive than the retrofitting for RC3-MR-LD (as stated above). This result shows that the EM-BMM algorithm presents valid and interpretable results with $K = 7$.

Table 6 presents the resulting clusters and their corresponding descriptions. To summarize, Cluster 1 mainly comprises the school buildings represented by the main attributes JRM/LR/LD, which are the *older reinforced masonry* buildings. Cluster 2 includes all the

Table 4
EM-BMM clusters' distribution and weight in the building portfolio.

	Number of clusters (*K*)					
	5	**6**	**7**	**8**	**9**	**10**
Clusters	51,4%	51,4%	25,9%	25,9%	25,9%	25,9%
	20,8%	20,8%	25,6%	25,7%	25,6%	25,6%
	17,0%	17,0%	20,8%	20,8%	20,8%	20,8%
	6,3%	6,1%	16,8%	16,8%	12,9%	12,9%
	4,5%	4,5%	6,1%	6,1%	6,1%	6,1%
		0,2%	4,5%	4,5%	4,5%	4,5%
			0,4%	0,2%	3,9%	3,9%
				0,0%	0,4%	0,4%
					0,0%	0,0%
						0,0%

Table 5
Clusters composition for K = 7.

Cluster ID	JRM-LR-LD	RM-LR-LD	RM-LR-PD	RC1-LR-HD	RC1-MR-HD	RC3-LR-LD	RC3-MR-LD	TF-LR-PD	Other	Total
1	100%	0%	0%	0%	0%	0%	0%	0%	0%	**6,1%**
2	0%	100%	0%	0%	0%	0%	0%	0%	0%	**20,8%**
3	0%	0%	100%	0%	0%	0%	0%	100%	27%	**4,5%**
4	0%	0%	0%	100%	100%	0%	0%	0%	67%	**16,8%**
5	0%	0%	0%	0%	0%	100%	0%	0%	0%	**25,6%**
6	0%	0%	0%	0%	0%	0%	100%	0%	0%	**25,9%**
7	0%	0%	0%	0%	0%	0%	0%	0%	6%	**0,4%**

RM/LR/LD buildings, corresponding to the *newer reinforced masonry*. Cluster 3 is composed of RM/LR/PD and TF/LR/PD, which are *non-engineered* buildings. Cluster 4 groups the RC1/LR/HD and RC1/MR/HD which are both *RC high seismic* design buildings. Cluster 5 considers all the RC3/LR/LD buildings, which are *low-rise short column*. Cluster 6 includes the RC3/MR/LD buildings, which are *mid-rise short column*. Finally, Cluster 7 includes different buildings in the category *others*. Considering this allocation, we named each cluster according to the general characteristics of the cluster.

4.4. *Retrofitting strategies*

The next step is to analyze the representative building for each cluster. This analysis is the basis for defining the retrofitting strategy and its economic valuation, which is later applied to all buildings in the same cluster. For the sake of conciseness, we present in detail the retrofitting strategy for clusters *newer reinforced masonry* and *mid-rise short column*, and only a summary for the rest. These strategies were developed based on structural drawings, surveys, and a field visit. It is important to note that the process presented below corresponds to a pre-feasibility study and does not replace the specific structural design, which should be done in a posterior step.

For the *newer reinforced masonry*, the following structural deficiencies are identified: low seismic design (low shear capacity, poor connections, out-of-plane failure due to inadequate internal reinforcement), flexible diaphragm, and low quality of materials. The proposed retrofitting of these buildings consists of installation of splints (form the splints and bandage technique [56, 67–69]) and tie columns in the shortest walls [61,62] (see Table 2). The splints should be installed mainly in the corners and connections between walls to improve the shear capacity and reduce the out-of-plane deformation. Fig. 6 presents a general scheme of the proposed retrofitting strategy. From a budget analysis, including the materials, the workmanship, the indirect costs, and the demolition costs, we estimate the cost of this retrofitting at US$ 120 per square meter, corresponding to about 25% of the total replacement cost (US$ 480 per square meter).

Fig. 7 presents the corresponding vulnerability functions for the original and retrofitted condition, normalized to the replacement value, developed with the GLOSI approach [72,73]. From these results, it is possible to note the reduction in vulnerability with the retrofitting. For example, for a spectral acceleration of 0.5 g, the mean damage ratio reduces from 60% to 20%, showing the tangible effect of retrofitting.

For the *mid-rise short column* cluster, we identify the following structural deficiencies: short column, excessive flexibility, weak story, and strong beam–weak column condition. The proposed structural retrofitting consists of the isolation from the structure of non-structural walls [47,48] and the installation of steel diagonal bracing on the first and second levels of the buildings [47,48,53,70] (infill walls isolation and steel bracing frames in Table 2). Fig. 8 presents the schematic representation of a retrofitted building for this cluster (in red the steel braced frame and in blue the gap between the infill and concrete frame). Similarly, as in the previous case, we estimate the retrofitting cost in US$ 140 per square meter, corresponding to 30% of the total replacement cost.

Table 6
Resulting cluster description.

Cluster ID	Cluster name	Main taxonomy strings	Representative image
1	Older reinforced masonry	JRM/LR/LD/RD/NI/SP/LO/RF/NP/OS/PC/NN	
2	Newer reinforced masonry	RM/LR/LD/FD/NI/SP/LO/RF/NP/OS/PC/VN	
3	Non-engineered	TF/LR/PD/FD/NI/SP/LO/FF/NP/OS/PC/VN RM/LR/PD/FD/NI/SP/LO/FF/NP/OS/PC/VN	
4	RC High-design	RC1/MR/HD/RD/NI/SS/RO/RF/NP/OS/GC/VN RC1/LR/HD/RD/NI/SS/RO/RF/NP/OS/GC/NN	
5	Low-rise short column	RC3/LR/LD/RD/NI/SS/SW/RF/NP/OS/PC/VN	
6	Mid-rise short column	RC3/MR/LD/RD/NI/SS/SW/RF/NP/OS/GC/VN	
7	Others	Others	

Fig. 9 presents the vulnerability functions for the original and retrofitted conditions. The reduction of vulnerability achieved with this retrofitting is even higher than the one obtained in the previous cluster. Indeed, for a spectral acceleration of 1.0 g, the mean damage ratio reduces from over 80% to around 20%. Also, for the retrofitted case, collapse is prevented for spectral acceleration up to 2 g.

For the *older reinforced masonry* cluster, we identified the following structural deficiencies: low seismic design (low shear capacity, poor connections, out-of-plane failure), and low quality of materials. The proposed retrofitting for strengthening these buildings consists of the installation of splints cement plaster elements applied on both sides (inner and outer) of the walls [60]. For the *low-rise*

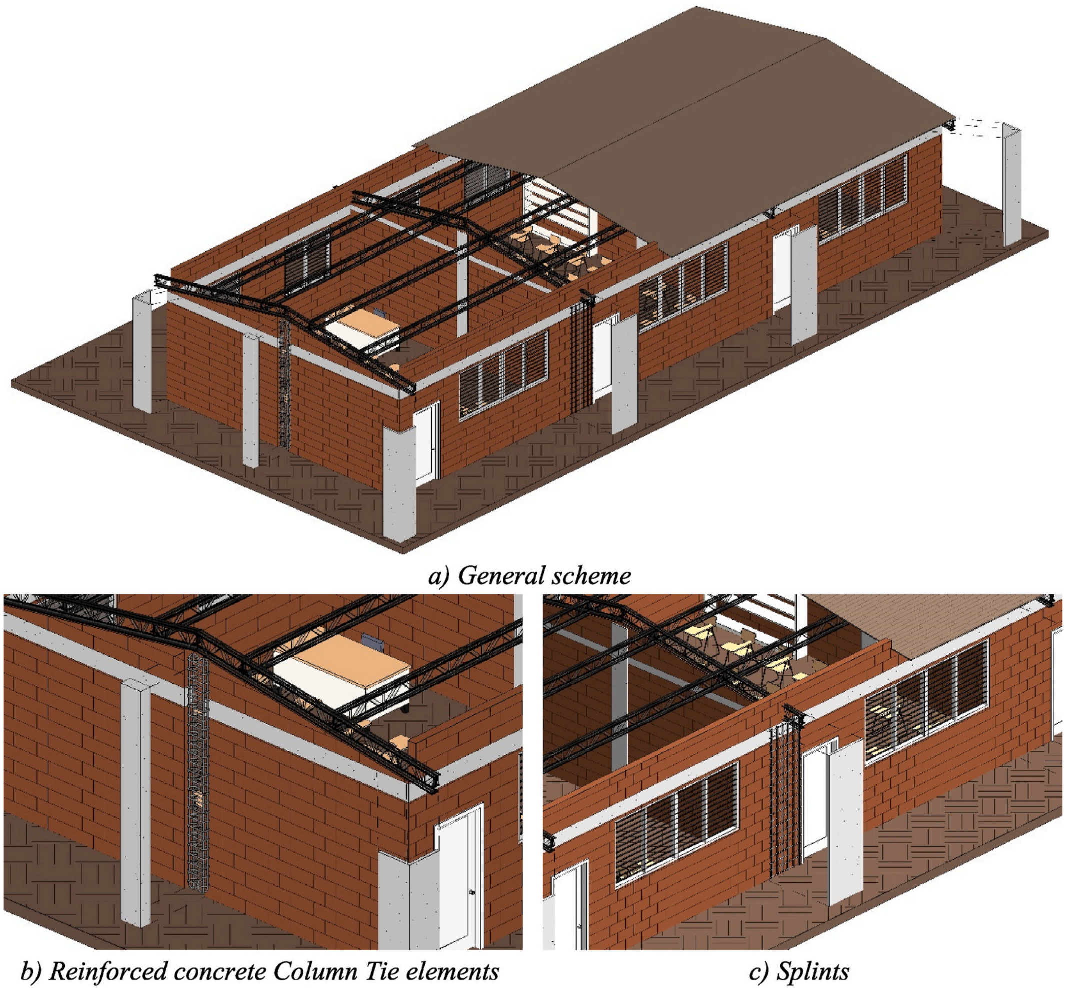

Fig. 6. Schematic representation of the retrofitting strategy for the *newer reinforced masonry* cluster.

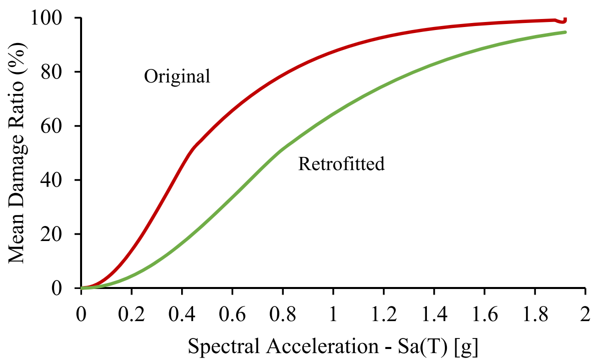

Fig. 7. Vulnerability functions for the *newer reinforced masonry* cluster (original vs. retrofitted).

short column cluster, we identified the short column effect as the main structural deficiency, which is due to the presence of masonry walls attached to the main structural system. The proposed retrofitting strategy consists of the isolation of the internal and external infills walls (infills isolation) [47,48]. For the *RC high-design* cluster, we did not identify structural deficiencies and therefore no intervention is planned. For the *non-engineered* cluster, the intervention is the replacement of the buildings. The reason behind this recommendation is because in this group there is high variability in the structural characteristics and the cost of retrofitting can be as high as the cost of the replacement. Lastly, for the o*thers* cluster, we do not assign any intervention since there are heterogeneous and

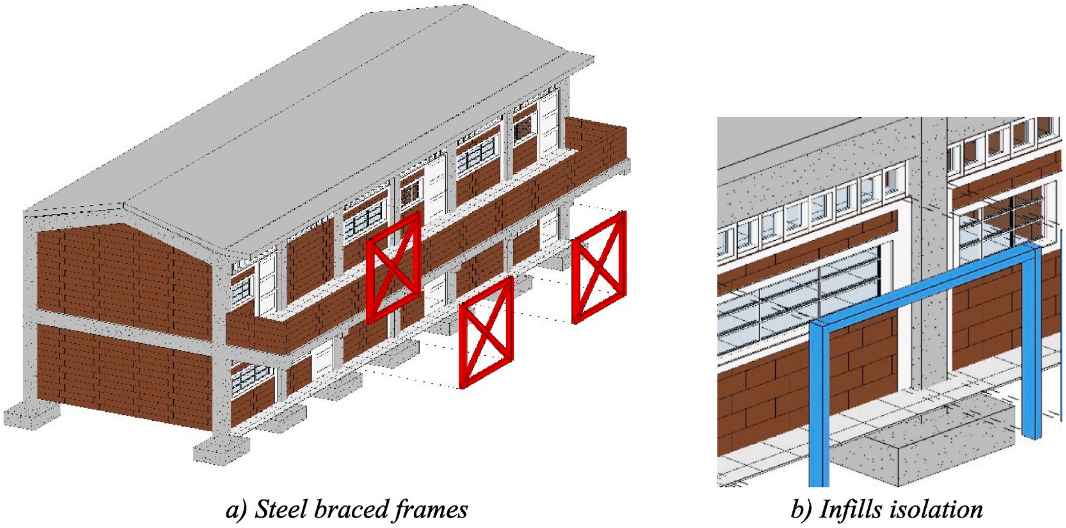

Fig. 8. Illustrative schemes for *mid-rise short column* retrofitting.

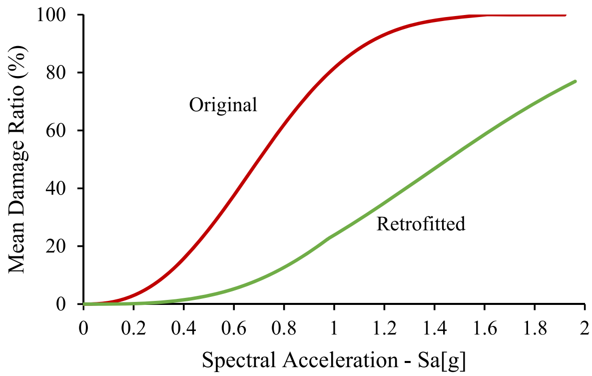

Fig. 9. Vulnerability functions for the *mid-rise short column* cluster (original vs. retrofitted).

no large-scale intervention can be planned. It is important to note that this does not mean that buildings in this cluster are in good condition, this means that further investigation needs to be done building by building to determine their structural requirements. To summarize, in Table 7 we present the structural deficiencies, the retrofitting strategy, and estimated costs for each cluster. It is relevant to note that the estimated costs are only applicable to this case study and cannot be globally extrapolated.

4.5. *Risk reduction programs*

The generic retrofitting strategies identified for each cluster can be implemented in all buildings at a national level. The cost of implementing the retrofitting strategies for the whole portfolio corresponds to approximately US$ 440 million (approximately US $25,000 per building). This blank strategy might not be cost-effective, since it does not account for geographic variation of hazard, and therefore of risk [74]. A more efficient investment could follow a prioritization strategy considering the risk reduction level quantifiable through a risk assessment [75]. A wide range of tools and open software for risk assessment are available, such as the CAPRA platform [76], OASIS loss modeling framework [77], and SELENA [78], among others. This case study uses the CAPRA platform, the national exposure model in §4.1, and the vulnerability functions in §4.4. The hazard model (stochastic catalogue of seismic events) is the one developed by the World Bank, Universidad de Los Andes, and University College London [21]. This catalogue includes 8,710 scenarios, with a minimum Mw of 5.0 and a maximum Mw of 8.2.

From the risk assessment, we calculate the expected annual losses (EAL) with at the resolution of the individual building. In the current condition, the sum of EAL for all buildings is US$ 12.16 million. After implementing the retrofitting strategy in all buildings nationwide, the EAL reduces to US$ 3.81 million. The difference, known as the expected annual benefits (EAB), is about $ 8.35 million.

The expected annual benefit (EAB) can be also calculated at the building level as the difference of the EAL between the original and retrofitted conditions. This information is used to determine the prioritization strategy, by ordering the buildings from the highest to the lowest EAB. Fig. 10 shows the relation between the cumulative cost of interventions and the cumulative benefits in terms of the EAB. These results are normalized by the total cost of interventions (US$ 440 million) in the horizontal axis and by the maximum EAB

Table 7
Vulnerability function and retrofitting strategy for each cluster.

Cluster ID	Cluster name	Structural deficiencies	Retrofitting strategy (Table 2)	Estimated cost (as percentage of the replacement cost)
1	Older reinforced masonry	➢ Low shear capacity ➢ Weak connections ➢ Out-of-plane failure ➢ Low material quality	• Splints	16%
2	Newer reinforced masonry	➢ Lack of rigid diaphragm ➢ Low shear capacity ➢ Weak connections ➢ Out-of-plane failure ➢ Low material quality	• Splints • Tie column	25%
3	Non-Engineered	➢ High seismic vulnerability	• Replacement	100%
4	RC High-design	➢ No deficiencies	• No intervention	–
5	Low-rise short column	➢ Short column ➢ Weak story	• Infills isolation	5%
6	Mid-rise short column	➢ Short column ➢ Low stiffness ➢ Weak story ➢ Strong beam weak column	• Infills isolation • Steel braced frames	30%
7	Others	–	• Not considered	–

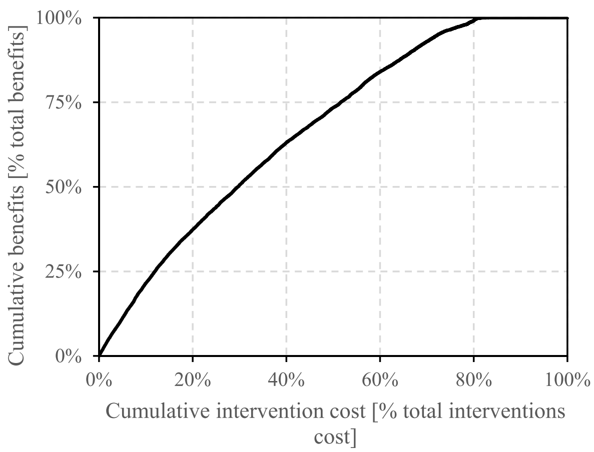

Fig. 10. Prioritizing interventions based on EAB.

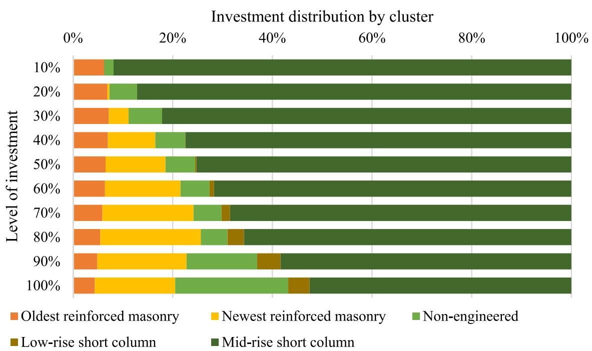

Fig. 11. Distribution of interventions by cluster and investment levels.

(US$ 8.35 million) in the vertical axis. It is shown that the relation between investment and benefit is not linear and that it is not worth investing more than 80% of the total intervention cost (around US$ 350 million), since retrofitting does not generate any additional benefits in terms of risk reduction beyond this threshold. This information can be used by decision makers to analyze what is the right level of budget allocation to be invested in a building portfolio to achieve a particular level of benefits.

Fig. 11 shows the investment distribution by cluster for different levels of investment. We observe that the *mid-rise short column* cluster takes most of the resources regardless of the investment level. For the *newer reinforced masonry* cluster, as more resources are available these buildings demand more investment, but are not as critical if the total investment is low (less than 30%). For the *older reinforced masonry* cluster, we note that its investment share reduces slightly as the total investment level grows, since this cluster is relatively small and eventually reaches a ceiling. The *low-rise short column* cluster investment share is relevant only with total investment levels of more than 50%. The *RC high-design* and the *others* cluster do not have any intervention assigned, therefore are not included in this figure.

4.6. Insights for decision makers

The results presented above are valuable for different types of decision makers. For instance, a national government agency can use this information to prioritize investment based on the available resources and yearly budget. With this information, it is possible to design a year-by-year program in which the interventions by clusters are prioritized based on risk reduction. It is also possible to combine the risk reduction program with the design of other disaster risk management applications, such as the development of emergency protocols and shelters distribution. For instance, if a set of buildings is known to be vulnerable and the intervention is planned in the medium or long term, the emergency management protocol for those facilities should be clearly defined. In addition, the knowledge of risk and intervention priorities could be useful to design a risk transfer system with an insurance company. Even though it is not common to insure these types of facilities, knowledge of comparative risk level could be a determinant to transfer risk and reduce the maximum probable losses over time, while retrofitting is implemented as part of an integrated disaster management strategy.

Other stakeholders that could benefit from these results are development banks, donors, and international multilateral organizations. These organizations are permanently working towards improving the condition of infrastructure, and in particular, they work towards improving school infrastructure with the laudable objective of improving education globally. Knowing the existing tradeoff between intervention cost and safety improvement could be a key metric to define the extent of a risk reduction program or to allocate tight budgets. The proposed methodology could be equally valuable for one-time intervention programs, as it is for central government agencies which work with tight annual budgets. Although national risk reduction programs should still focus on saving lives of future generations, they can also add additional variables such as maximizing the number of benefited students and preserving the value of the infrastructure. Also, this information can be used for programs of strategic investment for replacement and upgrading of infrastructure after an event. These programs can use the information of ranks by typology, deficiencies and risk to build back better and safer infrastructure. The collaborative work between international agencies and local governments could lead the way for state policies that integrate school improvement programs with maintaining the infrastructure, with the ultimate goal of having long-term achievements in education while keeping our future safe.

5. Concluding remarks

In this paper we presented a methodological approach for disaster risk mitigation on large-scale school intervention programs. The main objective of the proposed method is to assign a systematic retrofitting strategy to public school buildings and prioritize them based on the risk reduction level. The methodology identifies the seismic taxonomy string for each building to group them using an unsupervised machine learning clustering algorithm. For each one of the groups, we identify a suitable retrofitting strategy and, based on a risk assessment, we propose a prioritization of interventions that can be adapted to different budget levels. To illustrate this method, we implemented the methodology using infrastructure data from the Dominican Republic public school system. We identified seven clusters: *older reinforced masonry*, *newer reinforced masonry*, *non-engineered*, *RC high-design*, *low-rise short column, mid-rise short column,* and *others*. For each one of these groups, we developed a seismic retrofitting strategy to be implemented at a large scale, prioritizing interventions based on the probabilistic risk assessment. We found that buildings in the cluster *critical short column* should be prioritized first and about 50%–90% of the budget should be allocated in retrofitting these buildings, regardless of the amount of investment. Also, we noted how some interventions, which amount to roughly 20% of the total cost of interventions, should not be prioritized since the risk reduction in these cases is almost nil.

From the case study, we showed how the results serve as a basis for the development of a national retrofitting program that maximizes the impact of limited funds. Also, we demonstrated how the results can be used under different scenarios, for example, at different budget levels. This flexibility is particularly relevant in low- and middle-income countries where economic resources are scarce but donors and multilateral agencies can play a vital role if they can quantify the impact of their investment. Furthermore, the joint work between local government agencies and international organizations, such as development banks, could significantly improve infrastructure in the short and long term by the implementation of state policies and technical collaboration supported by the proposed methodology.

Although it is important to highlight that the proposed methodology requires experts' criteria in each step to ensure the results are applicable, it also reduces human bias by partially supporting the decision-making process. This process is needed to consider additional external factors, such as the building practices, resources availability and technical levels in each case study, making experts' criteria essential in the process. It is also worth to mention that other types of seismic taxonomies could be considered and implemented using this methodology. However, the impact in the process will be determined based on the parameters and information included in the selected classification system. For instance, more generic seismic taxonomies with fewer parameters could be used, but this certainly will affect the quality of the results, since results will be generated based on fewer information.

Besides the advantages of the proposed method, it is important to also mention its limitations. This method serves as a basis for the prioritization and to dimension the extent of a retrofitting project at a national level. However, the results should not be understood as a meticulous plan of interventions. Several uncertainties are embedded in the process, from the definition of the taxonomy string to the development of the retrofitting strategy in the representative building of each cluster. This means that as a reasonable next step it is necessary to verify the applicability of the retrofitting intervention in the selected buildings. Even though the taxonomy is sufficiently robust by including 12 parameters, the intrinsic parameters, like the geometric dimensions and the mechanical properties, generate the need of an individual design for each intervention. This should be done by structural engineers and should include information gathered in the field for each building (rather than for the representative building). This is particularly relevant in a more heterogeneous portfolio in which the cluster definition could be more time-consuming. In this type of building portfolios, the clustering should be analyzed in detail by structural engineer experts to ensure a correct classification. More analyses could be made in such cases, as clustering subsets of the buildings to verify the correct clusters.

The clustering procedure may be extended to other types of hazards, as well as other types of interventions. For example, we can apply the methodology on a set of buildings with a hurricane wind taxonomy. As an example, this taxonomy may include parameters such as the roof shape, structure type, health condition, connection to the walls, walls typology, openings size, and type or shutter provisions, among others. In this case, the interventions are to be designed to reduce hurricane wind risk in accordance with specific characteristics and deficiencies considered in the taxonomy. At another level, it could be possible to include the deterioration of structures over time and its effect on the vulnerability. Furthermore, functional interventions can also be assessed and planned to use the proposed methodology. The taxonomy in this case could include aspects such as student's density, bathrooms quality and quantity, accessibility, energy efficiency, information technology (IT) provision, and security fences in school facilities. Some of the functional interventions are the improvement of water and sanitation hygiene (WASH), ventilation, illumination, and student density, among others. Adding hurricane and functional interventions, directly related to education quality, open a wide set of future applications for large-scale school intervention programs.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.Dina D'Ayala reports financial support was provided by The World Bank Group. Dina D'Ayala reports financial support was provided by UNESCO Chair in Disaster Risk Reduction and Resilient Engineering. Rafael Fernandez reports financial support was provided by UNESCO Chair in Disaster Risk Reduction and Resilient Engineering.

Data availability

The authors do not have permission to share data.

Acknowledgments

Part of this study was supported by a project in the Dominican Republic funded by the Global Program for Safer Schools (GPSS) of The World Bank. The authors are thankful to the World Bank and the entire GPSS technical team. We would also like to thank the Dominican Republic Ministry of Education who facilitated the information for the project and were always keen to help. Prof. Dina D'Ayala and Dr. Rafael Fernández gladly acknowledge the support of the UNESCO Chair in Disaster Risk Reduction and Resilience Engineering, which funds some of their time. A special acknowledgment to the group that worked in the project and particularly to Juan Sebastián Echeverry, Gustavo Fuentes, and Rohit Adhikari for their help in the development of this research. Also, the team is grateful for the comments provided by Dr. Hori Tsuneki to improve the structure of the manuscript to reach a broader audience. This paper is dedicated to the memory of Prof. Dr. Luis E. Yamin, team leader of CIMOC at Universidad de Los Andes for over 30 years.

References

[1] Educate a Child, Barriers to Education, 2020 [Online]. Available:.
[2] The World Bank, Realizing the Future of Learning, The World Bank Group, Washington D.C., 2020.
[3] UNESCO, UNESCO Guidelines for Assessing Learning Facilities in the Context of Disaster Risk Reduction and Climate Change Adaptation - VISUS, vol. 1, 2019, p. 56. Paris.
[4] UNISDR, GADRRRES, Comprehensive School Safety, 2017, p. 8.
[5] R. Paci-Green, A. Varchetta, K. McFarlane, P. Iyer, M. Goyeneche, Comprehensive school safety policy: a global baseline survey, Int. J. Disaster Risk Reduc. 44 (2020), 101399.
[6] The World Bank, Global Program for Safer School, 2020 [Online]. Available:.
[7] Learning Idb, Twenty-First Century Schools, Interamerican Development Bank, 2014.
[8] M.K. Naja, H. Baytiyeh, Towards safer public school buildings in Lebanon: an advocacy for seismic retrofitting initiative, Int. J. Disaster Risk Reduc. 8 (2014) 158–165.
[9] S. Grimaz, P. Malisan, Multi-hazard visual inspection for defining safety upgrading strategies of learning facilities at territorial level: VISUS methodology, Int. J. Disaster Risk Reduc. 44 (2020), 101435.
[10] D.N. Grant, J.J. Bommer, R. Pinho, G.M. Calvi, A. Goretti, F. Meroni, A prioritization scheme for seismic intervention in school buildings in Italy, Earthq. Spectra 23 (2) (2007) 291–314.

[11] N.W. Chi, J.P. Wang, J.H. Liao, W.C. Cheng, C.S. Chen, Machine learning-based seismic capability evaluation for school buildings, Autom. ConStruct. 118 (2020), 103274.
[12] V.A. Muñoz, B. Carby, E.C. Abella, O.D. Cardona, T. López-Marrero, V. Marchezini, L. Meyreles, D. Olivato, R. Trajber, B. Wisner, Success, innovation and challenge: school safety and disaster education in South America and the Caribbean, Int. J. Disaster Risk Reduc. 44 (2020).
[13] H. Jaafaru, B. Agbelie, Bridge maintenance planning framework using machine learning, multi-criteria decision analysis and evolutionary optimization models, Autom. ConStruct. 143 (September) (2022), 104585.
[14] A. Gondia, M. Ezzeldin, W. El-Dakhakhni, Dynamic networks for resilience-driven management of infrastructure projects, Autom. ConStruct. 136 (2022), 104149.
[15] J. Dao, S.T. Ng, Y. Yang, S. Zhou, F.J. Xu, M. Skitmore, Semantic framework for interdependent infrastructure resilience decision support, Autom. ConStruct. 130 (2021), 103852.
[16] The World Bank, Global Library of School Infrastructure - Taxonomy, The World Bank Group, Washington D.C., 2019.
[17] R. Adhikari, D. D'Ayala, R. Fernández, L. Yamin, A. Nassirpour, A.P. Vatteri, C. Ferreria, F. Ramírez, GLOSI taxonomy: a tool for 'seismic risk assessment' oriented classification of school buildings, Int. J. Disaster Risk Reduc. 87 (2023).
[18] D. D'Ayala, C. Galasso, A. Nassirpour, R.R.K. Adhikari, L. Yamin, R. Fernandez, D. Lo, L. Garciano, A. Oreta, Resilient communities through safer schools, Int. J. Disaster Risk Reduc. 45 (2020), 101446.
[19] F. Clementi, E. Quagliarini, G. Maracchini, S. Lenci, Post-World War II Italian school buildings: typical and specific seismic vulnerabilities, J. Build. Eng. 4 (2015) 152–166.
[20] H. Aroquipa, A.I. Hurtado, Incremental seismic retrofitting for essential facilities using performance objectives: a case study of the 780-PRE school buildings in Peru, J. Build. Eng. 62 (2022), 105387.
[21] The World Bank, Evaluación de riesgo por sismo y viento huracanado y planes de reducción del riesgo para la infraestructura educativa pública de República Dominicana, Bogotá, 2020.
[22] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning. Data Mining, Inference, and Prediction, second ed., Springer, New York, NY, 2009.
[23] L. Yamin, F. Ghesquiere, O. Cardona, M. Ordaz, Modelación probabilista para la gestión del riesgo de desastre, El caso de Bogotá, Colombia, 2013, p. 183.
[24] F. Aleskerov, A.I. Say, A. Toker, H.L. Akin, G. Altay, A cluster-based decision support system for estimating earthquake damage and casualties, Disasters 29 (3) (2005) 255–276.
[25] J.S. Prasad, Y. Singh, A.M. Kaynia, C. Lindholm, Socioeconomic clustering in seismic risk assessment of urban housing stock, Earthq. Spectra 25 (3) (2009) 619–641.
[26] R. Gunasekera, O. Ishizawa, C. Aubrecht, B. Blankespoor, S. Murray, A. Pomonis, J. Daniell, Developing an adaptive global exposure model to support the generation of country disaster risk profiles, Earth Sci. Rev. 150 (2015) 594–608.
[27] IDB, Perfil de riesgo de desastres por evento sísmico de Ecuador, Interamerican Development Bank, 2022.
[28] IDB, Análisis de la inversión pública para la reducción de riesgo de desastres: evaluación del riesgo de deslizamiento en el sector vial y posibles medidas para la mitigación del riesgo -Estudio de caso en Ecuador-, Interamerican Development Bank, 2022.
[29] V. Röthlisberger, A.P. Zischg, M. Keiler, Identifying spatial clusters of flood exposure to support decision making in risk management, Sci. Total Environ. 598 (2017) 593–603.
[30] S.E. Chang, J.Z. Yip, T. Conger, G. Oulahen, M. Marteleira, Community vulnerability to coastal hazards: developing a typology for disaster risk reduction, Appl. Geogr. 91 (2018) 81–88.
[31] G. Casella, S. Fienberg, I. Olkin, An Introduction to Statistical Learning, 2013, p. 618.
[32] A.K. Jain, R.C. Dubes, Algorithms for Data Clustering, Prentice Hall, New Jersey, 1988.
[33] Z. Huang, Extensions to the k-means algorithm for clustering large data sets with categorical values, Data Min. Knowl. Discov. 2 (3) (1998) 283–304.
[34] D. Chen, D.-w. Cui, C.-x. Wang, Z.-r. Wang, A rough set-based hierarchical clustering algorithm for categorical data, Int. J. Inf. Technol. 12 (3) (2006) 149–159.
[35] W.H. Day, H. Edelsbrunner, Efficient algorithms for agglomerative hierarchical clustering methods, J. Classif. 1 (1) (1984) 7–24.
[36] N. Rajalingam, K. Ranjini, Hierarchical clustering algorithm - a comparative study, Int. J. Comput. Appl. 19 (3) (2011) 42–46.
[37] Y. Rani, H. Rohil, A study of hierarchical clustering algorithms, Int. J. Inf. Comput. Technol. 3 (10) (2013) 1115–1122.
[38] F. Murtagh, P. Contreras, Algorithms for hierarchical clustering: an overview, Wiley Interdisciplinary Reviews: Data Min. Knowl. Discov. 2 (1) (2012) 86–97.
[39] P. Agarwal, M.A. Alam, R. Biswas, A hierarchical clustering algorithm for categorical attributes," 2010 2nd International Conference on Computer Engineering and Applications, ICCEA 2 (2010) 365–368.
[40] Z. Šulc, H. Řezanková, Comparison of similarity measures for categorical data in hierarchical clustering, J. Classif. 36 (2019) 58–72.
[41] A. Juan, E. Vidal, On the use of Bernoulli mixture models for text classification, Pattern Recogn. 35 (12) (2002) 2705–2710.
[42] A. Juan, E. Vidal, Bernoulli mixture models for binary images, Int. Conf. Pattern Recogn. 3 (2) (2004) 367–370.
[43] G. Govaert, M. Nadif, An EM algorithm for the block mixture model, IEEE Trans. Pattern Anal. Mach. Intell. 27 (4) (2005) 643–647.
[44] A. Zheng, A. Casari, Feature Engineering for Machine Learning, O'Reilly, 2018.
[45] M. Saeed, K. Javed, H. Atique Babri, Machine learning using Bernoulli mixture models: clustering, rule extraction and dimensionality reduction, Neurocomputing 119 (2013) 366–374.
[46] G. Govaert, M. Nadif, Block clustering with Bernoulli mixture models: comparison of different approaches, Comput. Stat. Data Anal. 52 (6) (2008) 3233–3245.
[47] ASCE, SEI, Seismic Evaluation and Retrofit of Existing Buildings - ASCE 41-17, 2017.
[48] FEMA, ATC, Reducing the Risks of Nonstructural Earthquake Damage – A Practical Guide - FEMA E-74, 2012, p. 885.
[49] FEMA, ATC, Repair of earthquake damaged concrete and masonry wall buildings, - FEMA 308 (1998).
[50] European committee for standardization, Eurocode 8: Design of Structures for Earthquake Resistance - Part 3: Assessment and Retrofitting of Buildings, The European Union, Brussels, 2005.
[51] International Code Council, International Building Code (IBC), 2018 [Online], https://www.iccsafe.org/products-and-services/i-codes/2018-i-codes/ibc/.
[52] A. Muñoz, D. Quiun, M. Tinman, Repair and seismic retrofitting of hospital and school buildings in Peru, in: 13th World Conference on Earthquake Engineering, 2004. Vancouver, BC, Canada, August.
[53] R.I. Fernández, R. Rincón, L.E. Yamin, Incertidumbre en el Beneficio Obtenido para Opciones de Reforzamiento Sísmico, Cali, Colombia, 2019.
[54] H.D. Shrestha, R. Yatabe, N.P. Bhandary, J. Subedi, Vulnerability assessment and retrofitting of existing school buildings: a case study of Aceh, Int. J. Disaster Resilience Built Environ. 3 (1) (2012) 52–65.
[55] C. Chrysostomou, N. Kyriakides, A. Kappos, L. Kouris, E. Georgiou, M. Millis, Seismic retrofitting and health monitoring of school buildings of Cyprus, Open Construct. Build Technol. J. 7 (1) (2014) 208–220.
[56] H. Shrestha, S. Pradhan, R. Guragain, Experiences on retrofitting of low strength masonry buildings by different retrofitting techniques in Nepal, in: 15th World Conference on Earthquake Engineering, Portugal., Lisbon, 2012.
[57] F. Porco, A. Fiore, G. Uva, D. Raffaele, The influence of infilled panels in retrofitting interventions of existing reinforced concrete buildings: a case study, Struct. Infrastruct. Eng. 11 (2) (2015) 162–175.
[58] M.L. Segovia-Verjel, M.V. Requena-García-Cruz, E. De-Justo-Moscardó, A. Morales-Esteban, Optimal seismic retrofitting techniques for URM school buildings located in the southwestern Iberian peninsula, PLoS One 14 (10) (2019) 1–18.
[59] UNESCO, Manual for Restoration and Retrofitting of Rural Structures in Kashmir, 2007.
[60] Nepal Reconstruction Authority, Seismic Retrofitting Guidelines for Buildings in Nepal, Nepal., Kathmandu, 2016.
[61] S. Bhattacharya, S. Nayak, S.C. Dutta, A critical review of retrofitting methods for unreinforced masonry structures, Int. J. Disaster Risk Reduc. 7 (2014) 51–67.
[62] F. Yavartanoo, T.H. Kang, Retrofitting of unreinforced masonry structures and considerations for heritage-sensitive constructions, J. Build. Eng. 49 (2022), 103993.
[63] A. S. Arya, T. Boen and Y. Ishiyama, "Guidelines for Earthquake Resistant Non-engineered Construction," UNESCO, France.

[64] G.M. Calvi, G. Magenes, Experimental results on unreinfornced masonry walls damaged and repaired, in: 10th International Brick Masonry Conference, 1994.
[65] M. Elgawady, P. Lestuzzi, M. Badoux, A review of conventional seismic retrofitting techniques for URM, in: 13th International Brick and Block Masonry Conference, 2004.
[66] IAEE, Guidelines for Earthquake Resistant Non-engineered Construction, International Association for Earthquake Engineering, 2004.
[67] N. Anwar, S. Adhikari, M. Sahid, S. Shrestha, Making Schools Safer from Earthquakes - Effective Retrofitting: Case Studies from Nepal, Asian Institute of Technology, Thailand, 2016.
[68] Nepal Reconstruction Authority, Repair and Retrofitting Manual for Masonry Structure," Nepal Reconstruction Authority, Government of Nepal, Kathmandu, Nepal, 2017.
[69] The World Bank, Vulnerability Reduction Solutions - Splint and Bandage Technique," 03 10, 2018.
[70] R.I. Fernández, R.K. Adhikari, L.E. Yamin, D. D'Ayala, G.A. Fuentes, Seismic vulnerability assessment and risk reduction strategy of low-rise school buildings, in: 17th World Conference on Earthquake Engineering, 2020. Sendai.
[71] L.T. Guevara, L.E. García, The captive- and short-column effects, Earthq. Spectra 21 (1) (2005) 141–160.
[72] The World Bank, Global Library for School Infrastructure - Fragility and Vulnerability Assessment Guide, Washignton D.C., The World Bank Group, 2019.
[73] R. Fernández, L. Yamin, D. D'Ayala, R. Adhikari, J.C. Reyes, J. Echeverry, G. Fuentes, A simplified component-based methodology for the seismic vulnerability assessment of school buildings using nonlinear static procedures: application to RC school buildings, Bull. Earthq. Eng. 20 (2022) 6555–6585.
[74] MOPC-DGRS, "R001, Reglamento para el Análisis y Diseño Sísmico de Estructuras, Santo Domingo, 2011.
[75] R. Fernández, L.E. Yamin, J.C. Reyes, A.P. García, G. Fuentes, J. Echeverry, Mitigación del riesgo sísmico de la infraestructura escolar, in: Investigaciones en gestión del riesgo de desastres para Colombia: Avances, perspectivas y casos de estudio, Bogotá, Unidad Nacional de Gestión del Riesgo de Desastres, 2021.
[76] CAPRA, Comprehensive Approach for Probabilistic Risk Assessment, 2022 [Online]. Available:.
[77] OASIS, OASIS Loss Modelling Framework, 2022 [Online]. Available:.
[78] S. Molina, D.H. Lang, C.D. Lindholm, SELENA – an open-source tool for seismic risk and loss assessment using a logic tree computation procedure, Comput. Geosci. 36 (3) (2010) 257–269.