International Journal of Disaster Risk Reduction 96 (2023) 103959

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

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

# Damage classification after the 2009 L'Aquila earthquake using multinomial logistic regression and neural networks

Angelo Aloisio [a,*], Marco Martino Rosso [b], Andrea Matteo De Leo [a], Massimo Fragiacomo [a], Maria Basi [c]

[a] Department of Civil, Construction-Architectural and Environmental Engineering, Università degli Studi dell'Aquila, L'Aquila, Italy
[b] Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Turin, Italy
[c] Abruzzo Region Risk Prevention of Civil Protection, L'Aquila, Italy

ARTICLE INFO

*Keywords:*
Seismic risk
Post-earthquake survey
Multinomial logistic regression
Neural network

ABSTRACT

Post-earthquake surveys represent a fundamental tool for managing the emergency phase after a strong earthquake. In Italy, the evaluation of the post-earthquake functionality of ordinary buildings is based on the AeDES forms (*Agibilitá e Danno nell'Emergenza Sismica*, or equivalently, Rapid Post-Earthquake Damage evaluation forms). This form includes information on the building and records of the observed damage classified according to type and intensity in 60 subclasses. Based on the observed damage and expert judgment, the buildings are clustered into six risk classes, from A to F. The assigned class is used to calculate the maximum economic reimbursement owed for the reconstruction or repair of the building. However, often the cluster assignment is not entirely objective due to the inherent responsibility associated with a less conservative assessment. This paper uses the data from the 2009 L'Aquila earthquake to develop classification models based on multinomial logistic regression (MLR) and artificial neural networks (ANN) calibrated with data theoretically less influenced by personal biases. The proposed models, particularly the MLR, are intended to support the decision-making of the evaluation team in future updates of the AeDES forms. This approach cannot substitute expert evaluation, which is always necessary for complex scenarios but may mitigate the impact of subjectivity and can provide an indication of the expected outcome of the survey.

## 1. Introduction

Post-earthquake surveys are crucial in managing the emergency phase following a strong earthquake. The post-earthquake surveys have two main objectives: (i) providing information on the functionality of the building and (ii) giving a summary assessment of the damage, which is then used to calculate the maximum economic reimbursement that the state would provide for the reconstruction or repair of the building. In Italy, this evaluation is based on the AeDES forms. This form is used for field damage and safety assessments of buildings in the aftermath of earthquakes [1]. The AeDES forms drive the evaluator to a synthetic class assignment, from A to F, expressing the risk level associated with the risk state of the building. Nonetheless, the above classification is not automatic and somewhat subjective because based mainly on expert judgment. The complexity of post-earthquake scenarios cannot always be pigeonholed into short forms. The emergency phase demands rapid evaluations. Therefore, the AeDES forms, keeping a simple structure useful for quick assessments, cannot comprehend all scenarios that require the licensed practitioner's

* Corresponding author.
*E-mail address:* angelo.aloisio1@univaq.it (A. Aloisio).

https://doi.org/10.1016/j.ijdrr.2023.103959
Received 16 May 2023; Received in revised form 8 August 2023; Accepted 21 August 2023
Available online 23 August 2023
2212-4209/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

evaluation. Nonetheless, leaving room for subjectivity can favour non-uniformity in the evaluations. The reasons that lead to different, subjective evaluations are mostly the followings:

- Different evaluators could interpret the same damage patterns differently, according to their working experience and understanding of the structural behaviour of the building.
- The class assignment is directly related to the maximum economic reimbursement provided by the state for the reconstruction/repair [2]. This economic interest might bias the evaluation towards higher-risk classes in borderline cases.

Despite the possible bias, the data related to Italian earthquakes over the last 50 years, collected and organized on the Da.D.O. web-based platform, represent a unique exposure and vulnerability information source that could provide guidance for Disaster Risk Reduction (DRR) and prevention activities. Specifically, empirical damage data collected in post-earthquake surveys represent a valuable source of loss statistics, helpful in predicting the vulnerability of specific classes of assets to ground motion intensities that could occur in future events [3]. Additionally, the analyses of these databases are fundamental for defining the taxonomy of evaluation forms [4,5] and understanding the vulnerability of existing buildings in seismic-prone areas highlighting specific damage clusters [6,7]. For these reasons, most papers mining into these databases attempts to relate the typological characteristics information to the observed damage.

The first studies on empirical damage data led to the definition of damage probability matrices (DPMs) [4], describing the conditional probability of reaching a certain damage level due to a ground motion. This taxonomy was first applied by Braga et al. [5] after the 1980 Irpinia earthquake. Dolce et al. [8] added a new vulnerability class to consider improvements in the seismic behaviour of buildings retrofitted or designed according to recent seismic design codes; Di Pasquale et al. [9] introduced a different parameter to measure ground motion intensity. Several studies, based on the L'Aquila [10–15], Emilia [16–18], Centre Italy [19–21] and Ischia [22,23] earthquakes, exploited the databases related to such earthquakes for empirical large-scale vulnerability assessments and, more recently to derive the Global Seismic Risk Maps [24]. However, such vulnerability assessments are highly heterogeneous because they use several intensity measures to characterize the seismic intensity [25–28] and because the quality of the used loss or damage databases is highly variable [29–35].

Despite the significant effort in understanding the vulnerability of Italian buildings from the enormous amount of post-earthquake damage data collected after the seismic events, fewer studies attempted to understand the limits of the AeEDES forms, investigating the uncertainty and subjectivity in the synthetic risk class assignment. Nicodemo at al. [36] attempted to mitigate the impact of subjective, expert-based judgment in the class assignment and provided a solid characterization of the underlying uncertainties using a fuzzy formulation. Specifically, they presented a methodology based on the fuzzy compatibility score to assign a risk-oriented class to each surveyed building, using the EMS-98 typologies as reference classification. Despite the high value of such research, the authors deem such a methodology challenging to apply in post-earthquake scenarios, where a rapid and straightforward evaluation is needed. Human judgment cannot be entirely replaced by algorithmic approaches based on damage matrices due to the intrinsic complexity of damage patterns, which are difficult to express in a synthetic form. According to the authors, the extensive databases of post-earthquake damage should be used to understand the limitations of the current clustering approach, trying to limit the subjectivity without substituting the human-based decision based on the damage matrices. The AeDES forms proved to be an efficient tool in the post-earthquake phase.

This paper uses a database of 878 school buildings damaged by the 2009 L'Aquila earthquake to predict the synthetic class assignment provided by the evaluator, from A to E, using the damage information recorded in the AeDES form. The use of the sole data coming from the AeDES forms filled out for school buildings is based on the assumption that, being compiled by academic experts, such forms are less affected by personal biases.

The database is used for the following purposes:

- Calibration and validation of multinomial logistic regression (MLR) and artificial neural network (ANN) models to predict the risk class assignment from the filled damage matrix data;
- Use of the developed models to investigate the possible issues behind the proposed class assignment and the proposal of possible improvements, such as the reduction of the number of risk classes;
- Proposal of enhancing the current risk classes in the AeDES forms, providing the coefficients for MLR to support the evaluator's decision process. The calibrated model might be considered in future digital updates of the AeDES forms since it can be easily implemented in spreadsheets to be filled on-site.

As anticipated, among the several possibilities available in the literature, two possible approaches are used in this paper to develop classification models. One approach is based on linear models or semi-linear models such as MLR [37], which have been widely used to study engineering problems [38,39]. These models have their strength in their high explainability, which can lead to identifying the main drivers of the phenomenon being studied and application speed. For this case, specifically, using such models entails identifying the critical drivers of the classification process of the AeDES form that would be needed to propose a possible simplification of such form. However, linear approaches require manual feature engineering, a good understanding of the modelled phenomenon, and the choice of predictors that are linearly related to quantity to be predicted, which may be difficult to find. The second approach is considering a deep learning method such as an ANN. Although such models are less explainable, their versatility allows them to cope with classification or regression tasks, and many ground-breaking variants have been developed in the last decades. Additionally, depending on the number of output layers, ANNs are highly adaptable in working with binary and multinomial classification tasks. The architecture of the ANN determines the model complexity, the accuracy rate, and the

| | Severe ($D_4$-$D_5$) | | | Medium ($D_2$-$D_3$) | | | Low ($D_1$) | | | Null |
|---|---|---|---|---|---|---|---|---|---|---|
| | >2/3 | >1/3 <2/3 | <1/3 | >2/3 | >1/3 <2/3 | <1/3 | <2/3 | >1/3 <2/3 | <1/3 | |
| Bearing structure | X | | | | | | | | | |
| Floor | | | | | | | | X | | |
| Stairs | | | | | | | | | | X |
| Roof | | | | | | | | | | X |
| Infill | X | | | | | | | | | |
| Pre-exsisting | | | | | | | | | | X |

**Fig. 1.** Illustration of the damage matrix in the AeDES form.

inclination to underfitting or overfitting issues. Therefore the ANN architecture should be carefully evaluated [40]. Nowadays, within engineering research, the adoption of machine learning and pattern recognition techniques such as the ANN is continuously growing, providing, also in civil engineering, new research perspectives [41–43]. In the current study, a multi-layer perceptron (MLP) ANN with an optimized architecture is employed and compared to the before-mentioned MLR for analysing the class assignment in the AeDES form.

The paper is organized as follows. After the general introduction to the problem and the mathematical background of MLR and ANN, the authors provide insights into the considered database. The last sections present the class predictions' results with the proposed models. The discussion of the results led to the proposal of a class condensation discussed in the last section.

## 2. Problem background

This section first describes the most relevant sections of the AeDES form, then it introduces the characteristics of the post-earthquake survey.

### 2.1. The AeDES form

The AeDES form consists of nine sections organized into three pages, with an additional fourth page that includes explanatory notes. Specifically, Section 1 - BUILDING IDENTIFICATION contains information on the survey identification and building. Section 2 - BUILDING DESCRIPTION synthetically describes the structure (i.e., the total number of storeys including basements, the number of basements, the average storey height and the average storey surface), the age of construction and the use of the building. Section 3 - TYPOLOGY details the characteristics of the structure, with a particular focus on masonry buildings. Section 4 - DAMAGE TO STRUCTURAL ELEMENTS contains information on the observed damage, organized by damage location and intensity. Section 5 - DAMAGE TO NON-STRUCTURAL ELEMENTS contains information on the observed damage on non-structural elements, organized by damage location and intensity. Section 6 - OUTSIDE DANGER indicates possible danger induced by other buildings, networks, slopes and emergency measures performed. Section 7 - SOIL AND FOUNDATIONS details the site's morphology and possible instability of the foundations. Section 8 - RISK CLASS indicates the assigned risk class. Section 9 - OTHER REMARKS contains a blank page to be filled with possible additional notes of the evaluator.

Section 4 - DAMAGE TO STRUCTURAL ELEMENTS contains the information relevant to this study. The damage levels used in such section of the AeDES form are discrete (EMS-98) and range from $D_0$ (no damage) to $D_5$ (total collapse); see Fig. 1. If a building falls in damage class $D_1$, the earthquake did not affect the capacity of the building nor the safety of the occupants due to failures of non-structural components. The damage levels $D_2 - D_3$ correspond to a significant capacity reduction, although the building did not approach the collapse limit state. The damage classes $D_4 - D_5$ correspond to a manifest capacity reduction, where the building approached the collapse limit state.

Each damage class is divided into three subsections of the damage regarded as more than 2/3, the range 2/3-1/3, or less than 1/3 of the building extent. Since the evaluations are based on visual inspections, the damage matrix only accounts for visible damage. According to the damage matrix, i.e., according to the type and extent of damage found in the building, it is possible to classify the building as follows:

- A: Building accessible;
- B: Building temporarily unusable or useable only following emergency interventions;
- C: Partially Unusable building
- D: Temporarily unusable building to be re-examined;
- E: Unusable building;

- F: Unusable building due to external risk only.

It is essential to note that while classes A, B, C, and E are directly related to the level of damage, indicating an increasing severity from A to E, classes D and F do not directly represent damage levels. Class D signifies that a repeated survey is required, making it irrelevant to predict this outcome based solely on the damage matrix, as it depends on other circumstances. Class F represents external risk, which means that this outcome is not directly related to the observed damage of the building itself but rather to the damage to nearby buildings that may influence the safety of the surveyed building. As a result, the damage-related classes are A, B, C, and E. These classes will be predicted based on the information provided by the damage matrix. It is worth mentioning that no analytical expression exists that directly relates the entries in the matrix to the final evaluation outcome. This paper calibrates two algorithms, MLR and ANN, which can predict whether the outcome will be A, B, C, or E based on the information filled in the damage matrix by the engineer.

The risk classes, derived as the output of the AeDES form, are indeed related to the maximum reimbursement the state provides for the reconstruction or repair of buildings, as previously mentioned. However, it is essential to acknowledge that this direct relationship between the AeDES outcome and funding only occurred for the L'Aquila earthquake. The funding process differed for subsequent earthquakes like Emilia, Centre Italy, and Ischia. In these later earthquakes, an additional parameter known as the operational level, which considers the levels of damage and vulnerability, drives the conventional parametric cost assessment within each risk class provided by the AeDES forms.

### 2.2. The post-earthquake survey

On April 6, 2009, a magnitude Mw = 6.3 earthquake struck central Italy in the proximity of L'Aquila [44]. The seismic event was recorded by the digital strong-motion stations operated by the Italian Strong Motion Network (RAN), managed by the Italian Department of Civil Protection (DPC) and by the broadband stations of the Italian National Institute of Geophysics and Volcanology (Istituto Nazionale di Geofisica e Vulcanologia, INGV) [45,46]. Visual surveys performed by experts immediately after the earthquake [47] led to identifying the perimeter of the seismic crater, including 316 municipalities. The buildings in the seismic crater, exhibiting macroseismic intensities greater than VI on the MCS scale, were inspected to assess their safety and functionality level. The AeDES form [1,48] is the first-level survey form for such post-earthquake assessment. Such a form is meant to provide a trade-off between the accuracy and rapidity of the evaluation. After the L'Aquila earthquake, the AeDES form was filled out for 74.254 buildings following in situ inspections.

Of the 74,254 buildings in the dataset, 80% are reinforced concrete (RC) or masonry buildings (12,223 RC buildings and 47,077 masonry buildings), and the remaining part corresponds to other structural types. In Section 3 of the AeDES form, there is a provision to identify various structural typologies for buildings. For masonry structures, multiple options are available for different horizontal and vertical bearing structures, including isolated columns and mixed and reinforced masonry. On the other hand, the form offers four options for non-masonry structures: reinforced concrete (RC) frames, RC walls, steel frames, and timber frames/walls. In this context, when referring to "other structures", the authors are combining the categories of steel frames and timber frame/walls.

As highlighted in the introduction and seldom evidenced in past research, the accuracy of the AeDES forms filled out for such buildings is heterogeneous because the AeDES forms are especially prone to subjectivity when they are filled for private buildings and in the case of minor damage, the consequences and economic impact of the damage are more difficult to evaluate.

Unlike the Central Italy earthquake, in the L'Aquila earthquake, the assessment of strategic buildings like the schools, which are the focus of the dataset used in this research, was performed by academics who allegedly are more qualified to provide a nearly-objective evaluation. Thus, in this investigation, the authors will only consider the AeDES forms related to schools that are a significant part of the public facilities in the seismic crater. This choice is considered prudential and intended to avoid using data based on biased classifications resulting from excessively conservative assessments.

Fig. 2(a) shows a bubble plot of the AeDES forms filled out for the schools falling within the seismic crater. The total number of schools is 878, where more than 30% belong to the municipality of L'Aquila, see the seismic crater in Fig. 3(a). Fig. 2(b) shows a pie chart of the evaluation results. More than 50% of the schools had no damage and fell in class A. Two other largely populated classes are B and E, comprising 26% and 15% of the dataset. A few buildings fell into the intermediate risk classes, i.e., C and D. Fig. 3(b) displays a geo-referenced scatter plot of the evaluated buildings and the outcomes of the AeDES forms according to the legend.

## 3. Classification models used for the damage class prediction

This section briefly overviews the MLR and ANN models used for damage classification.

### 3.1. Multinomial logistic regression

As remarked in the introduction, MLR is a well-established and interpretable algorithm for multiclass classification problems. It can handle categorical dependent variables, making it suitable for the damage classification task with multiple classes. MLR also provides coefficients that allow understanding the relationship between the predictors (damage regressors) and the outcome (damage classes).

In this paper, both nested and non-nested MLR models are developed using the damage data. MLR is used for classification problems when the dependent variable is categorical [37], as occurring in the specific case where the outcome of an inspection

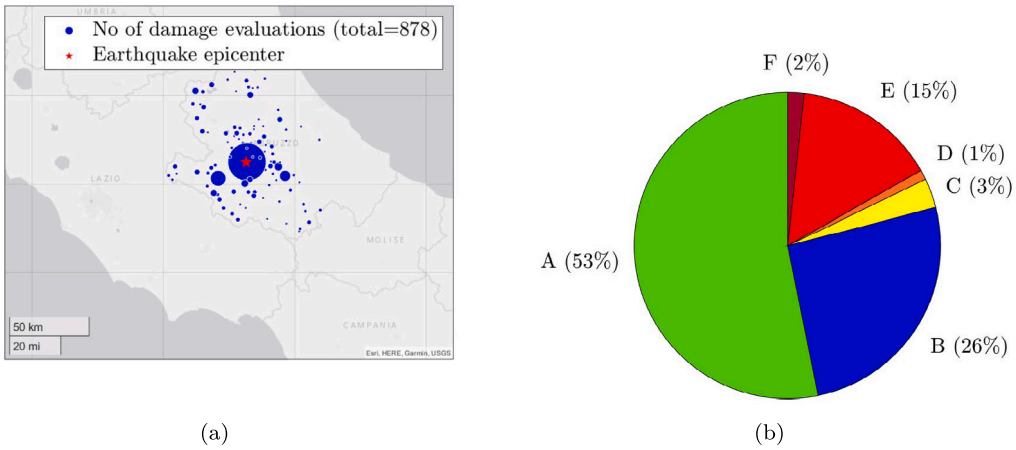


**Fig. 2.** (a) Bubble chart of the AeDES forms filled for schools falling in the seismic crater; (b) Pie chart of the evaluation results.

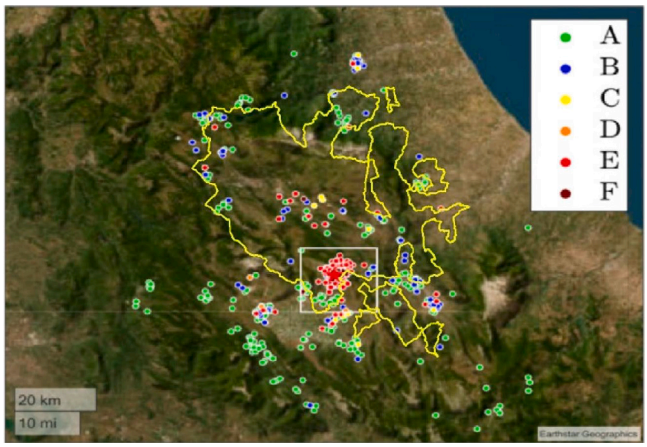


**Fig. 3.** Geo-referenced scatter plot of the evaluated buildings and the outcomes of the AeDES form according to the legend. The yellow contour delimits the seimic crater. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

can be classified from class A to class F. The model assumes the independence of irrelevant alternatives, which can be considered generally satisfied in this application. Such assumptions entail that the odds of preferring one damage class over another do not depend on the presence or absence of other "irrelevant" alternatives. Since the response variable $Y$ is categorical and has $r$ possible outcomes, $Y = \{y_1, \dots, y_r\}$, ordered for increasing level of damage, a generalized logit model handles the $r$ categorical outcomes. Accordingly, $r-1$ logits for the response variable are compared with a baseline category, i.e., $r-1$ independent binary logistic regression models are calibrated, where one outcome is chosen as a pivot, and the other $r-1$ outcomes are separately regressed against the pivot outcome. For simplicity and coherence, the pivot outcome is herein chosen as the one related to the highest damage class $y_r$. The outcomes, initially six, are reduced to four (i.e., $r = 4$) ordered from 1 to 4 and corresponding to the damage classes A, B, C, and E, respectively. Classes D and F are removed since they do not represent damage levels. Specifically, D denotes an incomplete survey, which demands further inspections, whereas F indicates an external risk to the building occupants (e.g., the collapse of a nearby building). Accordingly, the data corresponding with classes D and F are eliminated from the database, resulting in 844 remaining outcomes.

The vector of the regressors $\mathbf{X}$, corresponding to the damage data, consists of six categorical variables (i.e., damage levels of Bearing structure, Floor, Stairs, Roof, Infill, and pre-existing damage) with ten categories each (see Fig. 1). To simplify the implementation, the six variables are transformed into a vector of 60 corresponding dummy variables $\mathbf{x}$ with binary outcomes to indicate the absence or presence of specific categorical information, i.e., the presence of a given damage characteristic. The $r-1$ regression equations for a generic outcome $y_j$ are written as

$$\ln \frac{\Pr(Y_i = y_j)}{\Pr(Y_i = y_r)} = \boldsymbol{\beta}_j \cdot \mathbf{x}_i \quad , \quad j < r, \tag{1}$$

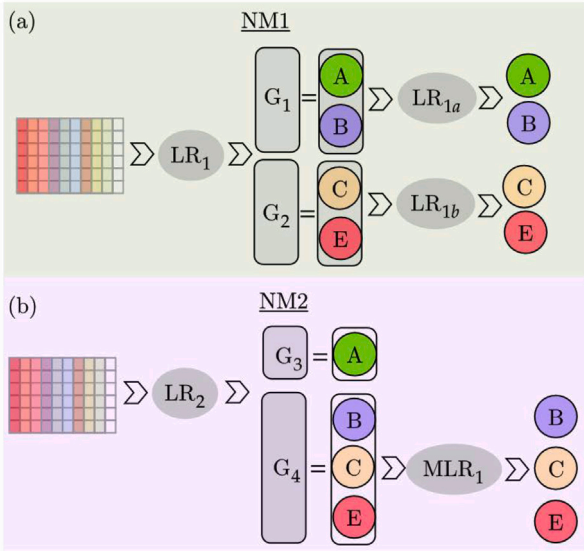


**Fig. 4.** Workflow of the (a) NM1 and (b) NM2 nested MLR.

where $\boldsymbol{\beta}_j$ is the coefficient vector for the $j$th category of the response variable and $\mathbf{x}_i$ is the $i$th realization of the variables used as regressors. Generally, in an MLR model, the generic $k$th element of $\boldsymbol{\beta}_j$ can be interpreted as the increase in log-odds of falling into category $j$ versus category $r$, resulting from a one-unit increase in the $k$th regressor while holding the other regressors constant. In this specific case, the $k$th element of $\boldsymbol{\beta}_j$ is the increase in log-odds due to the presence of the $k$th damage characteristic. Any category could be chosen as the baseline, and the model will have the same fitting accuracy. However, the values and the interpretation of the coefficients in $\boldsymbol{\beta}_j$ would change. Following from Eq. (1), the probability of falling into non-baseline categories can be calculated as follows:

$$\Pr(Y_i = j) = \frac{e^{\boldsymbol{\beta}_j \cdot \mathbf{x}_i}}{1 + \sum_{j=1}^{r-1} e^{\boldsymbol{\beta}_j \cdot \mathbf{x}_i}} \quad , \quad j < r, \tag{2}$$

whereas the baseline category

$$\Pr(Y_i = r) = \frac{1}{1 + \sum_{j=1}^{r-1} e^{\boldsymbol{\beta}_j \cdot \mathbf{x}_i}}. \tag{3}$$

The calibration of the presented model is based on the maximization of the log-likelihood that is performed following the procedure for calibrating distributional regression models presented in [49]. Such a procedure is among the most efficient methods for models that have a concave log-likelihood function as the one at hand because it is based on the Newton–Raphson method. After the model calibration, the predicted class $\hat{Y}_i$ can be chosen as the one associated with the highest probability of occurrence among the $r$ categorical alternatives

$$\hat{j}_i = \underset{j}{\operatorname{argmax}} \left[\Pr(Y_i = y_j)\right] \quad , \quad j \in [1, r] \tag{4}$$

where $\hat{j}_i$ is the order of the predicted class corresponding to the $i$th set of values of the regressors $\mathbf{x}_i$.

Using this formulation, an initial MLR model (NN in the following, meaning Non-Nested) is developed to predict the odds of a specific case falling in the A, B, C and E classes. The authors also implemented two nested MLR models; such models are called NM1 (Nested Model No. 1) and NM2 (Nested Model No. 2) in the following. Their conceptual schemes based on two layers of logistic or multilinear logistic regressions are illustrated in Fig. 4.

Both NM1 and NM2 provide two-step predictions, one for each layer of the models. In the first layer, logistic regressions (LR) are used to classify between two groups of damage classes, whereas the second prediction defines the specific damage class. The two models, NM1 and NM2, differ in how the damage classes are grouped in the first layer: in NM1, Fig. 4(a), two groups include the damage classes $G_1$ = {A,B} and $G_2$ = {C,D}, respectively; in NM2, Fig. 4(b), the first group only contains class A (i.e., $G_3$ = {A}), whereas the second group contains the three remaining damage classes (i.e., $G_4$ = {B,C,D}).

In the case of the nested models, the probability of a given outcome can be straightforwardly obtained by applying the rules of conditional probability as a consequence of the procedure used to calibrate the nested models. For example, by using the nested models NM1, the probability of classification in CLASS A can be found as $\mathrm{P}(Y = A) = \mathrm{P}(Y = A | Y \in A \cup B)\mathrm{P}(Y \in A \cup B)$ where the first logistic regression $\mathrm{LR}_1$ provides $\mathrm{P}(Y \in A \cup B)$ (i.e., $\mathrm{P}(Y \in G_1)$) and $\mathrm{LR}_{1a}$ is used to get $\mathrm{P}(Y = A | Y \in A \cup B)$.

### 3.2. Artificial neural networks

An ANN model is developed for two reasons. Firstly, the ANN model helps to assess the accuracy and effectiveness of the ANN in capturing complex relationships in the data compared to the more traditional MLR. Secondly, the ANN model can detect nonlinear interactions among the regressors that could lead to potential improvements in damage classification. By evaluating the ANN's performance, the authors can determine whether it could offer advantages over MLR in predictive accuracy.

The authors implemented a feed-forward fully-connected multi-layer perceptron architecture [50–52]. This acknowledged widespread model can be described as a computational graph whose building blocks are represented by perceptron units [51].

A perceptron unit is the building block of a neural network. A perceptron unit takes multiple input signals, each with an associated weight, and computes the weighted sum of these inputs. The weighted sum is then passed through an activation function, which introduces non-linearity to the model. Mathematically, for a single perceptron unit with inputs $x_1, x_2, \dots, x_n$ and corresponding weights $w_1, w_2, \dots, w_n$, the output $h$ is computed as follows:

$$h = w_1 \cdot x_1 + w_2 \cdot x_2 + \cdots + w_n \cdot x_n \tag{5}$$

The output $h$ is then passed through the activation function $\Phi(\cdot)$ to obtain the final output of the perceptron unit. The activation function determines whether a neuron should be "activated" or "fired" based on its input. For example, some activation functions like the sigmoid or tanh function map the input values to a range between 0 and 1 or −1 and 1, respectively, making the neurons produce binary-like outputs. Common activation functions include:

- Sigmoid function: $\sigma(x) = \frac{1}{1+e^{-x}}$
- Tanh function: $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
- ReLU (Rectified Linear Unit): $\mathrm{ReLU}(x) = \max(0, x)$
- Softmax function: $\mathrm{softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$

Perceptron units can be arranged in layers with weight parameters to form the network's architecture. These networks can then be trained using various optimization algorithms to learn patterns and relationships in the input data, enabling them to perform classification, regression, and pattern recognition tasks.

In this paper, the model architecture is composed of an input layer, an output layer and a user-defined number of hidden layers [50–52]. The hidden layers fulfil the role of providing recursive non-linear transformations of the input feature vector, in this case, represented by the 60 dummy binary variables $\mathbf{x}$ for each of the 844 school data. The output of the first hidden layer is the result of softmax activation function $\Phi(\cdot)$ applied to the linear combination of the input feature vector $\mathbf{x}$ and a weight matrix $\mathbf{W}_1^T$ of learnable parameters:

$$\boldsymbol{h}_1 = \Phi(\mathbf{W}_1^T \mathbf{x}) \tag{6}$$

The hidden state $\boldsymbol{h}_1$ is subsequently fed to the next hidden layer or a final output layer. In this study, the dimensions of the weight matrix $W_1$ is $61 \times n_1$, being, respectively, the number of input features increased by one for accounting for the bias term, and the number of hidden units $n_1$. The bias term is an additional neural network parameter allowing the model to make predictions even when all the input features are zero. Mathematically, the bias term is added to the weighted sum of inputs and acts as a constant term in the activation function.

Implicitly, this kind of implementation employs the same activation for each building block lying in the same hidden layer [52]. The input layer has no weights because it is assumed to only transmit the data to every neuron of the next hidden layer (fully connected). For the last $k$th hidden layer before the output layer, the hidden state is given by

$$\boldsymbol{h}_k = \Phi(\mathbf{W}_k^T \boldsymbol{h}_{k-1}) \tag{7}$$

In the current study, since the ANN is demanded to perform a multinomial classification task, the output layer is structured with four units, respectively, for the four damage classes A, B, C, and E. In this last layer, the activation function is a softmax function $\Phi_s(\cdot)$, which provides probabilities of each feature vector $\mathbf{x}$ belonging to a specific output class.

$$\boldsymbol{o} = \Phi_s(\mathbf{W}_{k+1}^T \boldsymbol{h}_k) \tag{8}$$

To train the ANN and thus define the weight matrices, the stochastic gradient descendent back-propagation algorithm was adopted to minimize the categorical cross-entropy loss function [52].

The stochastic gradient descent (SGD) is an optimization algorithm used to update the neural network weights during the training process. Instead of computing the gradients of the entire dataset, SGD uses random subsets, called mini-batches, of the data to estimate the gradients, making it computationally more efficient and suitable for large datasets. It iteratively updates the weights, aiming to find the optimal set of weights that minimizes the error between the model's predictions and the actual data. The gradients are obtained from the backpropagation method, a technique used in neural networks to compute the gradients of the loss function with respect to the model's parameters (weights). The authors used the categorical cross-entropy loss as the objective function, commonly used in multi-class classification problems. This objective function compares the model's predicted probability distribution (output) with the true encoded labels (ground truth) for each data point. In this paper, the authors used the Adam (Adaptive Moment Estimation) implementation for solving the optimization algorithm, which is an extension of the SGD. It adapts the learning rate for each weight based on the average of the past gradients and their squared gradients [52].

The architecture of the ANN should be carefully evaluated since it strongly affects the model complexity, the classification accuracy rate, and the model's tendency to underfitting or overfitting issues [40,52].

Underfitting occurs when a model is too simple to capture the underlying patterns in the training data. Thus, the model cannot learn from the data and fails to generalize well to new, unseen data. Conversely, overfitting occurs when a model is too complex and captures noise or random fluctuations in the training data rather than the true underlying patterns. In this case, the model performs extremely well on the training data but poorly on the validation or test data.

Therefore, to minimize the likelihood of overfitting issues, it is advisable to reduce the number of learnable parameters compared to the number of limited data at the disposal while simultaneously attempting to avoid underfitting issues. Various attempts have been made in the literature to define an optimal architecture for ANN as a single-objective optimization problem [40,53–57]. Therefore, to define an optimal ANN architecture in a more systematic and structured way, the authors implemented an unconstrained multi-objective optimization problem. The two opposing goals are maximizing the classification accuracy $a_{[\%]}$ (defined as the percentage of correct predictions) on one side and minimizing the total number of hidden units $\sum_{j=1}^{k} n_j$ on the other side, i.e. reducing the model complexity to an acceptable level attempting to avoid both overfitting and underfitting issues. To formulate the optimization problem as a double minimization problem, the maximization of accuracy was rephrased as the minimization of the classification error, assumed as the complement value of the accuracy to the ideal level of 100%.

$$\begin{cases} \min_{\theta} f_1 = 100 - a_{[\%]} \\ \min_{\theta} f_2 = \sum_{j=1}^{k} \theta_j = \sum_{j=1}^{k} n_j \end{cases} \tag{9}$$

where $\theta_j$ is the $j$th element of the design vector $\boldsymbol{\theta}$ of the multi-objective problem. To effectively solve the aforementioned multi-objective optimization problem, the authors adopted the well-acknowledged meta-heuristic algorithm non-dominated sorting genetic algorithm (NSGA-II) provided in the Python implementation of [58]. In the NSGA-II, similarly to the other genetic-algorithm-based versions, a population of candidate solutions is iteratively improved toward the optimum according to the Darwinian theory of biological evolution [59]. In the genetic algorithm jargon, the iterations are named generations.

The design parameters considered in the problem (9) are initially encoded as in [40]. The parameter vector length accounts for the number of hidden layers, and each vector component is an integer value that establishes the number of units within that hidden layer. Additionally, the last component is an integer $p$ from one to three, which selects different activation functions, specifically the Rectified Linear Unit (ReLU), the hyperbolic tangent function (tanh), and the sigmoid function, respectively [52]. The same activation function is applied to all the hidden layers.

$$\boldsymbol{\theta} = [\underbrace{n_1, \overbrace{n_2}^{\text{No. units}}, \ldots, n_k}_{\text{No. hidden layers}}, \overbrace{p}^{\text{Act. fn.}}]^T \tag{10}$$

However, as shown by a preliminary analysis and expected by the authors, the tanh activation function is preferable in dealing with such multinomial classification tasks due to its more accessible training properties, e.g., the sigmoid function. The tanh represents a stretched and re-scaled version of the sigmoid function, thus providing more significant gradients [60]. Furthermore, it is regular, and its derivative is smooth everywhere without any discontinuities compared to the ReLU. Several numerical preliminary tests confirmed that the algorithm almost always prefers the tanh function. Therefore, the design parameter definition in Eq. (10) can be simplified by dropping the activation function term and posing the tanh for every hidden neuron.

$$\boldsymbol{\theta} = [\underbrace{n_1, \overbrace{n_2}^{\text{No. units}}, \ldots, n_k}_{\text{No. hidden layers}}]^T \tag{11}$$

Once a population of candidate design parameters is heuristically generated from the NSGA-II procedure, every individual was trained with cross-entropy loss [50,60], optimized with the Adam algorithm [50,60], posing a maximum number of 100 epochs and a batch size of 256.

In the ML lexicon, one epoch means the model has seen and learned from all the training samples once. During each epoch, the model updates its weights and biases based on the training data to improve its performance on the given task. Training for multiple epochs allows the model to iteratively refine its parameters and learn from the data more effectively.

For every individual, five models are trained in the cross-validation fashion with $k = 5$ fold.

Specifically, the process involves dividing the dataset into five equal parts, known as folds. For each cross-validation iteration, one of these folds is set aside as the test set, while the remaining four folds are used as the training set. The model is then trained on the training set and evaluated on the test set to measure its accuracy. This process is repeated five times, with each fold as the test set once. The reason for doing this is to ensure that a specific set of data does not bias the model's performance, and it provides a more robust assessment of how well the model can generalize to unseen data. The results from all five iterations are averaged to compute the average accuracy. The data in each fold is randomly shuffled before training the model to ensure that the cross-validation process is unbiased. This means that the order of the data points in the test set is different for each cross-validation iteration. The random shuffling helps prevent patterns or biases in the original data.

When the stopping criteria of the maximum allowable number of generations are reached, the multi-objective optimization procedure is interrupted, and the final Pareto front of the non-dominated founded solutions is obtained. Non-dominated solutions

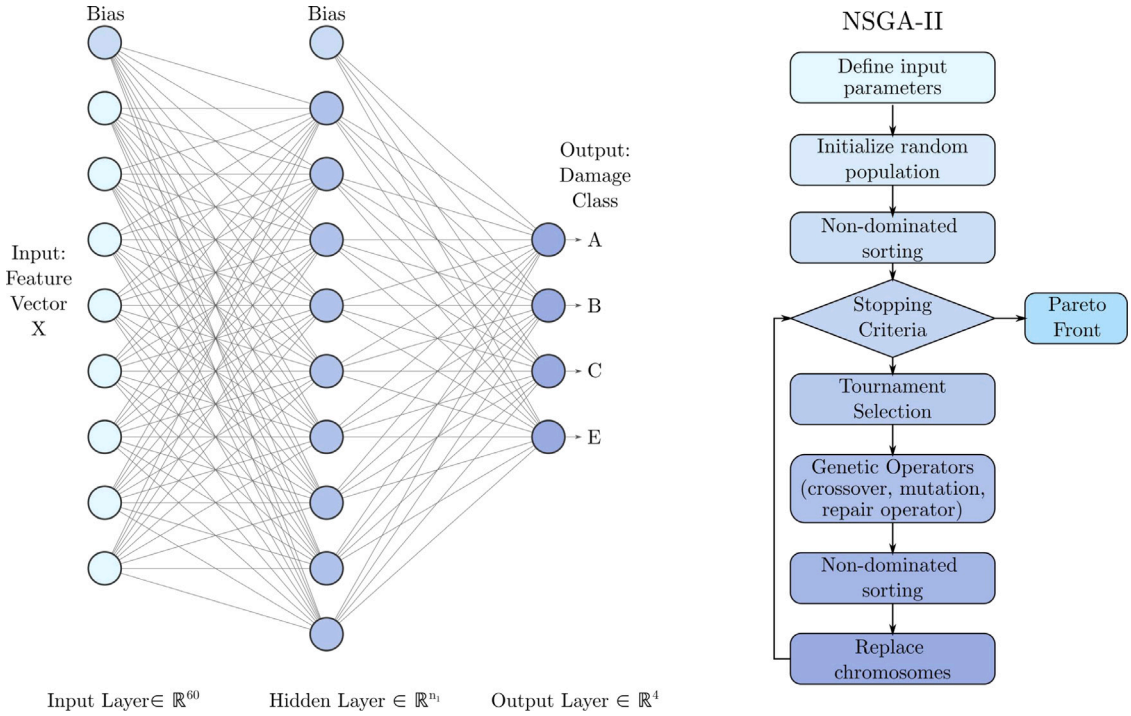


**Fig. 5.** Workflow of the optimal ANN architecture.

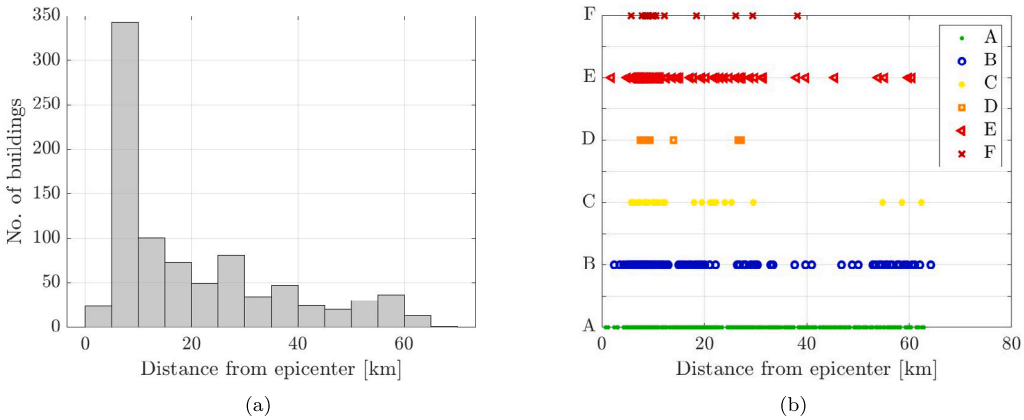


**Fig. 6.** (a) Histogram plots of the number of buildings as a function of the distance from the epicentre; (b) Scatter plot of the outcome of the evaluation as a function of the distance from the epicentre.

represent all the admissible, efficient, and feasible solutions found by the algorithm that do not present any apparent relationship among them besides belonging to the Pareto optimal set, and thus composing the frontier of the Pareto decisional front [61].

A schematic view of the current ANN application is illustrated in Fig. 5.

Fig. 5 presents a visualization of the ANN architecture and the flowchart of the optimization algorithm. On the left side, the architecture of the ANN is depicted, consisting of three layers: the input layer, hidden layers, and the output layer, with four categorical outputs representing the different damage classes. Each circle in the architecture represents a neuron, with the top one being the bias term.

The right side of the figure displays the flowchart of the optimization algorithm employed to optimize the ANN's architecture, based on the NSGA-II (Non-dominated Sorting Genetic Algorithm II). NSGA-II is a well-known evolutionary algorithm for multi-objective optimization, as mentioned before [62]. The flowchart consists of several sequential steps:

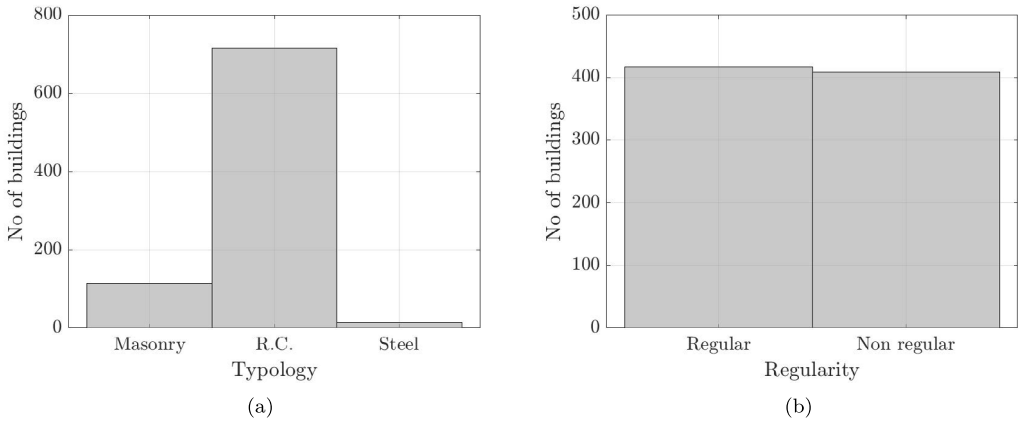


**Fig. 7.** (a) Histogram plot of the building typology and (b) of the regularity features.

- Define input parameters: This step specifies the parameters required for the optimization process, such as the population size and the number of generations.
- Initialize random population: The algorithm starts by creating a population of random candidate solutions (chromosomes) representing different ANN architectures.
- Non-dominated sorting: NSGA-II performs a sorting process to identify non-dominated solutions that are not outperformed by any other solution in the population.
- Stopping criterion: The algorithm checks if the termination condition is met.

Within the while loop, the following four steps are executed repeatedly until the stopping criterion is satisfied:

- Tournament selection: A tournament selection process is performed to choose parent chromosomes based on their fitness values.
- Genetic operators: These are the crossover and mutation operators applied to the selected parents to create new offspring. Crossover involves exchanging information between two parent chromosomes to produce new child chromosomes, while mutation introduces small random changes to the offspring. The repair operator ensures that the generated offspring conforms to the architectural constraints.
- Non-dominated sorting: The newly generated offspring and the current population are sorted again based on non-dominated criteria.
- Replace chromosomes: The least fit solutions from the current population are replaced by the new offspring.

The algorithm then loops back to the stopping criterion, repeating the steps above until the termination condition is met. The final output of the NSGA-II algorithm is the Pareto front, which represents a set of non-dominated solutions, where each solution is optimal for one or more of the objectives while not being dominated by any other solution in the set. The Pareto front provides a range of trade-off solutions, allowing the selection of the most appropriate ANN architecture based on different evaluation criteria.

## 4. Analysis of the dataset

This section presents the analysis of the considered database that includes data from the AeDES forms compiled after the 2009 earthquake in L'Aquila for the school buildings. The initial number of the considered forms is 878; however, after removing the data associated with the damage classes F and D, as discussed in the previous section, the total number of the forms reduces to 844. The Department of Civil Protection of the Abruzzo Region only shared selected data from the original database related to the geometric-typological data and the observed damage.

Right after the earthquake, teams of engineers were set up to conduct site inspections on public and private buildings. The goal of assigning this task to high-qualified teams, especially for public and strategic buildings, aimed at reducing the sources of uncertainty in the evaluation. Nonetheless, as explained before, the need for objectivity classes goes against the complexity of the evaluation only based on visual inspections and the direct relationship between the outcome of the assessment and economic reimbursement for the damage. Thus, the assessment is unavoidably slightly biased by the sensitivity of the evaluator. The analysis in this section highlights the existence of this personal bias and its influence on the classification of similar classes: A and B.

The selected database comprises buildings inside and outside the seismic crater; see Fig. 3. As a result, there is an unbalance between the no-damaged and damaged cases and more than 50% of the database corresponded to buildings classified as A. Fig. 6 plots the number of buildings as a function of the distance from the epicentre. Most of the buildings belong to the range of 10–20 km. However, as highlighted in Fig. 6(b), there is no evident relationship between the damage class and the distance from the epicentre.

**Table 1**
Input matrix for defining the features of a masonry building.

| Horizontal structure | Vertical structure | | | | |
|---|---|---|---|---|---|
| | Not identified | Irregular masonry | | Regular masonry | |
| | | Without curbs | With curbs | Without curbs | With curbs |
| Not identified | A1 | B1 | C1 | D1 | E1 |
| Vaults without steel chain | A2 | B2 | C2 | D2 | E2 |
| Vaults with steel chain | A3 | B3 | C3 | D3 | E3 |
| Beams with deformable floor | A4 | B4 | C4 | D4 | E4 |
| Beams with semi-deformable floor | A5 | B5 | C5 | D5 | E5 |
| Beams with rigid floor | A6 | B6 | C6 | D6 | E6 |

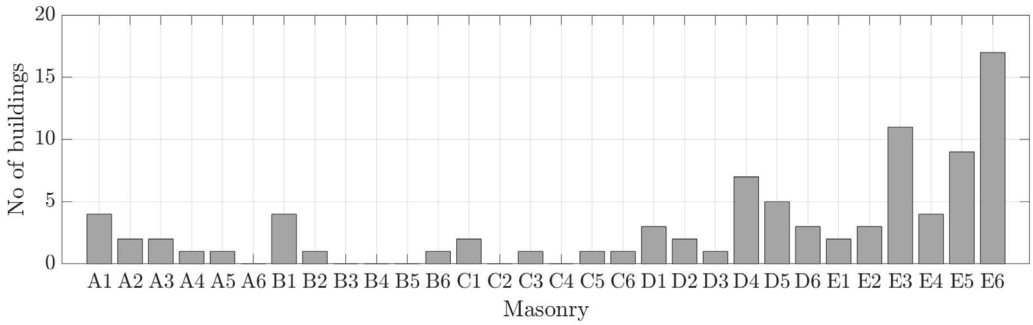


**Fig. 8.** Histogram plot of the features of masonry buildings, following the nomenclature in Table 1.

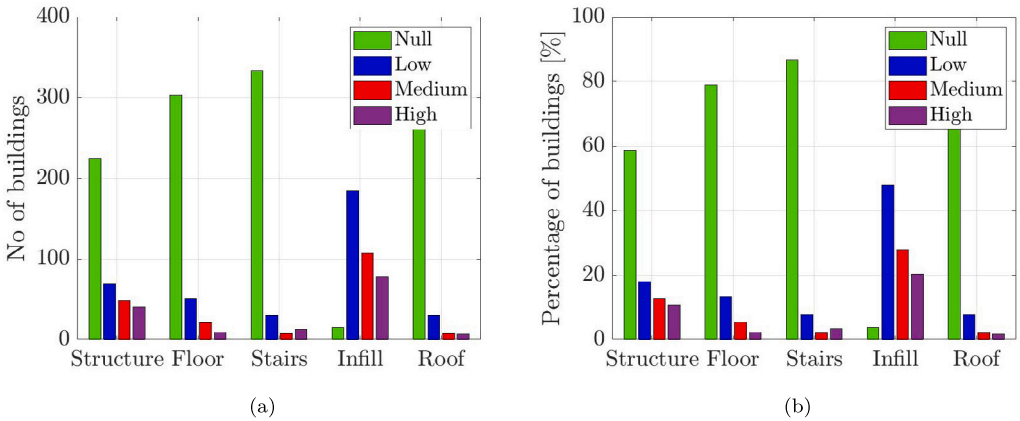


**Fig. 9.** Histogram plots of the observed damages to the structure, floor, stairs, infill and roof in terms of (a) absolute and (b) relative number buildings.

Heavily damaged school buildings can be found significantly far from the epicentre (>50 km), and conversely, undamaged school buildings can be found in their relative proximity.

However, if we look at each damage class, we can see that the number of school buildings classified in the heavier damage classes, i.e., classes C and E, decreases farther from the epicentre. Most school buildings are made of reinforced concrete (see Fig. 7(a)), a lesser number are built with masonry, and a few are steel structures. There is a substantial balance between regular and non-regular buildings in height and plane (see Fig. 7(a)). The AeDES forms include a few options for the typological classification of RC buildings. At the same time, a significant number of alternatives are provided for masonry buildings due to their wider typological variety, as shown in Table 1.

Fig. 8 shows the histogram plots of the masonry buildings falling in the categories described in Table 1. Except in a few cases where the visual inspection did not allow identifying the floor or bearing structures (i.e., those classified as A1-A6), the school buildings mostly exhibit a regular masonry structure.

Looking at the damage records, generally, floors, stairs, and roofs did not manifest significant damage, with almost 80% of the buildings with no damage to these sub-structures, see Fig. 9. Consequently, most inspected buildings were classified as A. Conversely, most damages were concentrated in the bearing structure and infill. Table 2 summarizes the information illustrated in the histogram plots in Fig. 9.

**Table 2**
Percentages of the buildings exhibiting the damage typologies identified in Fig. 1.

| Location | Null | Low | Medium | High |
|---|---|---|---|---|
| Structure | 58.59% | 17.97% | 12.76% | 10.68% |
| Floor | 78.91% | 13.28% | 5.47% | 2.34% |
| Stairs | 86.72% | 7.81% | 2.08% | 3.39% |
| Infill | 3.91% | 47.92% | 27.86% | 20.31% |
| Roof | 88.28% | 7.81% | 2.08% | 1.82% |

## 5. Results

The results are organized into two subsections, showing the results of the MLR and ANN models.

### 5.1. Multinomial logistic regressions

This section presents and compares the results obtained with the following models:

- NN model (Not-nested model): This model is calibrated with all available data and provides in a single step the predicted damage class among all the considered damage classes: A, B, C, and E.
- NM1 model (Nested model No 1): This model (see Fig. 4) predicts the damage class in two steps. In the first step, an LR ($LR_1$) finds the probability of the damage matrix being associated with a no/lower damage (A and B) or higher damage class (C and E). In the second step, two MLRs classify the damage between classes A and B and C or E, respectively. Therefore, three MLRs are separately calibrated to predict the final damage class.
- NM2 model (Nested model No 2): The second nested model is similar to the first except for defining the intermediate damage classes. In this case, the no-damage class comprises the sole class A, while the higher-damage class include classes B, C, and E. Therefore, in the first layer, an LR predicts whether the damage to the school building is negligible. In contrast, in the second layer, an MLR predicts the level of non-negligible damage among classes B, C, and E. This model requires the calibration of two MLR models.

The three models have a similar performance in capturing the variability of the damage classification in the AeDES forms. Specifically, the NN and NM2 models have an almost identical prediction accuracy, nearly equal to 84%. In comparison, NM1 performs poorly with an 82% prediction accuracy. The analysis of the prediction accuracy of the presented models shown through confusion matrices can provide insights into possible issues related to the current classification approach of the AeDES forms.

Confusion matrices summarize the model's predictions and the actual ground truth for each class. The diagonal elements (top-left to bottom-right) represent the number of correct predictions for each class. The off-diagonal elements show the misclassifications, where predictions did not match the true class labels. In the depicted confusion matrices, there are additional metrics reported: recall, precision and accuracy. The right column, filled in grey, shows the recall, also known as sensitivity or true positive rate, which measures the model's ability to correctly identify positive instances for each class. The bottom row, filled in grey, displays the precision, representing the proportion of true positive predictions out of all predicted positive instances for each class. Furthermore, the bottom-left cell in the matrix shows the model's overall accuracy, which measures the proportion of correctly classified instances out of the total predictions. Accuracy provides a general overview of the model's overall performance.

In particular, the comparison between the confusion matrices of the NN and the first layer of the NM1 models (Fig. 10) clarifies the reasons behind the limits in the accuracy of the models. While the accuracy of the NN model in classifying the damage is 84% (Fig. 10(a)), in the first layer of the NM1 model, the LR performs excellently in classifying between no-damage vs. damage classes, with an accuracy higher than 97% (Fig. 10(b)). This evidence proves that the loss in accuracy in the final classification of the NM1 model depends on the performance of the LRs in the second layer of the model, i.e., on the difficulty in distinguishing between similar damage classes, classes A-B (Fig. 11(a)) and classes C-E (Fig. 11(b)), where the green and red percentages indicate the correct and incorrect classification percentages, respectively.

The confusion matrices of the LRs in the second layer of NM1 show that the classification between A and B is less accurate (82% accuracy) than the classification between C and E (98% accuracy). Therefore, the accuracy loss in the final classification of NM1 (and the 20% of wrong predictions) can be reasonably attributed to the logistic regression $LR_1 a$ (in Fig. 4) and to the classification between class A and class B.

The conceptual reason behind this issue is the difficulty in understanding whether minor damages require some actions or whether the structure can be considered fully operational. For this reason, in some cases, similar damage patterns might have led to the selection of different damage classes. In addition to the "complexity" of the choice between classes A and B, a further issue is the responsibility of the choice. It must be remarked that choosing between classes A and B involves more responsibility than choosing between classes C and E. In the case of the choice of class A, the investigation team states that the building is fully operational and requires no intervention. Likely, the complexity of the case-by-case circumstance makes an automatic classification between classes A and B difficult and always demands the expert's judgment. Additionally, this responsibility grows in the presence of private buildings. A less conservative classification can lead to an appeal by the owner, who might require further investigations to confirm the results of the first assessment. This can lead to lengthening the administrative procedures in an emergency phase.

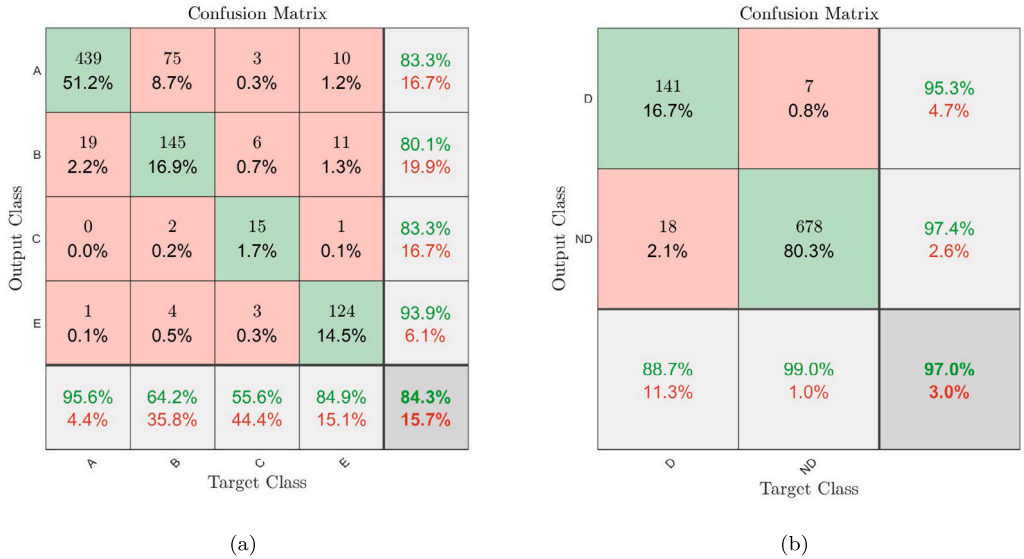


**Fig. 10.** Confusion matrices of the non-nested model and the nested No one regarding damage/no-damage outputs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

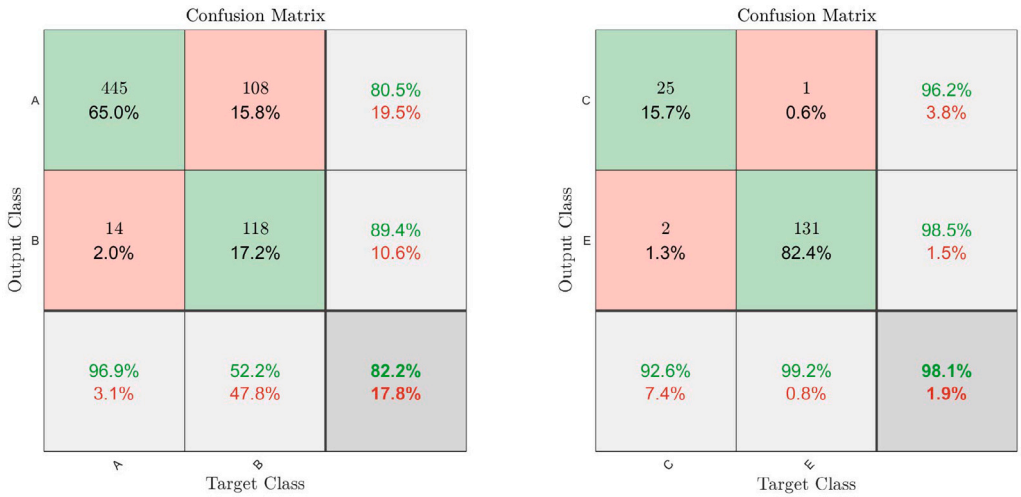


**Fig. 11.** Confusion matrices of the nested no.1 model in teams of (a) A–B, and (b) C–E predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

So, in the case of private buildings, the expert's judgment can be more conservative to avoid dialectics that could compromise the speed of the procedure. This further justifies the reasons for selecting the sample of only public buildings to calibrate the models in this paper. In this case, the expert's judgment is less conditioned by a possible debate with the owner.

The comparison among the nested models in Fig. 12 proves that binning classes B, C and E in a damage macro-class leads to a better performance than binning classes A and B in a no-damage/light damage class since the two nested models yield 82% and 84% accuracy, respectively. Given the substantial similarity between the three MLR models' performance, the authors selected the first no-nested model, due to its simplicity of implementation, as an MLR model for the AeDES form enhancement.

The coefficient matrix, reported in Table 3, can be used to explicitly write the model and predict the expected damage class given a damaged matrix. The MLR model provides the probability of falling in classes A, B, C and E, thus driving the examiner towards a more objective evaluation consistent with the observed damage. Although using this model will not substitute the personal responsibility of the expert team, which will make the final decision about the damage class, it can still be used to support the decision-making process towards a more coherent and informed assessment.

In MLR, the $\beta$ coefficients (also known as logit coefficients) in Table 3 represent the relationship between the independent variables (predictors) and the log-odds of the different categories (classes) of the dependent variable. To assess the significance

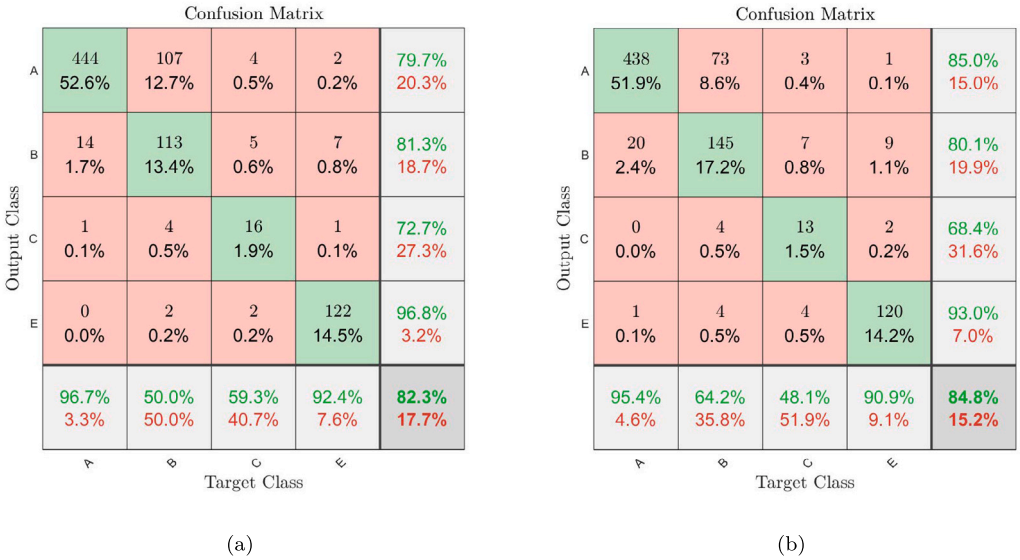


**Fig. 12.** Confusion matrices of the two nested models.

of a regressor and compare its influence relative to other regressors, the authors examined the magnitude of the coefficient and its statistical significance. Larger magnitudes indicate a more substantial impact on the odds, while smaller magnitudes suggest a weaker influence. Additionally, the authors performed a t-test to determine whether the coefficient significantly deviates from zero. A coefficient with a $p$-value below a chosen significance level (assumed equal to 0.05) is considered statistically significant. Table 4 provides an overview of the absolute values of all $\beta$s coefficients for the NN model and their corresponding p-values. The colour formatting used in the table highlights the magnitudes of the betas on a red scale and identifies p-values below 0.05. This combined approach allows a visually intuitive understanding of the most relevant damage regressors.

The most impactful predictors for classification are those associated with A versus E baseline and B versus E baseline. Interestingly, there is statistical significance for high preexisting damage, particularly in the less than 1/3 category for B versus E, implying that for B outcomes with preexisting damage, the earthquake has potentially exacerbated the existing damage. The colour formatting predominantly emphasizes the damages to structures, floors, roofs, infill, and pre-existing damages, excluding stairs. However, the statistical significance of the coefficients is mainly limited to the higher damage levels, high and medium. The coefficients for stairs do not appear as strong regressors, indicating that although they may have affected seismic performance, the damage to stairs itself may not have driven the final assessment. Structural damage classification is paramount among the mentioned damage locations, with all coefficients for high and medium classes averaging higher than 10. Some classes exhibit higher effects, possibly due to an unbalanced dataset where specific options were preferred more than others during compilation. Furthermore, it should be noted that assessing the extent of damage in a fractional way (e.g., 2/3, 1/3) can be challenging in rapid assessments. Higher coefficients for certain classes might reflect the impact of missing or underrepresented data in the form compilation process. This can result in an overestimation of the significance and effect of these predictors on the classification process. The observed good match between p-values lower than 0.05 and higher $\beta$ values indicates that statistically significant predictors generally have larger effects on the classification process.

The confusion matrix for the selected NN model in Fig. 10(a) is also important to address challenges posed by subjective evaluations, where different evaluators might interpret damage patterns differently based on their experience. In an unbalanced dataset, accuracy, good for model comparisons, may not be suitable for evaluation, as it can be misleading when the majority class dominates predictions. Therefore, focusing on recall becomes crucial. High recall indicates the model's ability to correctly identify instances of the minority class, such as severe damage, which is of greater interest or concern. Precision, on the other hand, measures the model's ability to avoid false positives. To strike a balance between capturing true positives and avoiding false positives, the F1-score is also calculated, accounting for both recall and precision.

Table 5 resumes the recall, precision and F1-score for the four classes based on the NN model.

- Class A: Approximately 83.13% of instances that truly belong to class A are correctly identified by the model as class A. This indicates the model's effective ability to capture most instances of class A. When the model predicts an instance as class A, it is correct about 94.91% of the time, demonstrating high precision for class A predictions. To determine the higher tendency between classifying A as B or vice versa, the authors compared the false positive rates (FP rates) for each class. The False Positive Rate (FP Rate) is calculated as FP/(FP + TN), where FP represents the number of instances classified as B but truly belonging to class A (instances incorrectly classified as B), and TN is the number of instances correctly classified as not B (instances truly belonging to class B and accurately identified). Comparing the FP rates, it is evident that the false positive

**Table 3**
Coefficient matrix of the NN model.

| Damage regressors | | | Coefficient vectors $\beta_j$ | | |
|---|---|---|---|---|---|
| Location | Intensity | Extension | $\beta_1$ (A/E) | $\beta_2$ (B/E) | $\beta_3$ (C/E) |
| Intercept | | | 3.326 | 3.623 | 1.285 |
| Structure | High | >2/3 | −10.349 | −14.176 | 0.226 |
| | | 1/3-2/3 | −11.878 | −20.925 | −11.570 |
| | | <1/3 | −10.375 | −14.872 | −2.646 |
| | Medium | >2/3 | −12.835 | −5.569 | −15.850 |
| | | 1/3-2/3 | −5.934 | −5.422 | −6.112 |
| | | <1/3 | −28.410 | −3.422 | −1.439 |
| | Low | >2/3 | −10.457 | 1.436 | −2.057 |
| | | 1/3-2/3 | −2.202 | −0.817 | −20.158 |
| | | <1/3 | −0.757 | 0.126 | −3.506 |
| | Null | | 0.439 | 0.732 | −2.943 |
| Floor | High | >2/3 | 7.075 | 0.620 | 3.888 |
| | | 1/3-2/3 | 6.181 | −2.743 | −6.052 |
| | | <1/3 | 12.939 | 3.326 | −3.034 |
| | Medium | >2/3 | −27.881 | −19.400 | 2.544 |
| | | 1/3-2/3 | −8.264 | −7.432 | −5.841 |
| | | <1/3 | −10.419 | 1.258 | −0.735 |
| | Low | >2/3 | −2.612 | 5.396 | −8.346 |
| | | 1/3-2/3 | 1.371 | −0.311 | −3.800 |
| | | <1/3 | 2.875 | 2.128 | −1.609 |
| | | Null | 3.956 | 2.686 | −0.666 |
| Stairs | High | >2/3 | −16.746 | 16.597 | −0.842 |
| | | 1/3-2/3 | −3.230 | 3.105 | 0.221 |
| | | <1/3 | −4.247 | 0.699 | 4.276 |
| | Medium | >2/3 | −0.090 | 3.238 | 3.428 |
| | | 1/3-2/3 | −0.521 | −1.392 | 0.593 |
| | | <1/3 | −4.101 | −11.579 | 2.990 |
| | Low | >2/3 | −1.213 | −21.374 | 10.417 |
| | | 1/3-2/3 | −0.064 | −1.012 | 0.137 |
| | | <1/3 | −0.338 | 0.263 | 0.265 |
| | | Null | 0.600 | 0.302 | 3.438 |
| Roof | High | >2/3 | −19.229 | −8.115 | −6.440 |
| | | 1/3-2/3 | −15.763 | −11.252 | 2.032 |
| | | <1/3 | 9.388 | 15.728 | 5.900 |
| | Medium | >2/3 | −0.499 | −8.800 | −14.813 |
| | | 1/3-2/3 | −9.602 | −2.450 | 3.178 |
| | | <1/3 | −0.017 | 2.050 | 6.099 |
| | Low | >2/3 | 5.781 | 1.428 | −3.460 |
| | | 1/3-2/3 | −12.432 | −6.255 | −7.577 |
| | | <1/3 | −4.618 | −2.738 | −1.453 |
| | Null | | −3.773 | −2.833 | 0.008 |
| Infill | High | >2/3 | −22.820 | −13.903 | −14.184 |
| | | 1/3-2/3 | −19.906 | −3.748 | −2.747 |
| | | <1/3 | −15.604 | −1.196 | −0.394 |
| | Medium | >2/3 | −10.837 | −3.102 | −8.721 |
| | | 1/3-2/3 | −21.658 | −2.470 | −1.820 |
| | | <1/3 | −4.549 | −0.231 | −1.370 |
| | Low | >2/3 | −1.293 | 12.415 | 9.269 |
| | | 1/3-2/3 | −0.866 | −0.193 | 1.070 |
| | | <1/3 | −1.113 | −2.097 | −1.464 |
| | Null | | −0.953 | −3.143 | −2.134 |
| Existing | High | >2/3 | −12.646 | −13.549 | −8.127 |
| | | 1/3-2/3 | −16.160 | −8.776 | −7.407 |
| | | <1/3 | 14.831 | −44.331 | −13.555 |
| | Medium | >2/3 | 18.847 | −7.357 | 4.117 |
| | | 1/3-2/3 | −0.153 | 1.342 | 3.214 |
| | | <1/3 | 1.944 | −1.295 | −0.344 |
| | Low | >2/3 | 0.620 | 9.191 | −0.932 |
| | | 1/3-2/3 | 20.697 | 10.861 | −0.635 |
| | | <1/3 | 3.592 | 3.490 | −0.877 |
| | Null | | 1.523 | 0.944 | 0.511 |

rate for class A (13.77%) is higher than that for class B (0.83%). This suggests a higher tendency for the model to incorrectly classify instances of class A as class B (false positives) compared to misclassifying instances of class B as class A. Such a tendency

**Table 4**
Absolute values of the coefficient matrix of the NN model and the corresponding p-values.

| Damage regressors | | | Absolute coefficients and p-values | | | | | |
|---|---|---|---|---|---|---|---|---|
| Location | Intensity | Extension | (A/E) | | (B/E) | | (C/E) | |
| | | | $\lvert\beta_1\rvert$ | p-value | $\lvert\beta_2\rvert$ | p-value | $\lvert\beta_3\rvert$ | p-value |
| Intercept | | | 3.326 | 0.001872 | 3.623 | 5.34E−05 | 1.285 | 0.281539 |
| Structure | High | 2/3; | 10.349 | 0.035814 | 14.176 | 0.00414 | 0.226 | 0.9634 |
| | | 2/3-1/3 | 11.878 | 2.32E−05 | 20.925 | 2.93E−08 | 11.570 | 0.003116 |
| | | 1/3; | 10.375 | 6.07E−06 | 14.872 | 2.58E−06 | 2.646 | 0.052101 |
| | Medium | 2/3; | 12.835 | 1.79E−05 | 5.569 | 0.000203 | 15.850 | 2.02E−06 |
| | | 2/3-1/3 | 5.934 | 1.61E−05 | 5.422 | 2.42E−06 | 6.112 | 0.004402 |
| | | 1/3; | 28.410 | 2.98E-74 | 3.422 | 0.000161 | 1.439 | 0.197066 |
| | Low | 2/3; | 10.457 | 0.152756 | 1.436 | 0.865938 | 2.057 | 0.819942 |
| | | 2/3-1/3 | 2.202 | 0.23146 | 0.817 | 0.545018 | 20.158 | 3.75E−07 |
| | | 1/3; | 0.757 | 0.438765 | 0.126 | 0.877155 | 3.506 | 0.004286 |
| | Null | | 0.439 | 0.680526 | 0.732 | 0.426139 | 2.943 | 0.017756 |
| Floor | High | 2/3; | 7.075 | 0.600748 | 0.620 | 0.958371 | 3.888 | 0.720005 |
| | | 2/3-1/3 | 6.181 | 0.312353 | 2.743 | 0.653393 | 6.052 | 0.395285 |
| | | 1/3; | 12.939 | 0.07167 | 3.326 | 0.520208 | 3.034 | 0.716158 |
| | Medium | 2/3; | 27.881 | 2.42E−06 | 19.400 | 2.52E−06 | 2.544 | 0.745782 |
| | | 2/3-1/3 | 8.264 | 0.013288 | 7.432 | 0.037185 | 5.841 | 0.245792 |
| | | 1/3; | 10.419 | 0.000153 | 1.258 | 0.278991 | 0.735 | 0.641975 |
| | Low | 2/3; | 2.612 | 0.748328 | 5.396 | 0.013049 | 8.346 | 0.262031 |
| | | 2/3-1/3 | 1.371 | 0.466672 | 0.311 | 0.856715 | 3.800 | 0.554553 |
| | | 1/3; | 2.875 | 0.005809 | 2.128 | 0.018396 | 1.609 | 0.309061 |
| | | Null | 3.956 | 7.77E−05 | 2.686 | 0.002972 | 0.666 | 0.627519 |
| Stairs | High | 2/3; | 16.746 | 0.006398 | 16.597 | 6.37E−05 | 0.842 | 0.890451 |
| | | 2/3-1/3 | 3.230 | 0.5503 | 3.105 | 0.434638 | 0.221 | 0.971274 |
| | | 1/3; | 4.247 | 0.78302 | 0.699 | 0.964 | 4.276 | 0.782906 |
| | Medium | 2/3; | 0.090 | 0.987298 | 3.238 | 0.052768 | 3.428 | 0.661572 |
| | | 2/3-1/3 | 0.521 | 0.897963 | 1.392 | 0.540831 | 0.593 | 0.900087 |
| | | 1/3; | 4.101 | 0.246003 | 11.579 | 0.000315 | 2.990 | 0.141546 |
| | Low | 2/3; | 1.213 | 0.942209 | 21.374 | 0.141442 | 10.417 | 0.527939 |
| | | 2/3-1/3 | 0.064 | 0.9906 | 1.012 | 0.576595 | 0.137 | 0.980206 |
| | | 1/3; | 0.338 | 0.747643 | 0.263 | 0.772587 | 0.265 | 0.866955 |
| | | Null | 0.600 | 0.440833 | 0.302 | 0.674907 | 3.438 | 0.005467 |
| Roof | High | 2/3; | 19.229 | 0.04222 | 8.115 | 0.359846 | 6.440 | 0.474771 |
| | | 2/3-1/4 | 15.763 | 0.00116 | 11.252 | 0.009851 | 2.032 | 0.383073 |
| | | 1/3; | 9.388 | 0.625792 | 15.728 | 0.393906 | 5.900 | 0.755741 |
| | Medium | 2/3; | 0.499 | 0.97566 | 8.800 | 0.242258 | 14.813 | 0.153974 |
| | | 2/3-1/4 | 9.602 | 0.006787 | 2.450 | 0.15446 | 3.178 | 0.122806 |
| | | 1/3; | 0.017 | 0.997304 | 2.050 | 0.685131 | 6.099 | 0.215695 |
| | Low | 2/3; | 5.781 | 0.632387 | 1.428 | 0.847563 | 3.460 | 0.791735 |
| | | 2/3-1/4 | 12.432 | 0.096326 | 6.255 | 0.00089 | 7.577 | 0.207939 |
| | | 1/3; | 4.618 | 0.000105 | 2.738 | 0.007599 | 1.453 | 0.377175 |
| | Null | | 3.773 | 0.00018 | 2.833 | 0.001719 | 0.008 | 0.995221 |
| Infill | High | 2/3; | 22.820 | 1.52E-12 | 13.903 | 3.79E−05 | 14.184 | 0.006092 |
| | | 2/3-1/5 | 19.906 | 2.87E-20 | 3.748 | 9.42E−05 | 2.747 | 0.06001 |
| | | 1/3; | 15.604 | 4.91E-19 | 1.196 | 0.104649 | 0.394 | 0.698559 |
| | Medium | 2/3; | 10.837 | 5.49E−05 | 3.102 | 0.003252 | 8.721 | 0.002994 |
| | | 2/3-1/5 | 21.658 | 9.69E-33 | 2.470 | 0.000768 | 1.820 | 0.118567 |
| | | 1/3; | 4.549 | 1.39E−06 | 0.231 | 0.722671 | 1.370 | 0.169275 |
| | Low | 2/3; | 1.293 | 0.881872 | 12.415 | 0.107499 | 9.269 | 0.197345 |
| | | 2/3-1/5 | 0.866 | 0.497478 | 0.193 | 0.859318 | 1.070 | 0.462089 |
| | | 1/3; | 1.113 | 0.188236 | 2.097 | 0.001451 | 1.464 | 0.116159 |
| | Null | | 0.953 | 0.335728 | 3.143 | 0.000211 | 2.134 | 0.059545 |
| Existing | High | 2/3; | 12.646 | 0.031412 | 13.549 | 3.37E−05 | 8.127 | 0.018939 |
| | | 2/3-1/6 | 16.160 | 0.00016 | 8.776 | 0.081803 | 7.407 | 0.208178 |
| | | 1/3; | 14.831 | 0.387194 | 44.331 | 0.008119 | 13.555 | 0.422011 |
| | Medium | 2/3; | 18.847 | 0.20109 | 7.357 | 0.638014 | 4.117 | 0.805649 |
| | | 2/3-1/6 | 0.153 | 0.938515 | 1.342 | 0.48004 | 3.214 | 0.119378 |
| | | 1/3; | 1.944 | 0.091662 | 1.295 | 0.13719 | 0.344 | 0.717103 |
| | Low | 2/3; | 0.620 | 0.941695 | 9.191 | 0.021612 | 0.932 | 0.912669 |
| | | 2/3-1/6 | 20.697 | 2.46E-11 | 10.861 | 1.61E−05 | 0.635 | 0.910935 |
| | | 1/3; | 3.592 | 4.23E−05 | 3.490 | 4.37E−05 | 0.877 | 0.487135 |
| | Null | | 1.523 | 0.001963 | 0.944 | 0.049087 | 0.511 | 0.452422 |

**Table 5**
Recall, precision and F1-score for the NN model.

| Class | Recall | Precision | F1-score |
|---|---|---|---|
| A | 0.83 | 0.95 | 0.89 |
| B | 0.84 | 0.63 | 0.72 |
| C | 0.88 | 0.63 | 0.73 |
| E | 0.93 | 0.87 | 0.90 |

could lead to the overestimation of damage in ambiguous circumstances. The F1-score for class A is approximately 0.8866, indicating a balanced performance between recall and precision. The model demonstrates good accuracy in classifying class A instances, achieving a reasonable trade-off between false positives and false negatives.

- Class B: The model correctly identifies approximately 84.40% of instances belonging to class B, demonstrating a reasonable ability to capture class B instances. However, when the model predicts an instance as class B, it is accurate only about 63.39% of the time. The lower precision indicates a higher rate of false positives for class B predictions. The F1-score for class B is around 0.7231, reflecting a balanced evaluation of recall and precision. While the model shows decent recall for class B, it also exhibits a notable number of false positives.
- Class C: The model performs well for class C, correctly identifying about 88.24% of instances that belong to class C. When the model predicts an instance as class C, it is accurate approximately 62.50% of the time. Similar to class B, class C also exhibits a lower precision. The F1-score for class C is approximately 0.7297, indicating a balanced performance between recall and precision. The model effectively captures most class C instances but has some false positives.
- Class E: The model shows a high recall of about 93.08% for class E, indicating that it correctly identifies most instances that belong to class E. When the model predicts an instance as class E, it is accurate approximately 86.54% of the time. This high precision suggests that the model has a low false positive rate. The F1-score for class E is around 0.8965, demonstrating a balanced recall and precision evaluation. The model performs well in classifying class E instances, with relatively few false positives.

The model performs well in capturing instances of each class, with relatively high recall values for all classes. However, the precision values show some variability. Class A and E have higher precision, indicating lower false positive rates, while classes B and C have lower precision, suggesting higher false positives in their predictions.

The authors provided a small demonstration of the model's performance in predicting the outcome of a classification, considering three scenarios illustrated in Table 6.

- Case No. 1 corresponds to a scenario with high damage to the bearing structures and no damage to other elements.
- Case No. 2 simulates significant damage to the structure, along with damage to the floor, roof, and stairs.
- Case No. 3 simulates the case of low damage to the structure, floor, and infill, along with medium pre-existing damage.

Table 7 reports the probabilities of the four outcomes associated with each scenario. In Case No. 1, where there is significant damage to the structure but not accompanied by other damages, the algorithm selects option C, which is generally neglected. On the other hand, in Case No. 2, when extensive damage occurs to the structure along with damage to other elements, the classification unambiguously selects option E. Finally, in Case No. 3, when moderate damage occurs, the algorithm selects option B as the most appropriate classification. The authors have provided a supplementary Excel spreadsheet containing the demonstrations found in Table 7.

### *5.2. Artificial neural networks*

As mentioned in Section 3.2, to better govern the model's growing complexity in terms of hidden layers, the authors solved a two-step optimization problem. The authors imposed a single hidden layer with a theoretically maximum number of units equal to 100. The Pareto front of the non-dominated solutions for this first case is reported in blue in Fig. 13. From a visual inspection, the Pareto front demonstrated that despite the reduced number of generations, the algorithm delivered a quite dense Pareto front. Among the competing solution, the one in the knee of the Pareto front is chosen since it provides the best trade-off between accuracy and complexity, thus determining the number of hidden units equal to $n_1 = 16$.

From Fig. 13, it is evident that the further increase in the number of units does not correspond to a significant increase in accuracy. After fixing the optimal number of units for the first hidden layer to $n_1 = 16$, a second optimization is performed to find the optimal number of units of a second hidden layer. The Pareto front of the non-dominated solutions for this second case is reported in orange in Fig. 13; the solution at the knee point considers $n_2 = 7$ hidden units for the second layer. However, the addition of a second layer produces an increase in model complexity without any substantial increase in average accuracy. Therefore, it is not advisable to consider a further hidden layer. Moreover, as stated in [52], an ANN with a single hidden layer is directly comparable to tan MLR models. Therefore, the model with a single hidden layer is with $n_1 = 16$ optimal units is preferred.

Fig. 14 shows the confusion matrix of such a model. The confusion matrix in Fig. 14 illustrates the average classification performances from the $k = 5$ folds cross-validation procedure of the optimized single hidden layer ANN model. The ANN global classification accuracy is comparable to the MLR model, standing at about 80%. For this model, the most problematic class to predict

**Table 6**
Damage matrices of the three considered scenarios.

| Damage regressors | | | Damage matrices | | |
|---|---|---|---|---|---|
| Location | Intensity | Extension | Case No. 1 | Case No. 2 | Case No. 3 |
| Intercept | | | | | |
| Structure | High | 2/3; | 1 | 1 | 0 |
| | | 2/3-1/3 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Medium | 2/3; | 0 | 0 | 0 |
| | | 2/3-1/3 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Low | 2/3; | 0 | 0 | 1 |
| | | 2/3-1/3 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Null | | 0 | 0 | 0 |
| Floor | High | 2/3; | 0 | 0 | 0 |
| | | 2/3-1/3 | 0 | 0 | 0 |
| | | 1/3; | 0 | 1 | 0 |
| | Medium | 2/3; | 0 | 0 | 0 |
| | | 2/3-1/3 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Low | 2/3; | 0 | 0 | 1 |
| | | 2/3-1/3 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | | Null | 0 | 0 | 0 |
| Stairs | High | 2/3; | 0 | 0 | 0 |
| | | 2/3-1/3 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Medium | 2/3; | 0 | 1 | 0 |
| | | 2/3-1/3 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Low | 2/3; | 0 | 0 | 0 |
| | | 2/3-1/3 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | | Null | 0 | 0 | 0 |
| Roof | High | 2/3; | 0 | 1 | 0 |
| | | 2/3-1/4 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Medium | 2/3; | 0 | 0 | 0 |
| | | 2/3-1/4 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Low | 2/3; | 0 | 0 | 0 |
| | | 2/3-1/4 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Null | | 0 | 0 | 0 |
| Infill | High | 2/3; | 0 | 0 | 0 |
| | | 2/3-1/5 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Medium | 2/3; | 0 | 0 | 0 |
| | | 2/3-1/5 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Low | 2/3; | 0 | 0 | 1 |
| | | 2/3-1/5 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Null | | 0 | 0 | 0 |
| Existing | High | 2/3; | 0 | 0 | 0 |
| | | 2/3-1/6 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Medium | 2/3; | 0 | 0 | 1 |
| | | 2/3-1/6 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Low | 2/3; | 0 | 0 | 0 |
| | | 2/3-1/6 | 0 | 0 | 0 |
| | | 1/3; | 0 | 0 | 0 |
| | Null | | 0 | 0 | 0 |

**Table 7**
Probabilities associated with the three considered scenarios.

| Probability | Case No. 1 | Case No. 2 | Case No. 3 |
|---|---|---|---|
| P(A) | 0.0002 | 0.0000 | 0.0005 |
| P(B) | 0.0000 | 0.0000 | 0.9995 |
| P(C) | 0.8192 | 0.0106 | 0.0000 |
| P(E) | 0.1807 | 0.9894 | 0.0000 |
| Sum | 1 | 1 | 1 |

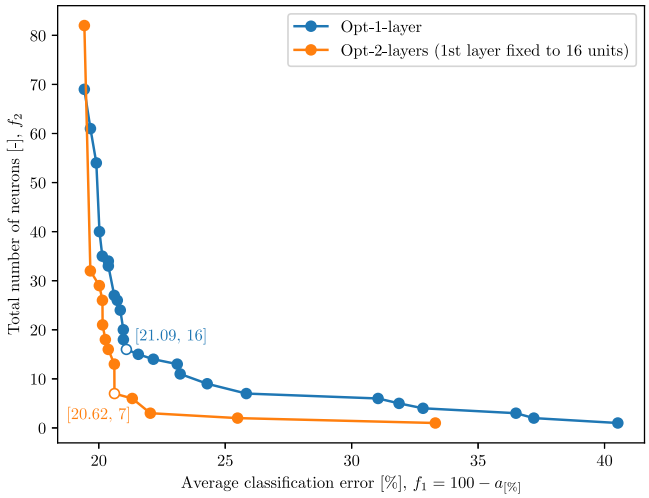


**Fig. 13.** Pareto front results of the two-step multi-objective ANN architecture optimization. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

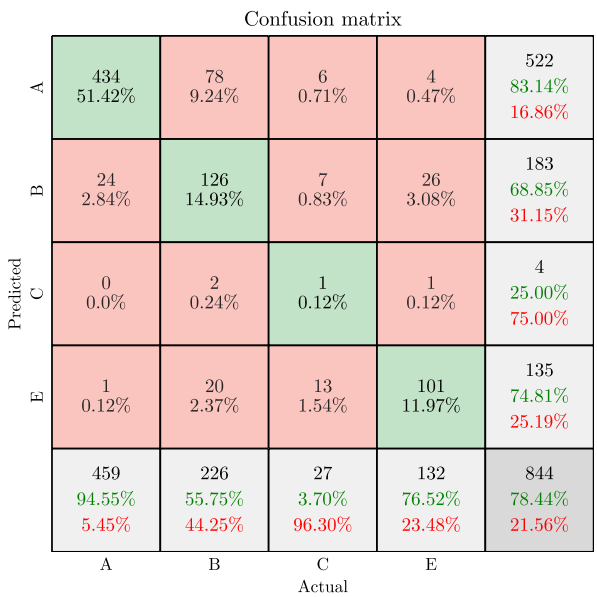


**Fig. 14.** Average confusion matrix of the optimal ANN architecture with a single hidden layer and 16 hidden neurons.

is class C. The main reason for the reduced accuracy lies in the unbalance between the number of records in this class C and the number of records in the other classes, which detrimentally affects the learning ability of the classifier. Furthermore, it is evident from the confusion matrix how class B tends to be partially confused with class A. Therefore, as already discussed in the comments to Fig. 10, to improve the classifiers' learning capabilities, it could be convenient to consider only two major classes, thus binning

**Table 8**
Comparison between recall, precision and F1-score from the MLR (NN) and the ANN models. The relative difference expresses the difference between the ANN performance compared to MLR.

| Class | Recall | | | Precision | | | F1-score | | |
|---|---|---|---|---|---|---|---|---|---|
| | MLR (NN) | ANN | Rel. Diff. [%] | MLR (NN) | ANN | Rel. Diff. [%] | MLR (NN) | ANN | Rel. Diff. [%] |
| A | 0.83 | 0.82 | −1% | 0.95 | 0.93 | −2% | 0.89 | 0.87 | −1% |
| B | 0.84 | 0.57 | −32% | 0.63 | 0.74 | 16% | 0.72 | 0.65 | −11% |
| C | 0.88 | 0.03 | −96% | 0.63 | 0.05 | −92% | 0.73 | 0.04 | −95% |
| E | 0.93 | 0.79 | −15% | 0.87 | 0.73 | −15% | 0.90 | 0.76 | −15% |

classes B, C and E in a unique damage macro-class. However, since this choice may appear quite simplistic from an engineering point of view, a possible more efficient reformulation of the seismic damage ratings may involve three classes only, i.e. referred to high structural damage, low structural damage, and no damage.

Table 8 shows the comparison between recall, precision and F1-score from the MLR (NN) and the ANN models. The relative difference expresses the difference between the ANN performance compared to MLR.

The ANN model slightly underperforms compared to MLR in terms of recall, precision, and F1-score for class A, but the differences are relatively small. The ANN model shows a significant decrease in recall for class B (32% decrease) but compensates with higher precision (16% increase), resulting in a slightly lower F1-score compared to MLR. The ANN model performs slightly worse than MLR for class E, with a 15% decrease in recall, precision, and F1-score.

There could be several reasons why MLR might perform better than an ANN. Firstly, MLR is a linear model that makes fewer assumptions and requires fewer parameters to estimate compared to ANN, which is a complex and nonlinear model. In cases where the relationship between predictors and outcomes is relatively simple, a simpler model like MLR might be sufficient and perform well. Additionally, if the relationships between predictors and outcomes are mostly linear, MLR can effectively capture these relationships. On the other hand, ANN excels in modelling complex nonlinear relationships and might not perform as well when the relationships are predominantly linear. Moreover, if the features in the dataset have little or no interactions with each other, MLR can adequately capture the individual contributions of each predictor to the outcome. The relationships between predictors and outcomes are likely linear in the present case, with minor interactions among regressors. Thus, the MLR, despite being simpler, is recommended due to its simplicity and accuracy.

## 6. Conclusions

This paper discusses the risk classification of the AeDES forms compiled following the earthquake in L'Aquila. The AeDES form is a first-level form for post-earthquake damage and functionality assessment. The inspection team can choose among six categorical risk classes from A to F based on the observed damage and the status of the nearby buildings, which might compromise the safety of the building under investigation. The risk class is related to the sum provided by the government for the reconstruction or repair of the building.

The authors developed two models, a multinomial logistic regression (MLR) and artificial neural networks (ANN), to predict the categorical outcome of the evaluation (classes A, B, C, or E), given the recorded damage. Classes D and F were removed since they do not represent the observed damage in the building. The database used for calibration comprises all school buildings evaluated after the earthquake, consisting of 878 buildings. The selected database includes only public buildings to avoid possible classification bias more likely to occur in private buildings due to the economic relevance of choice.

The MLR model exhibits an average accuracy of 84% in classifying the four categorical outputs and performs excellently when distinguishing among damage (C and E) and no-damage (A-B) cases, yielding a 97% accuracy. The accuracy loss between the two cases depends on the difficulty in choosing between class A and class B. The two outcomes, corresponding to null or minor damages, respectively, involve, in fact, more responsibility than choosing between classes C and A. The choice between classes A and B entails deciding whether the building needs some interventions or is fully operational. The complexity and responsibility behind this choice cannot be easily expressed by a damaged matrix despite its extent and accuracy.

The ANN model is developed to provide a comparison of the MLR model. Since the architecture of the ANN must be carefully evaluated, a multi-objective optimization problem was solved aiming to maximize the accuracy and minimize the model complexity, i.e. finding the minimum number of hidden neurons. A single hidden layer with 16 units has been found to be as optimal architecture to provide a fair comparison with the MLR. The ANN model exhibited global average accuracy of 78.4% in classifying the four categorical outputs and confirmed the same learning difficulties evidenced by the MLR.

This paper mainly aims to provide an MLR model and an ANN, calibrated on actual data that can be easily implemented in future updates of the AeDES forms. The model gives the odds of a given damage pattern falling in a specific damage class (A, B, C, and E). This MLR will not replace the personal responsibility of the expert team, which will have the last decision about the damage class. Still, it supports the decision-making process towards a more coherent and informed assessment.

## Declaration of competing interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

## Data availability

The data that has been used is confidential.

## Acknowledgement

The authors acknowledge the Department of Civil Protection of the Abruzzo Region for providing us with the data used in this investigation.

## References

[1] C. Baggio, A. Bernardini, R. Colozza, L. Corazza, M. Della Bella, G. Di Pasquale, M. Dolce, A. Goretti, A. Martinelli, G. Orsini, F. Papa, G. Zuccaro, Field manual for post-earthquake damage and safety assessment and short term countermeasures (aedes), in: Field Manual for Post-Earthquake Damage and Safety Assessment and Short Term Countermeasures (AeDES), 2007, Cited By 181.

[2] G. De Martino, M. Di Ludovico, A. Prota, C. Moroni, G. Manfredi, M. Dolce, Empirical damage and actual repair costs on rc private buildings after l'aquila earthquake, in: Proceeding of the 16th World Conference on Earthquake Engineering, Santiago, Chile, 2017, pp. 9–13.

[3] C. Del Gaudio, G. De Martino, M. Di Ludovico, G. Manfredi, A. Prota, P. Ricci, G.M. Verderame, Empirical fragility curves for masonry buildings after the 2009 l'aquila, italy, earthquake, Bull. Earthq. Eng. 17 (2019) 6301–6330.

[4] R. Whitman, J. Reed, S. Hong, Earthquake damage probability matrices, in: Proceedings of the Fifth World Conference on Earthquake Engineering, 1973, pp. 2531–2540, Cited By 143.

[5] F. Braga, M. Dolce, D. Liberatore, A statistical study on damaged buildings and an ensuing review of the msk-76 scale, in: Proceedings of the Seventh European Conference on Earthquake Engineering, 1982, pp. 431–450, Cited By 150.

[6] C. Del Gaudio, P. Ricci, G.M. Verderame, A class-oriented mechanical approach for seismic damage assessment of rc buildings subjected to the 2009 l'aquila earthquake, Bull. Earthq. Eng. 16 (2018) 4581–4605.

[7] C. Drago, R. Ferlito, Z. M, Clustering of damage variables for masonry buildings measured after l'aquila earthquake, SIS: Stat. Demogr. (2015).

[8] M. Dolce, A. Masi, M. Marino, M. Vona, Earthquake damage scenarios of the building stock of potenza (southern italy) including site effects, Bull. Earthq. Eng. 1 (1) (2003) 115–140, http://dx.doi.org/10.1023/A:1024809511362, cited By 118.

[9] G. Di Pasquale, G. Orsini, R. Romeo, New developments in seismic risk assessment in italy, Bull. Earthq. Eng. 3 (1) (2005) 101–128, http://dx.doi.org/10.1007/s10518-005-0202-1, cited By 99.

[10] P. Ricci, F. De Luca, G.M. Verderame, 6Th april 2009 l'aquila earthquake, italy: reinforced concrete building performance, Bull. Earthq. Eng. 9 (2011) 285–305.

[11] A.B. Liel, K.P. Lynch, Vulnerability of reinforced-concrete-frame buildings and their occupants in the 2009 l'aquila, italy, earthquake, Natl. Hazards Rev. 13 (1) (2012) 11–23.

[12] M. D'Amato, R. Laguardia, G. Di Trocchio, M. Coltellacci, R. Gigliotti, Seismic risk assessment for masonry buildings typologies from l'aquila 2009 earthquake damage data, J. Earthq. Eng. 26 (9) (2022) 4545–4579.

[13] M. Di Ludovico, A. Prota, C. Moroni, G. Manfredi, M. Dolce, Reconstruction process of damaged residential buildings outside historical centres after the l'aquila earthquake: part ii—heavy damage reconstruction, Bull. Earthq. Eng. 15 (2017) 693–729.

[14] M. Zucconi, R. Ferlito, L. Sorrentino, Simplified survey form of unreinforced masonry buildings calibrated on data from the l'aquila earthquake, Bull. Earthq. Eng. 16 (2018) 2877–2911.

[15] S.A. Scala, C. Del Gaudio, G.M. Verderame, Influence of construction age on seismic vulnerability of masonry buildings damaged after 2009 l'aquila earthquake, Soil Dyn. Earthq. Eng. 157 (2022) 107199.

[16] G. Manfredi, A. Prota, G.M. Verderame, F. De Luca, P. Ricci, Emilia earthquake, italy: reinforced concrete buildings response, Bull. Earthq. Eng. 12 (2014) (2012) 2275–2298.

[17] N. Buratti, F. Minghini, E. Ongaretto, M. Savoia, N. Tullini, Empirical seismic fragility for the precast rc industrial buildings damaged by the 2012 emilia (italy) earthquakes, Earthq. Eng. Struct. Dyn. 46 (14) (2017) 2317–2335.

[18] L. Praticò, M. Bovo, N. Buratti, M. Savoia, Large-scale seismic damage scenario assessment of precast buildings after the may 2012 emilia earthquake, Bull. Earthq. Eng. 20 (15) (2022) 8411–8444.

[19] L. Sorrentino, S. Cattari, F. Da Porto, G. Magenes, A. Penna, Seismic behaviour of ordinary masonry buildings during the 2016 central italy earthquakes, Bull. Earthq. Eng. 17 (10) (2019) 5583–5607.

[20] M.R. Valluzzi, L. Sbrogiò, Y. Saretta, H. Wenliuhan, Seismic response of masonry buildings in historical centres struck by the 2016 central italy earthquake, impact of building features on damage evaluation, Int. J. Archit. Herit. 16 (12) (2022) 1859–1884.

[21] F. De Luca, G.E. Woods, C. Galasso, D. D'Ayala, Rc infilled building performance against the evidence of the 2016 eefit central italy post-earthquake reconnaissance mission: empirical fragilities and comparison with the fast method, Bull. Earthq. Eng. 16 (2018) 2943–2969.

[22] G. De Martino, P. Marotta, M. Di Ludovico, S. Iannella, V. Albanese, A. Prota, The reconstruction process of the buildings damaged by the 2017 ischia earthquake, Proc. Struct. Integr. 44 (2023) 1816–1823.

[23] C. Del Gaudio, M. Di Domenico, P. Ricci, G.M. Verderame, Preliminary prediction of damage to residential buildings following the 21st august 2017 ischia earthquake, Bull. Earthq. Eng. 16 (2018) 4607–4637.

[24] V. Silva, D. Amo-Oduro, A. Calderon, J. Dabbeek, V. Despotaki, L. Martins, A. Rao, M. Simionato, D. Viganó, C. Yepes, A. Acevedo, N. Horspool, H. Crowley, K. Jaiswal, M. Journeay, M. Pittore, Global earthquake model (gem) seismic risk map, 2018, Global Earthquake Model (GEM) Seismic Risk Map Cited By 31.

[25] R. Spence, A. Coburn, A. Pomonis, S. Sakai, Correlation of ground motion with building damage: The definition of a new damage-based seismic intensity scale, in: Proceedings of the 10th World Conference on Earthquake Engineering, Vol. 1, 1992, pp. 551–556, cited By 77.

[26] G. Orsini, A model for buildings' vulnerability assessment using the parameterless scale of seismic intensity (psi), Earthq. Spectra 15 (3) (1999) 463–483, http://dx.doi.org/10.1193/1.1586053, cited By 62.
[27] 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, Cited By 63.
[28] P. Mouroux, B. Le Brun, Presentation of risk-ue project, Bull. Earthq. Eng. 4 (4) (2006) 323–339, http://dx.doi.org/10.1007/s10518-006-9020-3, cited By 138.
[29] T. Rossetto, Vulnerability curves for the seismic assessment of reinforced concrete building populations, in: Vulnerability Curves for the Seismic Assessment of Reinforced Concrete Buildings Populations, 2004, Cited By 20.
[30] S. Lagomarsino, S. Giovinazzi, Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings, Bull. Earthq. Eng. 4 (4) (2006) 415–443, http://dx.doi.org/10.1007/s10518-006-9024-z, cited By 516.
[31] M. Colombi, B. Borzi, H. Crowley, M. Onida, F. Meroni, R. Pinho, Deriving vulnerability curves using italian earthquake damage data, Bull. Earthq. Eng. 6 (3) (2008) 485–504, http://dx.doi.org/10.1007/s10518-008-9073-6, cited By 82.
[32] M. Rota, A. Penna, C. Strobbia, Processing italian damage data to derive typological fragility curves, Soil Dyn. Earthq. Eng. 28 (10–11) (2008) 933–947, http://dx.doi.org/10.1016/j.soildyn.2007.10.010, cited By 200.
[33] A. Liel, K. Lynch, Vulnerability of reinforced-concrete-frame buildings and their occupants in the 2009 l'aquila, italy, earthquake, Natl. Hazards Rev. 13 (1) (2012) 11–23, http://dx.doi.org/10.1061/(ASCE)NH.1527-6996.0000047, cited By 36.
[34] G. Zuccaro, F. Cacace, Revisione dell'inventario a scala nazionale delle classi tipologiche di vulnerabilità ed aggiornamento delle mappe nazionali di rischio sismico, in: Revisione Dell'Inventario a Scala Nazionale Delle Classi Tipologiche Di Vulnerabilità Ed Aggiornamento Delle Mappe Nazionali Di Rischio Sismico, in: Atti Del XIII Convegno ANIDIS, 2009, Cited By 9.
[35] C. Del Gaudio, G. De Martino, M. Di Ludovico, G. Manfredi, A. Prota, P. Ricci, G. Verderame, Empirical fragility curves from damage data on rc buildings after the 2009 l'aquila earthquake, Bull. Earthq. Eng. 15 (4) (2017) 1425–1450, http://dx.doi.org/10.1007/s10518-016-0026-1, cited By 105.
[36] G. Nicodemo, M. Pittore, A. Masi, V. Manfredi, Modelling exposure and vulnerability from post-earthquake survey data with risk-oriented taxonomies: Aedes form, gem taxonomy and ems-98 typologies, Int. J. Disaster Risk Reduct. 50 (2020) 101894.
[37] D.G. Kleinbaum, L.L. Kupper, A. Nizam, E.S. Rosenberg, Applied Regression Analysis and Other Multivariable Methods, Cengage Learning, 2013.
[38] A. Contento, P. Gardoni, A.D. Egidio, A. de Leo, Probabilistic models to assess the seismic safety of rigid block-like elements and the effectiveness of two safety devices, J. Struct. Eng. 145 (11) (2019) 04019133.
[39] Y. Wang, P. Gardoni, C. Murphy, S. Guerrier, Worldwide predictions of earthquake casualty rates with seismic intensity measure and socioeconomic data: A fragility-based formulation, Natl. Hazards Rev. 21 (2) (2020) 04020001.
[40] A. Aloisio, A. Contento, R. Alaggio, G. Quaranta, Physics-based. models, Surrogate models and experimental assessment of the vehicle–bridge interaction in braking conditions, Mech. Syst. Signal Process. 194 (2023) 110276.
[41] G.C. Marano, M.M. Rosso, A. Aloisio, G. Cirrincione, Generative adversarial networks review in earthquake-related engineering fields, Bull. Earthq. Eng. (2023) 1–52.
[42] G.C. Marano, M.M. Rosso, J. Melchiorre, Optimization as a tool for seismic protection of structures, in: Seismic isolation, in: Energy Dissipation and Active Vibration Control of Structures: 17th World Conference on Seismic Isolation (17WCSI), Springer, 2023, pp. 100–113.
[43] M.M. Rosso, Comparative deep learning studies for indirect tunnel monitoring with and without fourier pre-processing, Integr. Comput.-Aided Eng. (2023) 1–20, http://dx.doi.org/10.3233/ICA-230709.
[44] M. Dolce, A. Goretti, Building damage assessment after the 2009 abruzzi earthquake, Bull. Earthq. Eng. 13 (8) (2015) 2241–2264, http://dx.doi.org/10.1007/s10518-015-9723-4, cited By 82.
[45] A. Michelini, L. Faenza, V. Lauciani, L. Malagnini, Shakemap implementation in italy, Seismol. Res. Lett. 79 (5) (2008) 688–697, http://dx.doi.org/10.1785/gssrl.79.5.688, cited By 135.
[46] A. Michelini, L. Faenza, G. Lanzano, V. Lauciani, D. Jozinović, R. Puglia, L. Luzi, The new shakemap in italy: Progress and advances in the last 10 yr, Seismol. Res. Lett. 91 (1) (2020) 317–333.
[47] P. Galli, R. Camassi, R. Azzaro, F. Bernardini, S. Castenetto, D. Molin, E. Peronace, A. Rossi, M. Vecchi, A. Tertulliani, April 6, 2009 l'aquila earthquake: Macroseismic survey, surficial effects and seismotectonic implications [il terremoto aquilano del 6 aprile 2009: Rilievo macrosismico, effetti di superficie ed implicazioni sismotettoniche], Alpine Mediterr. Quatern. 22 (2) (2009) 235–246, cited By 77.
[48] A. Goretti, G. Di Pasquale, An overview of post-earthquake damage assessment in italy, EERI invitational workshop an action plan to develop earthquake damage and loss data protocols, 2002, Cited By 24.
[49] F.E. Harrell, Regression Modeling Strategies, Springer-Verlag, Berlin, Heidelberg, 2006.
[50] A. Géron, Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media, Inc, 2022.
[51] S. Raschka, Python Machine Learning, Packt publishing ltd, 2015.
[52] C.C. Aggarwal, et al., Neural Networks and Deep Learning, Vol. 10, Springer, 2018, p. 3, (978).
[53] J.V. Domashova, S.S. Emtseva, V.S. Fail, A.S. Gridin, Selecting an optimal architecture of neural network using genetic algorithm, Procedia Comput. Sci. 190 (2021) 263–273.
[54] T.A. Pham, V.Q. Tran, H.-L.T. Vu, H.-B. Ly, Design deep neural network architecture using a genetic algorithm for estimation of pile bearing capacity, PLoS One 15 (12) (2020) e0243030.
[55] Y. Azimi, S.H. Khoshrou, M. Osanloo, Prediction of blast induced ground vibration (bigv) of quarry mining using hybrid genetic algorithm optimized artificial neural network, Measurement 147 (2019) 106874.
[56] H. Ramchoun, Y. Ghanou, M. Ettaouil, M.A. Janati Idrissi, Multilayer perceptron: Architecture optimization and training, 2016.
[57] H. Ramchoun, M.J. Idrissi, Y. Ghanou, M. Ettaouil, Multilayer perceptron: Architecture optimization and training with mixed activation functions, in: Proceedings of the 2nd International Conference on Big Data, Cloud and Applications, 2017, pp. 1–6.
[58] J. Blank, K. Deb, Pymoo: Multi-objective optimization in python, IEEE Access 8 (2020) 89497–89509.
[59] H. Holland John, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975.
[60] C.C. Aggarwal, Machine Learning for Text, Vol. 848, Springer, 2018.
[61] C.A.C. Coello, G.B. Lamont, D.A. Van Veldhuizen, et al., Evolutionary Algorithms for Solving Multi-Objective Problems, 5, Springer, 2007.
[62] K. Deb, H. Jain, An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: solving problems with box constraints, IEEE Trans. Evol. Comput. 18 (4) (2013) 577–601.