POLITECNICO DI TORINO
Repository ISTITUZIONALE

Machine-learning-aided regional post-seismic usability prediction of buildings: 2016–2017 Central Italy earthquakes

Original
Machine-learning-aided regional post-seismic usability prediction of buildings: 2016–2017 Central Italy earthquakes / Aloisio, A., Rosso, M.M., Di Battista, L., Quaranta, G.. - In: JOURNAL OF BUILDING ENGINEERING. - ISSN 2352-7102. - 91:(2024), pp. 1-33. [10.1016/j.jobe.2024.109526]

Availability:
This version is available at: 11583/2991633 since: 2024-08-09T17:14:37Z

Publisher:
Elsevier

Published
DOI:10.1016/j.jobe.2024.109526

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
Elsevier postprint/Author's Accepted Manuscript

© 2024. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/.The final authenticated version is available online at:
http://dx.doi.org/10.1016/j.jobe.2024.109526

(Article begins on next page)

25 November 2026

Machine-learning-aided regional post-seismic usability prediction of buildings: 2016-2017 Central Italy earthquakes

Angelo Aloisio*[a], Marco Martino Rosso[b], Luca Di Battista[c], Giuseppe Quaranta[d]

[a] *Department of Civil, Construction-Architectural and Environmental Engineering, University of L'Aquila, L'Aquila, Italy*
[b] *Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Turin, Italy*
[c] *Ministry of Infrastructure and Transport, Public Works Department of Lazio, Abruzzo, and Sardinia, L'Aquila, Italy*
[d] *Department of Structural and Geotechnical Engineering, Sapienza University of Rome, Rome, Italy*

Abstract

This study addresses the development of machine learning models for predicting the post-seismic buildings usability within regions prone to frequent earthquakes. The analysis leverages on field data from 2016-2017 Central Italy earthquakes. Several machine learning techniques are employed for this task, namely K-nearest neighbors, linear support vector machine, radial basis function support vector machine, decision tree, random forest, neural network, adaptive boosting, naive Bayes, quadratic discriminant analysis, logistic regression, and linear discriminant analysis. The input variables include both building attributes and seismic intensity measures. Since the database turns out to be strongly imbalanced, the potential influence of two preprocessing techniques is examined, namely principal component analysis and synthetic minority oversampling technique. Several metrics are considered to evaluate the performance of the resulting predictive machine learning models. Moreover, this study investigates the optimal machine learning model's robustness against uncertainties, quantifies the importance of its features, and investigates how usability classes clustering can impact its performance. Every step of the implemented procedure is deeply explained and discussed to provide useful guidelines for similar applications.

Keywords: Building, Classification, Earthquake, Feature importance, Machine learning, Usability

1. Introduction

The goals of damage and usability assessment at territorial scale of buildings exposed to seismic hazard are twofold. In fact, they are essential for a rational elaboration of regional or national seismic risk mitigation

*Corresponding author
Email address: angelo.aloisio1@univaq.it (Angelo Aloisio*)

plans through scenario simulations and cost-benefit evaluations. Furthermore, damage and usability assessment are also among the priority tasks in the aftermath of an earthquake. In such case, they support the rapid losses quantification as well as the definition of the actions needed to facilitate the return of residents to their homes and the recovery of socio-economical activities whenever possible, or the establishment of the necessary repair interventions and temporary housing otherwise. Different approaches have been adopted in the past years to cope with damage and usability assessment of buildings at territorial scale. They can be broadly classified into mechanics-based, data-driven, or hybrid methods. Several studies have offered a thorough overview of the current approaches along with their advantages and disadvantages [e.g., 1–3].

Mechanics-based approaches are especially attractive because they rely on structural models to predict buildings damage and usability [e.g., 4–13]. The major concern is related to synthetic data generation, which requires a good knowledge about the buildings within the study area in order to prepare structural models able to simulate their expected behaviors. The availability of more and more field data collected from past earthquakes is promoting the development of data-driven approaches for seismic damage and usability assessment. The main advantage of this approach relies on the use of field data that virtually reflect actual building performance. Existing studies about data-driven approaches for damage and usability assessment implement deterministic, statistical, or machine learning (ML) methods.

As far as the post-seismic building usability prediction is concerned, Zucconi et al. [14] illustrate a predictive model for unreinforced masonry buildings that was calibrated on the basis of post-seismic surveys data collected in the aftermath of the 2009 L'Aquila earthquake. A linear regression analysis has been performed to develop the predictive model. The regressions yield a coefficient ranging from zero (indicating full usability) to one (denoting complete unusability). This index is generated by weighting the usability coefficients derived from a principal component analysis. The model performance is assessed in terms of accuracy only: to this end, the average model usability index is compared with the observed buildings usability, taking into account different levels of macroseismic intensity through the Mercalli-Cancani-Sieberg scale. This approach has been further refined and validated recently [15, 16]. Bertelli et al. [17] calculated the empirical fragility curves for residential buildings based on data gathered from structures affected by the 2009 L'Aquila earthquake. The main innovation of this study lies in adopting a nonlinear regression approach to derive fragility functions. The shift in focus from damage to usability represents a significant methodological change compared to past research. In this context, the fragility provides the probability of having a usability class higher than a given one as a function of the intensity measure.

Lately, the use of ML algorithms has been also considered. Beyond the applications to buildings, the

work by Kameshwar et al. [18] stands as one of the first efforts to delve into this approach. Specifically, a decision-trees-based model has been developed to predict both duration and type of traffic restrictions following extreme events on bridges. Such research relies on empirical data and insights gathered through expert opinion surveys. The predictive model for bridges functionality involves three decision trees, each one devoted to identify a different type of traffic restriction, such as bridge or lane closures and speed or load limitations. Tocchi et al. [19] apply a few ML techniques to predict the usability of both masonry and reinforced concrete buildings. Once again, data from 2009 L'Aquila earthquake have been employed. Indeed, a growing mass of research is demonstrating that ML techniques are especially promising to support seismic risk assessment and mitigation tasks [e.g., 20–27].

It is worth noting that all the available data-driven usability predictive models for buildings under earthquake have been developed by means of field data gathered from the 2009 L'Aquila earthquake [14–17, 19]. As a matter of fact, considering the effects due to a single seismic event is beneficial to enhance data-driven damage and usability assessment. Nevertheless, several active seismic regions can be struck by multiple earthquakes over time. This is a significant challenge when predicting the post-seismic usability of existing structures because the effects of multiple earthquakes can accumulate and amplify, especially in those constructions that do not comply with modern technical codes and good construction practices.

Hence, the present work aims at tackling the problem of predicting the post-seismic usability of buildings at regional scale through a data-driven approach. In particular, the main goal of the present research is to explore the effectiveness of ML methods in predicting the usability of buildings after the occurrence of multiple earthquakes over time and across a large area. The overall workflow is illustrated in Fig. 1 and the selected case study deals with the 2016-2017 Central Italy earthquakes. Data about buildings and seismic ground motion are presented in Sec. 2. The calculation of the seismic intensity measures over the study area is discussed in Sec. 3. The development of machine learning models is presented in Sec. 4. Such presentation encompasses the implemented algorithms and data preprocessing techniques as well as the considered performance metrics and hyperparameters optimization. Next, impact of the involved uncertainties and features importance are quantified. Ultimately, this study would like to propose some guidelines about the implementation of ML algorithms and the critical examination of the corresponding results in the attempt to secure their proper application in similar contexts.

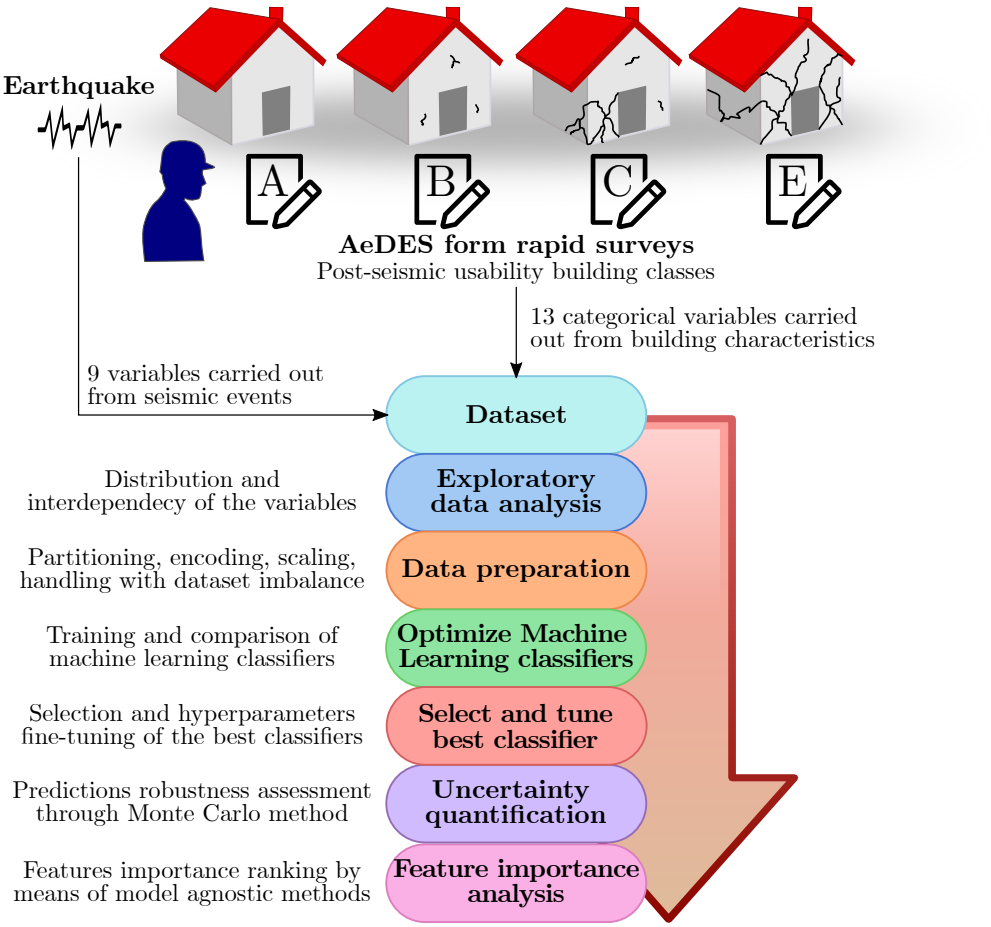


Figure 1: Flowchart of the implemented data-driven approach for predicting the post-seismic usability of buildings at regional scale.

2. Field data

2.1. 2016-2017 Central Italy earthquakes

The data employed in the present study refers to the 2016–2017 Central Italy earthquakes. Compared to recent Italian earthquakes, the 2016–2017 Central Italy seismic events stand out due to the significant number of high-intensity earthquakes distributed over two years, affecting a large area that encompasses four regions. (namely Abruzzo, Lazio, Marche and Umbria). The main seismic events are listed hereafter, see also Fig. 2.

- A first major earthquake with M_w 6.0 and epicentre close to Accumoli (Lazio) hit the centre of Italy on 2016-08-24 at 01:36:32 UTC. This event resulted in approximately 300 casualties (mostly in the regions of Lazio and Marche).

- On 2016-10-26, two further seismic events occurred with M_w 5.4 and M_w 5.9 at 17:10:36 UTC and



Figure 2: Epicentre of the most significant seismic events of the 2016-2017 Central Italy earthquakes. The legend shows the name of the municipality where the epicenter is located, together with magnitude and date of the seismic event.

19:18:06 UTC, respectively. Another seismic event with M_w 6.5 occurred four days later on 2016-10-30 at 06:40:18 UTC, with epicentre close to the Sibillini mountains (Marche and Umbria).

- A short sequence of four seismic events with M_w larger than 5 struck 25 km northwest of L'Aquila (Abruzzo) on 2017-01-18.

The seismogenetic characteristics of the 2016-2017 Central Italy earthquakes have been deeply analyzed in previous researches. The active faults and their segmentation are described by Galadini et al. [28]. Barchi et al. [29] studied the influence of subsurface geology while Chiarabba et al. [30] highlighted the role of the pore pressure in the footwall. Calderoni et al. [31] and Grelle et al. [32] investigated rupture directivity effects and topographic effects, respectively. A comprehensive suite of earthquake catalogues for the 2016-2017 Central Italy seismic sequence has been presented recently by Chiaraluce et al. [33].

The seismic events of the 2016-2017 Central Italy earthquakes have somehow interacted to produce the observed damage. However, diffuse building collapses in the municipalities of Arquata del Tronto (Marche), Amatrice and Accumoli (Lazio) are mainly attributable to the first seismic sequence. The second seismic sequence mostly involved the area close to the Sibillini mountains, thereby determining significant damage in several cities in Marche and Umbria. The third seismic sequence mostly impacted Abruzzo, especially the northern part of the L'Aquila province. Several researches analyzed the seismic demand of the 2016-2017 Central Italy earthquakes [34] as well as the performance of buildings [35], schools [36], churches [37], healthcare facilities [38], and bridges [39].

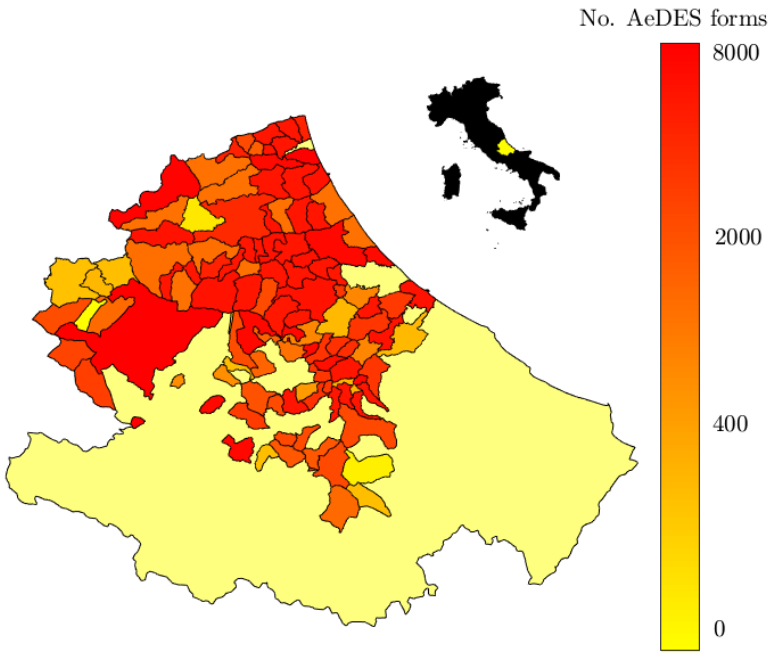


Figure 3: Number of available AeDES forms within the considered study area.

2.2. Survey data about seismic damage on buildings

Data from 12,662 buildings located in Abruzzo and collected after the third seismic sequence are processed in the present study. Indeed, most of the major damage in Abruzzo were observed in the aftermath of the third sequence. It is evident from Fig. 3 that buildings data are distributed over a large area that encompasses several municipalities. Particularly, the study area includes a total of 11 municipalities and extend over an area of 4,513 km^2 (about 42% of Abruzzo's total area). Hence, it is expected a significant scattering about the observed damage even for buildings having a fairly similar features because of the diverse construction practices and quality level of the construction materials in use within different zones and at different times, especially for old and non-code-conforming buildings.

Buildings data for the present study have been retrieved from so-called AeDES forms that were filled during post-earthquake surveys. AeDES forms are the main information source for post-seismic assessment of buildings in Italy [40]. They have been filled for the buildings within the area struck by the 2016-2017 Central Italy earthquakes (as identified by the Italian Department of Civil Protection) in order to address two specific needs: i) to provide information about the post-earthquake usability of the buildings, and ii) to assess expediently and qualitatively the severity of the seismic damage, in such a way to quantify the

maximum refund for post-earthquakes interventions.

The AeDES form consists of nine sections splitted into four pages. While Sections 1-3 and Section 7 collect variables about the buildings and their site, the evidence of the post-seismic damage survey are included into Sections 4-6 and Section 9. The outcome of the survey process is available in Section 8. Particularly, Section 1 provides general information on the building position. Section 2 presents essential data about geometry, age and use of the structure. Section 3 includes the characteristics of the structure (although masonry, reinforced concrete, steel and timber constructions are considered, minimum data are collected for masonry structures only). Section 4 and Section 5 contain information about the observed damage in structural and non-structural components, respectively. The damage levels for structural components range from no damage total collapse according to EMS-98 [41]. Section 6 highlights dangerous situations attributable to adjacent constructions, pipelines or landslides. Section 7 describes the site morphology and possible damage to the foundation. Based on all these information, one of the following consequence classes (and the corresponding label) is assigned to the building within Section 8:

- accessible building (A);
- temporarily unusable building or building usable after emergency interventions (B);
- partially unusable building (C)
- building to be reviewed, but temporarily unusable (D);
- unusable building (E);
- unusable building due to external causes only (F).

Attachments and notes (including relevant information and recommendations about post-seismic emergency interventions), if any, are collected into Section 9.

Fig. 4 shows the set of the variables retrieved from Sections 1-3 and Section 7 through the AeDES form. In detail, the main variables into Section 1 and Section 2 are the geographic coordinates, the position within the buildings aggregate (i.e., isolated building, building inside the aggregate or building at either one edge or one corner of the aggregate), the number of floors, the average inter-storey height, the average floor area, age, occupancy rate and use of the building. Section 3 deals with the typology of the structural system, namely masonry, reinforced concrete, steel or timber structure. As far as masonry constructions are concerned, information regarding horizontal structure, masonry texture, curbs/chains, isolated pillars, and mixed/reinforced masonry are specified. Three cases are considered for mixed masonry structures,

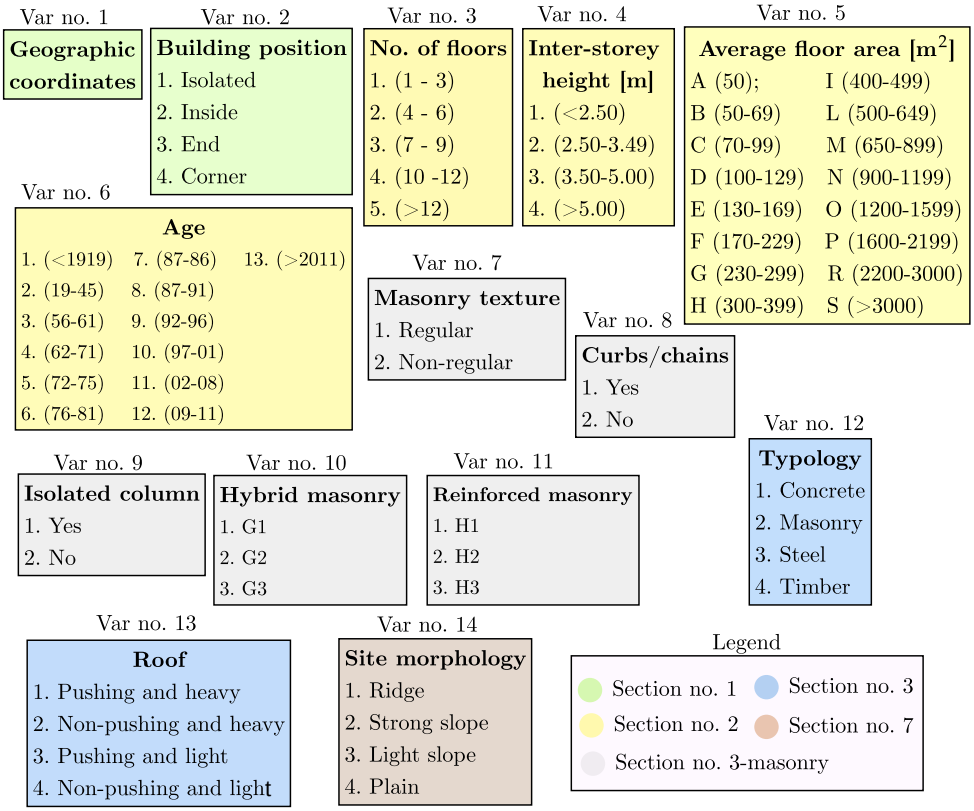


Figure 4: List of explanatory variables available from Sections 1-3 and Section 7 of the AeDES form.

namely reinforced concrete or other framed structures on a masonry structure (G1), masonry structure on reinforced concrete or other framed structures (G2), masonry structure mixed at the same floor in parallel with reinforced concrete or other framed structures (G3). Three cases are considered for reinforced masonry structures, namely masonry reinforced with injections or non-reinforced plasters (H1), reinforced masonry or masonry with reinforced plaster (H2), masonry with other or unidentified reinforcement (H3). Roof characteristics (i.e., pushing or non-pushing roof, heavy or light roof) are also available from Section 3. Information about the site in Section 7 are limited to its morphology (i.e., ridge, strong slope, light slope, or plain site).

The distribution of the variables within the considered database is detailed into Tab. 1 and Tab. 2. Particularly, Tab. 1 shows the data taken from Sections 1-2 of the AeDES form while Tab. 2 illustrates those extracted from Section 3 and Section 7 (for a more effective and compact presentation of the data, the variables into Tab. 1 and Tab. 2 do not follow the layout reported in Fig. 4). The percentage value into Tab. 1 and Tab. 2 is the ratio between the number of buildings belonging to the specified category and the total number of buildings, except for the hybrid masonry buildings and reinforced masonry buildings into

Tab. 2, where the percentage value is obtained by divining by the number of masonry buildings.

Table 1: Distribution of the explanatory variables within the available database retrieved from Sections 1-2 of the AeDES forms.

Variable	Category	Number	Percentage	Variable	Category	Number	Percentage
No. of floors	1	1,030	8.57%	Average floor area [m^2]	A (<50)	2,342	18.53%
	2	3,637	30.26%		B (50-69)	2,181	17.25%
	3	5,211	43.35%		C (70-99)	2,599	20.56%
	4	2,125	17.68%		D (100-129)	2,107	16.67%
	5	446	3.71%		E (130-169)	1,459	11.54%
	6	164	1.36%		F (170-229)	882	6.98%
	7	73	0.61%		G (230-299)	500	3.96%
	8	27	0.22%		H (300-399)	296	2.34%
	9	17	0.14%		I (400-499)	134	1.06%
	10	3	0.02%		L (500-649)	71	0.56%
	11	3	0.02%		M (650-899)	45	0.36%
	Total	12,736	98.29%		N (900-1199)	16	0.13%
Age	1 (<1919)	1,582	12.66%		O (1200-1599)	6	0.05%
	2 (19-45)	1,712	13.70%		P (1600-2199)	2	0.02%
	3 (46-61)	2,436	19.50%		Q (2200-3000)	1	0.01%
	4 (62-71)	2,387	19.11%		R (>3000)	1	0.01%
	5 (72-75)	1,422	11.38%		Total	12,642	97.57%
	6 (76-81)	957	7.66%	Building position	Corner	1,347	11.21%
	7 (82-86)	745	5.96%		Edge	3,875	32.24%
	8 (87-91)	384	3.07%		Internal	2,132	17.74%
	9 (92-96)	304	2.43%		Isolated	4,667	38.82%
	10 (97-01)	200	1.60%		Total	12,021	92.78%
	11 (02-08)	245	1.96%	Inter-storey height [m]	1 (<2.50)	1,225	9.81%
	12 (09-11)	69	0.55%		2 (2.50-3.49)	10,749	86.04%
	13 (>2011)	51	0.41%		3 (3.50-5.00)	459	3.67%
	Total	12,494	96.43%		4 (>5.00)	60	0.48%
					Total	12,493	96.42%

The following considerations can be drawn from the analysis of Tab. 1 and Tab. 2.

- Some data are missing for all the variables. Moreover, the database is strongly imbalanced. In fact, certain records or categories of variables are over-represented compared to others.

- The considered database mostly includes low-rise buildings (about 30% and 40% of the surveyed buildings have two and three floors, respectively). Many buildings (nearly 76% of the buildings stock) were built before 1975. The number of new buildings in the study area has decreased steadily since the 80ies. The average floor area generally spans from 50 m^2 to 200 m^2 (almost 91% of the buildings stock lies within this range). Most buildings are isolated (about 38% of the buildings stock) and have an average inter-storey height between 2.50 m and 3.49 m (about 86% of the buildings stock).

- Most buildings are made of masonry (nearly 80% of the surveyed buildings). A smaller, yet significant, fraction of buildings has a concrete structure (nearly 20% of the buildings stock). Most of the buildings

Table 2: Distribution of the explanatory variables within the available database retrieved from Section 3 and Section 7 of the AeDES forms.

Variable	Category	Number	Percentage	Variable	Category	Number	Percentage
Typology	Concrete	2,545	19.65%	Masonry texture	Irregular	3,913	41.65%
	Masonry	10,296	79.49%		Regular	5,483	58.35%
	Timber	81	0.63%		Total	9,396	91.26%
	Steel	31	0.24%	Chains/curbs	No	3,105	31.20%
	Total	12,953	99.97%		Yes	6,848	68.80%
Roof	Pushing and heavy	1,622	13.29%		Total	9,953	96.67%
	Non-pushing and heavy	4,835	39.62%	Isolated pillars	No	8,764	91.37%
	Pushing and light	2,169	17.77%		Yes	828	8.63%
	Non-pushing and light	3,578	29.32%		Total	9,592	93.16%
	Total	12,204	94.19%	Hybrid masonry	H1	210	20.04%
Site morphology	Ridge	573	4.52%		H2	123	11.74%
	Strong slope	4,390	34.65%		H3	715	68.23%
	Light slope	5,879	46.40%		Total	1,048	10.18%
	Plain	1,827	14.42%	Reinforced masonry	G1	14	8.86%
	Total	12,669	97.78%		G2	76	48.10%
					G3	68	43.04%
					Total	158	1.53%

have a heavy or light non-pushing roof (almost 40% and 30% of the buildings, respectively), but the number of buildings with pushing roof is not negligible.

- Most of the masonry buildings have a regular texture, but those with irregular texture are many. The number of masonry buildings with chains or curbs is more than two times that of those with neither chains nor curbs. The number of masonry buildings with isolated pillars is negligible (only 8% of the buildings stock). There are a few cases of hybrid and reinforced masonry buildings.

- A plane site is not a common condition, which means that most of the buildings is in mountain zones (a strong or slight slope has been detected for more than 80% of the buildings).

Fig. 5 illustrates the number of buildings as function of the minimum distance from the earthquakes epicentre. It can be inferred from Fig. 5 that more than 80% of the buildings is located within 40 km from the epicentre. A summary of the consequence class assignment to the buildings surveyed through AeDES form is also provided in Fig. 5. Fig. 6 shows the consequence classes distribution over the study area (for buildings belonging to classes A, B, C or E). The colormap reflects the dominant consequence class within each municipality (i.e., green, blue, orange or red are adopted when the prevalent outcome in each municipality is A, B, C or E, respectively).

Fig. 5 and Fig. 6 highlight once again that the available database is strongly imbalanced since the buildings are not fairly distributed among the consequence classes. On the one hand, more than one

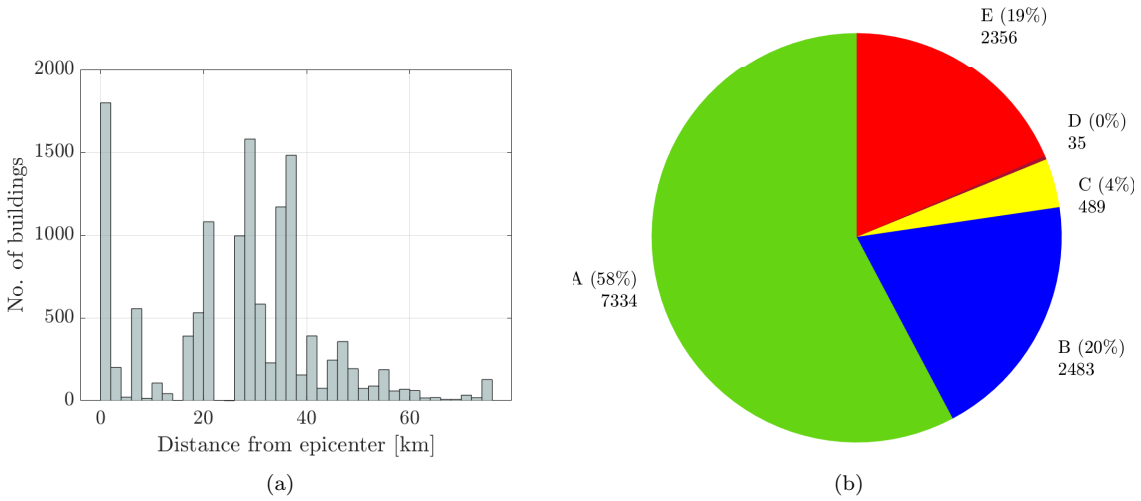


Figure 5: Number of buildings as function of the minimum epicentral distance (a). Consequence class assignment to the buildings surveyed through AeDES form within the study area (b).

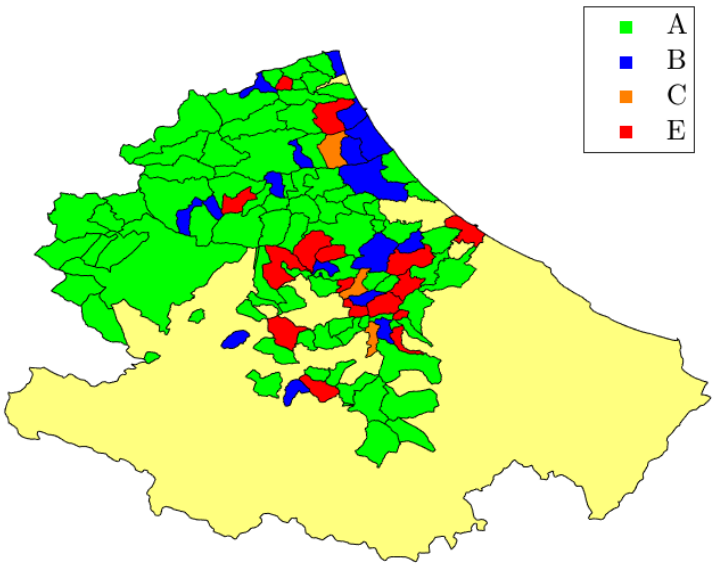


Figure 6: Consequence classes distribution over the study are (for buildings belonging to classes A, B, C or E).

half of the surveyed building was accessible after the earthquake (class A was assigned to about 60% of the buildings). On the other hand, almost an equal number of buildings resulted either temporarily or permanently unusable (class B or class E was assigned to about 20% of the buildings). The number of buildings that was assigned to the intermediate consequence class is low (class C was assigned to about 4% of the buildings), while a negligible number of not conclusive outcomes has been reported at the end of the surveys (class D was assigned to a few buildings only).

The degree of interdependency among the variables is also investigated [42–44]. To this end, the categorical explanatory variables in Fig. 4 are further divided according to their typology. Specifically, number of floors, inter-storey height, average floor surface, and age are polytomous ordinal categorical variables, whereas the remaining variables are nominal ones [42, 43]. The ordered categorical variables have been represented in Fig. 7 using a correlations graphs matrix, mapping with different colors the possible building usability outputs herein considered. The subplots along the main diagonal illustrate non-parametric kernel density estimation diagrams and they show the distribution of the variable. It is worth noting in Fig. 7 that the majority of unusable buildings (i.e., classes B, C, and E) were built before 1975, and are characterized by an average floor area below 229 m^2 and an inter-storey height within the range 2.50-3.49 m. Contingency tables have been computed to explore the degree of interdependency of nominal categorical variables, thereby providing cross-tabulation of feature pairs, and they have been finally depicted through mosaic graphs as shown in Fig. 8. Although most of the data refers to masonry buildings, it is evident from Fig. 8 that relevant crucial structural details about masonry texture and curbs or chains are often missing.

3. Seismic intensity assessment

3.1. Selected intensity measures

The prediction of the buildings usability is sometimes performed by neglecting the spatial variation of the seismic intensity [e.g., 19]. However, while it might be reasonable to neglect the spatial variability of the seismic intensity when predicting the buildings usability within a limited area, such an assumption can be unsuitable when the study area scales up. In light of this complication, the present study accounts for the spatial variation of the seismic intensity, since it might be correlated with the observed damage at territorial scale.

Several measures have been proposed in the past years to quantify the intensity of the seismic ground motion. They can be broadly classified into structure- and non-structure-specific intensity measures as well as into acceleration, velocity, displacement or energy-based intensity measures. Both scalar and vector-

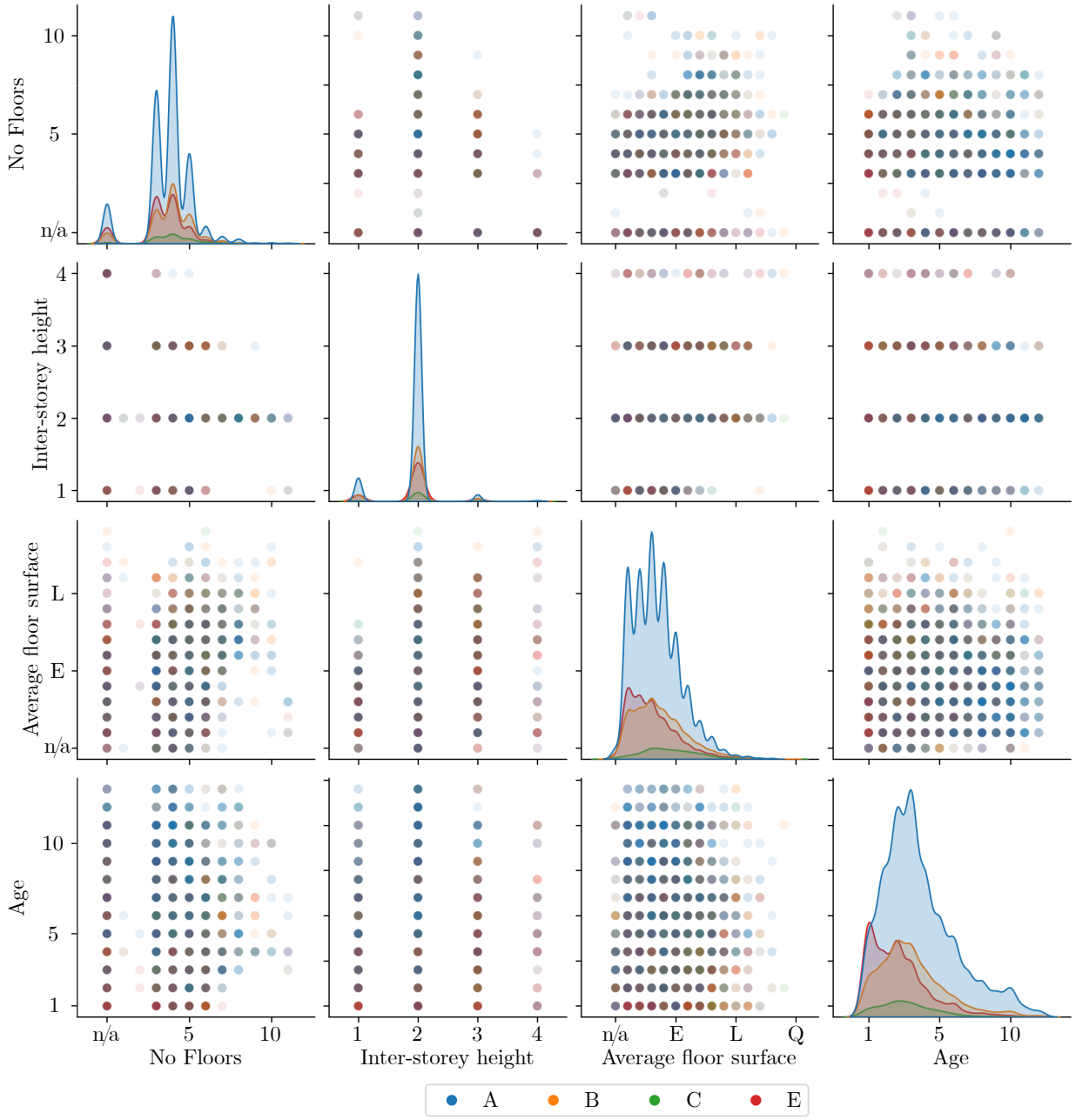


Figure 7: Correlations graphs matrix for ordinal categorical explanatory variables.

valued intensity measures have been proposed. A comprehensive list of common intensity measures for seismic assessment of buildings has been reported among the others by Fiore et al. [20].

Apart from buildings data collected through AeDES forms, the following set of seismic intensity measures is thus considered in the present study:

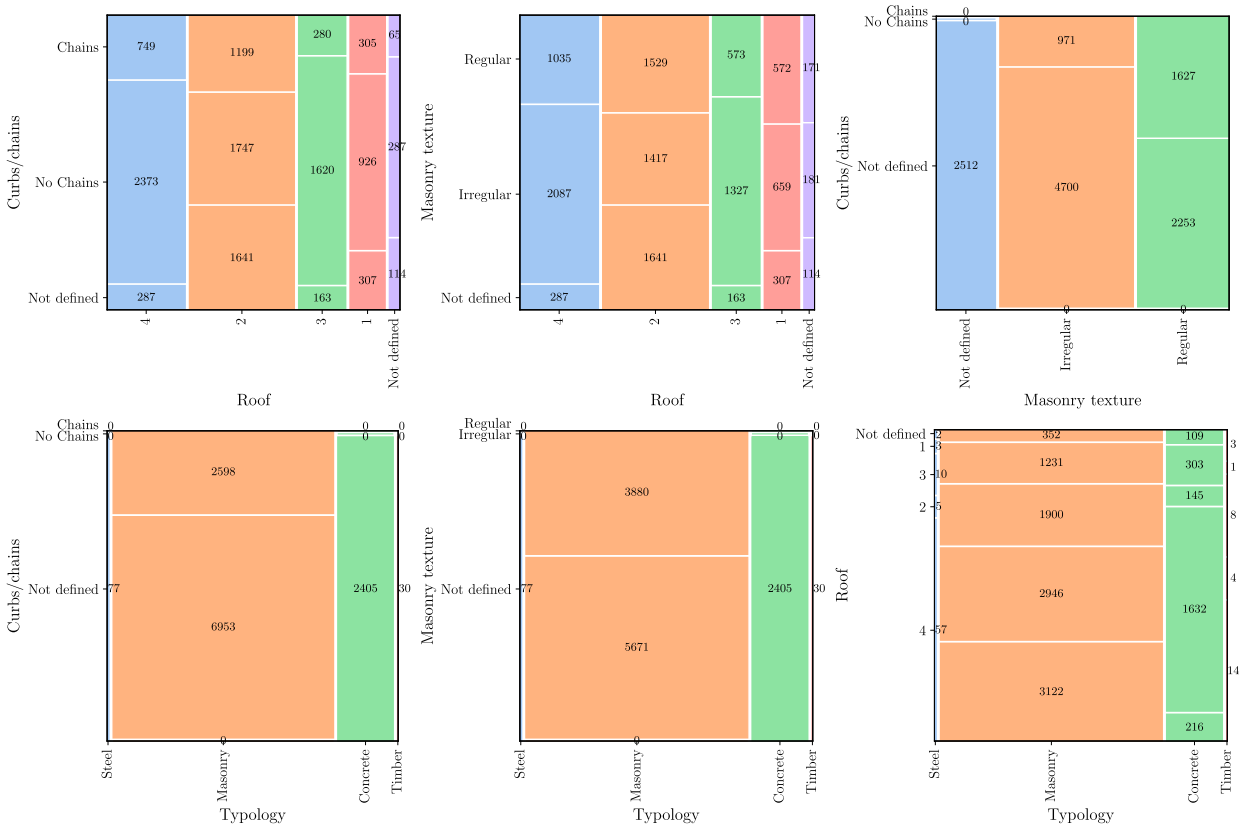


Figure 8: Mosaic graphs for nominal categorical explanatory variables.

- horizontal and vertical peak ground acceleration;
- horizontal pseudo-spectral acceleration;
- Arias intensity;
- Housner intensity;
- earthquake duration.

They are selected among the most common ones, and often result among those that relevant agencies provide automatically after an earthquake. Additionally, there are well-established empirical laws that can be applied over the considered study area in order to estimate the attenuation of these seismic intensity measures from the causative fault given the epicentral distance, the earthquake magnitude and the soil characteristics.

The attenuation laws for the horizontal and vertical peak ground acceleration adopted in the present study were calibrated by Sabetta and Pugliese [45] using the Italian earthquakes catalogue. They read as

follows:

$$\log \mathrm{PGA}_h = -1.845 + 0.363M - \log\left[(d^2+5^2)^{0.5}\right] + 0.195s, \tag{1}$$

$$\log \mathrm{PGA}_v = -2.637 + 0.443M - \log\left[(d^2+4.1^2)^{0.5}\right] + 0.209s, \tag{2}$$

where PGA_h and PGA_v are the peak ground accelerations in [g] for the horizontal and vertical directions, respectively, M is the (surface-wave/local) magnitude, d is the epicentral distance in [km]. Moreover, s is a coefficient equal to 0 or 1 for stiff and soft soil, respectively.

Given the fundamental period of the building T in [s], it is possible to estimate the psuedo-spectral velocity in the horizontal or vertical direction by means of the following attenuation law derived by Sabetta and Pugliese [45] for Italian earthquakes:

$$\log S_{pv} = a + bM - \log\left[(d^2+h^2)^{0.5}\right] + e_1 s + e_2 s, \tag{3}$$

where S_{pv} is the pseudo-spectral velocity in [cm/s] whereas the coefficients a, b, h, e_1 and e_2 have been tabulated by Sabetta and Pugliese [45] individually for the horizontal and the vertical direction as function of the fundamental period for a 5% damped elastic pseudo-spectrums, see Tab. 3. The pseudo-spectral acceleration in the horizontal direction $S_{pa,h}$ is obtained by multiplying the result of Eq. (3) for the natural pulsation corresponding to the fundamental vibration mode T.

Table 3: Numerical values for the coefficients involved in the attenuation law of the pseudo-spectral velocity.

T [s]	Coefficient for S_{pv}				
	a	b	e_1	e_2	h
4	-2.5	0.73	0	0.1	3
3	-2.25	0.72	0	0.11	3
2	-1.9	0.69	0	0.15	4
1.5	-1.65	0.66	0.01	0.18	4
1.0	-1.28	0.61	0.05	0.21	4
0.8	-1	0.57	0.12	0.19	5
0.5	-0.6	0.5	0.23	0.12	5
0.4	-0.28	0.45	0.22	0.08	5
0.3	0.1	0.38	0.19	0.02	5
0.2	0.296	0.32	0.16	0	6
0.2	0.222	0.31	0.16	0	6
0.1	-0.02	0.3	0.16	0	6
0.1	-0.31	0.3	0.16	0	6
0	-0.82	0.33	0.16	0	5

The attenuation laws for Arias intensity I_a, Housner intensity I_h and earthquake duration D employed in the present study were derived for Italian earthquakes by Pacor et al. [46]. Their general form is the following:

$$\log(Y) = e_1 + F_D(R,M) + F_M(M) + F_S + F_{sof}, \tag{4}$$

$$F_D(R,M) = [c_1 + c_2(M - M_{ref})] \log \left[\sqrt{R_{JB}^2 + +h^2} / R_{ref} \right] - c_3 \left(\sqrt{R_{JB}^2 + +h^2} - R_{ref} \right), \quad (5)$$

$$F_M(M) = \begin{cases} b_1(M - M_h) + b_2(M - M_h)^2 & \text{if } M \leq M_h \\ b_3(M - M_h) & \text{otherwise} \end{cases}, \quad (6)$$

where Y is I_a, I_h or D while M is the (moment) magnitude and R is the Joyner-Boore distance in [km] when available, otherwise the epicentral distance d in [km] is assumed. The term F_S in Eq. (4) represents the site amplification. It is given by $F_S = s_i C_i$ for $i = 1, \ldots, 5$, where s_i are coefficients while C_i are dummy variables that denote the considered site classes according to EC8 [47]. The term F_{sof} in Eq. (4) represents the style of faulting correction, and it is given by $F_{sof} = f_j E_j$ for $j = 1, \ldots, 4$, where f_j are coefficients and E_j are dummy variables that denote the different fault mechanisms. The variables M_{ref}, M_h, R_{ref} are equal to 5.0, 6.75 and 1 km, respectively. Tables 4 displays the numerical values of the coefficients involved in the attenuation laws of Arias intensity I_a, Housner intensity I_h and earthquake duration D.

Table 4: Numerical values for the coefficients involved in the attenuation laws of Arias intensity, Housner intensity and earthquake duration.

Coefficient	I_h	I_a	D
e_1	2.6712	3.9337	0.6976
b_1	0.352	0.1322	0.7775
b_2	-0.0233	-0.0539	0.075
b_3	0	0	0
c_1	-1.3519	-3.1615	0.8685
c_2	0.2845	0.6484	-0.2975
h	8.7497	10.8563	10.1238
c_3	0	0	0
f_1	0.0573	0.0492	0.1262
f_2	0.1589	0.271	-0.0149
f_3	0.064	0.0752	0.1728
f_4	0	0	0
s_1	0	0	0
s_2	0.2393	0.4049	0.0418
s_3	0.3345	0.5926	0.0563
s_4	0.535	0.4451	0.2034
s_5	0.2543	1.1316	-0.1012

The use of the attenuation laws for the selected seismic intensity measures requires information about the soil type and the fundamental period of the structure. As far as the soil type is concerned, it is obtained from the seismic soil classification of Italy carried out by Forte et al. [48] using information about surface geology and shear-wave velocity measurements. The fundamental building period T in [s] is estimated by means of the following empirical relationship:

$$T = \alpha H^{\beta}, \quad (7)$$

where H is the building height in [m] (which is obtained by multiplying the average inter-storey height and the number of floors as reported into the AeDES forms, excluding underground floors). For unreinforced masonry buildings and reinforced concrete buildings, the numerical value of the coefficients α and β recommended by Gallipoli et al. [49] are adopted. Based on ambient vibration data from European buildings, Gallipoli et al. [49] proposed $\alpha = 0.0212$ and $\beta = 0.887$ for unreinforced masonry buildings while $\alpha = 0.0243$ and $\beta = 0.864$ are suggested for reinforced concrete buildings. Given the lack of proposals for other building typologies, code confirming values included within NTC2008 [50] are adopted otherwise. Accordingly, $\beta = 3/4$ is assumed whereas $\alpha = 0.085$ and $\alpha = 0.05$ is considered for steel structures and other structures, respectively.

3.2. Estimation of the seismic intensity

The prevailing soil type according to EC8 within the study area is B. There are a few cases close to the Adriatic coast where the soil type according to EC8 is C. For some areas of the Apennines, the buildings are built on outcropping bedrock, and the soil class according to EC8 is A. More details about soil classification are reported into Tab. 5.

Table 5: Distribution of the soil types within the study area.

Soil class	Number	Percentage
A	147	1.13%
B	11,414	88.11%
C	1,393	10.75%
Total	12,954	

The maximum estimated fundamental period T of the buildings is about 1 s. Fig. 9 shows that most of the buildings' fundamental periods are between 0.2 s and 0.3 s, which fairly corresponds to the region of the acceleration spectra where the maximum amplifications of the seismic ground motion occur.

The reliability of the selected attenuation laws is evaluated by comparing predicted and actual values of the selected seismic intensity measures. To this end, it is considered the strongest seismic event with its epicenter closest to the study area. It is the M_w 5.5 Capitignano earthquake, which occurred on 2017-01-18. Hence, all accelerograms recorded at a distance of no more than 200 km from the epicenter of this seismic event are examined. A direct comparison between predicted and actual values is straightforward for all selected intensity measures, except for the horizontal pseudo-spectral acceleration. In such a case, predicted and actual values of the pseudo-spectral acceleration have been determined for the fundamental periods range relevant for the buildings in the study area.

Fig. 10 illustrates actual and predicted value of horizontal peak ground acceleration PGA_h, horizontal pseudo-spectral acceleration $S_{pa,h}$, Arias intensity I_a, and Housner intensity I_h. Moreover, Fig. 11 shows

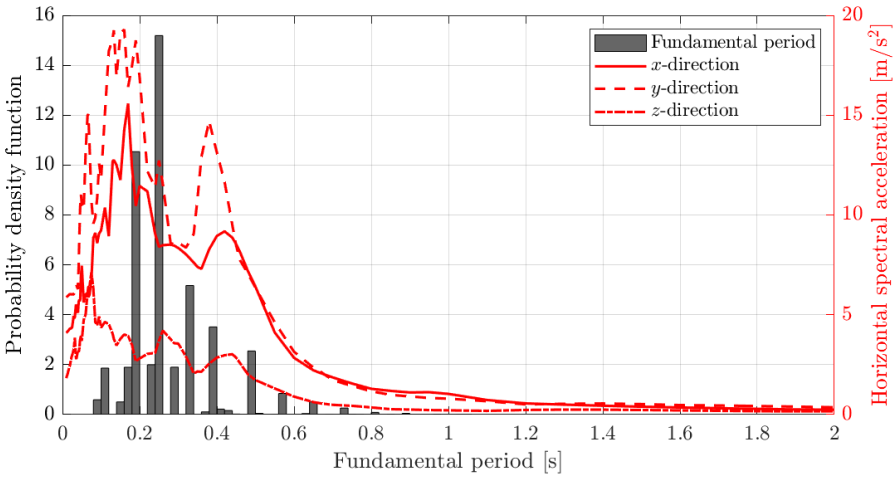


Figure 9: Histogram of the buildings' fundamental period values togheter with the 5% damped acceleration elastic spectra for the M_w 5.5 Capitignano earthquake occurred on 2017-01-18.

the histograms corresponding to the estimated values of horizontal peak ground acceleration PGA_h, vertical peak ground acceleration PGA_v, and horizontal pseudo-spectral acceleration $S_{pa,h}$.

The analysis of Fig. 10 substantiates a general very good accuracy of the considered attenuation laws for the present study area. There is a systematic underestimation of all seismic intensity measures in the neighbourhood of the epicentre. As soon as the distance of the buildings from the epicentre is larger than a few kilometers, Fig. 10 confirms that considered attenuation laws provide a satisfactory prediction of the selected seismic intensity measures. In this regard, although the attenuation law by Sabetta and Puglisi [45] is not very recent, it was calibrated by using data from Italian earthquakes mostly occurred in the Appennines. Therefore, it is especially suitable for the considered study area, as confirmed by Fig. 10.

Figures 12-14 provide a more comprehensive overview about the seismic intensity measures over the study area and how they correlate with the observed damage. Particularly, the spatial distribution of horizontal peak ground acceleration PGA_h, horizontal pseudo-spectral acceleration $S_{pa,h}$ and Arias intensity I_a are shown in Fig. 12, Fig. 13 and Fig. 14, respectively.

4. Machine learning models

4.1. Definition of input explanatory variables and outcome

The main goal of the current study is to investigate the suitability of ML techniques in developing a predictive model able to foresee the expected post-seismic building usability starting from a set of input explanatory variables related to the constructions and their site (as reported within AeDES form) together with some information related to the intensity of the seismic ground motion.

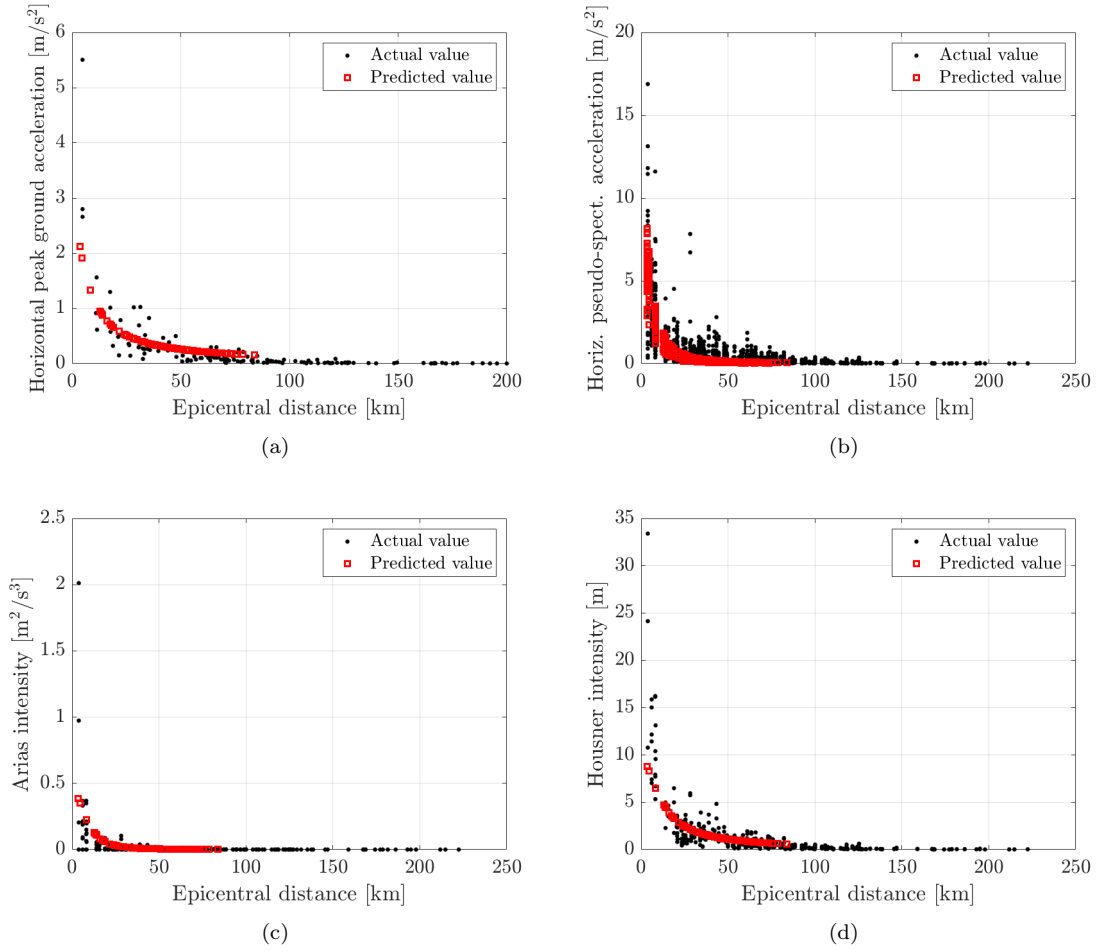


Figure 10: Comparison between actual and predicted values for some selected seismic intensity measures: horizontal peak ground acceleration PGA_h (a), horizontal pseudo-spectral acceleration $S_{pa,h}$ (b), Arias intensity I_a (c), and Housner intensity I_h (d).

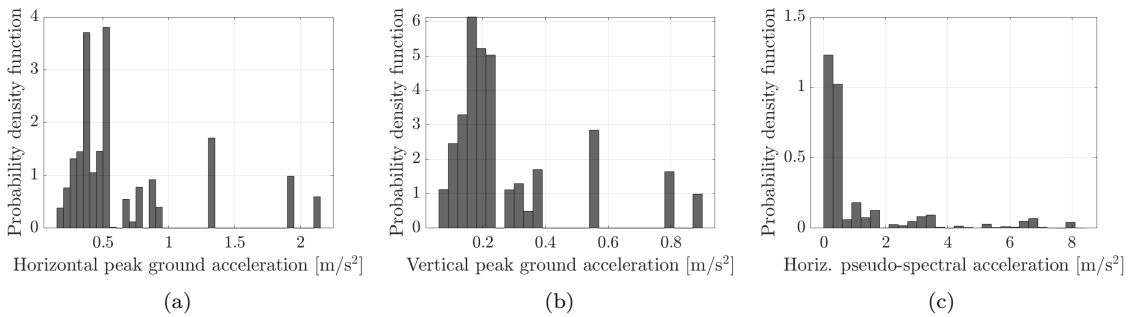
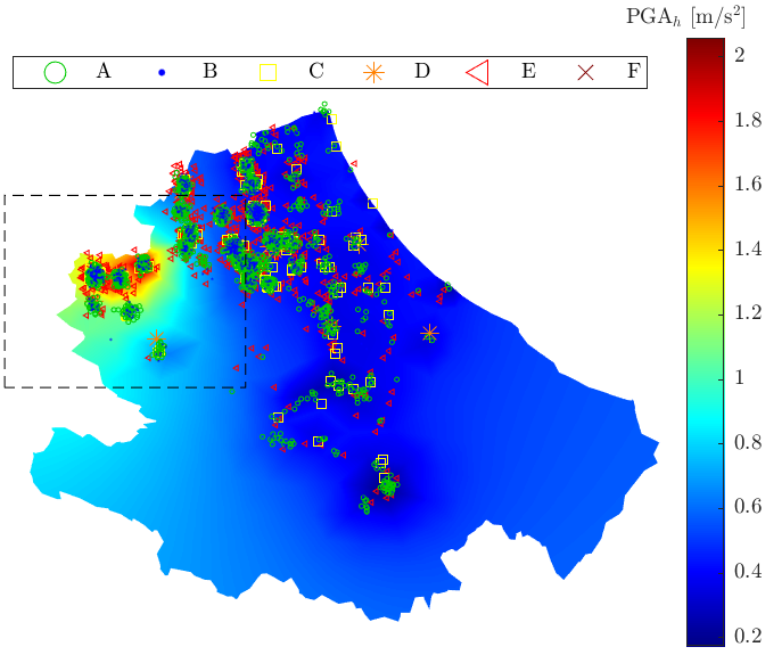
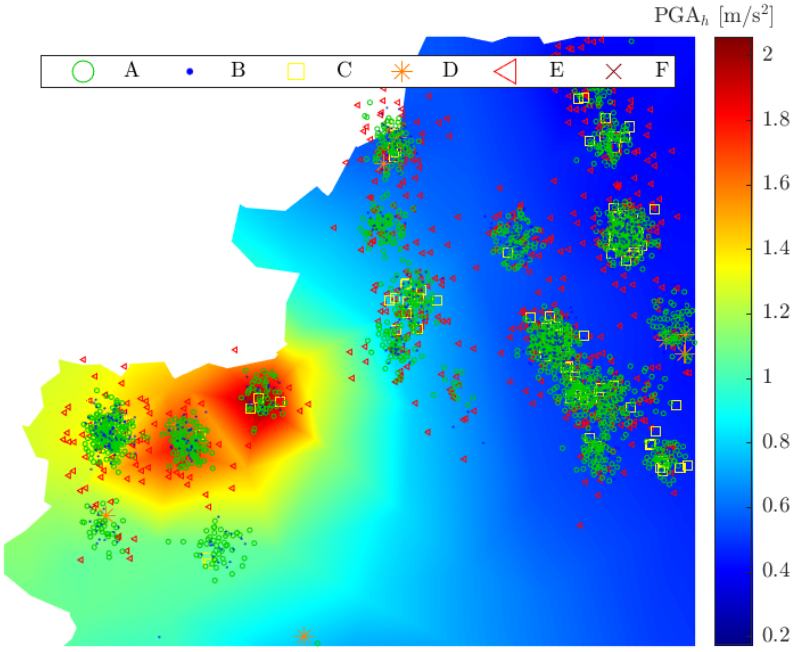


Figure 11: Histograms for some selected seismic intensity measures: horizontal peak ground acceleration PGA_h (a), vertical peak ground acceleration PGA_v (b), and horizontal pseudo-spectral acceleration $S_{pa,h}$ (c).

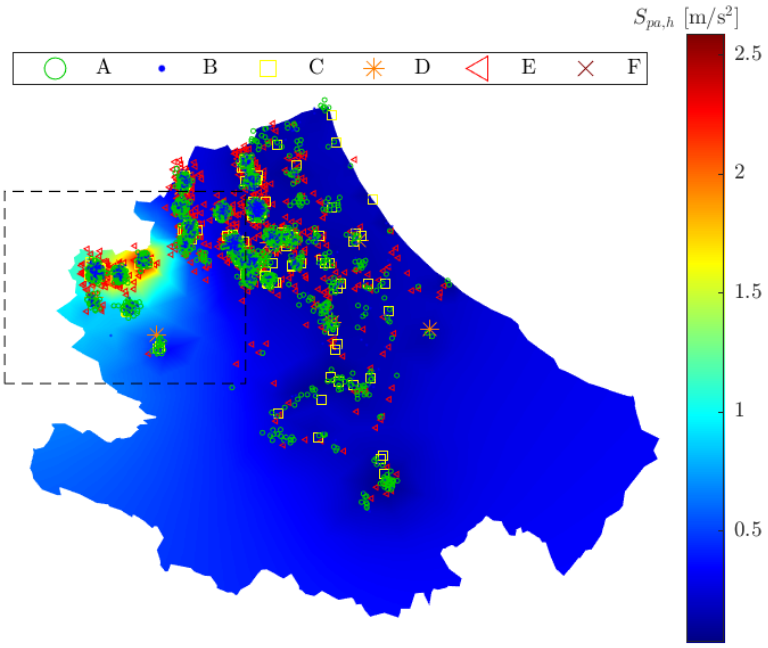


(a)

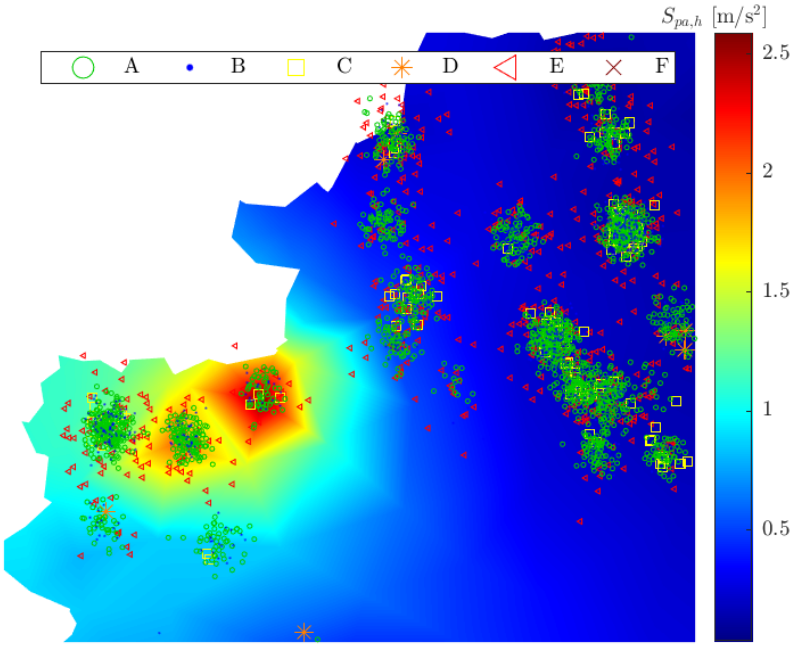


(b)

Figure 12: Contour density plots of the horizontal peak ground acceleration PGA$_h$ over the study area: general view (a) and close-up view on the outlined area (b).

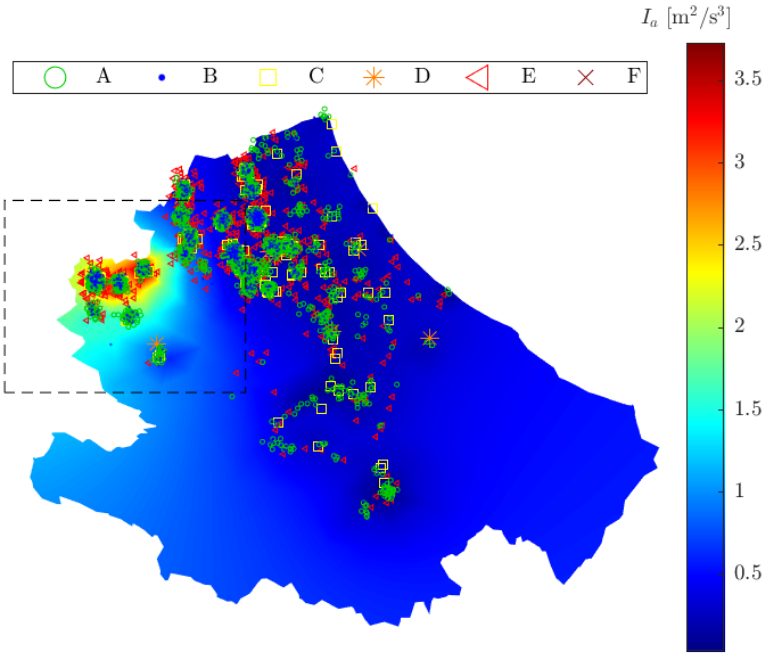


(a)

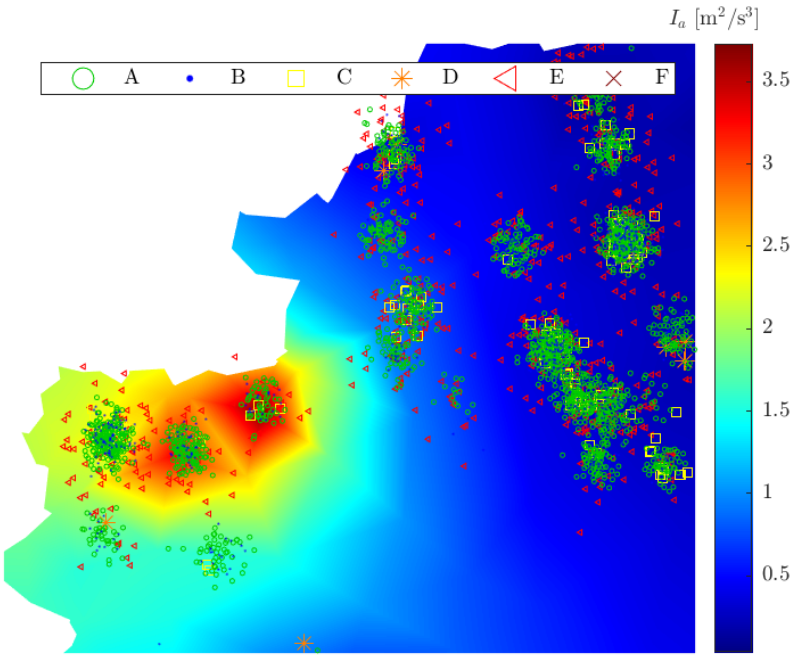


(b)

Figure 13: Contour density plots of the horizontal pseudo-spectral acceleration $S_{pa,h}$ over the study area: general view (a) and close-up view on the outlined area (b).



(a)



(b)

Figure 14: Contour density plots of the Arias intensity I_a over the study area: general view (a) and close-up view on the outlined area (b).

To this end, a database including relevant data from buildings within the study area has been prepared. The set of input explanatory variables within such database includes 13 variables related to the buildings and their site as reported within the AeDES form, and 9 variables derived from the assessment of the seismic intensity. All the input explanatory variables and the target outcome are listed into Tab. 6.

The outcome is the post-seismic usability building class, which is defined through a rational elaboration of the consequence classes reported within the AeDES forms. From the full dataset including a total of 12,662 buildings, those that belong to class D (inconclusive survey) and class F (unusable building because of external causes only) have been initially removed. Buildings with missing data have been also removed from the dataset delivering a final size of 12,063 data. Then, class A and class B are merged into a single seismic building class AB (total number of samples equal to 9,361), which collects buildings that can be used immediately or after emergency interventions because they experienced no or light damage. This is motivated by the fact that the assignment of a surveyed building to class A or class B is affected by large uncertainties and subjectivity since the proper assessment of both extent and intensity of light damage, if any, was not always straightforward [27]. Additionally, the assignment of a surveyed building to class A or class B was also influenced, to a significant extent, by social and economical factors. Actually, in case of ambiguous assignment to class A or class B, the latter was sometimes selected because it allowed the building owner to receive refunds for emergency interventions. The former was selected instead when the building owner was available to bear the economic costs in such a way to carry out minor repairs in a short time, without waiting for the approval of refund request. Finally, class C and class E are also merged into a single seismic building class CE (total number of samples equal to 2,702), which collects buildings that are partially or totally unusable because they experienced severe damage. It is highlighted that considering two classes only (namely AB and CE) is also beneficial to improve the reliability of the classification. It is worth stressing that, when dealing with binary classification problems for imbalanced datasets, the positive class is commonly associated with the minority one [51]. In this case, the minority class is CE, representing about 22.40% of the entire available dataset.

4.2. Data preparation and machine learning techniques

All categorical variables underwent label encoding to transform them into numerical information. A min-max scaling is performed afterward the label encoding in order to normalize their value within a common range and avoid a possible biased learning [52, 53]. Two distinct normalization approaches are employed for continuous numerical variables. While epicentral distance values are standardized in such a way that mean and standard deviation are equal to zero and one, respectively, min-max scaling is applied to other

Table 6: Input explanatory variables and target outcome.

No.	Variable	Type	No.	Variable	Type
1	Building position	Categorical	14	Soil category	Categorical
2	No. of floors	Discrete numerical	15	Epicentral distance	Continuous numerical
3	Inter-storey height	Discrete numerical	16	Fundamental period	Continuous numerical
4	Average floor surface	Discrete numerical	17	Horiz. peak ground acceleration	Continuous numerical
5	Age	Discrete numerical	18	Vert. peak ground acceleration	Continuous numerical
6	Typology	Categorical	19	Horiz. pseudo-spectral acceleration	Continuous numerical
7	Roof	Categorical	20	Arias Intensity	Continuous numerical
8	Masonry texture	Categorical	21	Housner Intensity	Continuous numerical
9	Curbs/chains	Categorical	22	Duration	Continuous numerical
10	Isolated column	Categorical	23	Post-seismic usability class	Categorical
11	Hybrid masonry	Categorical	Legend		Building and site data
12	Reinforced masonry	Categorical			Seismic intensity data
13	Site morphology	Categorical			Outcome

continuous numerical variables. The pre-processed database has been divided into a training set and a validation set with 85% and 15% relative proportions (10,253 and 1,810 instances, respectively).

A suite of 11 different ML techniques has been implemented using the Python Scikit-learn library [54] in the attempt to find the best classifier for the present study. The motivation behind this extensive comparative assessment stems from the recognition that every ML algorithm comes with its own set of advantages and disadvantages, and it is not possible to know a priori how it will perform. The adopted ML techniques are listed hereafter.

- K-nearest neighbors (KNN) [55, 56]: it is a supervised ML algorithm that classifies data points based on the majority class among their K-nearest neighbours.
- Linear support vector machine (Linear SVM) [57]: it is a classification algorithm that finds a hyperplane that separates data into different classes. It is especially effective for linearly separable datasets.
- Radial basis function support vector machine (RBF SVM) [58]: it is a variant of the corresponding linear version that implements a radial basis function kernel to handle non-linearly separable data. It maps data to a higher-dimensional space in such a way to improve the separation among the classes.
- Decision tree (DT) [59]: it is a supervised algorithm in which the learning process implements a divide-and-conquer strategy by conducting a greedy search that identifies the optimal split points within a tree.
- Random forest (RF) [60]: it is an ensemble learning method that combines multiple decision trees to improve predictive accuracy and to reduce overfitting.
- Neural network (Neural Net) [61, 62]: it consists of layers of interconnected nodes (neurons) that mimic

how the human brain works. The standard feedforward neural network is considered in the present work, also denoted in literature as multi-layer perceptron (MLP). Specifically, the default shallow MLP architecture was adopted with one hidden layer composed of 100 neuron units [54].

- Adaptive boosting (AdaBoost) [63]: it is an ensemble learning method that combines weak classifiers to create a strong classifier. It focuses on improving the classification of difficult-to-classify examples.

- Naive Bayes [64]: it is a probabilistic classification algorithm based on Bayes' theorem where the features are assumed to be conditionally independent each other.

- Quadratic discriminant analysis (QDA) [65]: it is a classification method that models the distribution of each class as a quadratic shape. It's a more flexible variant of the corresponding linear version and is especially suitable when the classes have different covariance matrices.

- Logistic regression (LR) [65, 66]: it is a supervised statistical-based method in which the posterior probability of one class is formulated as a logistic sigmoid function taking as argument a linear function of the features. Contrary to its name, the LR is a classification model born for binary classification tasks.

- Linear discriminant analysis (LDA) [67]: it is an analytical closed-form classifier that does not require any hyperparameter tuning. It represents a generalization of Fisher's linear discriminant method [65], and it aims to find a linear combination of features to linearly separate classes in a classification scheme.

Generally, ML classifiers should be trained using as much as possible of the available training data. Nevertheless, if a validation set extracted from the training data is small, then this can provide a noisy and non-representative estimate of the predictive performance of a ML model [65]. Accordingly, the training dataset is divided into a number k of equally sized folds. Each fold then serves as training set and validation set throughout a process known as k-fold cross-validation [68]. It has been demonstrated in previous studies that the stratified cross-validation with 10 folds delivers smaller bias whilst limiting the variance even when computational capabilities virtually permits for more folds [69]. Therefore, the dataset has been split into $k = 10$ equal subsets, and each fold was used for training and validation. Specifically, one fold is chosen as validation set while the remaining $k - 1$ folds are used for training. This process leads to k different models, and each one is trained on a distinct subset of data. Classification metrics across the various folds are then collected and averaged to enable a fair comparison of such different ML models. In practice, the

cross-validation procedure offers a means to compare entirely distinct ML models, thereby pinpointing the predictors that consistently deliver the best performance on average and exhibit robust responses across different segments of the training set [70]. Initially, all ML algorithms have been applied by adopting their default hyperparameters provided by SciKit-learn [54], and the corresponding performances are then compared to deduce which one is the most appropriate for the present case study. Next, the predictive capacity of such subset of ML algorithms will be further enhanced by optimizing their hyperparameters.

It is worth pointing out that the use of the stratified cross-validation scheme in the present work differs from some previous studies that opted for a repeated scheme when using ML algorithms for buildings usability assessment [e.g., 19]. Actually, a repeated cross-validation scheme does not account for the relative proportion of every class. Moreover, it has been shown that multiple repetitions may introduce a significant bias, thereby masking the real performance of the trained classifier [71]. This motivated the implementation of the stratified cross-validation scheme in the present study.

Several well-known classification metrics are computed, including those specialized for handling with imbalanced datasets [51, 72]. The considered metrics are listed and discussed hereafter.

- Accuracy: out of the total classifications, accuracy reflects what percentage is truly positive. Accuracy has some well-known limitations when dealing with imbalanced datasets. This is because high accuracy values are achieved when the classifier is completely biased toward the majority class.
- Precision: out of all the positive classifications, precision reflects what percentage is truly positive.
- Recall: out of the total positive classifications, recall gives the true positive rate. This is typically used to measure the coverage of the minority class of major interest in imbalanced learning problems.
- Balanced accuracy: it computes the average of the percentage of positive and negative class instances correctly classified, i.e. the average of the recall of the majority and minority classes. This is commonly recognized as one of best metrics to synthesize the classification performance over both classes in terms of recall values.
- F1-score: it is the harmonic mean of precision and recall. Its generalization is also known as Fβ-score or F-measure.
- AUC-ROC: it is the area under the receiver operating characteristic (ROC) curve. It is based on evaluating the models at different error rates, delivering the proportions of instances correctly classified for a given false positive rate (FPR).

Accuracy, balanced accuracy, precision on the minority class, recall on the minority class, F1-score and AUC-ROC are thus monitored during the cross-validation procedure. Additionally, after retraining the ML model on the entire dataset, hold-out test set performance metrics have been finally computed from the confusion matrix [73].

The importance of examining alternative classification metrics should not be overlooked. Indeed, the considered metrics are not all equally important. In fact, it is widely recognized that that accuracy is a poor metric when dealing with imbalanced datasets. Although AUC-ROC has been combined with the accuracy into previous studies to cope with such limitation in buildings usability assessment via ML models [e.g., 19], recent studies [74] have demonstrated that it can be an overly optimistic and extremely volatile metric. This is especially relevant for the considered class of applications, since it would imply an uneven response of the classifier that reflects into too many non-conservative assignments. Several studies recommended to replace the AUC-ROC with the precision-recall curve and the area beneath [75]. Therefore, while all classification metrics are examined in the present study, recall and balanced accuracy will deserve more consideration following well-established recommendations available in the existing literature for strongly imbalanced datasets [51].

Data preprocessing techniques can play a somewhat important role when handling with imbalanced data. This usually depends on the specific application and the implemented algorithm. Although such task is often skipped in the pertinent literature, the explicit quantification of the impact of data preprocessing techniques is a good practice because their potential beneficial effects is not known a priori. Hence, two different preprocessing techniques are taken into account in the attempt to alleviate possible detrimental effects due to the lack of uniform distribution of data, namely the principal component analysis (PCA) and the synthetic minority over-sampling technique (SMOTE). PCA is a dimensionality reduction method that aims at reducing the number of feature dimensions while preserving as much valuable information as possible while still able to describe the dataset with a controlled information loss [56, 76, 77]. Since the entropy is a measure of the information content, the basic assumption of PCA is that the total variance of the dataset is not explained equally by all dimensions of the dataset. Thus, through a singular value decomposition (SVD) of the dataset matrix, this latter is projected with a linear transformation onto the eigenvector space [76–78]. If some components have low variance, it is possible to truncate the SVD at a lower dimension delivering a projection onto a reduced subspace without a detrimental information loss. In order to contain the information loss associated with the dimensionality reduction, it is advisable to select a proper number of components providing a sufficient explained variance ratio (EVR). The EVR is the ratio of the variance

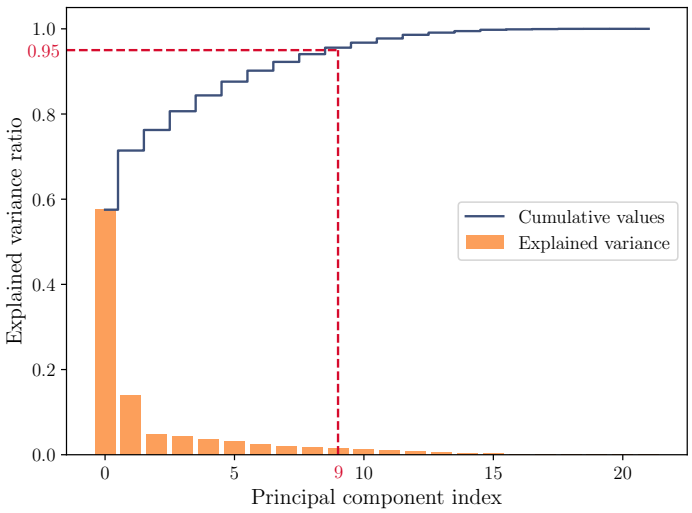


Figure 15: Pareto chart of the PCA for dimensionality reduction.

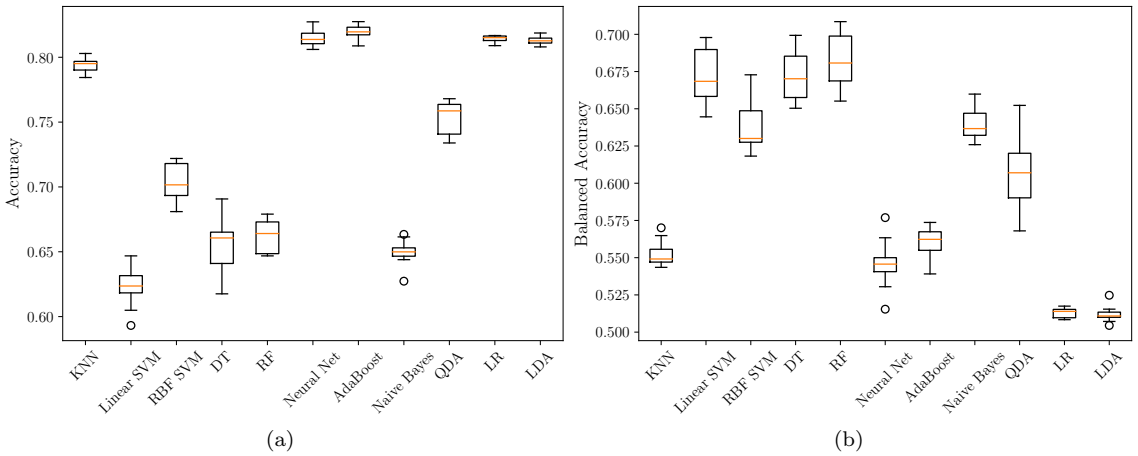
explained by the single component over the total variance [76]. It is shown in Fig. 15 that, by retaining an intrinsic dimension of 9 out of 22, it is possible to provide a dimensionality reduction whilst limiting the information loss within the 5% of the EVR.

The second preprocessing technique herein employed to cope with imbalanced data is the SMOTE. Although alternative variants have been proposed [79–82], the core of this method is based on the generation of synthetic instances for the minority class, thereby fostering a more balanced dataset [79]. Moreover, it is recognized to be more stable than naive oversampling methods that relies on randomly chosen replicates of the minority class samples because this can induce possible overfitting issues [51, 79]. The new instances are denoted as synthetic because, once a minority instance is picked up randomly, the SMOTE attempts to find the nearest samples based on their similarity within the feature space, and then it extrapolates a new instance by a convex combination of the selected ones.

4.3. Machine learning models based on default hyperparameters

Initially, each candidate ML algorithm was implemented using the default hyperparameters from SciKit-learn [54]. Their performances were then compared to determine the most suitable one for this specific case study. In this regard, Fig. 16 confirms that accuracy is not a reliable performance metric for imbalanced datasets. Indeed, it is observed that ML models having the highest accuracy often display the lowest balanced accuracy score, which more faithfully reflects the model's performance on both minority and majority classes.

After the training phase, the hold-out test confusion matrix and the related metrics have been computed for all the 11 ML models taking also into account the effect of the preprocessing techniques (viz., PCA and SMOTE). The results are reported in Tab. 7.



(a) (b)

Figure 16: Cross-validation average results in terms of accuracy (a) and balanced accuracy (b) for the considered ML models.

Table 7: Comparison of hold-out test metrics for different ML algorithms and preprocessing techniques.

		Preprocessing technique							
		No PCA No SMOTE		No PCA SMOTE		PCA No SMOTE		PCA SMOTE	
Algorithm	Metric	Class AB	Class CE	Class AB	Class CE	Class AB	Class CE	Class AB	Class CE
KNN	Acc.	75.03		67.24		76.24		66.46	
	Bal. acc.	54.58		60.41		57.02		59.74	
	Precision	79.35	37.77	82.89	33.80	80.32	44.12	82.58	32.88
	Recall	91.67	17.49	72.79	48.03	91.88	22.17	71.94	47.54
	F1-score	85.06	23.91	77.51	39.67	85.71	29.51	76.89	38.87
	ROC AUC	65.57		66.01		64.40		65.62	
Linear SVM	Acc.	62.32		64.81		62.04		64.42	
	Bal. acc.	65.91		65.32		65.90		64.20	
	Precision	88.16	34.03	86.84	34.98	88.26	33.91	86.05	34.26
	Recall	59.40	72.41	64.39	66.26	58.90	72.91	64.60	63.79
	F1-score	70.98	46.30	73.95	45.79	70.65	46.29	73.80	44.58
	ROC AUC	71.29		70.95		69.55		68.88	
RBF SVM	Acc.	67.29		68.62		65.30		68.01	
	Bal. acc.	61.58		61.21		63.28		63.27	
	Precision	83.61	34.55	83.17	35.27	85.14	34.28	84.58	35.98
	Recall	71.94	51.23	74.64	47.78	66.95	59.61	71.87	54.68
	F1-score	77.34	41.27	78.68	40.59	74.96	43.53	77.71	43.40
	ROC AUC	67.48		67.46		67.57		67.95	
DT	Acc.	60.94		64.42		63.70		63.09	
	Bal. acc.	64.84		64.90		61.72		61.07	
	Precision	87.68	32.99	86.61	34.59	84.36	32.64	84.01	32.01
	Recall	57.76	71.92	64.03	65.76	65.31	58.13	64.74	57.39
	F1-score	69.64	45.24	73.63	45.33	73.63	41.81	73.13	41.09
	ROC AUC	70.01		69.05		67.04		66.10	
RF	Acc.	64.14		67.68		65.03		69.61	
	Bal. acc.	65.77		64.72		64.94		61.59	
	Precision	87.41	34.83	85.64	36.46	86.47	34.93	83.26	36.31
	Recall	62.82	68.72	70.09	59.36	65.10	64.78	76.14	47.04
	F1-score	73.10	46.23	77.09	45.17	74.28	45.38	79.54	40.99
	ROC AUC	72.60		72.39		69.92		69.24	
Neural Net	Acc.	76.35		69.17		77.46		66.35	
	Bal. acc.	53.86		63.06		51.42		61.94	
	Precision	79.01	41.41	84.22	36.76	78.07	47.22	84.00	34.17
	Recall	94.66	13.05	74.15	51.97	98.65	4.19	69.94	53.94
	F1-score	86.13	19.85	78.86	43.06	87.16	7.69	76.33	41.83
	ROC AUC	71.50		70.68		70.76		69.03	

Table 7: (continued from previous page)

Method	Metric	preprocessing techqniue							
		No PCA No SMOTE		No PCA SMOTE		PCA No SMOTE		PCA SMOTE	
		Class AB	Class CE	Class AB	Class CE	Class AB	Class CE	Class AB	Class CE
AdaBoost	Acc.	78.01		68.01		76.57		66.85	
	Bal. acc.	54.84		64.15		51.81		63.49	
	Precision	79.35	54.17	85.17	36.42	78.23	37.84	84.96	35.30
	Recall	96.87	12.81	71.15	57.14	96.72	6.90	69.59	57.39
	F1-score	87.24	20.72	77.53	44.49	86.50	11.67	76.51	43.71
	ROC AUC	72.69		70.77		69.01		68.94	
Naive Bayes	Acc.	64.42		52.71		76.69		61.27	
	Bal. acc.	63.85		62.34		52.49		62.43	
	Precision	85.78	34.09	88.48	29.51	78.48	40.70	85.47	31.99
	Recall	64.89	62.81	44.87	79.80	96.37	8.62	60.33	64.53
	F1-score	73.88	44.19	59.55	43.09	86.51	14.23	70.73	42.78
	ROC AUC	68.35		67.38		68.17		65.51	
QDA	Acc.	70.50		58.84		73.70		61.05	
	Bal. acc.	59.54		65.68		54.86		62.99	
	Precision	81.99	35.78	89.37	32.58	79.52	35.29	85.99	32.18
	Recall	79.42	39.66	53.28	78.08	89.03	20.69	59.47	66.50
	F1-score	80.68	37.62	66.76	45.98	84.01	26.09	70.32	43.37
	ROC AUC	69.99		69.06		69.21		66.98	
LR	Acc.	77.57		66.35		77.51		65.19	
	Bal. acc.	52.28		64.83		50.66		64.17	
	Precision	78.38	50.00	86.04	35.64	77.80	47.06	85.83	34.66
	Recall	98.15	6.40	67.59	62.07	99.36	1.97	66.03	62.32
	F1-score	87.16	11.35	75.71	45.28	87.27	3.78	74.64	44.54
	ROC AUC	71.47		71.10		69.59		68.83	
LDA	Acc.	77.40		65.97		77.57		64.36	
	Bal. acc.	51.82		66.07		50.70		64.16	
	Precision	78.22	46.81	87.10	35.96	77.81	50.00	86.04	34.21
	Recall	98.22	5.42	65.88	66.26	99.43	1.97	64.53	63.79
	F1-score	87.09	9.71	75.02	46.62	87.30	3.79	73.75	44.54
	ROC AUC	71.13		71.08		69.56		68.97	

High precision scores for class AB suggest that all ML models are able to identify accurately those buildings that will experience no or minimal damage after the earthquake. Precision scores for class CE are lower than those for class AB, thus indicating a high incidence of false positives in the classification of the buildings that will experience a severe seismic damage. The impact of preprocessing techniques on precision scores seems somehow limited. The recall value, which quantify the model's ability to capture positive instances, is high for class AB. This further confirms that ML models are able to recognize effectively those buildings that will result either undamaged or slightly damaged after the earthquake. While precision scores are comparable across ML models for class AB, recall scores are largely different. Additionally, it is noted that preprocessing techniques have significant influence the recall values. On the one hand, the Linear SVM stands out with the highest recall value (equal to 72.91%) for class CE when it is applied with PCA dimensionality reduction. However, the RF algorithm demonstrates the most consistent recall scores between class AB and class CE (recall values equal to 62.82% and 68.72%, respectively). Tab. 7 suggests that preprocessing is especially beneficial for those ML models that exhibit a substantial difference in terms

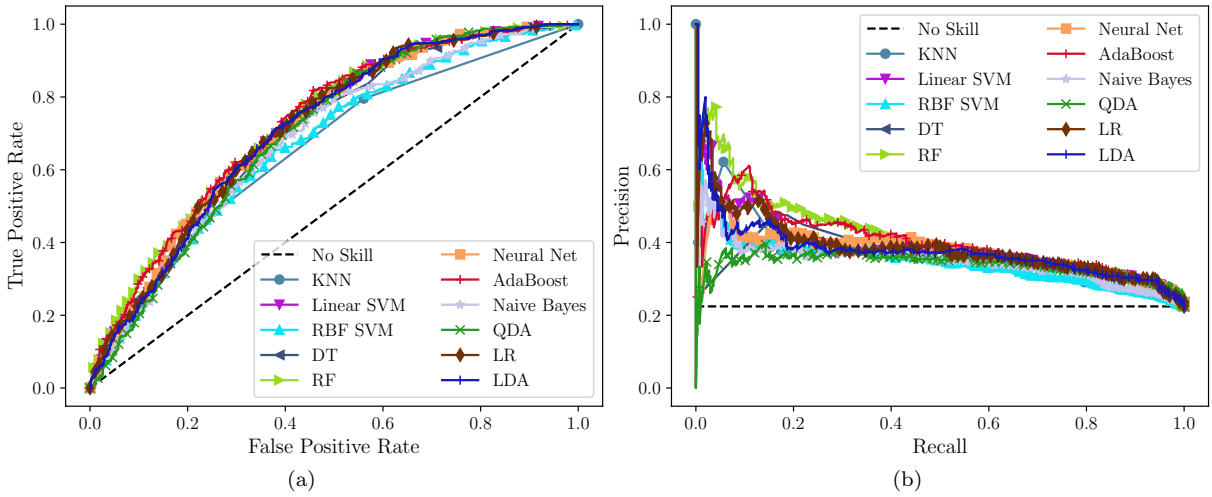


Figure 17: Comparisons of ROC curves (a) and precision-recall curves (b) for the considered ML models.

of F1-scores between class AB and class CE, while a moderate influence is observed otherwise.

The ROC curves and the precision-recall curves have been reported in Fig. 17. The ROC curves in Fig. 17a are almost overlapped and tend to be quite optimistic since all the ML models display virtually the same behavior and a performance better than that of the no-skill learner, which is represented by a diagonal dashed line [74]. The precision-recall curve in Fig. 17b is useful to further discriminate the ML models [75]. It demonstrates that the RF algorithm exhibits the highest value of the area enclosed by this curve. In such a case, the no-skill learner is represented by the horizontal dashed line corresponding to the minority class proportion (i.e., 22.40%).

Balanced accuracy is shown in Fig. 18 and is a key metric to guide the selection of ML algorithms and preprocessing techniques when handling imbalanced datasets. It can be deduced from Fig. 18 that LDA method achieved the highest balanced accuracy (equal to 66.07%) when SMOTE is applied. However, LDA also shows inconsistent results since its performance drops to about 50% (akin to a no-skill learner) when no preprocessing or PCA only is employed. The Linear SVM demonstrates good robustness (achieving a balanced accuracy of 65.91%) without any preprocessing technique, and similar scores are obtained with PCA and/or SMOTE. The RF algorithm was already found able to provide satisfactory and closely matched recall values without the application of any preprocessing technique, and Fig. 18 confirms a consistent good performance also in terms of balanced accuracy (which results about 65% regardless of the implemented preprocessing technique).

It is not straightforward to decide which ML model is the best since different goals, metrics, and pre-

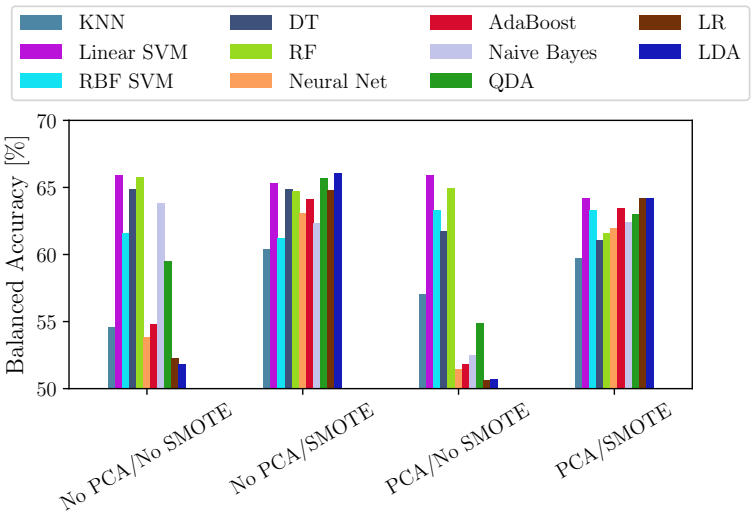


Figure 18: Comparisons of balanced accuracy metric values for different ML models considering alternative preprocessing methods.

processing techniques can lead to somewhat different conclusions. For the goals of the present study, it is advisable to choose the ML model that ensures the most stable performance and the optimal balance among various criteria, rather than the best classifier based on a single criterion.

Taking all into account, the present comparative assessment among alternative ML techniques has demonstrated that the implementation of Linear SVM and RF without the preliminary application of preprocessing techniques is most appropriate for the goals of the present research. Ultimately, they represent the optimal trade-off over all classification metrics for both seismic classes with minimum computational effort. Therefore, these two ML models only will be considered henceforth for the hyperparameter tuning process.

4.4. Optimal machine learning model

The ML-based approaches that resulted in the most suitable classifiers, namely the Linear SVM and the RF without data preprocessing, are now optimized by means of hyperparameters fine-tuning.

Only the regularization parameter $\mathcal{C}$ can be optimized in the Linear SVM. It basically controls the balance between model complexity and the misclassification tolerance, thus controlling the occurrence of overfitting phenomena. However, Linear SVM is quite sensitive to class imbalance issues, and it might result in a skewed behavior toward the minority class [51]. For small $\mathcal{C}$ values, Linear SVM applies a low regularization penalty to misclassified data points delivering a soft margin. This indicates a higher tolerance for classification errors, which turns out into a less flexible decision boundary with a broader margin between classes. Small $\mathcal{C}$ values can lead to better performance for the minority class in imbalanced datasets by reducing the risk of overfitting for the majority class. In contrast, large $\mathcal{C}$ values impose a

Table 8: Mean and standard deviation of the classification metrics for the Linear SVM corresponding to different values of the regularization parameter $\mathcal{C}$.

	Accuracy		Balanced acc.		Precision		Recall		F1-score		AUC-ROC	
Linear SVM	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
$C = 0.001$	55.37	1.59	61.76	1.99	29.82	1.29	73.35	3.27	42.40	1.79	66.31	2.31
$C = 0.01$	58.80	1.09	64.94	1.38	32.22	0.88	76.05	2.82	45.25	1.26	70.30	1.53
$C = 0.1$	61.91	1.18	65.93	1.22	33.83	0.93	73.22	2.37	46.27	1.17	71.22	1.74
$C = 1$	62.53	1.38	66.22	1.41	34.22	1.16	72.91	2.27	46.57	1.39	71.38	1.83
$C = 10$	62.22	1.40	66.10	1.49	34.03	1.17	73.13	2.43	46.44	1.45	71.49	1.91
$C = 100$	62.12	1.35	66.02	1.39	33.96	1.11	73.09	2.23	46.36	1.35	71.50	1.95
$C = 1000$	62.29	1.29	66.21	1.29	34.11	1.04	73.30	2.15	46.55	1.26	71.50	1.96
$C = 10000$	62.14	1.37	66.25	1.43	34.06	1.11	73.70	2.42	46.58	1.38	71.51	1.98

stronger regularization penalty on misclassified data, in such a way as to achieve a high precision and an accurate classification for all training data. This implies a hard and narrower margin between classes, thus resulting in a more complex decision boundary. In this case, there is a heightened risk of overfitting, as the model may perform exceptionally well on the training dataset but struggle when faced with new, unseen data, owing to its diminished generalization capacity.

Tab. 8 displays mean and standard deviation of the classification metrics (i.e., accuracy, balanced accuracy, precision, recall, F1-score, and AUC-ROC) for the grid-search stratified cross-validation procedure of the Linear SVM for different values of the regularization parameter $\mathcal{C}$ (it is noted that the default value is $\mathcal{C} = 10$). Precision, recall, F1-score refers to the minority class only. Tab. 8 demonstrates that the impact of the regularization parameter is almost negligible for $\mathcal{C} > 1$.

Furthermore, Fig. 19 illustrates the balanced accuracy for different values of $\mathcal{C}$, together with the training computational time (expressed in seconds). In summary, intermediate values of $\mathcal{C}$ should be preferred to avoid both underfitting and overfitting phenomena while saving computational time. Indeed, the optimal value of $\mathcal{C} = 1$ has been selected for the optimized Linear SVM model.

As far as the hyperparameters tuning of the RF algorithm is concerned, it is worth reminding that it is an ensemble ML algorithm based on multiple weak classifiers decision trees and whose final prediction is governed by a majority voting scheme [60, 83]. This working mechanism makes the RF algorithm very robust against possible overfitting issues, reducing both bias and variance, and it sometimes does not even require dedicated hyperparameters tuning [83]. Nonetheless, several hyperparameters can be possibly optimized. They deal with the single DT (e.g., the number of learners, how to control the DT growth, information gain splitting criterion, etc.) and the bagging scheme (bootstrap aggregation) [83, 84]. In agreement with the most common recommendations [85], the following hyperparameters have been selected for the grid-search cross-validation procedure of the RF algorithm.

- Criterion: it defines the objective function adopted to assess the information gain (IG) at each leaf

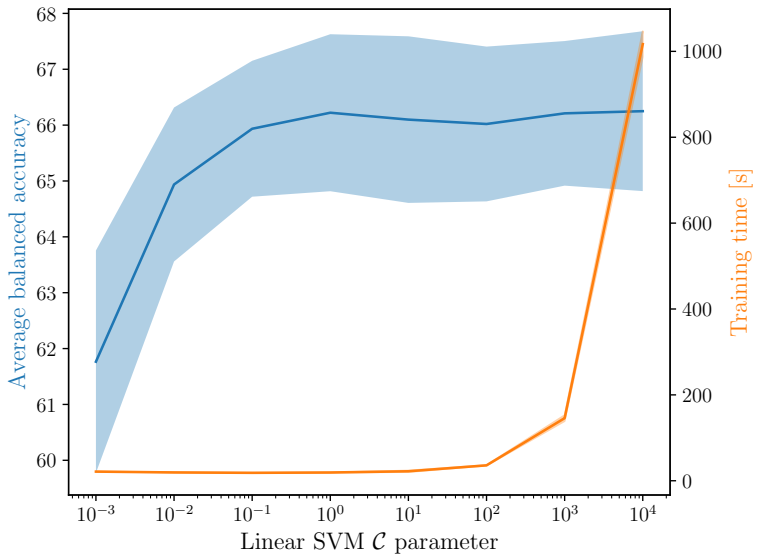


Figure 19: Average balanced accuracy and training elapsed time during grid-search cross-validation for hyperparameter tuning of the Linear SVM by varying the regularization factor $\mathcal{C}$.

node. Three possible definitions have been considered, i.e. the entropy, the Gini impurity index (default criterion), and the cross entropy (or Log Loss) [83, 84, 86].

- Max depth: it establishes the maximum number of splits of every DT in the forest. By default, no constraints are fixed on the DT's depth, thus achieving pure terminal leaf nodes [83, 84]. In the grid-search scheme, this hyperparameter was set equal to 5, 10, 20, 50, and 100.

- Max leaf nodes: it governs the maximum number of splits by constraining the total number of leaf nodes, both intermediate and terminal ones [83, 84]. Besides the default value outlining no constraints applied on this hyperparameter, a maximum number of splits equal to 5, 10, 20, 50, and 100 has been considered.

- Min samples leaf: it defines the minimum number of samples that have to remain in a leaf node [83, 84]. Besides the default value of 1, a minimum number of samples equal to 2 and 4 has been investigated.

- Min samples split: it defines the minimum number of instances necessary to split an internal node [83, 84]. The values of 2 (default value), 5, and 10 have been considered.

- Number of estimators: it defines the number of trees to consider in the forest [83, 84]. Since it should be large enough [85], six possible values have been explored, i.e. 100 (default value), 200, 300, 400, 500, and 1000.

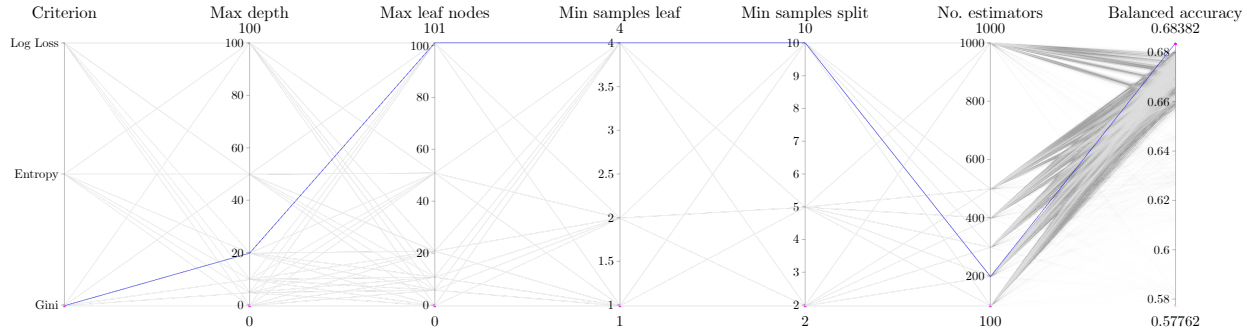


Figure 20: Parallel coordinates graph that illustrates the grid-search cross-validation performed for hyperparameters tuning of the RF algorithm (the blue line indicates the best hyperparameters delivering the highest balanced accuracy).

Table 9: Hold-out test set performances of the optimized Linear SVM and RF models.

Metric	Optimal ML model (No PCA/No SMOTE)			
	Linear SVM		RF	
	Class AB	Class CE	Class AB	Class CE
Accuracy	62.32		67.73	
Balanced accuracy	65.91		65.72	
Precision	88.16	34.03	86.35	36.95
Recall	59.40	72.41	69.37	62.07
F1-score	70.98	46.30	76.94	46.32
AUC-ROC	71.29		74.33	

The bagging procedure is maintained with the default arguments, thus adopting the classic randomly drawing with replacement scheme that accounts for a subset of features to train each DT with dimension equal to the square root of the number of features [83–85]. Globally, 5,832 different combinations of hyperparameters have been explored for the RF algorithm with the grid-search method and, considering a stratified 10-folds cross-validation scheme, about 58,320 RF models have been actually trained. In order to select the best combination of the hyperparameters for the RF algorithm, the average of balanced accuracy among the folds has been monitored during the grid-search cross-validation procedure. All the inspected solutions are reported as grey lines within the parallel coordinate plot in Fig. 20, whereas the best combination is highlighted with a blue line. The following optimal hyperparameters for the RF algorithm have been obtained: Criterion – Gini impurity index, Max depth – 20, Max leaf nodes – 100, Min samples leaf – 4, Min samples split – 10, Number of estimators – 200.

Once the optimal hyperparameters based on the grid-search cross-validation scheme are obtained, the optimized ML models have been trained on the entire training set and the final performance metrics on the hold-out test set are listed in Tab. 9. Both models exhibit a similar balanced accuracy value almost equal to 66%. The Linear SVM still exhibits a slightly better recall value for the minority class than the

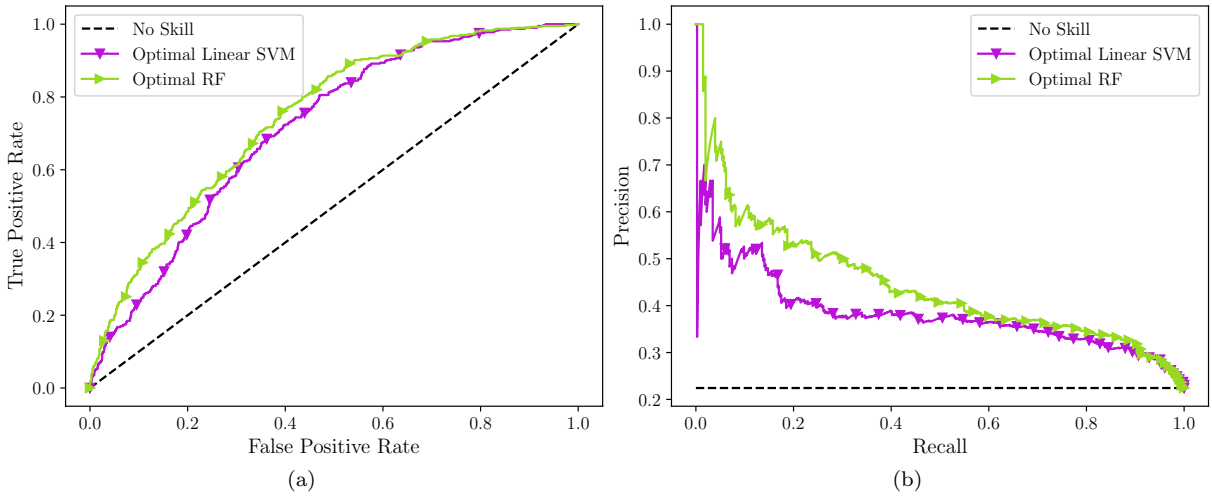


Figure 21: Comparisons of ROC curves (a) and precision-recall curves (b) for Linear SVM and RF models after hyperparameters tuning.

RF algorithm, which instead learned equally from both classes. Nevertheless, small differences between the two optimized classifiers are discernible from Fig. 21. Indeed, RF manifests slightly higher generalization capabilities both in terms of ROC curve and precision-recall curve.

In conclusion, a slightly evident superiority of the optimized RF algorithm emerges with respect to the optimized Linear SVM model. Therefore, the optimized RF model is designated as the best ML tool for addressing the current classification problem.

4.5. Uncertainty quantification based on Monte Carlo simulations

A comprehensive evaluation of the optimized RF model also requires proper consideration of the involved uncertainties. A notable concern for real-word applications is the extent to which the predictive model can effectively handle uncertainties and errors that affect both input explanatory variables and assigned classes. All sources of uncertainties are likely to affect the available database, including human errors (such as errors in data transcription, whether during the transition from in-situ observations to digital records or within paper-based documentation). It should be stressed that looking for a good classifier is as important as searching for a robust classifier. Thus, it is of utmost importance to also evaluate the robustness of the ML model against uncertainties. Nonetheless, robustness against the uncertainties has never investigated in the current studies about buildings usability prediction.

There exists some probability-based metrics that allow to evaluate the uncertainty level for general ML classifiers, for instance the cross-entropy loss or the Brier score [87]. Regrettably, these metrics do not facilitate an engineering interpretation of the effects due to uncertainties. Therefore, a Monte Carlo based

procedure is proposed to quantify the impact of uncertainties on the final building seismic class predicted through the optimized RF model. This is accomplished by means of a mutation probability that defines the probability of transitioning from state S_i to state S_j as follows:

$$P_{ij} = P\left(S_i \rightarrow S_j\right) \text{ for } i \neq j, \tag{8}$$

under the condition $|i-j|=1$ in order to ensure the transition between two adjacent states only. This condition serves at considering uncertainties and/or errors with moderate magnitude only. Tab. 10 has been compiled accordingly through personal communications with those who managed AeDES forms within the Italian Department of Civil Protection. This survey suggested that a reasonable value for the mutation probability $\hat{P}$ lies between 0.05 and 0.1. A mutation probability of $\hat{P}=10\%$ is finally selected for a conservative estimation of the model robustness. When an input variable in Tab. 10 is labelled as "Not defined", it means that no mutation occurs. This label applies to all variables that can exhibit a mutation, with the exception of the seismic class because it is always defined.

Table 10: Numerical values of the mutation probability $\hat{P}$.

No.	Variable	States (see Fig. 4)	Probability of transition
1	Building position	{1,2,3,4, Not defined}	$P_{ij} = \hat{P}$ for $\{i,j\} \in \{3,4\}$; P_{ij}=0 for $i \in \{1,2\}$
2	No. of floors	{1,2,..,13, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,...,13\}$
3	Inter-storey height	{1,2,3,4, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,...,4\}$
4	Average floor surface	{A,B..S, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in$ {A,..,S}
5	Age	{1,2,..13, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,...,13\}$
6	Typology	{1,2,3,4, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,2\}$; P_{ij}=0 $\{i,j\} \in \{3,4\}$
7	Roof	{1,2,3,4, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,...,4\}$
8	Masonry texture	{1,2, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,2\}$
9	Curbs/chains	{1,2, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,2\}$
10	Isolated pillar	{1,2, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,2\}$
11	Hybrid masonry	{1,2,3, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,...,3\}$
12	Reinforced masonry	{1,2,3, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,...,3\}$
13	Site morphology	{1,2,3,4, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,...,3\}$
14	Soil category	{1,2,3, Not defined}	P_{ij}=$\hat{P}$ for $\{i,j\} \in \{1,...,3\}$
15	Epicentral distance		
16	Fundamental period		
17	Horiz. peak ground acceleration		
18	Vert. peak ground acceleration		
19	Horiz. pseudo-spectral acceleration		Deterministic
20	Arias Intensity		
21	Housner Intensity		
22	Duration		
23	Seismic class	{AB, CE}	P_{ij}=$\hat{P}$ for $\{i,j\} \in$ {AB,CD}

Besides balanced accuracy, precision and recall, a new metric is introduced to quantify the robustness against the misclassification. It is the False Negative Rate (FNR), which is the conditional probability of a negative result given that the condition being looked for is present. Monte Carlo simulations are stopped once

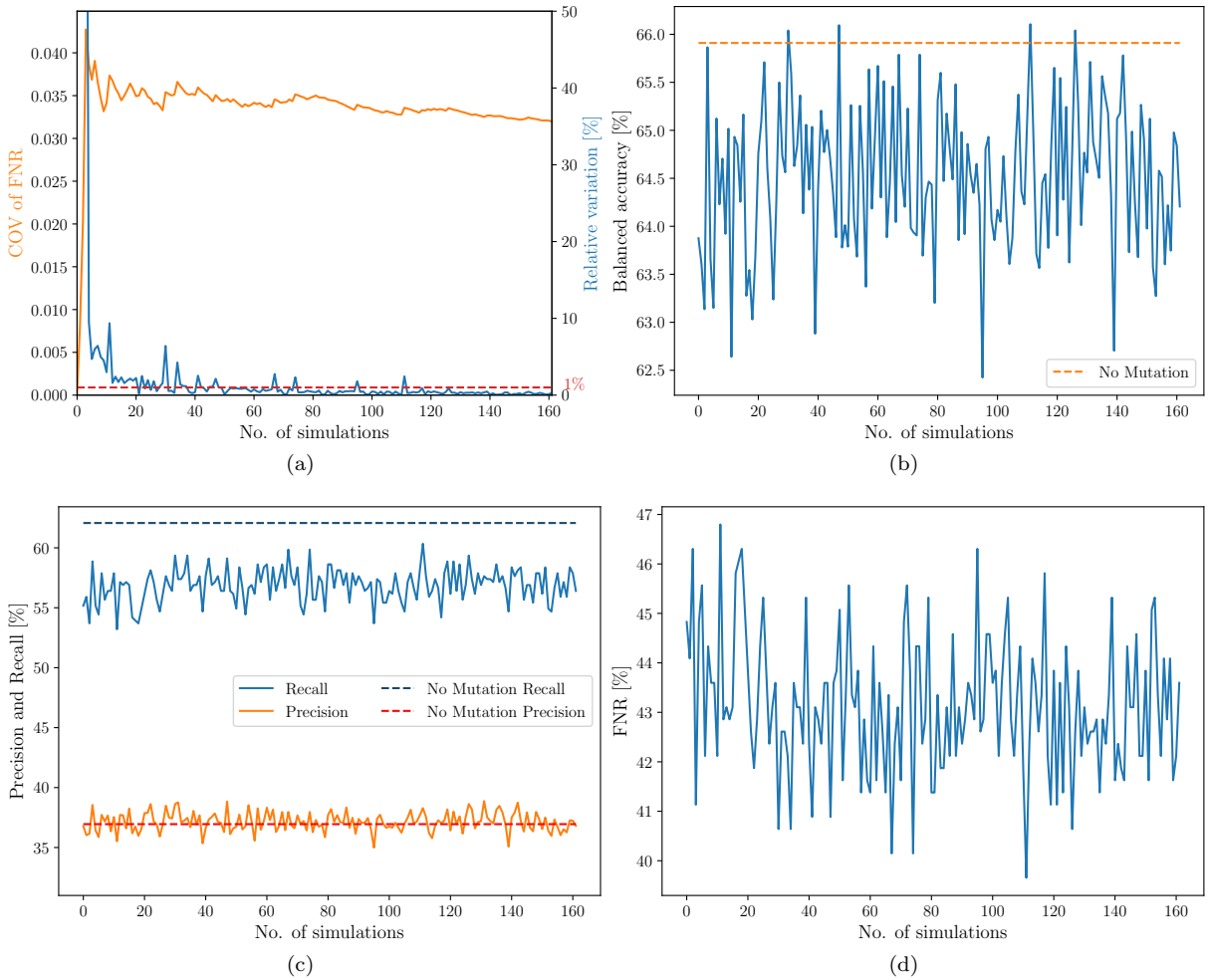


Figure 22: Coefficient of variation of the FNR (orange solid line) and its relative variation (blue solid line) (a). Effects of the uncertainties on the seismic class assignment in terms of balanced accuracy (b), precision and recall (c), and FNR (d).

the coefficient of variation (COV) of the FNR is lower than 1%. The acceptable shifting convergence band rule (ASCBR) [88, 89] determines the termination point of the Monte Carlo simulations. More precisely, it discerns the point at which the COV of the FNR stabilizes, meeting the convergence criterion and persisting for a minimum of 50 simulations.

A total of 161 simulations was completed to achieve the convergence. The main outcomes of the Monte Carlo simulations are provided in Fig. 22. The COV of the FNR is about 0.032 at the end of the Monte Carlo simulations. The average value of the balanced accuracy over all Monte Carlo simulations is 64.52%, and it is very close to the reference value of 65.72% corresponding to the actual database (i.e., without mutation). It is interesting to note in Fig. 22 that the uncertainties have slight influence on the precision, whereas their impact on the recall of the minority class is rather large. In particular, the mean value of the recall over

the Monte Carlo simulation is 56.87% and results lower than the actual value of 62.07% corresponding to the original database. The FNR shows moderate fluctuations due to the uncertainties around its average value, which is equal to 43.13%. This Monte Carlo simulations-based assessment of the optimized RF model highlights a significant robustness against the uncertainties since most of the classification metrics do not deteriorate drastically for a reasonable, yet conservative, uncertain level. However, these results also indicate that the accurate assignment to the minority class may be further compromised by the uncertainties.

4.6. Features importance

Understanding the relevance of the input variables involved in the classification process is crucial because such information can assist the optimization of the survey forms in order to achieve a high confidence on the class assignment while reducing time and cost of the examination process. Furthermore, it is useful to assist model's interpretation [90, 91].

The following three methods are herein adopted to assess the importance of the features involved within the optimized RF model:

- feature importance scores from the optimal RF model [59];
- permutation feature importance [60];
- Shapley additive explanations (SHAP) method [92].

The inherent implementation of the RF algorithm allows to compute automatically the feature importance scores by employing the impurity criterion. This provides a way to rank input features according to their relative significance in affecting the outcomes of the RF algorithm [90]. The Gini impurity scores are computed for each feature based on the average of the error function decrease at each split node. Therefore, those features exhibiting the most significant average decrease are the ones that exert a strong influence on the overall predictions of the RF model. Fig. 23a outlines the RF feature importance scores, demonstrating that the current classification problem is mainly governed by the age of the buildings, the presence of curbs/chains, the roof typology, and the masonry texture. Several seismic intensity measures – especially the vertical peak ground acceleration and the horizontal spectral acceleration – ranks among the most important explanatory variables and are more relevant than several building features. Although this scoring method is common when using a RF algorithm, some studies have emphasized that it may be biased when dealing with mixed features (i.e., both continuous and categorical) [91]. Alternative model-agnostic explainability schemes have been thus proposed in the recent years.

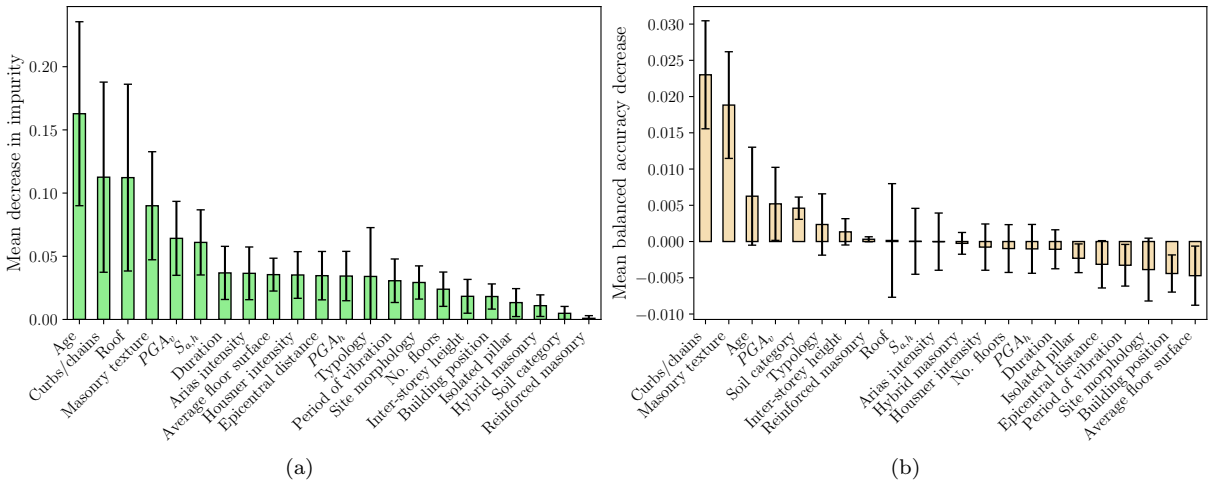


Figure 23: Feature importance scores for RF model based on Gini impurity decrease (a) and permutation feature importance for balanced accuracy decrease (b).

The permutation feature importance method was originally proposed by Breiman et al. [59, 60] and is a model-agnostic explainability method based on decoupling the relationship established between a feature column and their outcome label by means of a random shuffling of the entire feature column. This process is carried out individually for each feature column by monitoring the variation of a certain control metric with respect to its baseline value without any shuffling. This process is repeated several times, and the average decrease in the control metric for each feature is computed. Those features characterized by the highest decrease values of the control metric represent the most relevant ones. Since this study deals with a strongly imbalanced database, the balanced accuracy is selected as control metric. A total number of 100 repetitions is considered [90]. The feature ranking according to the permutation method for the hold-out test set is reported in Fig. 23b. This model-agnostic approach confirms that building age, curbs/chains and masonry texture are the most important building features whereas the vertical peak ground acceleration emerges once again as an important seismic intensity measure for the present case study. It is worth noting that permutation is still a global approach because the feature column shuffling affects the entire dataset.

A rigorous model-agnostic explainability approach is represented by the SHAP method [92]. It is a local explanation tool with a mathematical basis grounded in the cooperative game theory [90, 92]. Basically, the method uses an interpretable approximation of the original model to explain the expected prediction of every single input (local property), thereby retrieving the importance of each feature, which is assumed independent, according to the additive feature attribution principle [92]. The results of the SHAP interpretation are depicted in Fig. 24. Specifically, Fig. 24 provides a global summary of the mean SHAP values in order to

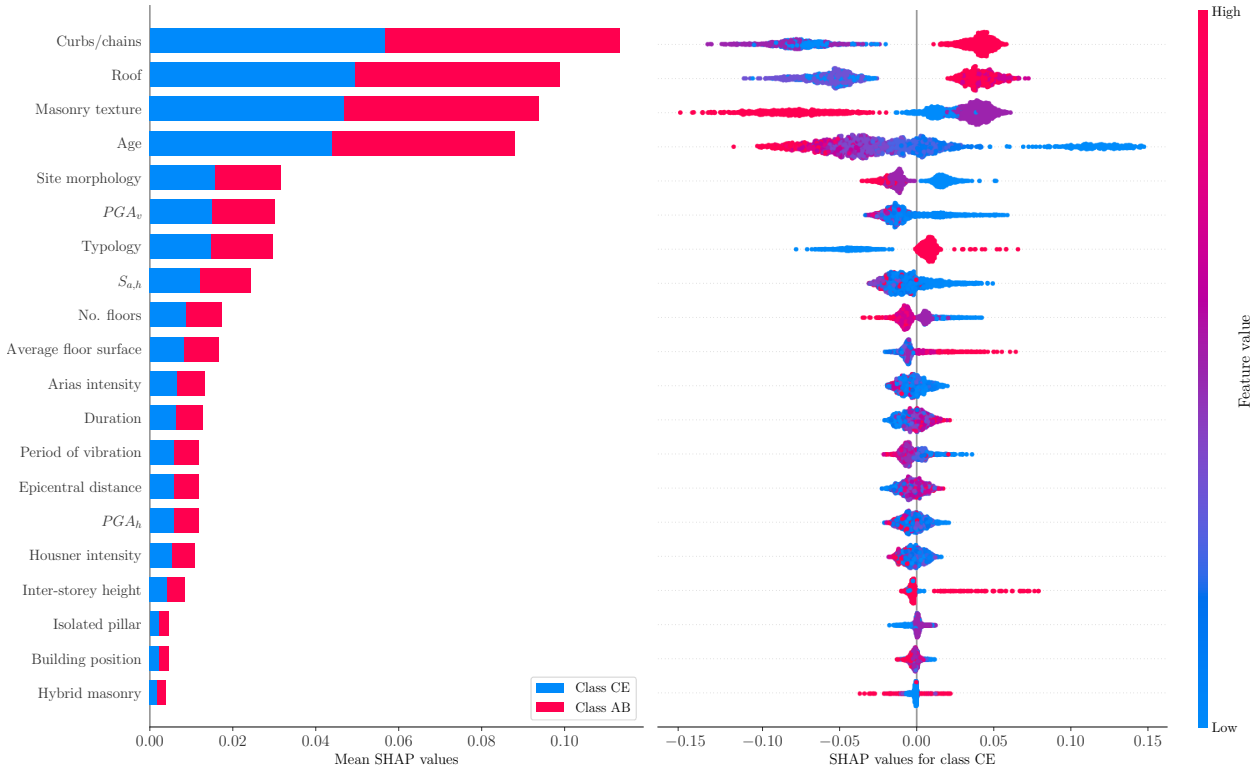


Figure 24: SHAP interpretability method for the optimized RF classifier: SHAP values summary graph (a) and beeswarm SHAP values plot for the minority class (b).

evaluate the average impact on the RF model of every feature affecting the predictions for both classes. The outcomes of this plot are somehow consistent with those in Fig. 23. In fact, the presence of curbs/chains, roof topology, masonry texture and age emerges as the most important building features. Among the seismic intensity measures, the vertical peak ground acceleration and horizontal pseudo-acceleration rank once again among the most important features. Furthermore, Fig. 24 shows the SHAP values for the minority class only in the format of a beeswarm graph. It illustrates how the values of each individual feature either positively or negatively contribute to the prediction of the minority class assignment, respectively. Fig. 24 corroborates the physical consistency of the classifier. For instance, it can be noted in Fig. 24 that the lack of curbs/chains, pushing and heavy roofs, and non-regular masonry textures are detrimental for post-earthquake buildings usability. Since old buildings are typically designed and built without fulfilling stringent requirements and specific prescriptions about seismic loads, they are also likely to be unusable in the aftermath of an earthquake. It is also evident from Fig. 24 that large values of vertical peak ground acceleration and horizontal pseudo-acceleration lower the chance of post-earthquake buildings usability.

There is a general agreement among all these results plotted in Figs. 23-24 about the most important

building features. They are mainly associated to masonry buildings, also because they are by far the most representative construction system in the database. The most important masonry building features according to the ML model align well with the observed performance. Extensive technical reports based on field surveys within the municipalities of Norcia and Amatrice confirm that the main causes of the observed damage for masonry buildings are the poor masonry texture, the lack of proper connections between walls and/or with slabs, the presence of heavy roofs or concrete curbs, missing of adequate retrofitting systems such as tie rods and chains. All these factors have prompted buildings unusability and structural failures, especially due to out-of-plane mechanisms [34, 93, 94]. Figures 23-24 also confirm that some features of the seismic ground motion are more important than several buildings' attributes. This emphasizes the need of taking into account the seismic intensity when predicting the post-earthquake usability of buildings. Particularly, the vertical component of the earthquake seems somehow important to predict the buildings usability for the considered earthquake-prone region. This is in agreement with existing studies, which have highlighted an exceptional vertical seismic ground motion during the 2016–2017 Central Italy earthquakes and its damaging effect on buildings [95].

4.7. Final remarks

The analysis carried out so far rests on a specific clustering of the consequence classes into the AeDES form. Although suitable motivations support the adopted clustering, a different criterion can lead to a diverse grouping strategy. In order to provide some insights about this aspect, an alternative clustering strategy for the consequence classes into the AeDES form is now considered. It results into a cascade binary classification, where class E is first recognized within the database (the remaining classes are then collected into a single class ABC), and class A is next recognized from the remaining classes (which are merged into a single class BC) [19]. All ML techniques have been employed for this task. Neither PCA nor SMOTE are applied, whereas the stratified cross-validation scheme is implemented.

Tab. 11 and Fig. 25 show that, even after changing the classification strategy, the best models are once again Linear SVM, RF and DT. The results related to the first step of this alternative classification approach (i.e., class E versus class ABC) are similar to those obtained previously. Fairly better results are obtained in the second step (i.e., class A versus class BC). These results also confirm the robustness of the obtained findings against variations in the adopted clustering approach and classification strategy.

Table 11: Comparisons of hold-out test metrics using a cascade binary classification scheme.

Method	Metric	E vs ABC		A vs BC	
		Class E	Class ABC	Class A	Class BC
KNN	Acc.	79.72		69.06	

Table 11: (continued from previous page)

Method	Metric	E vs ABC		A vs BC	
		Class E	Class ABC	Class A	Class BC
	Bal. acc.	55.04		55.51	
	Precision	83.08	38.69	73.91	43.10
	Recall	94.30	15.77	87.43	23.58
	F1-score	88.34	22.41	80.10	30.49
	ROC AUC	65.68		60.11	
Linear SVM	Acc.	62.43		59.63	
	Bal. acc.	67.28		60.91	
	Precision	91.27	29.72	79.89	38.01
	Recall	59.57	75.00	57.90	63.92
	F1-score	72.09	42.57	67.15	47.67
	ROC AUC	73.00		64.66	
RBF SVM	Acc.	70.44		59.91	
	Bal. acc.	62.55		57.30	
	Precision	86.82	31.40	76.29	36.11
	Recall	75.10	50.00	63.43	51.18
	F1-score	80.54	38.58	69.27	42.34
	ROC AUC	69.74		62.38	
DT	Acc.	66.80		57.19	
	Bal. acc.	66.75		58.98	
	Precision	89.79	31.42	78.66	36.07
	Recall	66.82	66.67	54.76	63.21
	F1-score	76.62	42.71	64.57	45.93
	ROC AUC	71.61		62.04	
RF	Acc.	66.57		60.79	
	Bal. acc.	67.07		60.24	
	Precision	90.05	31.45	78.78	38.23
	Recall	66.28	67.86	61.52	58.96
	F1-score	76.36	42.98	69.09	46.38
	ROC AUC	74.39		64.87	
Neural Net	Acc.	81.10		70.90	
	Bal. acc.	52.90		52.71	
	Precision	82.34	45.00	72.42	47.19
	Recall	97.76	8.04	95.52	9.91
	F1-score	89.39	13.64	82.38	16.37
	ROC AUC	74.36		65.50	
AdaBoost	Acc.	81.66		71.57	
	Bal. acc.	54.50		53.26	
	Precision	82.85	52.78	72.65	53.09
	Recall	97.69	11.31	96.38	10.14
	F1-score	89.66	18.63	82.85	17.03
	ROC AUC	74.16		66.54	
Naive Bayes	Acc.	66.02		64.99	
	Bal. acc.	65.01		58.34	
	Precision	88.87	30.21	76.18	39.87
	Recall	66.62	63.39	74.00	42.69
	F1-score	76.15	40.92	75.07	41.23
	ROC AUC	70.88		60.45	
QDA	Acc.	74.59		66.69	
	Bal. acc.	59.01		59.82	
	Precision	84.82	32.49	76.95	42.33
	Recall	83.79	34.23	76.00	43.63
	F1-score	84.30	33.33	76.47	42.97
	ROC AUC	71.19		64.07	
LR	Acc.	81.60		71.30	
	Bal. acc.	51.60		52.02	
	Precision	81.93	56.52	72.09	50.91
	Recall	99.32	3.87	97.43	6.60
	F1-score	89.79	7.24	82.87	11.69
	ROC AUC	73.20		65.08	
LDA	Acc.	81.60		71.10	
	Bal. acc.	51.48		52.30	

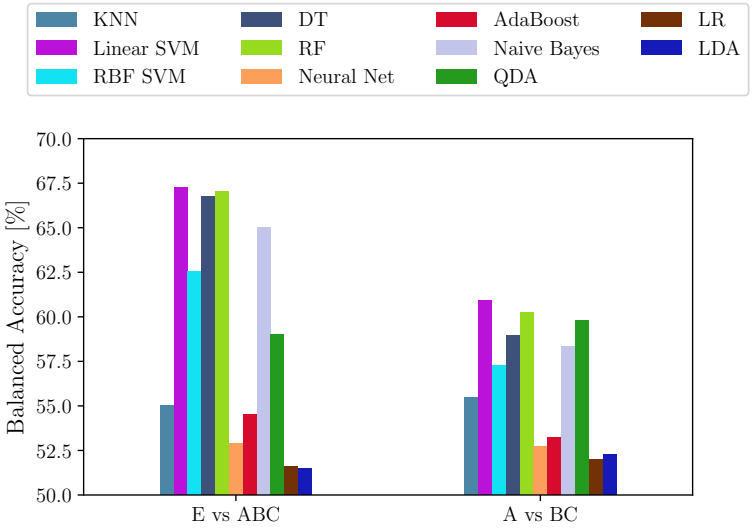


Figure 25: Comparisons of balanced accuracy metrics using a cascade binary classification scheme.

Table 11: (continued from previous page)

Method	Metric	E vs ABC		A vs BC	
		Class E	Class ABC	Class A	Class BC
	Precision	81.89	57.14	72.22	48.57
	Recall	99.39	3.57	96.57	8.02
	F1-score	89.79	6.72	82.64	13.77
	ROC AUC	72.65		64.71	

The last remark is about the suitability of the ML model to predict the post-seismic building usability, which performance in terms of balanced accuracy is close to about 70%.

Fig. 26 provides a comprehensive view at territorial scale of actual usability data and the predictions obtained by means of the optimized RF algorithm. This way of presenting the results flattens the degree of information provided by the confusion matrices and thus should be not considered to assess the classification. Nonetheless, it is a valuable output for the estimation of the economic resources needed to enhance the preparedness of the communities against future earthquakes through seismic retrofitting intervetions as well as for immediate losses quantification in the aftermath of an earthquake. Despite the different grouping strategies of usability data, Fig. 26 highlights a generalized conservative behavior of the optimized RF algorithm, since it tends to more safely estimate the minority unusable class.

Taking every aspect into careful consideration, the selected ML model is deemed adequate for the targeted applications, such as the rapid estimation of the consequences due to an earthquake at territorial scale (e.g., number of displaced people and unusable buildings, number of buildings that require emergency interventions and corresponding repair costs) as well as the preliminary planning of seismic risk mitigation policies (e.g., technical and economical analysis of measures to boost seismic retrofitting of buildings).

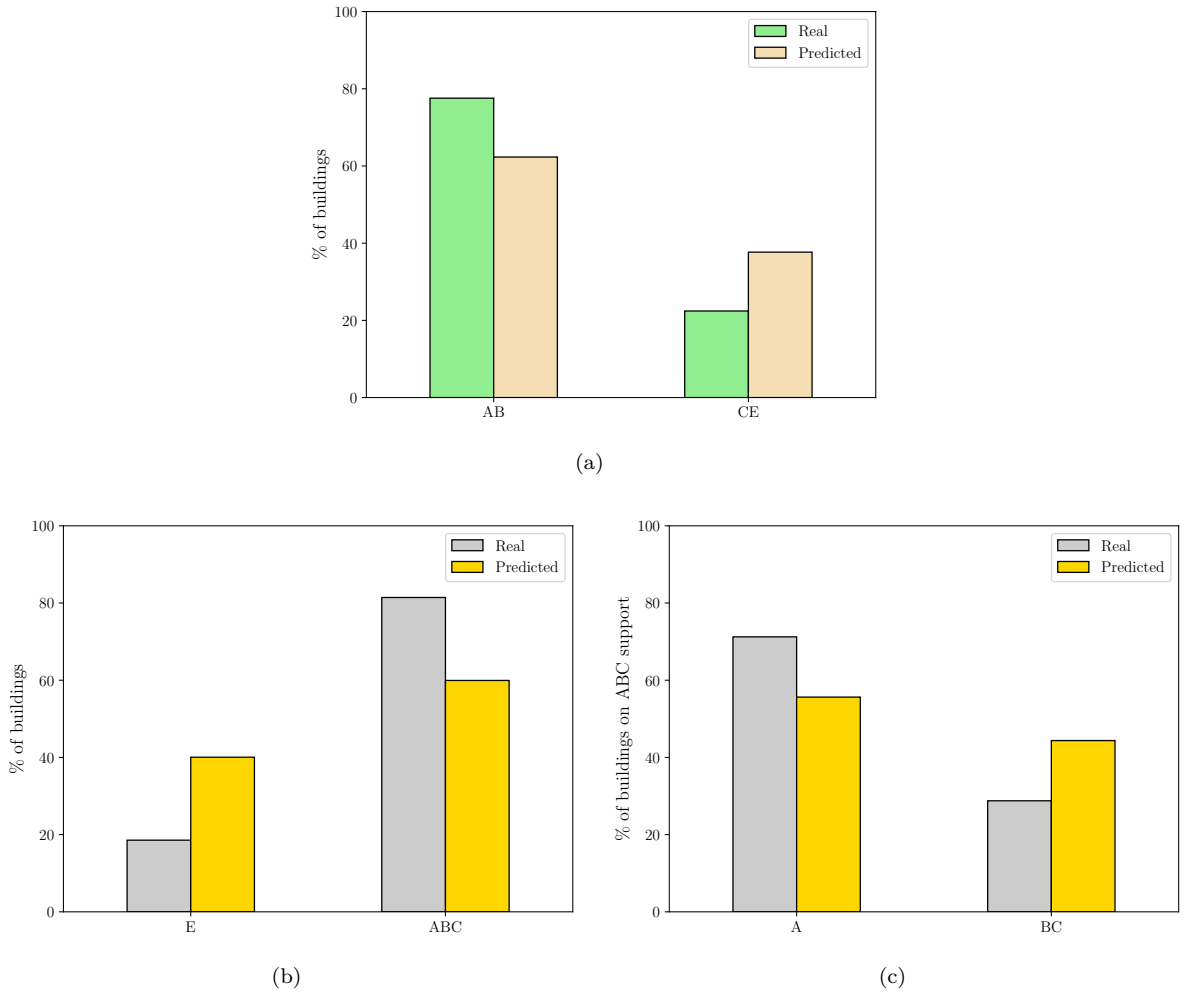


Figure 26: Comparison between actual usability data and the predictions made by the optimized RF algorithm considering the reference binary classification strategy AB vs CE (a) or the alternative cascade classification strategy E vs ABC (b) and A vs BC (c).

For the sake of completeness, the main issues and limitations about the use of ML techniques for post-seismic usability assessment deserve further considerations. The main concerns deal with the reliability and applicability of the final ML models as well as the transferability of their performances to future seismic events. It is too evident that the validity of all supervised ML methods depends on the trustworthiness of the labeled dataset used in the training phase, which also bounds the application domain. The careful analysis of buildings and earthquakes data as well as the comprehensive assessment of different preprocessing techniques, algorithms, and metrics play a pivotal role in securing the generalization capability of the predictive models for future applications, while uncertainty quantification, models interpretation, and sensitivity analysis for data labeling serve at determining the corresponding confidence level. The use of building data spread over

a very large area is useful, to some extent, for ensuring sufficient diversity throughout the training stage and for widening the application range. However, unsatisfactory results are likely to be obtained when predictive data-driven models are applied to areas where building characteristics are largely different from those considered for training. In particular, the correctness of the post-seismic building usability will reduce for areas with a dissimilar distribution of the buildings typology. The rate of non-conservative predictions is also expected to grow when such data-driven models are applied to areas where the attributes of weak buildings significantly diverge from those adopted for training, since they are an underrepresented group within the full dataset.

5. Conclusions

This study has tackled the challenge of forecasting the post-seismic usability of buildings at regional scale by means of machine learning (ML) models, with focus on areas prone to frequent earthquakes. The case study of the 2016-2017 Central Italy earthquakes has been considered. A comparative evaluations among various ML algorthms and preprocessing techniques based on a large suite of classification metrics has been performed. Suitable tools to quantify the features importance and to interpret the predictive models are also considered. Results reported in the present study confirm that ML methods can be a viable decision-making support to enhance the preparedness of the built environment against earthquakes and to speed up post-seismic recovery. However, the appropriate development of ML models is crucial for obtaining proper results, and their unbiased examination from different standpoints is of utmost importance to secure a correct implementation in real-word applications. In this perspective, the detailed presentation of the adopted workflow also served at outlining useful guidelines.

It has been pointed out in the present study that the accuracy is not the most appropriate performance metric for such kind of applications because of the availability of imbalanced datasets. Instead, the number of false positives and negatives should be equally distributed and minimized across the post-seismic usability classes. Taking this into account, it has been found that Linear Support Vector Machine and Random Forest yield satisfactory outcomes while exhibiting good robustness against errors in the training dataset and clustering strategy. Random Forest was designated as the most suitable classifier, and the corresponding value of the balanced accuracy value after hyperparameters fine-tuning was found close to 70%.

Acknowledgments

Marco Martino Rosso and Giuseppe Quaranta acknowledge the support received through the project 'Artificial Intelligence for SUstainable seismic risk reduction of STructures (AI-SUST)" (project code: 2022LEFKHS)

funded by European Union – NextGeneration EU through the PRIN 2022 program of the Italian Ministry of University and Research (MUR) (D. D. n. 104, 02-02-2022). This work reflects only the authors' views and opinions whereas the MUR cannot be considered responsible for them.

Declarations of competing interest

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

Availability of data and material

All data, models, and codes supporting this study's findings are confidential for legal reasons.

References

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

[2] D. D'ayala, A. Meslem, D. Vamvatsikos, K. Porter, T. Rossetto, H. Crowley, V. Silva, Guidelines for analytical vulnerability assessment of low/mid-rise buildings—methodology. vulnerability global component project (2014).

[3] F. da Porto, M. Donà, A. Rosti, M. Rota, S. Lagomarsino, S. Cattari, B. Borzi, M. Onida, D. De Gregorio, F. L. Perelli, et al., Comparative analysis of the fragility curves for italian residential masonry and rc buildings, Bulletin of Earthquake Engineering 19 (8) (2021) 3209–3252.

[4] C. Del Gaudio, P. Ricci, G. M. Verderame, G. Manfredi, Observed and predicted earthquake damage scenarios: the case study of pettino (l'aquila) after the 6th april 2009 event, Bulletin of Earthquake Engineering 14 (2016) 2643–2678.

[5] M. Faravelli, B. Borzi, D. Polli, M. Pagano, Calibration of a mechanics-based method for large-scale vulnerability assessment, Bulletin of Earthquake Engineering 17 (2019) 2485–2508.

[6] R. Rahmat Rabi, V. Bianco, G. Monti, Mechanical-analytical soil-dependent fragility curves of existing rc frames with column-driven failures, Buildings 11 (7) (2021) 278.

[7] R. Gentile, C. Galasso, Simplicity versus accuracy trade-off in estimating seismic fragility of existing reinforced concrete buildings, Soil Dynamics and Earthquake Engineering 144 (2021) 106678.

[8] S. Ruggieri, M. Calò, A. Cardellicchio, G. Uva, Analytical-mechanical based framework for seismic overall fragility analysis of existing rc buildings in town compartments, Bulletin of Earthquake Engineering 20 (15) (2022) 8179–8216.

[9] E. Cosenza, G. Manfredi, M. Polese, G. M. Verderame, A multilevel approach to the capacity assessment of existing rc buildings, Journal of earthquake engineering 9 (01) (2005) 1–22.

[10] S. Lagomarsino, S. Cattari, Fragility functions of masonry buildings, in: SYNER-G: typology definition and fragility functions for physical elements at seismic risk: buildings, lifelines, transportation networks and critical facilities, Springer, 2013, pp. 111–156.

[11] 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, Bulletin of Earthquake Engineering 16 (2018) 4581–4605.

[12] A. Masi, L. Chiauzzi, G. Santarsiero, V. Manfredi, S. Biondi, E. Spacone, C. Del Gaudio, P. Ricci, G. Manfredi, G. M. Verderame, Seismic response of rc buildings during the m w 6.0 august 24, 2016 central italy earthquake: The amatrice case study, Bulletin of Earthquake Engineering 17 (2019) 5631–5654.

[13] M. Donà, P. Carpanese, V. Follador, L. Sbrogiò, F. da Porto, Mechanics-based fragility curves for italian residential urm buildings, Bulletin of Earthquake Engineering 19 (2021) 3099–3127.

[14] M. Zucconi, L. Sorrentino, R. Ferlito, Principal component analysis for a seismic usability model of unreinforced masonry buildings, Soil Dynamics and Earthquake Engineering 96 (2017) 64–75.

[15] M. Zucconi, R. Ferlito, L. Sorrentino, Validation and extension of a statistical usability model for unreinforced masonry buildings with different ground motion intensity measures, Bulletin of earthquake engineering 18 (2020) 767–795.

[16] M. Zucconi, M. Di Ludovico, L. Sorrentino, Census-based typological usability fragility curves for italian unreinforced masonry buildings, Bulletin of Earthquake Engineering 20 (8) (2022) 4097–4116.

[17] S. Bertelli, T. Rossetto, I. Ioannou, Derivation of empirical fragility functions from the 2009 l'aquila earthquake, in: Prooceedings 16th European Conference on Earthquake Engineering, Vol. 16, European Association of Earthquake Engineering, 2018.

[18] S. Kameshwar, S. Misra, J. E. Padgett, Decision tree based bridge restoration models for extreme event performance assessment of regional road networks, Structure and Infrastructure Engineering 16 (3) (2020) 431–451.

[19] G. Tocchi, S. Misra, J. E. Padgett, M. Polese, M. Di Ludovico, The use of machine-learning methods for post-earthquake building usability assessment: A predictive model for seismic-risk impact analyses, International Journal of Disaster Risk Reduction 97 (2023) 104033. doi:https://doi.org/10.1016/j.ijdrr.2023.104033.

[20] A. Fiore, F. Mollaioli, G. Quaranta, G. C. Marano, Seismic response prediction of reinforced concrete buildings through nonlinear combinations of intensity measures, Bulletin of Earthquake Engineering 16 (2018) 6047–6076.

[21] Q. Tang, J. Dang, Y. Cui, X. Wang, J. Jia, Machine learning-based fast seismic risk assessment of building structures, Journal of Earthquake Engineering 26 (15) (2022) 8041–8062.

[22] P. Kourehpaz, C. Molina Hutt, Machine learning for enhanced regional seismic risk assessments, Journal of Structural Engineering 148 (9) (2022) 04022126.

[23] K. Demertzis, K. Kostinakis, K. Morfidis, L. Iliadis, An interpretable machine learning method for the prediction of r/c buildings' seismic response, Journal of Building Engineering 63 (2023) 105493.

[24] S.-H. Hwang, S. Mangalathu, J. Shin, J.-S. Jeon, Machine learning-based approaches for seismic demand and collapse of ductile reinforced concrete building frames, Journal of Building Engineering 34 (2021) 101905.

[25] H. Zhang, X. Cheng, Y. Li, D. He, X. Du, Rapid seismic damage state assessment of rc frames using machine learning methods, Journal of Building Engineering 65 (2023) 105797.

[26] E. Junda, C. Málaga-Chuquitaype, K. Chawgien, Interpretable machine learning models for the estimation of seismic drifts in clt buildings, Journal of Building Engineering 70 (2023) 106365.

[27] A. Aloisio, M. M. Rosso, A. M. De Leo, M. Fragiacomo, M. Basi, Damage classification after the 2009 l'aquila earthquake using multinomial logistic regression and neural networks, International Journal of Disaster Risk Reduction 96 (2023) 103959.

[28] F. Galadini, E. Falcucci, S. Gori, P. Zimmaro, D. Cheloni, J. P. Stewart, Active faulting in source region of 2016–2017 central italy event sequence, Earthquake Spectra 34 (4) (2018) 1557–1583.

[29] M. R. Barchi, F. Carboni, M. Michele, M. Ercoli, C. Giorgetti, M. Porreca, S. Azzaro, L. Chiaraluce, The influence of

subsurface geology on the distribution of earthquakes during the 2016—2017 central italy seismic sequence, Tectonophysics 807 (2021) 228797.
[30] C. Chiarabba, P. De Gori, M. Cattaneo, D. Spallarossa, M. Segou, Faults geometry and the role of fluids in the 2016–2017 central italy seismic sequence, Geophysical Research Letters 45 (14) (2018) 6963–6971.
[31] G. Calderoni, A. Rovelli, R. Di Giovambattista, Rupture directivity of the strongest 2016–2017 central italy earthquakes, Journal of Geophysical Research: Solid Earth 122 (11) (2017) 9118–9131.
[32] G. Grelle, L. Bonito, M. Rosalba, S. Iacurto, C. Madiai, P. Revellino, G. Sappa, Topographic effects observed at amatrice hill during the 2016–2017 central italy seismic sequence, Earthquake Engineering and Engineering Vibration 20 (2021) 63–78.
[33] L. Chiaraluce, M. Michele, F. Waldhauser, Y. J. Tan, M. Herrmann, D. Spallarossa, G. C. Beroza, M. Cattaneo, C. Chiarabba, P. De Gori, et al., A comprehensive suite of earthquake catalogues for the 2016-2017 central italy seismic sequence, Scientific Data 9 (1) (2022) 710.
[34] F. Mollaioli, O. AlShawa, L. Liberatore, D. Liberatore, L. Sorrentino, Seismic demand of the 2016–2017 central italy earthquakes, Bulletin of earthquake engineering 17 (10) (2019) 5399–5427.
[35] L. Sorrentino, S. Cattari, F. Da Porto, G. Magenes, A. Penna, Seismic behaviour of ordinary masonry buildings during the 2016 central italy earthquakes, Bulletin of Earthquake Engineering 17 (10) (2019) 5583–5607.
[36] M. Di Ludovico, A. Digrisolo, C. Moroni, F. Graziotti, V. Manfredi, A. Prota, M. Dolce, G. Manfredi, Remarks on damage and response of school buildings after the central italy earthquake sequence, Bulletin of earthquake engineering 17 (2019) 5679–5700.
[37] A. Penna, C. Calderini, L. Sorrentino, C. F. Carocci, E. Cescatti, R. Sisti, A. Borri, C. Modena, A. Prota, Damage to churches in the 2016 central italy earthquakes, Bulletin of Earthquake Engineering 17 (2019) 5763–5790.
[38] G. Santarsiero, L. Di Sarno, S. Giovinazzi, A. Masi, E. Cosenza, S. Biondi, Performance of the healthcare facilities during the 2016–2017 central italy seismic sequence, Bulletin of earthquake engineering 17 (2019) 5701–5727.
[39] M. G. Durante, L. Di Sarno, P. Zimmaro, J. P. Stewart, Damage to roadway infrastructure from 2016 central italy earthquake sequence, Earthquake spectra 34 (4) (2018) 1721–1737.
[40] 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), Field Manual for Post-earthquake Damage and Safety Assessment and Short Term Countermeasures (AeDES)Cited By 189 (2007).
[41] G. Grünthal, European macroseismic scale 1998 (ems-98) (1998).
[42] J. K. Kalita, D. K. Bhattacharyya, S. Roy, Fundamentals of Data Science: Theory and Practice, Elsevier, 2023.
[43] S. K. Mukhiya, U. Ahmed, Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand, summarize, and investigate your data, Packt Publishing Ltd, 2020.
[44] P. Llukan, Kendall's tau, International Encyclopedia of Statistical Science (2011) 713–715doi:https://doi.org/10.1007/978-3-642-04898-2_324.
[45] F. Sabetta, A. Pugliese, Estimation of response spectra and simulation of nonstationary earthquake ground motions, Bulletin of the Seismological Society of America 86 (2) (1996) 337–352.
[46] F. Pacor, L. Luzi, R. Puglia, M. D'Amico, D. Bindi, D2.3 - strong motion parameters of selected events (update the national database itaca v1.1), revised gmpe (07 2013). doi:10.13140/RG.2.1.1898.7365.
[47] Design of structures for earthquakes resistance part 1: General rules, seismic actions and rules for buildings. 2003, pr-en 1998-1 final draft, Comité Européen De NormalisationCited By 1 (2003).
[48] G. Forte, E. Chioccarelli, M. De Falco, P. Cito, A. Santo, I. Iervolino, Seismic soil classification of italy based on surface geology and shear-wave velocity measurements, Soil Dynamics and Earthquake Engineering 122 (2019) 79–93.
[49] M. R. Gallipoli, B. Petrovic, G. Calamita, N. Tragni, C. Scaini, C. Barnaba, M. Vona, S. Parolai, Towards specific t–h relationships: Fribas database for better characterization of rc and urm buildings, Bulletin of Earthquake Engineering (2023) 1–27.
[50] N. 2008, Decreto ministeriale 14/1/2008: Norme tecniche per le costruzioni, Ministry of Infrastructures and Transportations, GUSO n. 30, 4/2/2008 [in Italian] (2008).
[51] H. He, Y. Ma, Imbalanced learning: foundations, algorithms, and applications, John Wiley & Sons, 2013. doi:https://doi.org/10.1002/9781118646106.
[52] P. Cerda, G. Varoquaux, Encoding high-cardinality string categorical variables, IEEE Transactions on Knowledge and Data Engineering 34 (3) (2020) 1164–1176. doi:https://doi.org/10.1109/TKDE.2020.2992529.
[53] K. Takayama, Encoding categorical variables with ambiguity, in: Proceedings of the International Workshop NFMCP in conjunction with ECML-PKDD, Tokyo, Japan, Vol. 16, 2019.
[54] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in python, the Journal of machine Learning research 12 (2011) 2825–2830.
[55] E. Fix, J. L. Hodges, Discriminatory analysis: Nonparametric discrimination: Small sample performance (1952).
[56] T. Hastie, R. Tibshirani, J. H. Friedman, J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction, Vol. 2, Springer, 2009.
[57] V. N. Vapnik, The nature of statistical learning theory, Springer Science & Business Media, 1995.
[58] N. Cristianini, J. Shawe-Taylor, An introduction to support vector machines and other kernel-based learning methods, Cambridge university press, 2000.
[59] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, Classification and regression trees, CRC press, 1984.
[60] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32.

[61] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[62] G. Quaranta, W. Lacarbonara, S. F. Masri, A review on computational intelligence for identification of nonlinear dynamical systems, Nonlinear Dynamics 99 (2) (2020) 1709–1761.
[63] Y. Freund, R. E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of computer and system sciences 55 (1) (1997) 119–139.
[64] A. McCallum, K. Nigam, et al., A comparison of event models for naive bayes text classification, in: AAAI-98 workshop on learning for text categorization, Vol. 752, Madison, WI, 1998, pp. 41–48.
[65] C. M. Bishop, N. M. Nasrabadi, Pattern recognition and machine learning, Vol. 4, Springer, 2006.
[66] The origins of logistic regression (2002).
[67] R. A. Fisher, The use of multiple measurements in taxonomic problems, Annals of eugenics 7 (2) (1936) 179–188.
[68] M. Stone, Cross-validatory choice and assessment of statistical predictions, Journal of the royal statistical society: Series B (Methodological) 36 (2) (1974) 111–133.
[69] R. Kohavi, et al., A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Ijcai, Vol. 14, Montreal, Canada, 1995, pp. 1137–1145.
[70] A. Cernezel, I. Rozman, B. Brumen, Comparisons between three cross-validation methods for measuring learners' performances., in: EJC, 2014, pp. 77–87.
[71] G. Vanwinckelen, H. Blockeel, On estimating model accuracy with repeated cross-validation, in: Proceedings of the 21st Belgian-Dutch Conference on Machine Learning, 2012, pp. 39–44.
[72] H. He, E. A. Garcia, Learning from imbalanced data, IEEE Transactions on knowledge and data engineering 21 (9) (2009) 1263–1284.
[73] D. M. Powers, Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, arXiv preprint arXiv:2010.16061 (2020).
[74] Y. Liu, Y. Li, D. Xie, Implications of imbalanced datasets for empirical roc-auc estimation in binary classification tasks, Journal of Statistical Computation and Simulation (2023) 1–21.
[75] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets, PloS one 10 (3) (2015) e0118432.
[76] G. Bonaccorso, Mastering Machine Learning Algorithms: Expert techniques for implementing popular machine learning algorithms, fine-tuning your models, and understanding how they work, Packt Publishing Ltd, 2020.
[77] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometrics and intelligent laboratory systems 2 (1-3) (1987) 37–52.
[78] J. Shlens, A tutorial on principal component analysis, arXiv preprint arXiv:1404.1100 (2014).
[79] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, Smote: synthetic minority over-sampling technique, Journal of artificial intelligence research 16 (2002) 321–357.
[80] H. Han, W.-Y. Wang, B.-H. Mao, Borderline-smote: a new over-sampling method in imbalanced data sets learning, in: International conference on intelligent computing, Springer, 2005, pp. 878–887.
[81] C. Bunkhumpornpat, K. Sinapiromsaran, C. Lursinsap, Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem, in: Advances in Knowledge Discovery and Data Mining: 13th Pacific-Asia Conference, PAKDD 2009 Bangkok, Thailand, April 27-30, 2009 Proceedings 13, Springer, 2009, pp. 475–482.
[82] C. Bunkhumpornpat, K. Sinapiromsaran, C. Lursinsap, Dbsmote: density-based synthetic minority over-sampling technique, Applied Intelligence 36 (2012) 664–684.
[83] A. Géron, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow, " O'Reilly Media, Inc.", 2022.
[84] S. Raschka, V. Mirjalili, Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2, Packt Publishing Ltd, 2019.
[85] P. Probst, M. N. Wright, A.-L. Boulesteix, Hyperparameters and tuning strategies for random forest, Wiley Interdisciplinary Reviews: data mining and knowledge discovery 9 (3) (2019) e1301.
[86] L. Breiman, Classification and regression trees, Routledge, 2017.
[87] C. Ferri, J. Hernández-Orallo, R. Modroiu, An experimental comparison of performance measures for classification, Pattern recognition letters 30 (1) (2009) 27–38.
[88] M. Y. Ata, A convergence criterion for the monte carlo estimates, Simulation Modelling Practice and Theory 15 (3) (2007) 237–246.
[89] P.-H. Chan, J. Stebbins, A. B. Zavatsky, Efficacy of quantifying marker-cluster rigidity in a multi-segment foot model: a monte-carlo based global sensitivity analysis and regression model, Computer Methods in Biomechanics and Biomedical Engineering 25 (3) (2022) 308–319.
[90] A. Orlenko, J. H. Moore, A comparison of methods for interpreting random forest models of genetic association in the presence of non-additive interactions, BioData mining 14 (1) (2021) 1–17.
[91] C. Strobl, A.-L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: Illustrations, sources and a solution, BMC bioinformatics 8 (1) (2007) 1–21.
[92] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in neural information processing systems 30 (2017).
[93] G. Fiorentino, A. Forte, E. Pagano, F. Sabetta, C. Baggio, D. Lavorato, C. Nuti, S. Santini, Damage patterns in the town of amatrice after august 24th 2016 central italy earthquakes, Bulletin of Earthquake Engineering 16 (2018) 1399–1423.
[94] M. Acito, M. Buzzetti, C. Chesi, E. Magrinelli, G. Milani, Failures and damages of historical masonry structures induced by 2012 northern and 2016–17 central italy seismic sequences: Critical issues and new perspectives towards seismic prevention, Engineering Failure Analysis 149 (2023) 107257.
[95] L. Luzi, F. Pacor, R. Puglia, G. Lanzano, C. Felicetta, M. D'Amico, A. Michelini, L. Faenza, V. Lauciani, I. Iervolino,

et al., The central italy seismic sequence between august and december 2016: Analysis of strong-motion observations, Seismological Research Letters 88 (5) (2017) 1219–1231.