



OPEN

# Combined machine learning - 3D physics based approach for building damage evaluation: the case of L'Aquila 2009

F. Di Michele<sup>1,3</sup>✉, D. Pera<sup>2</sup>✉, I. Mazzieri<sup>6</sup>, C. Smerzini<sup>5</sup>, E. Stagnini<sup>3</sup>, V. Kastelic<sup>4</sup>, M. M. C. Carafa<sup>4</sup>, R. Aloisio<sup>3</sup>, B. Rubino<sup>2</sup> & P. Marcati<sup>3</sup>

This work combines artificial intelligence tools and physics-based simulations with the aim to develop a hybrid tool to support pre- and post-earthquake emergency phases. As a case study, we examine the 6.1 Mw earthquake that struck and partially destroyed the city of L'Aquila and many surrounding villages on 6 April 2009. Numerical simulations were performed using SPEED (SPectral Elements in Elastodynamics with Discontinuous Galerkin), which is a spectral element based software, for earthquake simulation on local and regional scale. The simulated values of peak ground velocity and peak ground acceleration, together with those provided by ShakeMaps, are included in a dataset containing information on 3.060 buildings damaged during the 2009 earthquake. Three different damage distributions were considered, obtained from the six damage levels assigned in the original dataset, ranging from D0 (light or no damage) to D5 (heavy damage or collapse). The results show that the peak ground velocity and peak ground acceleration obtained from numerical simulations have a significantly greater impact on the level of damage recorded than those derived from ShakeMaps. In fact, simulated intensity measures are among the four most important features for all tests considered, while variables obtained from ShakeMaps do not even appear in the top twelve, consequently, they may not serve as a reliable proxy for building damage assessment. The simulated intensity measures have been used to assess the building damage reaching a maximum accuracy of the 83% on a balanced dataset.

The continuous increase in computing resources has encouraged the development and application of high-performance computing (HPC) and artificial intelligence (AI) in a variety of scientific disciplines, ranging from medicine to economics and social sciences to engineering. The use of artificial intelligence is becoming increasingly important in the field of seismology. It is widely employed, for example, to analyze seismic signals, define early warning systems, and locate earthquake sources. Interested readers can refer to numerous review papers available in the literature (see, for example, <sup>1-3</sup>).

Physics based simulations (PBS) represent a valuable alternative to the Empirical Ground Motion Equations to compute the seismic wavefield and, thus, to calculate any ground motion Intensity Measure (IMs) of interest, such as peak ground velocity (PGV) and peak ground acceleration (PGA). IMs are essential parameters in the seismic hazard analysis and are often used for seismic risk assessment and prevention. Recently, in the same spirit of this work, an unconventional approach to defining fragility curves starting from IMs obtained from PBS has been proposed in <sup>4</sup>. However in the presented paper, extending the framework presented in <sup>4</sup>, the proposed approach relies predominantly on machine learning techniques rather than traditional fragility curve formulations, with the aim of enhancing the accuracy of building damage estimation. After a strong earthquake, once the emergency phase has passed, it is necessary to assess the impact of the earthquake on the building heritage and quickly identify buildings that are habitable or with limited damage, which can be made available to the population. On the other hand, another fundamental aspect of seismic risk prevention and assessment is

<sup>1</sup>Istituto Nazionale di Geofisica e Vulcanologia, 20133 Milan, Italy. <sup>2</sup>Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, 67100 L'Aquila, Italy. <sup>3</sup>Gran Sasso Science Institute (GSSI), 67100 L'Aquila, Italy. <sup>4</sup>Sezione Tettonofisica e Sismologia, Istituto Nazionale di Geofisica e Vulcanologia, 67100 L'Aquila, Italy. <sup>5</sup>Department of Civil and Environmental Engineering, Politecnico di Milano, 20133 Milan, Italy. <sup>6</sup>MOX-Dipartimento di Matematica, Politecnico di Milano, 20133 Milan, Italy. ✉email: federica.dimichele@ingv.it; donato.pera@univaq.it

the identification of the geographical, geological, and constructional characteristics that make a building type more vulnerable to the ground shaking. Currently, the assessment of possible impact of the earthquakes on the building damage level is carried out by the above-mentioned fragility curves, which assign the probability that a building will exceed a certain level of damage. The “classical methods” for the definition of fragility functions are usually based on IMs values obtained using Ground Motion Models (GMMs)<sup>5</sup> or on ShakeMap.<sup>6</sup> More recently an alternative approach based on machine learning have been proposed. In<sup>7</sup> the authors used a ML-based supervised classifiers to assess the damage level of 2276 buildings after the 2014 South Napa 6.1 Mw earthquake. Their working dataset contains many different features such as number of stories, buildings-fault distance,  $V_{S30}$ , spectral acceleration (SA) at 0.3 s, age of construction, value, cost, size and regularity of each building. The target variable is the level of damage, assigned as low, medium, and high. Other interesting applications can be found in<sup>8,9</sup> and<sup>10</sup>.

Concerning on the L'Aquila 2009 earthquake, in<sup>11,12</sup> the authors propose a comparative study using the same dataset as in this paper. In<sup>13</sup> a random forest based classifier has been employed to analyze the dataset introduced and described in<sup>14</sup>.

In this paper we employ a combination of machine learning tools and physics based simulations to assess the damage level on the built-up area after an earthquake, aiming to support both hazard mitigation efforts and the post-earthquake emergency response. As a case of study we considered the L'Aquila (6<sup>th</sup> April 2009) earthquake, that has been the first of a long and destructive series of seismic events which hit the central Italy between the 2009 and the 2017<sup>15–26</sup>.

There are at least a couple of reasons why L'Aquila earthquake is particularly suitable for the purpose. First of all, the main shock was significantly stronger than the aftershock having a magnitude 5.4 Mw. It makes reasonable the assumption that the damage recorded is mostly due to the 6.1 Mw event. The second reason relates to the availability of geological data. Thanks to the complex microzonation work carried out by the Italian Civil Protection Agency, in-depth knowledge of the territory's geology is available, which is essential for constructing an accurate computational domain. Furthermore, during the L'Aquila Opendata project (<http://www.opendatalaquila.it/>), data was collected and made available, allowing for the construction of a dataset containing approximately 3,000 buildings already used and validated in<sup>11</sup>. We also observe as the building typologies found in the historic centre of L'Aquila are representative of many other Apennine locations; therefore, the methodology presented in this study could be generalized to these areas, provided that the necessary data are available. However, extending this approach to seismological regimes or reconstruction frameworks other than those analyzed in this work remains more complex and would require further assessment. This paper consists of two main parts. In the first one, we used the numerical code SPEED (<https://speed.mox.polimi.it/>)<sup>27</sup> to simulate the 2009 earthquake and obtain synthetic values of PGV and PGA. Then, in the second part, we used simulated PGA and PGV as predictive variables for an AI-based model to assess the value of damage to buildings. In the absence of PBS numerical simulations, IMs can be obtained by interpolating recorded values (with a level of uncertainty that increases with distance from the detection station) or by using simplified models such as ground motion models<sup>28</sup>. The Shake-Maps repository (<https://shakemap.ingv.it>) combines observed ground motion values and predictive relations to provide regional and local shake maps<sup>29,30</sup>. Here, both simulated and inferred from Shake-Maps IMs are used as predictive variables, and their impact on determining the level of damage has been assessed with the aid of AI-based techniques.

The paper is organized as follows. In the next section, we describe the computational domain and validate the numerical results by comparing recorded and simulated waveforms. In Section 3, we present the working dataset and we introduce some information related to the dataset preparation and to the ML tools used for our analysis algorithm. In Section 4, we present and discuss our results. Finally, in the last section, we summarize our results and present further developments of this work.

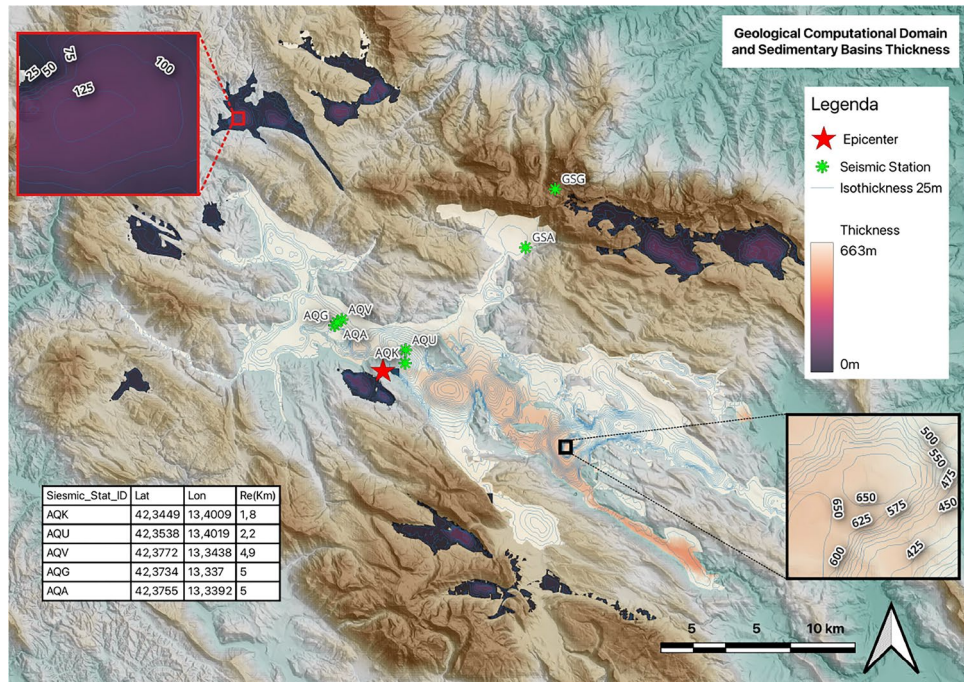
## Numerical model and sensitivity analysis

In this work, we consider a three-dimensional (3D) computational domain that extends 59.5 km in width, 57.5 km in length, and 19.8 km in depth, centered around the city of L'Aquila.

The domain has been built using a Python tool introduced in<sup>31</sup> and the TINITALY database for topographic data<sup>32–34</sup>. The CUBIT software (<https://coreform.com/coreform-cubit/>) was used to create a mesh containing 776,426 elements, with sizes ranging from 130 m (top layer) to 1 km (bottom layer). The maximum achieved frequency resolution is around 2–2.5 Hz, with polynomial degree equal to 3. The fault plane extends 28 Km in length and 20.9 Km in width with a strike equal to 133°, dip equal to 54°, and rake equal to -102°. The hypocenter position has been obtained using the epicenter position (42.34 Lat–13.38 Lon) and fault geometry. The adopted seismic source, reconstructed from the seismographic data after the earthquake<sup>19</sup>, has already been successfully employed in<sup>35–37</sup>.

In addition to the Plio-Quaternary sedimentary basin named *Media Valle dell'Aterno*, already included in<sup>36</sup>, we add nine smaller basins with a maximum depth of approximately 150 m. In Figure 1 we show the computational domain (top view) where the main basin and the additional nine basins are visible (all highlighted using a color scale ranging from purple to white, representing the thickness). Specifically, we refer to the main basin as the area that contains the measuring seismic stations named AQQ, AQQV, AQQK, AQA, AQU and GSA, whereas the new nine additional basins are the highlighted areas disjointed from the main basin.

The shallow structures of the secondary basins have a similar geological structure to that of the first 150 m of the main basin. Therefore, it is reasonable to assume that the velocities and density profiles of all considered basins are the same. For the numerical simulations, as in<sup>36</sup>, we considered a four layer computational domain. However, in the present study, we employed two distinct models for the mechanical properties of the topmost layer. Specifically, in the first model for topmost layer, as in<sup>35</sup>, we assumed that within the basins both  $V_S$  and  $V_P$  depend only on the depth  $z$  as follows:



**Fig. 1.** This figure shows the computational domain, including the main basin, named *Media Valle dell'Aterno*, and nine smaller sedimentary basins with a maximum depth of 150 m. We refer to the main basin as the highlighted area (color scale ranging from white to peach) that contains the seismic stations AQG, AQV, AQK, AQA, AQU and GSA, whereas the nine additional basins are the highlighted areas (in purple color) disjointed from the main basin. The (6<sup>th</sup> April 2009) L'Aquila earthquake epicenter is marked with a red star. (This figure has been realized using the software QGis (<https://qgis.org/>) version 3.40.3, license QGIS (GPL v.2+)).

Domain name	Number of layers	Topmost layer model	Basin
T1	4	$V_S, V_P$ and $\rho$ constants	Main
T2	4	$V_S, V_P$ and $\rho$ as in (2)	Main
T3	4	$V_S, V_P$ and $\rho$ constants	Main and additional 9
T4	4	$V_S, V_P$ and $\rho$ as in (2)	Main and additional 9

**Table 1.** Computational domains used for the numerical simulations.

$$V_S = 300 + 36 \cdot z^{0.43} (\text{m/s}), \quad V_P = 2.14V_S, \quad Q_S = 0.10V_S, \quad \rho = 1.9 (\text{g/cm}^3). \quad (1)$$

Here  $Q_S$  is the S-wave quality factor and  $\rho$  the soil mass density. In contrast, for the outcropping bedrock we considered constant  $V_S$  values as reported in <sup>35</sup>. Hereafter, we refer to this model as *non-improved bedrock* case. For the second topmost layer model, as described in <sup>37</sup>, we assume a depth-dependent shear velocity in the outcropping bedrock, in order to get more realistic  $V_{S30}$  values. Namely, we set:

$$V_S = 800 + 28.4 \cdot z^{0.5} (\text{m/s}), \quad V_P = 1.86V_S, \quad Q_S = 0.10V_S, \quad \rho = 2.2 + 9.5 \cdot z^{0.5} (\text{g/cm}^3). \quad (2)$$

Hereafter, we refer to this second model as *improved bedrock* case. For the other three layers, the mechanical properties were taken from <sup>35</sup> and are the same already used in <sup>36</sup>.

To evaluate the contribution of the nine small basins and the improved bedrock on the top layer, we considered four different computational domains. The first and second domains, here named T1 and T2 respectively, contain only the main basin, already included in <sup>36</sup>. In T1, we assume constant values of  $V_S, V_P$ , and  $\rho$  as in <sup>36</sup>, while in T2 we assume to work with the improved bedrock mechanical properties as in <sup>37</sup>. Domains T3 and T4 contain nine minor basins in addition to the main basin. In T3, we assume constant values of  $V_S, V_P$ , and  $\rho$  as in <sup>36</sup>, while in T4 we assume to work with the improved bedrock model. The above is summarised in Table 1 for the reader's convenience:

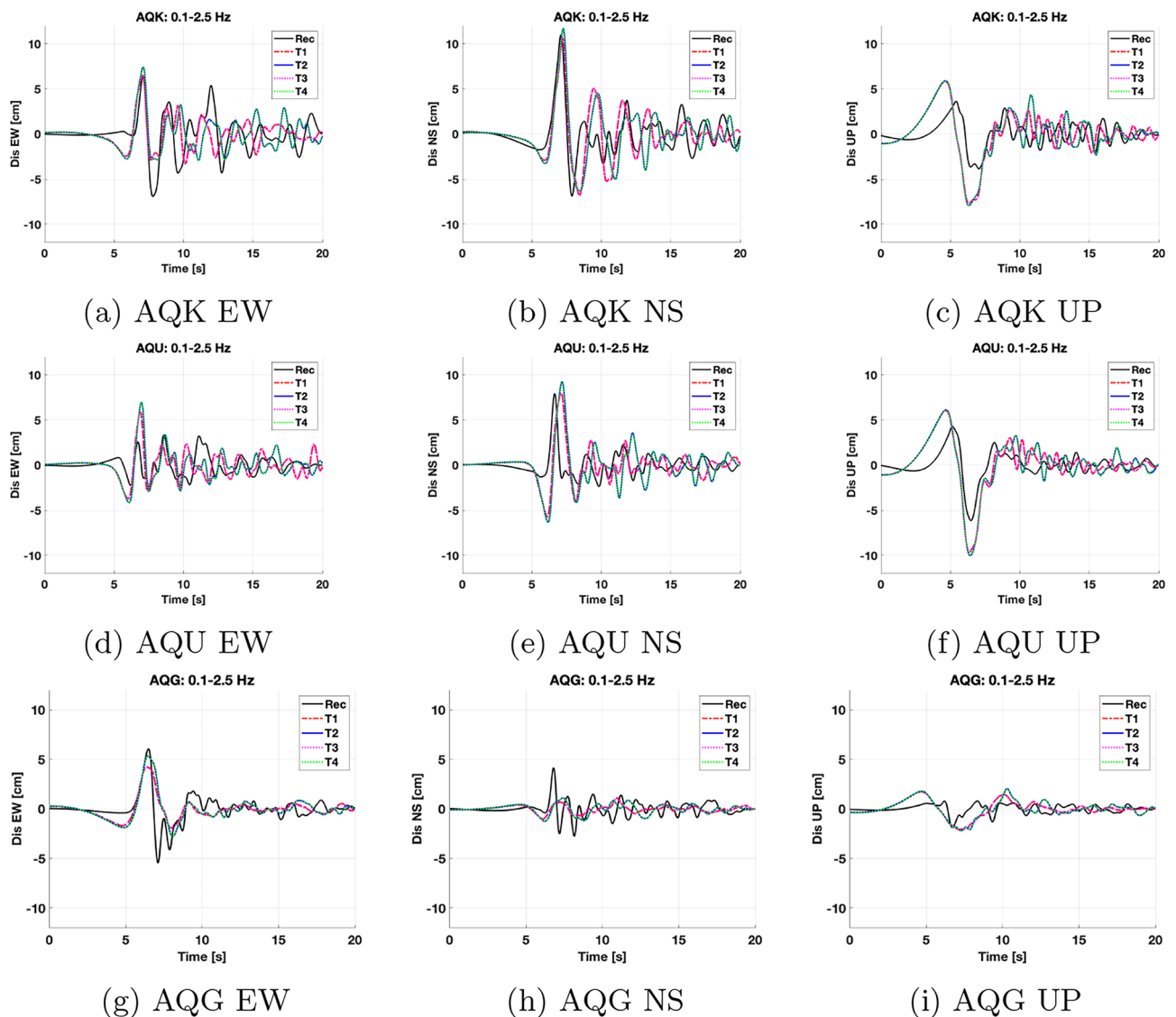
In the following, we compare the simulated waveform obtained for the four different computational domains with the available recorded data. Among the several seismic stations that recorded the 2009 event, just seven

of them fall in our computational domain, as shown in Figure 1. Three stations, AQK, AQU, and AQQ that are located in areas of high population density, are used to validate our computational models. In particular, AQK and AQU are situated close to the centre of L'Aquila (1.8 and 2.2 km from the epicentre, respectively), while AQQ is located on the western outskirts of the city at 5 km from the epicentre.

In Figure 2, the north-south (NS), east-west (EW), and up-down (UP) components of the synthetic seismographs are compared with the corresponding recorded data.

To estimate the agreement between the recorded and simulated waveforms, we computed the normalized cross-correlation (NCC) values reported in Table 2 for the stations AQK, AQU and AQQ and the components EW, NS and UP. The analysis of the NCC values reveals specific trends across the monitored stations. For AQK and AQU, the highest NCC values are achieved using domains  $T_2$  and  $T_4$  in the components EW and NS, while the best agreement for the vertical component (UP) is observed in domains  $T_1$  and  $T_3$ . In contrast, station AQQ consistently shows the highest NCC values for domains  $T_1$  and  $T_3$  across all three components (EW, NS and UP).

In view of the lack of a clearly better simulation scenario, we can observe that domains  $T_2$  and  $T_4$  at stations AQK and AQU yield the highest NCC values in two out of the three components (EW and NS). Given that the numerical differences between these values are minimal, domain  $T_4$  was selected for all subsequent analyses, as it represents the most comprehensive model.



**Fig. 2.** Comparison between the recorded waveform (in black) at AQK, AQU, and AQQ and the simulated one. The four different tests are indicated as follows: T1-red, T2-blue, T3-magenta, T4-green. In the horizontal axis, we reported the earthquake simulation time equal to 20 s, in the vertical axis, we reported the displacement (DIS-EW, DIS-NS, DIS-UP) in *cm* related to the three spatial components of the simulated seismograms.

Domain name	AQK (EW-NS-UP)	AQU (EW-NS-UP)	AQG (EW-NS-UP)
T1	0.4474-0.7011-0.7345	0.3424-0.7099-0.7974	0.6334-0.3178-0.4263
T2	0.5079-0.7331-0.7014	0.3496-0.7418-0.7747	0.6638-0.2947-0.4095
T3	0.4450-0.7012-0.7335	0.3423-0.7139-0.7975	0.6329-0.3174-0.4271
T4	0.5062-0.7321-0.7011	0.3482-0.7427-0.7746	0.6632-0.2946-0.4092

**Table 2.** Normalized cross correlation (NCC) computed between the simulated and recorded displacements for the stations AQK, AQU and AQG in the three components (EW, NS and UP) reported in Figure 2.

Finally, it should be emphasized that the agreement between recorded data and simulations is quite satisfactory, with the exception of the NS component at the AQG station. In line with the previous findings<sup>38</sup>, the mismatch at AQG-NS component may be due to inaccuracies in the local geological model which, although detailed, is not able to capture some localised site amplification effects.

As mentioned in the introduction, the PBS model can be useful for studying the behavior of seismic waves and simulating ground motion; however, its maximum frequency resolution is quite low, typically ranging from 1 to 3 Hz. This frequency limit leads to low-quality simulated data for high-frequency ground motion parameters, such as PGA. To obtain the high-frequency component of the spectrum, a hybrid approach can be used that combines PBS simulations with empirical or data-driven methods, such as Green's function, stochastic models, or deep learning techniques<sup>38–41</sup>. In this work, we employ the ANN2BB tool to generate broadband ground motions with a suitable frequency content (see<sup>42</sup>) and, thus, to compute the PGV and PGA datasets used as input for the AI-based tool. This approach, based on Artificial Neural Networks (ANN), produces a correlation between long and short period spectral ordinates trained on strong motion records. With this technique, starting from PBS outputs, one can produce synthetic signals having broadband content.

## Dataset preparation

In this work, we used supervised machine learning (ML) techniques to train, validate, and test a model that can assess the damage levels of buildings as a consequence of a major seismic event. The working dataset is the same as that used in<sup>11</sup> (see also<sup>43</sup> where similar dataset are presented), composed of 3060 buildings located in the area of the LAquila 2009 earthquake, but enriched by the PGA and PGV values provided by the physics-based simulations obtained from the domain T4, as described in the previous section. In the following each building is described by 20 predictive variables, divided into:

- *Buildings features*: construction techniques (C), aggregation type (C), position (with respect to the aggregate) (C), number of units (in the aggregate) (N), height (N), surface aggregate area (N), mean area (N), number of vertices (in the aggregate) (N), and age (C).
- *Geophysical features*: geographical coordinate (WGS84/UTM 33 N) (N), distance from the epicenter (N), distance from the hypocenter (N), peak ground velocity (N), peak ground acceleration (N), time-average shear wave velocity to 30 m of depth ( $V_{s30}$ ) (N), coefficient of stratigraphic amplification, coefficient of topographic amplification (N), slope (N), maximum design acceleration value (N).

Labels C and N indicate categorical and numerical variables, respectively.

For the peak ground velocity and peak ground acceleration, we considered both the values provided by the platform Shakemap<sup>29</sup>, named PGV and PGA, and by the PBS simulation, from now on  $PGV_{PBS}$  and  $PGA_{PBS}$ , respectively.

To ensure compatibility with the Random Forest algorithm, ordinal categorical variables (such as Building Age, Soil Morphology, and Damage Level) were transformed into numerical values. For the remaining categorical features, a one-hot encoding scheme was applied. The dataset is divided into training, validation and test sets, and the numerical variables are normalized. As usually, the normalization parameters are fitted on the training set only and then applied to the test set to prevent data leakage. The dataset must also be free of missing values ( $NaN$ ). In this case, the only variable that contains  $NaN$  is the age of construction (simply 'Age' in the list of characteristics). There are 528 missing values for this variable, which is a significant number compared to the size of the dataset. In<sup>11</sup> the buildings having missing values have been removed, significantly reducing the dataset size. On the contrary, in this study, the missing or uncorrected values have been estimated and included in the dataset using a technique known as data imputation. If the number of missing values is less than 5%, it is sufficient to substitute them with their mean or mode. If the number of missing data is between 5% and 20%, as in our case, a more complex approach should be that takes the entire dataset into account, rather than just the columns containing missing values.

On the other hand, when the percentage of missing data exceeds 20%, it is recommended to use more advanced techniques, such as those based on artificial neural networks. In this work, the KNN-Imputer, as implemented in Scikit-learn (<https://scikit-learn.org/stable/>), is used for data imputation. We emphasize that data imputation is only carried out on the training and validation sets, while the test set containing has not been modified to guarantee the reliability of the case study presented at the end of this section. Furthermore, only a restricted subset of predictive variables is included, which do not contain the target variables and all features related to the seismic event, such as epicenter distance and all the intensity measures. The latter are clearly not correlated with the construction era of the building, while the former are excluded to avoid generating artificial correlations that could potentially compromise the proper training of RF-based models. The complete list of the

features included in the data imputation procedure is reported in the Appendix, for the sake of completeness, where also the distribution of the imputed variable before and after is reported. After imputing missing values, the dataset used for training and validating the models contains 2754 data-points (dp).

The damage classification was originally based on six levels (from D0-no damage to D5 heavy damage or collapse). However, due to the size of the dataset, we will define three different combinations of damage levels, as listed below:

DS1: D0-D1 light damage (710 dp), D2-D3 moderate damage (1108 dp), D4-D5 heavy damage (936 dp)

DS2: D0-D1 light damage (710 dp), D2-D3-D4-D5 from moderate to heavy (2044 dp) damage

DS3: D0-D1-D2-D3 from light to moderate damage (1818 dp), D4-D5 heavy damage (936 dp) The first case is similar to the one considered in <sup>11</sup>. In the other two cases, instead, we aim to identify buildings with no or minor damage (DS2) and those with more serious damage or collapsed (DS3), with the scope of supporting the pre- and post-emergency phases.

To properly evaluate the classifier's performance, various metrics can be employed.

The parameter most often used to evaluate the performance of a supervised learning algorithm is the accuracy, defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Here, TP, FP, TN, and FN stand for true positives, false positives, true negatives, and false negatives, respectively.

Other possible metrics are recall, precision, and  $F_1$  score, which can be written as:

$$\text{Recall} = \frac{TP}{TP + FN}, \quad \text{Precision} = \frac{TP}{TP + FP}, \quad F_1 = \frac{TP}{TP + (FN + FP)/2} \quad (4)$$

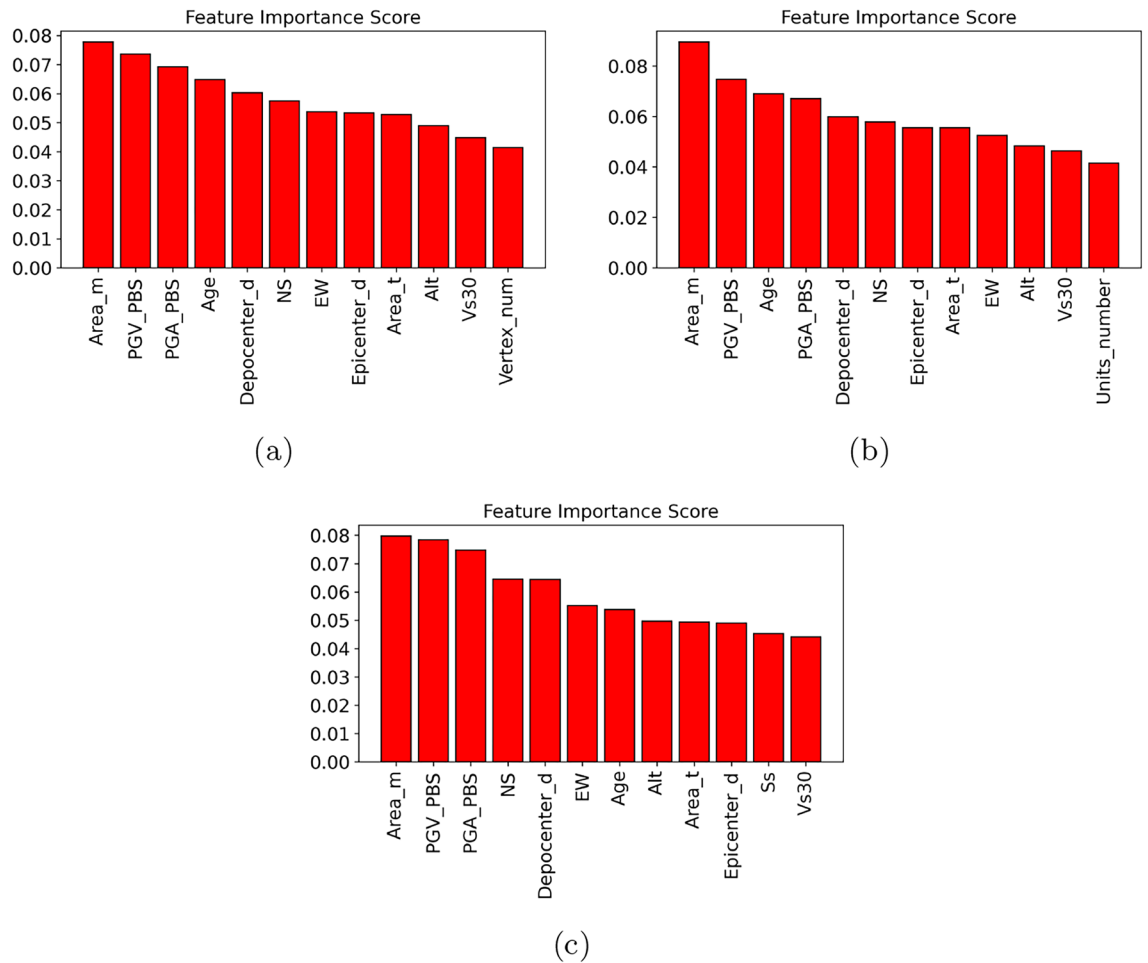
Accuracy is a suitable metric for well-balanced datasets, whereas for strongly or moderately unbalanced datasets, the  $F_1$ -score is a more appropriate metric for evaluating the classifier's performance. As mentioned above, in this work, we used a popular algorithm called Random Forest (RF), which has already been used for similar purposes in <sup>7,8,12</sup>. RF, first introduced by Ho <sup>44</sup> in 1995, is a robust algorithm based on an ensemble of decision trees (DT), which guarantees good quality performance on the treatment of tabular data. An accurate description of the used algorithm is beyond the scope of this paper, and the interested reader may refer to introductory texts on machine learning available in the literature, such as <sup>45</sup>.

In order to reduce the size of the datasets by removing unnecessary variables and saving computational resources, we performed a preliminary analysis to evaluate the impact of each variable on determining the damage level. Figure 3 shows the importance score (obtained using a balanced random forest (BRF) model with default hyperparameter values) of the top 12 features for the three different distributions of the target variables. We immediately notice that ten features, namely Mean Area, depocentre distance, epicenter distance,  $PGV_{PBS}$ ,  $PGA_{PBS}$ , NS and EW coordinates, age, height, total area, and  $V_{S30}$  are common to all three datasets. In all cases, the building (or aggregate) average surface area is the variable that most influences the level of damage, as stated in <sup>11</sup>. We also emphasize that, of the two metrics characterizing the site-source distance, the depocentre distance always registers a higher score than the epicenter one. Finally, we note that only the IMs calculated via PBS appear to have a significant effect on determining the damage level, at least for this dataset.

From now on, for sake of simplicity, the working datasets will include the eleven common features.

## Dataset Analysis and models validation

Our working datasets are moderately unbalanced, particularly for DS2 and DS3. This could affect the performance of the classifier. The simplest approaches to managing an unbalanced dataset are undersampling and oversampling. In the first case, the size of the majority class is reduced until it matches that of the minority class. This approach is practicable only if the dataset is large enough and it is not strongly unbalanced. In this spirit, a modified version of the RF algorithm, specifically designed for unbalanced datasets (BRF), will be trained and validated. Compared to the traditional RF, BRF trains each tree using a balanced dataset by undersampling the majority classes. Alternatively, one can increase the number of elements in the minority class, i.e., oversampling. Among the most popular algorithms developed for this purpose, we cite Synthetic Minority Over-sampling Technique (SMOTE) introduced in <sup>46</sup>. In recent years, a hybrid approach combining oversampling techniques with specific algorithms has emerged. This approach is particularly useful for treating strongly unbalanced datasets where the minority class contains fewer than 5% of the total elements. In this work, we apply the SMOTE algorithm to create a balanced version of our dataset, oversampling the minority classes. A classical RF algorithm is then trained and tested for all three datasets. The three models (RF, BRF, and RF+SMOTE) are compared. Repeated stratified K-Fold cross-validation approach (setting  $n_{\text{splits}}=10$  and  $n_{\text{repeats}}=5$ ) is applied to increase the robustness of the analysis and reduce the overfitting. For each dataset, we report in Table 3 and Table 4 the mean value and the standard deviation for accuracy and the  $F_1$ -score obtained for the three proposed models. In this preliminary phase, hyperparameters are not optimized; instead, default values are used for each test, covering both the training/validation procedure and the data augmentation via SMOTE. All machine learning procedures, including pre-processing, balancing and model evaluation, were implemented in Python using the Scikit-learn library, while the SMOTE algorithm was applied via the Imbalanced-learn toolbox. For the sake of completeness, it should be noted that the workflow is computationally efficient to be handled on standard workstations. It does not rely on high-performance computing (HPC) or complex optimization, making it highly accessible for a variety of applications.



**Fig. 3.** Feature importance scoring for the three datasets DS1(a), DS2(b) and DS3(c). To improve the clarity and readability of the figures, only the top 12 scores are shown.

Dataset	Acc-std (RF)	Acc-std (BRF)	Acc-std (RF+SMOTE)
DS1	0.555-0.027	0.534-0.026	0.614-0.030
DS2	0.738-0.020	0.731-0.022	0.829-0.021
DS3	0.728-0.020	0.697-0.025	0.784-0.025

**Table 3.** Accuracy for three datasets and different models, standard RF, balanced RF, and RF+SMOTE on the augmented dataset.

Dataset	F1-std (RF)	F1-std (BRF)	F1-std (RF+SMOTE)
DS1	0.546-0.028	0.546-0.036	0.615-0.028
DS2	0.852-0.018	0.810-0.019	0.830-0.021
DS3	0.571-0.030	0.607-0.030	0.785-0.024

**Table 4.** F1-score for three datasets and different models, standard RF, balanced RF, and RF+SMOTE on the augmented dataset.

We observe that introducing a balanced dataset significantly improves the accuracy of our results in all case studies. However, the F1 score does not show the same trend for the DS2 dataset. The F1 score is sensitive to the number of false negatives and false positives. The basic RF model achieved an artificially high F1 score due to over-specialisation in the dominant class, labelled as 1, essentially "ignoring" the minority class, labelled as 0, while maintaining high precision. By balancing both the training and validation sets, the RF+SMOTE model establishes

Dataset	RF (95% CI)	BRF (95% CI)	RF+SMOTE (95% CI)
DS1	[0.536, 0.574]	[0.515, 0.553]	[0.593, 0.635]
DS2	[0.724, 0.752]	[0.715, 0.747]	[0.814, 0.844]
DS3	[0.714, 0.742]	[0.679, 0.715]	[0.766, 0.802]

**Table 5.** 95% Confidence Intervals for Accuracy across different datasets and models (calculated with  $t = 2.262$ ,  $n = 10$ ,  $df = 9$ ).

Dataset	95% CI (RF)	95% CI (BRF)	RF+SMOTE (95% CI)
DS1	[0.526, 0.566]	[0.520, 0.572]	[0.595, 0.635]
DS2	[0.839, 0.865]	[0.796, 0.824]	[0.815, 0.845]
DS3	[0.550, 0.592]	[0.586, 0.628]	[0.768, 0.802]

**Table 6.** 95% Confidence Intervals for F1-score across different datasets and models (calculated with  $t = 2.262$ ,  $df = 9$ ), according to 5.

a fairer decision boundary, effectively trading distorted majority class accuracy for robust overall classification performance. For the sake of completeness, we report the 95% confidence intervals (CI) related to the metrics listed above. For accuracy and F1 score, we also report in Table 5-6 the 95% confidence intervals calculated according to (5)

$$CI = \mu \pm \left( t_{\alpha/2, n-1} \times \frac{\sigma}{\sqrt{n}} \right)$$

where:

- $\mu$  is the mean performance metric across the 10 folds as reported in Table 3-4;
- $t_{\alpha/2, n-1}$  is the critical value from the Student's t-distribution for a 95% confidence level ( $\alpha = 0.05$ ) and  $n - 1 = 9$  degrees of freedom ( $df$ ), which corresponds to 2.262 in our case;
- $\sigma$  is the standard deviation ;
- $n = 10$  is the number of folds in the cross-validation.

We would like to point out that in calculating the confidence intervals, we used the number of folds ( $n = 10$ ) rather than the total number of repetitions to ensure a more conservative estimate.

We observe that the confidence intervals achieved with RF+SMOTE are higher than those of the other models and do not overlap with them, except for the F1-score in dataset DS2, as previously discussed. These results confirm the reliability and robustness of the performance reached by the proposed approach.

Figure 4 shows the confusion matrices associated with a single fold, corresponding to 10% of the entire training set, the three datasets, and three models under consideration. Usually, the main diagonal contains the correctly classified buildings. We note that the balanced algorithm slightly improves the algorithm's ability to correctly assign the target to minority classes, especially for binary target distributions corresponding to cases DS2 and DS3. For example, in DS2, the classic algorithm incorrectly assigns a moderate-to-heavy damage level to 38 buildings that suffered instead, light damage. Using the balanced algorithm, this number drops to 24. Similarly, in DS3 (Figure 4-(b)) RF incorrectly classifies 52 buildings with heavy damage. Although the balanced algorithm improves performance slightly, it still incorrectly classifies 32 out of 94 buildings in the minority class. Despite its benefits for minority classes, BRF does not achieve acceptable levels of accuracy or F1-score, and the false positive rate remains unacceptably high.

However, significant performance improvements were observed both in terms of accuracy and F1-score, particularly for DS2 and DS3, using the augmented dataset. In fact, for both DS2 and DS3, the number of misclassified elements is significantly lower than the number classified correctly. This generates a significant improvement in Accuracy and F1 score values, as reported in Table 3 and Table 4.

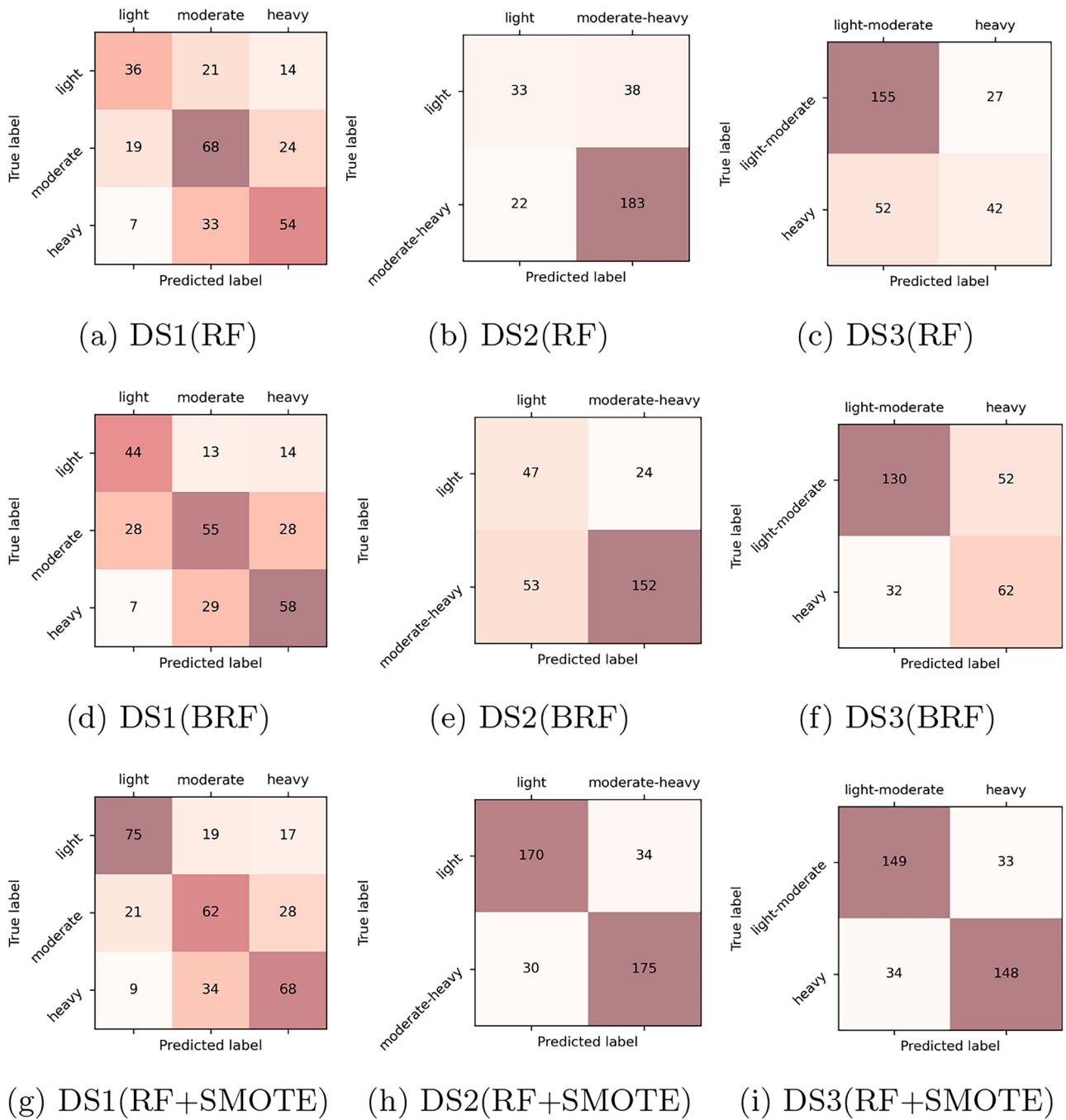
### Case of study

In the previous section, we trained and validated three different classifiers. A comparative analysis of their performance reveals the need to employ specific data augmentation techniques to achieve high accuracy, particularly for minority classes. We begin testing the model on a previously unseen dataset comprising 306 data points, which corresponds to 10% of the initial (non-imputed) and not augmented dataset. To this end, we consider the following case study:

- suppose we have completed the inspection of some of the buildings damaged by the earthquake, then we want to use the collected data to identify severely damaged buildings in a previously unseen dataset.

This case study is equivalent to the configuration of target variables as in DS3.

Before proceeding with the analysis of the test dataset, we removed all data points that had no information on the construction period (Age). The test dataset now contains 157 buildings with minor/moderate damage and 94 with heavy damage. Then, the RF model is trained on the entire (imputed and augmented) dataset and tested on the test set. The accuracy obtained for the test set, using the optimized hyperparameter values reported



**Fig. 4.** Confusion matrices referring to the validation dataset are reported for all the models and target variables considered. In particular, sub-figures (a) to (c) show the results of a classical RF model for classification. In (d)–(f), however, a balanced RF algorithm has been employed. Finally, in sub-figures (g)–(i), the augmented dataset and the RF+SMOTE algorithm have been used.

in the Appendix, is 0.76. This result is comparable to those obtained on the fully balanced dataset reported in Table 3 and, in any case, is higher than the accuracy achieved using an unbalanced training dataset Fixed *random\_state* = 42 is used to ensure reproducibility. A similar argument can be made for the F1 score, for which we have 0.76 (Weighted Average). The result in this case is significantly better than the results obtained with unbalanced datasets (0.571 (std 0.030) and 0.607(std 0.030) using RF and BRF, respectively). In Fig. 5, results are displayed in terms of confusion matrices for both the training and the dataset.

The performance obtained on the training dataset is better than that obtained on the unseen dataset. This was expected, given the relatively small size of the test dataset and the complexity of the problem under analysis. A larger dataset would certainly improve performance, and more advanced data augmentation techniques could help to limit the problem. This will be the subject of a future study. Finally we report the precision recall-curve for the test set (Fig. 6).

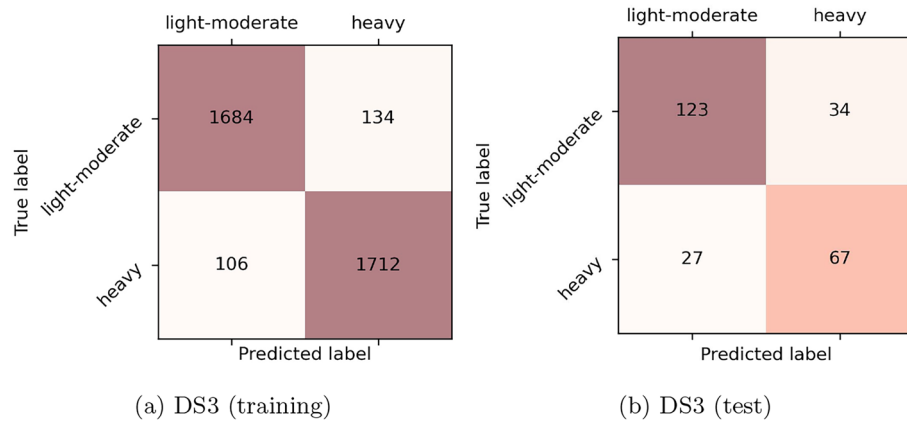


Fig. 5. Confusion matrix for the training (imputed and augmented) and the never-seen test set using RF.

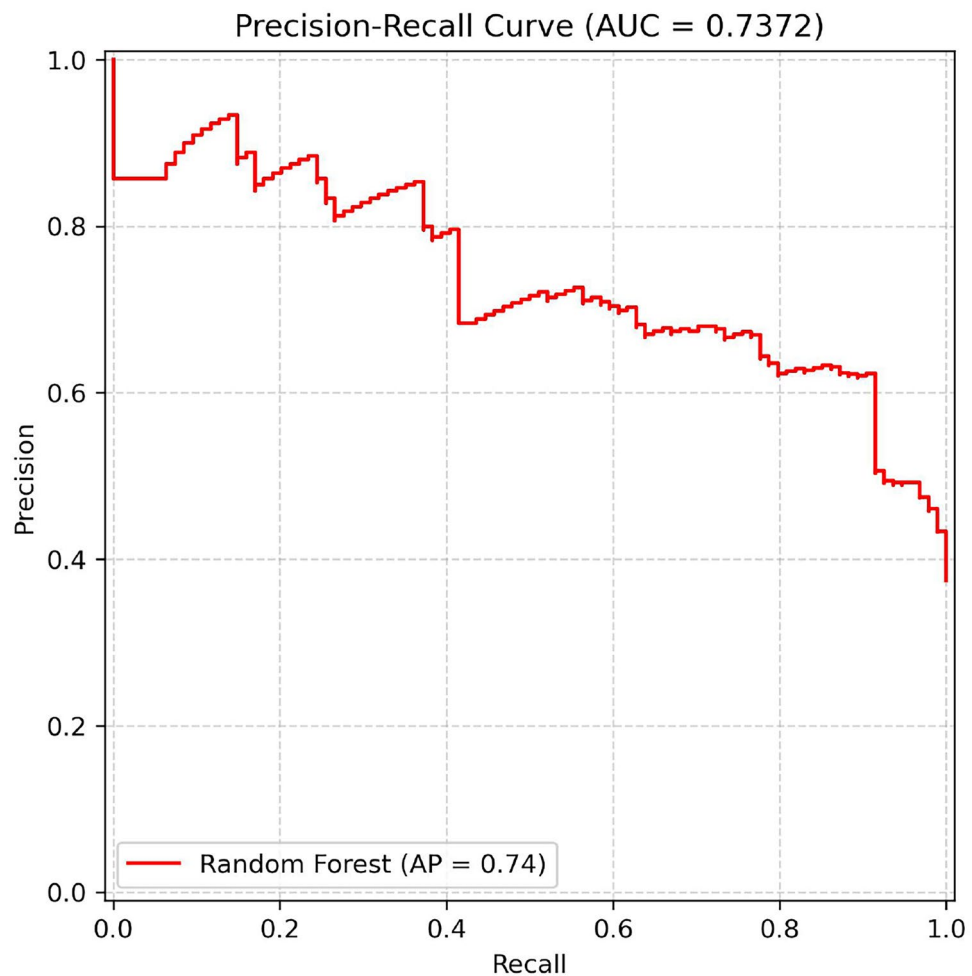


Fig. 6. Precision–recall curve for the Random Forest model with balanced training set and never seen test set.

The model achieved an average precision (AP) of 0.74, demonstrating a solid balance between precision and recall across different thresholds. Unlike the model evaluation phase, the hyperparameters for testing were tuned using a random search strategy. With this approach, different combinations of hyperparameters are evaluated and compared, with the latter being selected randomly within a range defined by the user. This approach is less precise than grid search, since it does not evaluate every possible combination, but it is far more computationally efficient and typically yields a satisfactory, though suboptimal, solution. All the information concerning the optimization procedure is reported in the Appendix.

## Conclusions

In this work, we developed a tool for assigning damage levels to buildings following a strong earthquake. The PGA and PGV values derived from physics-based numerical simulations were used as predictive variables in an artificial intelligence-based model to determine the level of damage suffered by buildings. A preliminary analysis was conducted using a random forest feature scoring algorithm, which showed that IMs calculated using PBS simulations have a significant impact on damage values. Using these variables, rather than those derived from ShakeMaps, we trained a random forest-based classifier. The use of appropriate data imputation and data augmentation techniques allowed us to significantly improve the performance of the classification algorithm on the validation dataset by about 80% for binary classification problems.

The classifier also performed well on a previously unseen test dataset, especially considering the small size of the test dataset. This tool, once properly validated and developed, can contribute to improving post-emergency procedures and implementing effective preventive measures.

However, it is necessary to carefully evaluate the model generality in the same spirit as <sup>12</sup>. In conclusion, while the model performed satisfactorily in the case study described, this may not be the case for other test sets or distributions of target variables. To this end, we are conducting a more in-depth study to examine the influence of predictive characteristics. Finally, we note that further developments of this work could include the use of larger datasets related to different earthquake events and/or a greater number of IMs obtained from PBS simulations <sup>4</sup>. This outcome highlights the importance of properly constraining input data. In this regard, the integration of satellite-derived imagery <sup>47</sup> offers a reliable and efficient means to quantify building damage at scale, providing essential constraints for identifying the most appropriate datasets.

Furthermore, it should be noted that the model's current generalizability is limited to structural typologies and site conditions similar to those in the training dataset (i.e., historic centers located in central Italy such as L'Aquila). Future research will focus on extending this methodology to diverse geological and urban contexts, as well as to multi-hazard frameworks, in the spirit of recent resilience modeling approaches <sup>48</sup>.

## Data availability

This work is part of the Open Data L'Aquila project, therefore the computational domain used will be available on the website <https://www.opendatalaquila.it/>, after the paper publication.

## Appendix

In this section, we provide some additional information on the tools used in the study, with the aim of making the results presented more clear and easily reproducible. First of all to enhance the readability and reproducibility of this work, we provide a summary of the Artificial Intelligence workflow below, covering the entire process from data preprocessing to the final results (Table 7).

**Input:** Dataset  $D$  (3060 buildings) from [11] + Physics-based simulations (PGA, PGV)

### Step 1: Pre-processing

- Enrich  $D$  with simulated PGA/PGV from Domain T4
- Map ordinal variables (Age, Morphology) to numerical scales
- Apply One-Hot Encoding to the other nominal categorical features
- Normalize all features and handle missing values via  $k$ -NN Imputer

### Step 2: Target Configuration & Feature Selection

- Regroup 6-level damage labels into 2 and 3 classes (DS1, DS2, DS3)
- Rank features via Random Forest Importance ranking
- Select 11 most relevant features (consistent across datasets)

### Step 3: Model Training & Internal Validation

**for** each dataset configuration (DS1, DS2, DS3) **do**

Apply **Repeated Stratified 10-Fold CV** (5 repeats)

Train models:  $RF$  (Standard),  $BRF$  (Balanced),  $RF+SMOTE$

Calculate Accuracy, F1-score and 95% Confidence Intervals (CI)

**end for** an unseen test set with DS3 label distribution

Apply **Repeated Stratified 10-Fold CV** (5 repeats)

Train models:  $RF$  (Standard),  $BRF$  (Balanced),  $RF+SMOTE$

- Perform **Hyperparameter Tuning** via **Random Search** for the  $RF+SMOTE$  configuration

- Using different training set for model with an without SMOTE for training

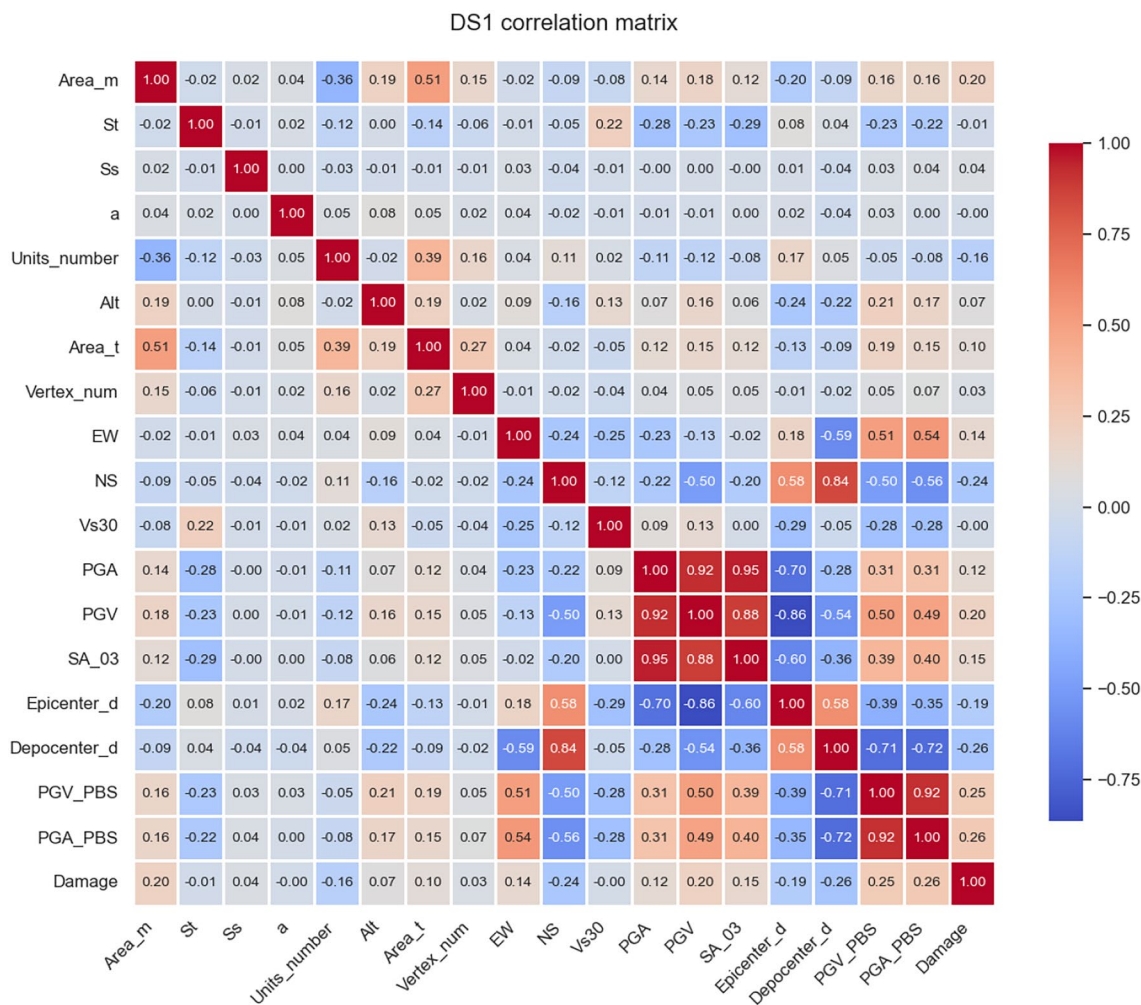
**Output:** Comparative performance metrics and generalization results

**Table 7.** Machine leaning workflow.

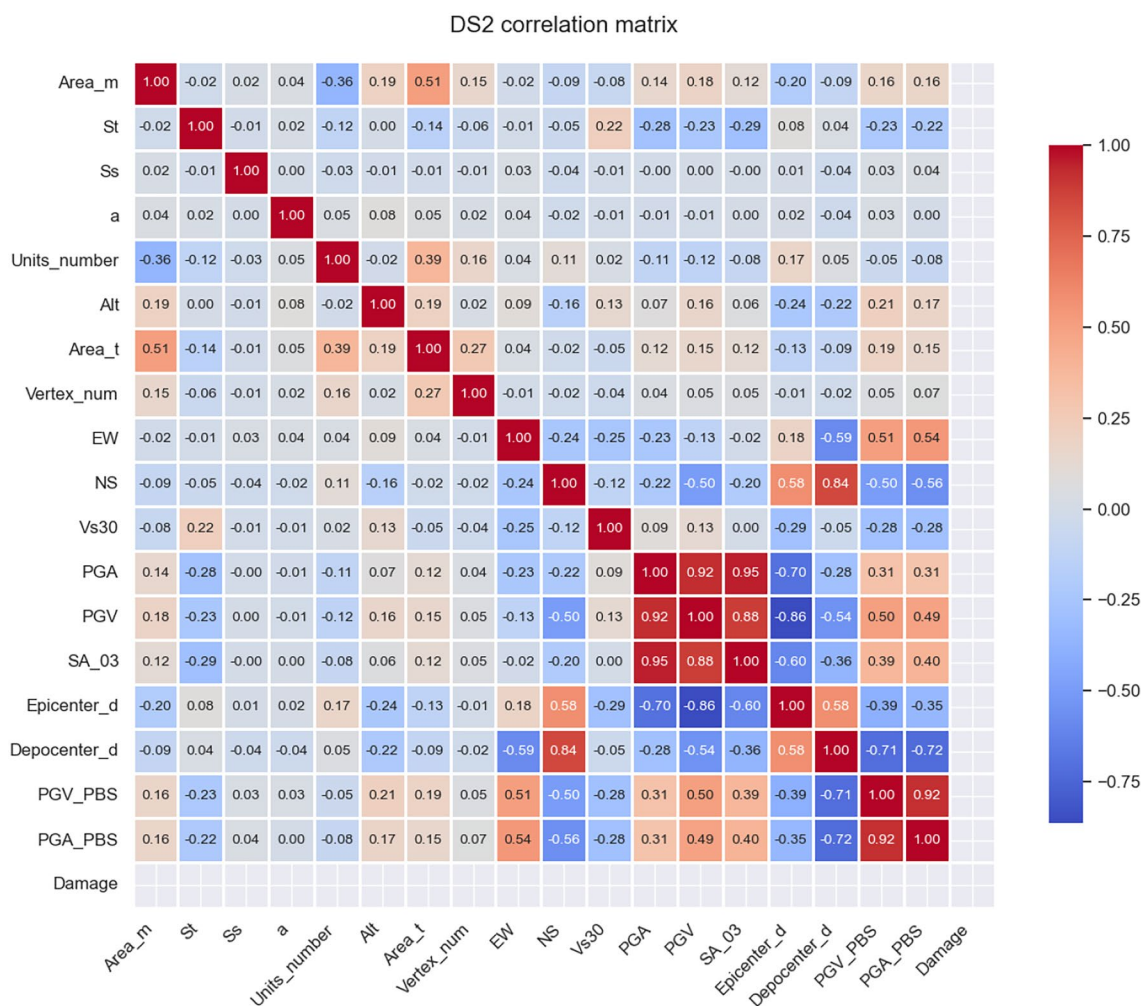
Subsequently, we provide further details regarding the feature correlation through cross-correlation matrices. We also describe the data imputation process (using the KNN Imputer), and the hyperparameter optimization strategy.

### Cross correlation matrices

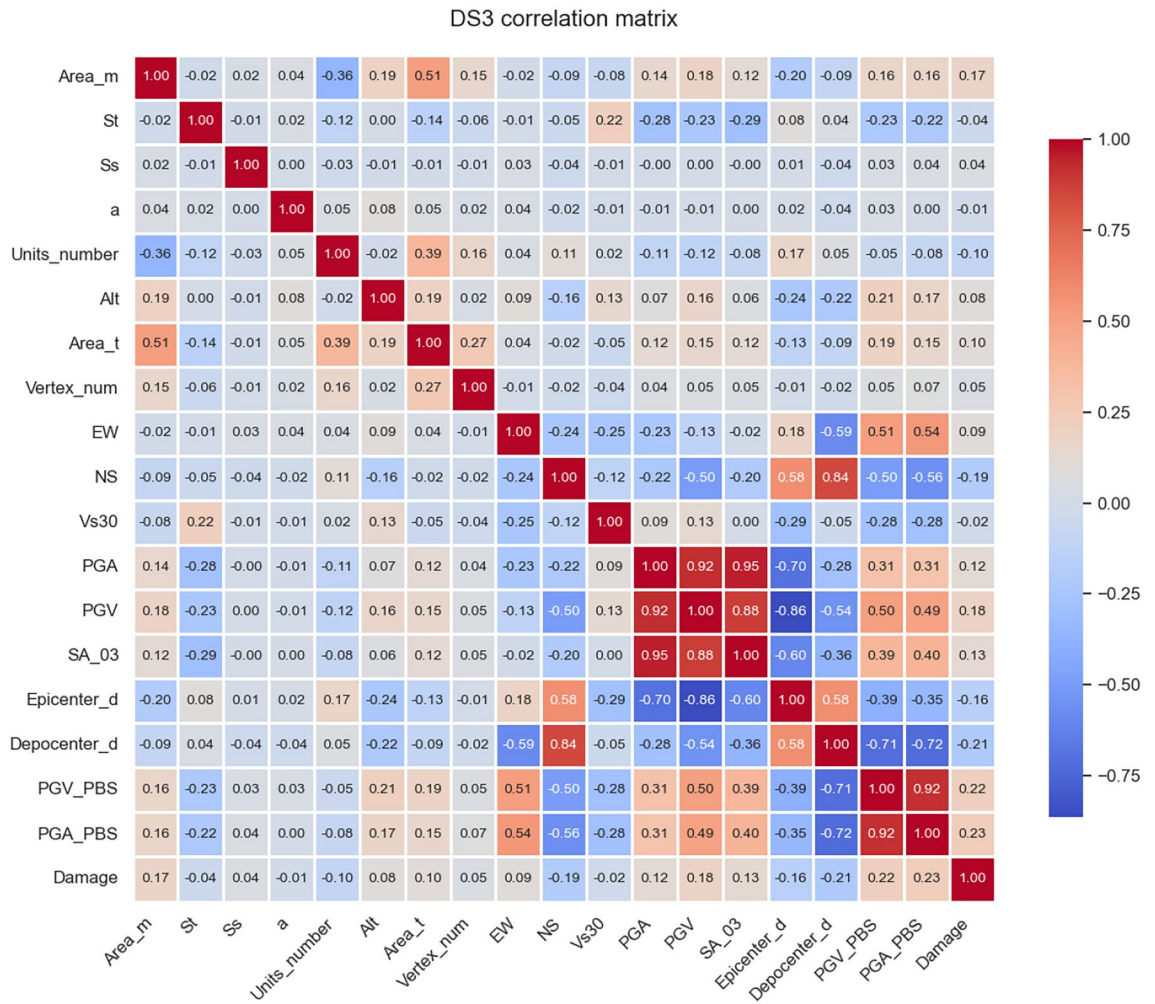
The feature importance scores reported in Fig. 3, focusing on the 12 top-ranking characteristics, provide valuable insights into the variables that most significantly influence building damage. To complement this analysis, we report the correlation matrices calculated using the Pearson correlation coefficient. For all damage distributions considered, the analysis was performed on the non-imputed dataset, while categorical variables were excluded to enhance readability. Although these matrices are based on linear correlations and may not fully capture the complex, non-linear interactions between variables, simulated PGA and PGV values consistently exhibit a higher correlation with observed damage compared to the corresponding Intensity Measures (IMs) derived from standard ShakeMaps. This confirms that physics-based simulations capture local seismic effects more effectively, providing more reliable predictors for the AI-based classification of structural damage (Figs. 7, 8, 9).



**Fig. 7.** Pearson correlation matrix for the DS1 damage distribution. Simulated PGA and PGV show the highest correlation with the damage level.



**Fig. 8.** Pearson correlation matrix for the DS2 damage distribution. Consistent with DS1, simulated intensity measures outperform ShakeMap data.



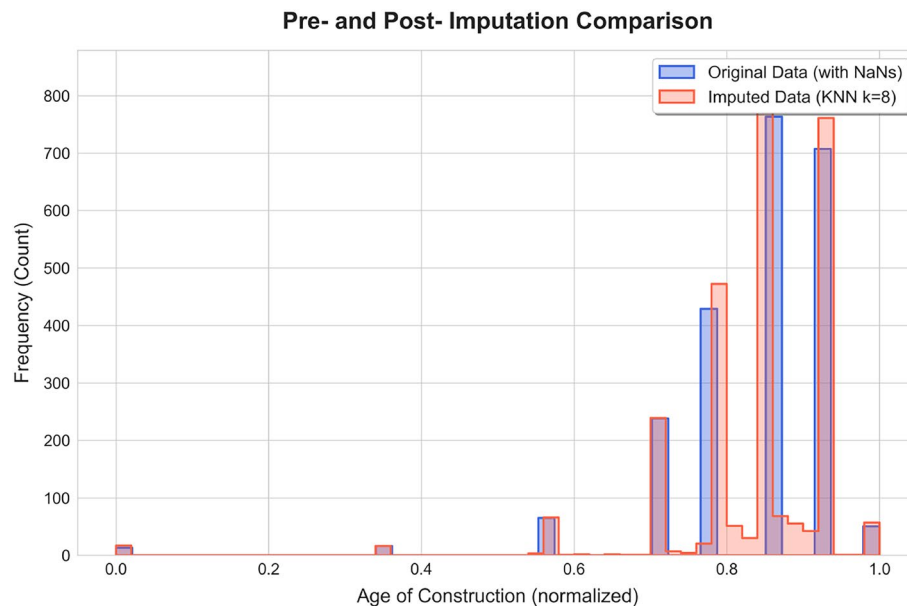
**Fig. 9.** Pearson correlation matrix for the DS3 damage distribution. The strong correlation of PBS-derived variables underscores their predictive power.

### KNN-imputer features

The features employed to construct the imputed dataset are: Area<sub>m</sub>, Age, Morphology, St, Ss, a, Units<sub>number</sub>, Alt, Area<sub>t</sub>, Vertex<sub>num</sub>, EW and NS coordinates and Vs<sub>30</sub>.

Moreover we set n<sub>neighbors</sub>=8, weights='distance', metric='nan\_euclidean'.

The choice of k=8 instead of the default value k=5 was made to reduce any anomalous or incorrect data that may be present in the dataset. Meanwhile, the choice to adopt weights based on distance (weights='distance') is common for problems that are in some way geo-referenced, such as ours. This means that points closer to the missing ones will have a much greater influence in determining the value to be inserted than those further away. The distribution of Age of Construction before and after the Imputation procedure is reported in Fig. 10. We observe that the distribution of the imputed data is consistent with the original dataset and there are no obvious distortions. The overlaying histograms demonstrate that the imputed values fill the gaps while respecting the original frequency peaks.



**Fig. 10.** Age of contraction distribution before and after the imputation procedure.

### Hyperparameter optimization

The hyperparameter optimization for the case study presented in the “Results” section was conducted using *randomized search with cross-validation* (RandomizedSearchCV). We evaluated 1000 different combinations of hyperparameters, randomly sampled from the search space defined below:

- `n_estimators`: [100, 2000]
  - `max_features`: [`'sqrt'`, `'log2'`]
  - `criterion`: [`'gini'`, `'entropy'`]
  - `max_depth`: [10, 200]
  - `min_samples_split`: [2, 10]
  - `min_samples_leaf`: [1, 10]
- The optimization process yielded a peak accuracy of 0.7853, achieved with the following optimal configuration on the train set: `{n_estimators: 234, min_samples_split: 2, min_samples_leaf: 1, max_features: 'sqrt', max_depth: 200, criterion: 'gini'}`. `max_features: sqrt, max_depth: 200, criterion: gini` } `random_state = 42` was added to ensure the reproducibility of the results.

Received: 16 September 2025; Accepted: 18 March 2026

Published online: 27 March 2026

### References

1. Sun, Z. et al. A review of earth artificial intelligence. *Comput. Geosci.* **159**, 105034 (2022).
2. Zhao, T. et al. Artificial intelligence for geoscience: Progress, challenges and perspectives. *Innovation* (2024).
3. Zhang, W. et al. Application of machine learning, deep learning and optimization algorithms in geoenvironment and geoscience: Comprehensive review and future challenge. *Gondwana Res.* **109**, 1–17 (2022).
4. Monsalvo Franco, I. E. et al. Seismic fragility curves with unconventional ground motion intensity measures from physics-based simulations. *Bull. Earthq. Eng.* **23**(5), 1885–1915 (2025).
5. Rossetto, T., Ioannou, I., Grant, D. & Maqsood, T. *Guidelines for empirical vulnerability assessment* (X, GEM Technical Reports, 2014).
6. Del Gaudio, C. et al. Empirical fragility curves from damage data on rc buildings after the 2009 L'Aquila earthquake. *Bull. Earthq. Eng.* **15**, 1425–1450 (2017).
7. Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z. & Burton, H. V. Classifying earthquake damage to buildings using machine learning. *Earthq. Spectra* **36**(1), 183–208 (2020).
8. Roeslin, S., Ma, Q., Juárez-García, H., Gómez-Bernal, A., Wicker, J. & Wotherspoon, L. A machine learning damage prediction model for the 2017 Puebla-Morelos, Mexico, earthquake. *Earthq. Spectra* **36**(2\_suppl), 314–339 (2020).
9. Harirchian, E. et al. A synthesized study based on machine learning approaches for rapid classifying earthquake damage grades to rc buildings. *Appl. Sci.* **11**(16), 7540 (2021).
10. Stojadinović, Z., Kovačević, M., Marinković, D. & Stojadinović, B. Rapid earthquake loss assessment based on machine learning and representative sampling. *Earthq. Spectra* **38**(1), 152–177 (2021).
11. Di Michele, F. et al. Comparison of machine learning tools for damage classification: The case of L'Aquila 2009 earthquake. *Nat. Hazards* **116**, 3521–3546 (2023).
12. Di Michele, F. et al. Machine learning for damage classification, risk mitigation and post-earthquake management. In *Proceedings of the 7th International Conference on Earthquake Engineering and Seismology. ICEES 2023. Lecture Notes in Civil Engineering*. Vol. 401. 181–190 (2023).
13. Di Michele, F., Stagnini, E., Pera, D., Aloisio, R. & Marcati, P. Random forest based estimate to assess the damages of a future earthquakes: Preliminary results. *Ann. Geophys.* **66** (2024).

14. Faenza, L., Michelini, A., Crowley, H., Borzi, B. & Faravelli, M. Shakedown: A data collection combining earthquake building damage and shakemap parameters for Italy. *Artif. Intell. Geosci.* **1**, 36–51 (2020).
15. A. Cirella, A. Piatanesi, M. Cocco, E. Tinti, L. Scognamiglio, A. Michelini, A. Lomax, E. Boschi. Rupture history of the L'Aquila (Italy) earthquake from non-linear joint inversion of strong motion and GPS data. *Geophys. Res. Lett.* **36**(19), 2009 (2009).
16. Chiarabba, C. et al. The 2009 L'Aquila (central Italy) MW6.3 earthquake: Main shock and aftershocks. *Geophys. Res. Lett.* **36**(18) (2009).
17. Giaccio, B. et al. Fault and basin depocentre migration over the last 2 Ma in the L'Aquila 2009 earthquake region, central Italian Apennines. *Quat. Sci. Rev.* **56**, 69–88 (2012).
18. S. Atzori, I. Hunstad, M. Chini, S. Salvi, C. Tolomei, C. Bignami, S. E. Stramondo, E. Trasatti, A. Antonioli, E. Boschi. Finite fault inversion of DInSAR coseismic displacement of the L'Aquila earthquake (central Italy). *Geophys. Res. Lett.* **36**(15), 2009 (2009).
19. Ameri, G., Gallovič, F. & Pacor, F. Complexity of the Mw 6.3 2009 L'Aquila (central Italy) earthquake: 2. Broadband strong motion modeling. *J. Geophys. Res. Solid Earth* **117**(B4) (2012).
20. Ameri, G. et al. The 6 April 2009 Mw 6.3 L'Aquila (central Italy) earthquake: Strong-motion observations. *Seism. Res. Lett.* **80**(6), 951–966 (2009).
21. Pucci, S. et al. Coseismic ruptures of the 24 August 2016, mw 6.0 amatrice earthquake (central Italy). *Geophys. Res. Lett.* **44**(5), 2138–2147 (2017).
22. Huang, M. H. et al. Coseismic deformation and triggered landslides of the 2016 mw 6.2 amatrice earthquake in Italy. *Geophys. Res. Lett.* **44**(3), 1266–1274 (2017).
23. Lavecchia, G. et al. Ground deformation and source geometry of the 24 August 2016 amatrice earthquake (central Italy) investigated through analytical and numerical modeling of dinosaur measurements and structural-geological data. *Geophys. Res. Lett.* **43**(24), 12–389 (2016).
24. Chiaraluce, L. et al. The 2016 central Italy seismic sequence: A first look at the mainshocks, aftershocks, and source models. *Seismol. Res. Lett.* **88**(3), 757–771 (2017).
25. Liu, C., Zheng, Y., Xie, Z. & Xiong, X. Rupture features of the 2016 mw 6.2 Norcia earthquake and its possible relationship with strong seismic hazards. *Geophys. Res. Lett.* **44**(3), 1320–1328 (2017).
26. Villani, F. et al. A database of the coseismic effects following the 30 October 2016 Norcia earthquake in central Italy. *Sci. Data* **5**(1), 1–11 (2018).
27. Mazziari, I., Stupazzini, M., Guidotti, R. & Smerzini, C. SPEED: SPectral Elements in Elastodynamics with Discontinuous Galerkin: A non-conforming approach for 3D multi-scale problems. *Int. J. Numer. Method Eng.* **95**(12), 991–1010 (2013).
28. Lanzano, G. et al. A revised ground-motion prediction model for shallow crustal earthquakes in Italy. *Bull. Seismol. Soc. Am.* **109**(2), 525–540 (2019).
29. Michelini, A., Faenza, L., Lauciani, V. & Malagnini, L. Shakemap implementation in Italy. *Seismol. Res. Lett.* **79**(5), 688–697 (2008).
30. Wald, D. J. et al. Trinet “shakemaps”: Rapid generation of peak ground motion and intensity maps for earthquakes in southern California. *Earthq. Spectra* **15**(3), 537–555 (1999).
31. May, J.B., Pera, D., Di Michele, F., Aloisio, R., Rubino, B. & Marcati, P. Fast cubit-python tool for highly accurate topography generation and layered domain reconstruction. In *Proceedings of the 29th International Meshing Roundtable* (2021).
32. Tarquini, S. et al. TINITALY/01: A new triangular irregular network of Italy. *Ann. Geophys.* **50**(3), 407–425 (2007).
33. Tarquini, S. et al. Release of a 10-m-resolution DEM for the Italian territory: Comparison with global-coverage DEMs and anaglyph-mode exploration via the web. *Comput. Geosci.* **38**(1), 168–170 (2012).
34. Tarquini, S. & Nannipieri, L. The 10 m-resolution TINITALY DEM as a trans-disciplinary basis for the analysis of the Italian territory: Current trends and new perspectives. *Geomorphology* **281**, 108–115 (2017).
35. Evangelista, L. et al. Physics-based seismic input for engineering applications: A case study in the Aterno River Valley, central Italy. *Bull. Earthq. Eng.* **15**(7), 2645–2671 (2017).
36. Di Michele, F. et al. Spectral element numerical simulation of the 2009 L'Aquila earthquake on a detailed reconstructed domain. *Geophys. J. Int.* **230**(1), 29–49 (2022).
37. Rosti, A., Smerzini, C., Paolucci, R., Penna, A. & Rota, M. Validation of physics-based ground shaking scenarios for empirical fragility studies: The case of the 2009 L'Aquila earthquake. *Bull. Earthq. Eng.* **21**(1), 95–123 (2023).
38. Smerzini, C. & Villani, M. Broadband numerical simulations in complex near-field geological configurations: The case of the 2009 m w 6.3 L'Aquila earthquake. *Bull. Seismol. Soc. Am.* **102**(6), 2436–2451 (2012).
39. Iwaki, A., Maeda, T., Morikawa, N., Miyake, H. & Fujiwara, H. Validation of the recipe for broadband ground-motion simulations of Japanese crustal earthquakes. *Bull. Seismol. Soc. Am.* **106**(5), 2214–2232 (2016).
40. Graves, R. W. & Pitarka, A. Broadband ground-motion simulation using a hybrid approach. *Bull. Seismol. Soc. Am.* **100**(5A), 2095–2123 (2010).
41. Mai, P. M., Imperatori, W. & Olsen, K. B. Hybrid broadband ground-motion simulations: Combining long-period deterministic synthetics with high-frequency multiple s-to-s backscattering. *Bull. Seismol. Soc. Am.* **100**(5A), 2124–2142 (2010).
42. Paolucci, R. et al. Broadband ground motions from 3d physics-based numerical simulations using artificial neural networks broadband ground motions from 3d pbss using anns. *Bull. Seismol. Soc. Am.* **108**(3A), 1272–1286 (2018).
43. Fondazione Eucentre. Da.D.O. (Database di Danno Osservato).
44. T. K. Ho. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol. 1. 278–282 (1995).
45. Raschka, S., & Mirjalili, V. *Python Machine Learning: Machine Learning and Deep Learning with Python. Scikit-Learn, and TensorFlow*. 2nd Ed. (Packt Publishing Limited, 2017).
46. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).
47. Anniballe, Roberta et al. Earthquake damage mapping: An overall assessment of ground surveys and vhr image change detection after L'Aquila 2009 earthquake. *Remote Sens. Environ.* **210**, 166–178 (2018).
48. Harati, M. & Van de Lindt, J. W. Methodology to generate earthquake-tsunami fragility surfaces for community resilience modeling. *Eng. Struct.* **305**, 117700 (2024).

## Acknowledgements

This paper received financial support from ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union - NextGenerationEU under the Italian Ministry of University and Research (MUR) National Centre for HPC, Big Data and Quantum Computing CN 00000013-CUP:E13C22001000006. This work was also partially supported by the GSSI “Centre for Urban Informatics and Modelling” (Italian Government-Presidenza Consiglio dei Ministri-CUIM project delibera CIPE n. 70/2017). All numerical simulations have been realized on the Linux cluster named Caliban of the High Performance Parallel Computing Laboratory of the Department of Information Engineering, Computer Science and Mathematics (DISIM) at the University of L'Aquila.

### Author contributions

Conceptualization: P. Marcati, F. Di Michele, B. Rubino, A. Aloisio. Data Curation: D. Pera, F. Di Michele, E. Stagnini, M. Carafa, V. Kastelic, I. Mazzieri and C. Smerzini. Formal Analysis: D. Pera, F. Di Michele. Funding Acquisition: B. Rubino, A. Aloisio. Writing and figures: F. Di Michele, D. Pera, I. Mazzieri and E. Stagnini. Supervision: P. Marcati, B. Rubino and A. Aloisio.

### Declarations

#### Competing interests

The authors declare no competing interests.

#### Additional information

**Correspondence** and requests for materials should be addressed to F.M. or D.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026