ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Explainable ensemble learning for structural damage prediction under seismic events

Michele Baldassini, Pierfrancesco Foglia, Beatrice Lazzerini, Francesco Pistolesi*, and Cosimo Antonio Prete

University of Pisa - Department of Information Engineering
Largo Lucio Lazzarino, 1 - 56122 Pisa (Italy)

Abstract. This paper presents an explainable ensemble learning framework using Bootstrap Aggregating to predict structural damage in masonry buildings during seismic events. It estimates the peak ground acceleration (PGA) leading to the damage control limit state (significant damage) based on structural parameters. The model achieves high accuracy (R^2=0.9536, MAE=0.0057) and interpretability through SHAP, aligning with structural engineering principles. Compared to finite element analyses, it offers faster computations (milliseconds) and scalability, enabling rapid intervention planning after earthquakes. Developed under the *MEDEA* project (EU Grant n. 10101236), it supports disaster response and enhances seismic resilience.

1 Introduction

Seismic events pose a significant threat to the structural integrity of buildings, which collapse when they can no longer sustain lateral loads during an earthquake. Accurate prediction of the structural response under seismic loading is thus key for risk assessment and mitigation. Building response is evaluated using *Engineering Demand Parameters (EDPs)*, such as internal forces and relative displacements, to estimate damage levels called *limit states*: D1 (minimal), D2 (moderate), D3 (significant), and D4 (severe) [1].

Traditional approaches, such as those based on the finite element method (FEM), are widely used but may have high computational costs and scalability limitations, particularly when applied to large datasets or real-time evaluations [2, 3]. Recent advancements in artificial intelligence (AI) have shown promising results in seismic damage prediction, offering faster and more scalable solutions to model the seismic behavior of buildings. For example, with artificial neural networks [4, 5], support vector machines [6], and decision trees [7].

*Corresponding author: Francesco Pistolesi `francesco.pistolesi@unipi.it`

This work is part of the European Project *MEDEA: Multidimensional seismic risk assessment combining structural damages and psychological consequences using explainable artificial intelligence* (n. 101101236), co-funded by the European Commission, Directorate-General for European Civil Protection and Humanitarian Aid Operations (ECHO) ECHO.B - Disaster Preparedness and Prevention B.2 - Prevention and Disaster Risk Management. This work was supported in part by the Italian Ministry of University and Research (MUR) in the frameworks of the *CrossLab* and *FoReLab* projects (Departments of Excellence), and by the PNRR-M4C2-Investimento 1.3, Partenariato Esteso PE00000013-FAIR-Future Artificial Intelligence Research-Spoke 1 Human-Centered AI.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

However, limitations include focusing on specific structural configurations and often disregarding key damage thresholds, such as the peak ground acceleration (PGA), i.e., the acceleration that leads a building to reach critical limit states that precede collapse [8]. Also, existing methods typically lack interpretability, a key requirement that helps experts work alongside these systems and validate predictions against structural engineering principles [9].

This paper presents an explainable ensemble learning framework that predicts the damage to masonry buildings during seismic events, based on structural parameters. The framework uses Bootstrap Aggregating and SHapley Additive exPlanations (SHAP) [10] to achieve high accuracy and interpretability. To design and train the framework, we created a large dataset of 1176 masonry buildings. Each sample comprised the structural parameters of a building associated with the PGA at the D3 limit state obtained via nonlinear FEM pushover analysis. Our framework accurately predicts the PGA at the D3 limit state while providing SHAP explanations consistent with structural engineering principles. Compared to FEM, our approach significantly reduces computational requirements, making it suitable for large-scale or real-time applications. The paper is organized as follows: Section 2 presents the dataset; Section 3 describes the experiments; Section 4 discusses the results; Section 5 draws the conclusions.

2 Dataset

We generated a dataset of 1176 benchmark masonry structures representing typical constructions built between 1945 and 1990 in regions targeted by the MEDEA project's pilot studies. The base configuration (see Fig. 1f and Fig. 1g) had a rectangular layout (12.4 m by 9.7 m) with a floor height of 2.85 m, two internal masonry walls (0.3 m thick), and slabs of reinforced concrete and clay elements. Buildings were generated by combining the number of levels (2-4), floor height (2.85-3.40 m), number of internal alignments (1 or 2), wall thickness (0.30-0.45 m), and opening area (6.2-21.6 m^2). Mechanical properties were sampled from the sets as follows: shear strength {0.028, 0.04, 0.09, 0.05} MPa, compressive strength {1.4, 2.0, 5.8, 2.6} MPa, gross density {16, 16, 22, 18} kN/m^3, elastic modulus {1080, 1410, 2850, 1500} MPa, and shear modulus {360, 450, 950, 500} MPa. We combined these mechanical properties to generate various building layouts: *disordered stone*, *regular ashlar*, *squared stone block*, and *clay brick masonry*. Figures 1a-1e show example building layouts considered.

Each dataset sample comprised 18 features from expert advice: number of levels; average floor height; side length ratio; total floor area; number of internal wall alignments in the x and y directions; area of the openings in external and internal walls; thickness of external and internal walls; shear strength; compressive strength; gross density; elastic modulus; shear modulus; effective shear area; seismic floor mass; and mass-to-shear area ratio. These features influence stability, stiffness, and resistance to seismic loads [11], determining the structure's ability to absorb and dissipate forces without collapsing (e.g., the shear modulus measures a material's resistance to shear deformation when subjected to a force

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

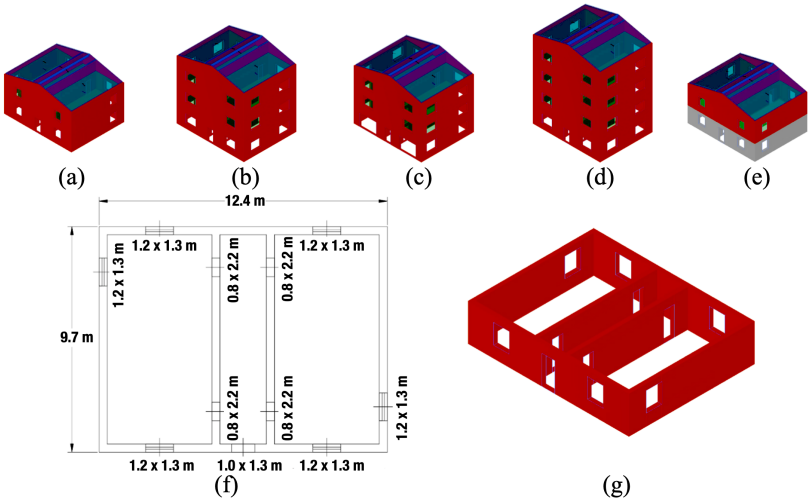


Fig. 1: Example structures: (a) two-level; (b) and (c) three-level; (d) four-level; (e) with two types of masonry; (f) plan view and (g) perspective view of the basic structural configuration used to generate masonry structures.

parallel to its surface). The first six features are *global* and describe the overall structure; the others are *local* and measured for each building level. Based on expert advice, we standardized the input to 30 values per building, regardless of the number of levels: the global features, and the local features of the ground and top levels. Features were numbered as follows: F_0-F_5 (global); $F_6, F_8, \ldots, F_{28}$ (local, ground floor); $F_7, F_9, \ldots, F_{29}$ (local, top floor). The dataset also included PGA values for D1-D4 damage states, derived from pushover analysis.

3 Experiments

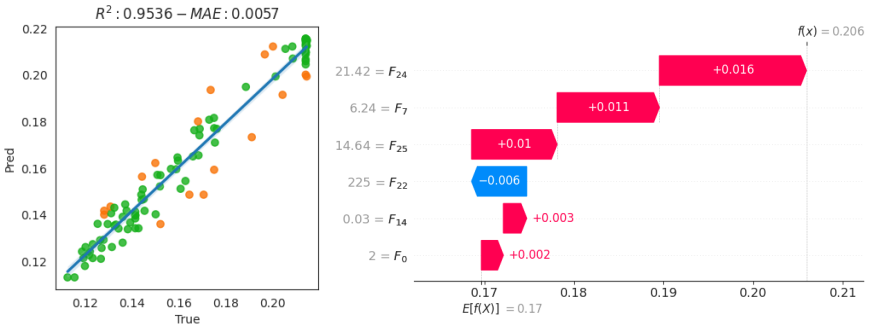
We developed and compared various AI models to predict the PGA at the D3 limit state, using *roof displacement* as an EDP. The D3 state indicates buildings needing structural intervention after a seismic event. We trained the decision tree, k-nearest neighbors (k-NN), random forest, bagging, adaptive boosting, gradient boosting, extreme gradient boosting (XGBoost), and light gradient boosting (LightGBM). Ensemble techniques were prioritized for reliability in high-dimensional, complex problems [12].

First, the dataset was randomly split into a training set (80%) and a test set (20%). Both sets underwent *z-score* normalization, using (μ, σ) derived from the training samples. Then, we jointly optimized the hyperparameters and selected relevant features, ensuring an unbiased evaluation. After defining a set of possible values for each hyperparameter via grid sampling, we generated hyperparameter configurations by combining these values in all possible ways. For each configuration, Sequential Feature Selection (SFS) identified the best features using samples from the training set. The SFS used the mean squared error (MSE) as loss function; performance was evaluated using the R^2 metric

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Table 1: Selected features and performance metrics on the test set.

Model	Selected Features (IDs)	R^2	MAE
Bagging	$F_0, F_7, F_{14}, F_{22}, F_{24}, F_{25}$	**0.9536**	**0.0057**
Random Forest	$F_0, F_4, F_6, F_7, F_{13}, F_{16}, F_{23}, F_{25}$	0.9474	0.0057
Light GBM	$F_0, F_6, F_{14}, F_{15}, F_{20}, F_{23}, F_{25}, F_{29}$	0.9531	0.0056
k-NN	$F_0, F_4, F_6, F_7, F_{14}, F_{20}, F_{22}, F_{24}, F_{27}$	0.9499	0.0056
Decision Tree	$F_0, F_7, F_{13}, F_{14}, F_{22}, F_{24}, F_{25}$	0.8769	0.0095
XGBoost	$F_0, F_7, F_{14}, F_{20}, F_{24}, F_{25}, F_{26}, F_{28}$	0.9534	0.0057
Gradient Boosting	$F_0, F_4, F_7, F_{10}, F_{11}, F_{14}, F_{22}, F_{23}, F_{25}$	0.9528	0.0058
Adaptive Boosting	$F_0, F_4, F_7, F_{11}, F_{14}, F_{25}, F_{28}$	0.9457	0.0059



(a) (b)

Fig. 2: Regression plots, R^2 and MAE on the test set of the bagging model (a) and SHAP local explanation for one sample (b).

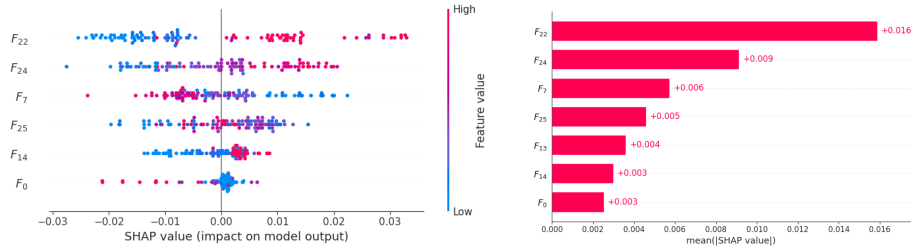
with the given feature set and hyperparameter configuration. After selecting the best features for a given configuration, 10-fold cross-validation was repeated 30 times to evaluate performance. The samples in each fold were normalized with (μ, σ) obtained from the training folds and folds were repopulated at each iteration. For each configuration, the average R^2 across all runs determined the effectiveness of the feature set and hyperparameter values.

We used the best-performing combination of features and hyperparameters to train the AI models on the training set. Table 1 summarizes the performances on the test set of all models. As can be seen, the best one was based on Bagging with decision trees as base learners. Fig. 2a shows its regression plot with R^2 and mean absolute error (MAE) metrics, where the points in orange are samples with errors higher than 1.5 $\times$ MAE. The hyperparameters were optimized by exploring the number of decision trees in {10, 20, 50, 100, 200} and the maximum depth in {5, 10, 15, 20} to balance computational efficiency and accuracy. Higher values did not increase performance.

4 Discussion

The framework allows experts to interpret results based on Shapley explanations. The SHAP plots in Fig. 3 show the most influential features and their values.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.



(a) (b)

Fig. 3: SHAP beeswarm plot (a) and bar plot (b).

The beeswarm plot in Fig. 3a provides a global explanation with the distribution of SHAP values for each feature across all samples. The x-axis position indicates the SHAP value, representing (feature impact on the predicted PGA); the color indicates the feature value. High (low) SHAP values push predictions towards higher (lower) PGAs, indicating better (worse) structural resilience. Key features such as F_{22} (shear modulus at the ground floor [GF]) and F_{24} (effective shear area at GF) dominate the plot. For F_{22}, higher values (red) consistently correspond to positive SHAP values, reflecting stiffer masonry materials that enhance lateral resistance and increase the predicted PGA. Likewise, wider effective shear areas (F_{24}) lead to higher PGA predictions, representing stronger walls that can carry greater seismic loads. Conversely, features like F_7 (area of openings in external masonry walls at the top floor [TF]) show that higher values (red) lead to lower predicted PGA values. This reflects a diminished wall stiffness due to large openings, which increases vulnerability to seismic forces.

The bar plot in Fig. 3b aggregates the average absolute SHAP values across samples, highlighting F_{22}, F_{24}, and F_7 as the most relevant features. The others, such as F_{25} (effective shear area at TF), F_{14} (average shear strength of masonry at GF), and F_0 (number of levels), represent interactions between material properties, structural geometry, and seismic behavior. For example, the average shear strength captures the masonry capacity to resist shear-induced failure, while the number of levels reflects the gravitational load and height-related vulnerabilities affecting the dynamic response of the structure.

The waterfall plot in Fig. 2b provides a local explanation with feature contributions for one sample (building) of those in Fig. 2a. As the plot shows, the feature values contribute to the predicted PGA (0.206, top-right) by increasing/decreasing the base value ($E[f(x)]$), i.e., the mean predicted PGA across all test samples. Features F_{24} (effective shear area at GF), F_7 (area of the openings of the external masonry walls at TF), and F_{25} (effective shear area at TF) have the largest positive impacts, namely, +0.016, +0.011, +0.01. This aligns with masonry mechanics, where severe seismic damage often results from shear failure in panels when the applied shear force exceeds their shear strength. A masonry structure can then resist lateral forces based on the effective shear area—reduced by openings and the material's shear strength. These factors determine the structure's overall lateral load-bearing capacity.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

SHAP explanations confirmed that the model captures critical engineering principles, such as the importance of material stiffness, load distribution, and the detrimental effects of excessive openings or insufficient shear resistance.

Regarding computational times, the Bagging model makes predictions in ∼8 milliseconds per building, while pushover analysis takes 5-15 minutes. Also, the model eliminates the need for detailed configuration required by pushover analyses, further accelerating the process and enabling quick, large-scale seismic risk assessments. This makes the framework ideal when assessing the risk of many buildings, as typical after an earthquake.

5 Conclusions

This paper has presented an explainable ensemble learning framework for predicting structural damage in masonry buildings under seismic events. Using Bagging with decision trees and SHAP analysis, the model showed high accuracy (R^2=0.9536, MAE=0.0057) while maintaining consistency with structural engineering principles. The framework offers efficient and interpretable predictions of the PGA at the damage control (D3) limit state, significantly reducing computational time compared to FEM analyses. This helps speed up post-earthquake safety assessment and intervention planning.

References

[1] C. Kang, et al. Evaluation of correlation between engineering demand parameters of structures for seismic system reliability analysis. *Structural Safety*, 93, 2021.

[2] J. M. Mayoral, et al. Site effects in mexico city basin: Past and present. *Soil Dynamics and Earthquake Engineering*, 121:369–382, 2019.

[3] FEMA P-58-1. *Seismic Performance Assessment of Buildings – Volume 1 – Methodology*. Federal Emergency Management Agency, Washington, D.C., 2nd edition, 2018.

[4] R. Falcone, et al. Soft computing techniques in structural and earthquake engineering: A literature review. *Engineering Structures*, 207:110269, 2020.

[5] N. Sharma, et al. Natural period of reinforced concrete building frames on pile foundation considering seismic soil-structure interaction effects. *Structures*, 27:1594–1612, 2020.

[6] H. A. Farfani, et al. Dynamic analysis of soil-structure interaction using neural networks and support vector machines. *Expert Systems with Applications*, 42(22):8971–8981, 2015.

[7] S. Liu, et al. Machine learning-based seismic prediction of building structures. In *Proc. of the 2024 International Conference on Machine Intelligence and Digital Applications*, pages 256–261, 2024.

[8] J. Won et al. Machine learning-based approach for seismic damage prediction of building structures considering soil-structure interaction. *Sustainability*, 13(8):1–14, 2021.

[9] N. K. Psyrras et al. Build-x: Expert system for seismic analysis and assessment of 3d buildings using opensees. *Advances in Engineering Software*, 116:23–35, 2018.

[10] S. M. Lundberg et al. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 4768–4777. Curran Associates Inc., 2017.

[11] Ž. Radovanović, et al. The mechanical properties of masonry walls - analysis of the test results. *Procedia Engineering*, 117:865–873, 2015.

[12] Z.-H. Zhou. *Ensemble Methods: Foundations and Algorithms*. Chapman & Hall, 2012.