Deliverable 4.1.: XAI Models

MULTIDIMENSIONAL SEISMIC RISK ASSESSMENT COMBINING STRUCTURAL DAMAGES AND PSYCHOLOGICAL CONSEQUENCES USING EXPLAINABLE ARTIFICIAL INTELLIGENCE

Co-funded by the
European Union

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Deliverable Title	**XAI Models**
Deliverable number	**4.1**
Deliverable Lead:	UNIPI
Related Work Package:	WP4: Benchmark structures, EDPs, and representative families
Author(s):	Francesco Pistolesi (UNIPI), Michele Baldassini (UNIPI)
Dissemination Level:	Public
Due Submission Date:	29/02/24
Actual Submission:	04/03/24
Project Number	101101236
Status	Version 1.0 (29/02/2024)
Reviewed (Authors)	Salvatore Verre, Riccardo Panico, Fabrizio Comodini, Francesco Focacci (eCampus University)
Start Date of Project:	01/12/22
Duration:	24 months
Abstract	This deliverable provides a comprehensive overview of the results achieved using *explainable artificial intelligence (XAI)* models to evaluate structural damage resulting from seismic events. A diverse set of XAI models has been formulated and fine-tuned through careful experimentation and iterative refinement to address the complexity of structural assessment post-seismic activity. The development process encompassed rigorous testing across an extensive range of parameter value configurations to explain the most effective methodologies in accurately estimating the extent of damage incurred by various structures. The results highlighted the efficacy of XAI techniques in delivering precise and reliable estimations of potential structural damage levels based on peak ground acceleration. The demonstrated high precision in damage estimation represents a noteworthy advancement in predictive analytics within structural engineering and holds profound implications for enhancing risk mitigation strategies and facilitating informed decision-making processes in disaster management scenarios.
Status changes history	Version 1.0 (29/02/2024) – First release

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

TABLE OF CONTENTS

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

ABBREVIATIONS

MEDEA	**Multidimensional Seismic Risk Assessment Combining Structural Damages And Psychological Consequences Using Explainable Artificial Intelligence**
eCampus	eCampus University
UNIPI	Università di Pisa
MED	Medjimurje County
GZS	Gasilska Zveza Slovenije
ANCI	Associazione Nazionale Comuni Italiani (National association of Italian municipalities)
SISST	Società Italiana per lo studio dello Stress Traumatico (Italian Society for the Study of Traumatic Stress)
XAI	eXplainable Artificial Intelligence
PTSD	Post-Traumatic stress disorder

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

1. INTRODUCTION

The overarching goal of the MEDEA project is to enhance cross-border disaster risk management by focusing on prevention and preparedness in Europe and neighboring EU countries. Specifically, the proposal aims to mitigate the impact of seismic events and enhance resilience, defined as the capacity to withstand, absorb, adapt to, and recover from earthquakes efficiently and promptly. To achieve this objective, the project proposes developing an intelligent system for multidimensional seismic risk assessment in cross-border regions. Using artificial intelligence, this system aims to estimate earthquake-induced losses by predicting structural damage, such as building collapses, while also forecasting the psychological ramifications for affected individuals. Integrating psychological consequences, the project will investigate familial and individual factors and relational and contextual aspects that may exacerbate psychological distress among family members in the aftermath of seismic events. By assessing potential medium and long-term psychological effects on those involved in earthquakes, the project seeks to identify high-risk families susceptible to psychological distress, thereby anticipating and preventing the onset of post-traumatic stress disorder (PTSD).

Within the framework of the MEDEA project, the specific objective of Work Package 4, "*XAI Models*", is to design and develop eXplainable Artificial Intelligence (XAI) models and data fusion techniques to predict the damages to structures as a consequence of seismic event. This was done by estimating a structure's peak ground acceleration to achieve the D3 limit state, starting from a set of technical parameters describing the structure and considering the roof displacement at the elevation floor as an engineering demand parameter. We achieved a precision higher than 95% in estimating this PGA value, and the explanations given by the models are in line with those given by experts in structural engineering.

This deliverable describes the results obtained and is organized as follows: Section 2 gives an overview of machine learning, regression, and classification problems; Section 3 presents the results achieved; Section 4 draws the conclusions.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

2. ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Machine Learning is the art and science of teaching computers to learn from data, uncover patterns, and make informed decisions without explicit programming instructions. It represents a departure from traditional rule-based programming paradigms, where developers meticulously craft algorithms to perform specific tasks. Instead, Machine Learning algorithms can generalize from examples, extrapolate insights, and improve performance over time as they ingest more data. The origins of Machine Learning can be traced back to the mid-20th century, with pioneers like Arthur Samuel laying the groundwork for what would become a transformative field. Initially rooted in artificial intelligence, Machine Learning experienced significant advancements fueled by computational advancements, algorithmic breakthroughs, and the explosion of data availability in the digital age. Today, Machine Learning permeates nearly every facet of our daily lives, from personalized recommendations on streaming platforms to autonomous vehicles navigating city streets. Its applications span diverse domains, including healthcare, finance, manufacturing, marketing, and beyond, revolutionizing processes, enhancing decision-making, and unlocking unprecedented insights from complex datasets.

The algorithms underpinning its functionality are central to the efficacy of Machine Learning. These algorithms encompass a spectrum of techniques, from classical methods like linear regression and decision trees to cutting-edge deep learning architectures such as convolutional neural networks and recurrent neural networks. Each algorithmic approach has strengths, weaknesses, and suitability for particular tasks, enabling practitioners to tailor solutions to specific problems effectively. The proliferation of Machine Learning increased with the availability of robust frameworks, libraries, and tools that democratize its application. Platforms like TensorFlow, PyTorch, and scikit-learn provide accessible interfaces for developing, deploying, and scaling Machine Learning models, lowering barriers to entry and empowering a broader community of researchers, engineers, and enthusiasts to harness its potential. However, amidst its remarkable advancements and transformative potential, Machine Learning also grapples with inherent challenges and ethical considerations. Issues surrounding bias, fairness, transparency, and accountability underscore the importance of responsible AI development and governance frameworks to mitigate unintended consequences and ensure equitable outcomes for all stakeholders. Looking ahead, the trajectory of Machine Learning promises continued innovation and

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

evolution, driven by interdisciplinary collaboration, interdisciplinary collaboration, and the convergence of fields such as computer science, statistics, mathematics, and domain-specific expertise. As algorithms grow more sophisticated, data more abundant, and computational resources more powerful, the boundaries of what's achievable with Machine Learning continue to expand, heralding a future where intelligent systems work in tandem with humans to tackle complex problems and advance society.

2.1 Classification and Regression

Classification and regression are two fundamental tasks in machine learning. These tasks enable systems to make sense of data, infer relationships, and make informed decisions in various applications. Classification and regression represent distinct but interconnected paradigms within machine learning, each tailored to address specific types of problems and data. While *classification* involves predicting discrete class labels or categories, *regression* focuses on estimating continuous numerical values. Together, they form the foundation of supervised learning (see Figure 1), where models are trained on labeled data to make predictions based on input features. In *classification tasks*, the goal is to assign input data points to predefined categories or classes based on their features. This could involve distinguishing between spam and legitimate emails, identifying handwritten digits in images, or predicting a patient's disease likelihood based on medical test results.

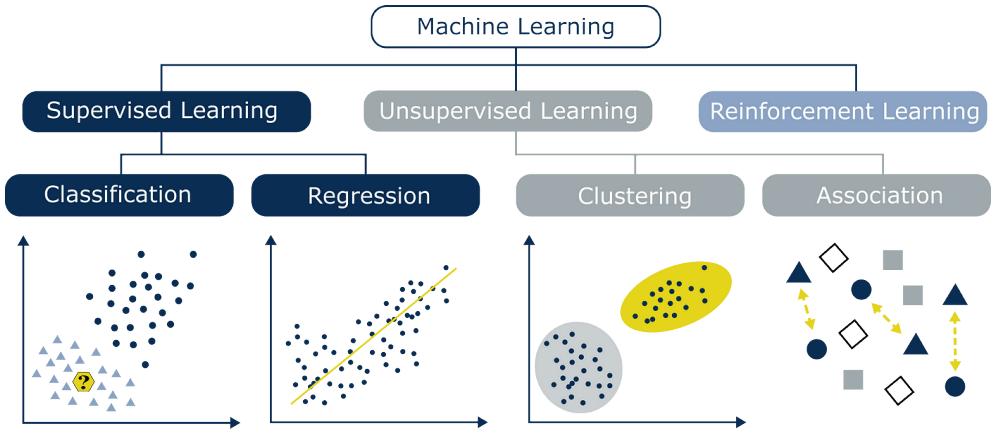


Figure 1: Machine learning tasks.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Classification algorithms learn decision boundaries in the feature space that separate different classes, allowing them to classify new, unseen data points accurately. Common algorithms for classification include *logistic regression*, *decision trees*, *support vector machines (SVM)*, *k-nearest neighbors (KNN)*, and, of course, *neural networks*. Each algorithm has strengths, weaknesses, and suitability for different data types and problem domains. For instance, decision trees excel at handling categorical data and capturing nonlinear relationships, while SVMs are effective in high-dimensional spaces with clear separation between classes. On the other hand, *regression tasks* involve predicting a continuous numerical value based on input features. This could include forecasting stock prices, estimating house prices based on property features, or predicting the temperature based on weather variables. Regression algorithms learn to model the relationship between input features and output values, allowing them to make accurate predictions for new data points. Popular regression algorithms include *linear regression*, *polynomial regression*, *decision trees*, *random forests*, *gradient boosting machines (GBM)*, and *neural networks*. Like classification algorithms, each regression technique has its strengths and weaknesses, making it important to select the most appropriate algorithm based on the data's characteristics and the problem's nature. In both classification and regression, the performance of a machine learning model is evaluated using metrics such as *accuracy*, *precision*, *recall*, *F1 score*, *mean squared error (MSE)*, and *R-squared*. These metrics provide insights into how well the model generalizes to unseen data and can help guide the model selection and optimization process. While classification and regression represent powerful tools for predictive modeling, they are not without challenges. Overfitting, where a model learns to memorize training data rather than generalize to new data, and underfitting, where a model is too simplistic to capture the underlying patterns in the data, are common pitfalls that must be addressed through techniques such as regularization, cross-validation, and ensemble learning. Moreover, the choice of features, data preprocessing, and feature engineering play crucial roles in the performance of classification and regression models. Feature selection techniques such as principal component analysis (PCA), feature scaling, and normalization can help improve model performance and efficiency by focusing on the most relevant information in the data.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

2.2 Explainable AI

In an era where machine learning models play an increasingly pivotal role in decision-making across diverse domains, the demand for transparency, accountability, and trustworthiness has never been greater. As these models permeate critical sectors such as healthcare, finance, criminal justice, and autonomous systems, understanding how they arrive at their predictions and recommendations is paramount for ensuring fairness, mitigating biases, and fostering user acceptance. The explainability of machine learning models refers to the ability to understand and interpret the inner workings of these complex algorithms, elucidating the factors and features that drive their decisions. It encompasses the technical mechanisms underlying model predictions and the broader socio-ethical implications of algorithmic decision-making on individuals and society. The pursuit of explainable AI depends on the balance between achieving high performance and maintaining interpretability.

While the state-of-the-art machine learning models, particularly deep neural networks, often achieve remarkable accuracy and predictive power, their black-box nature can challenge understanding how they arrive at their conclusions. This lack of transparency can hinder adoption, especially in high-stakes applications where trust and accountability are paramount. The importance of explainability extends beyond mere curiosity about how algorithms work. In contexts such as healthcare, where decisions impact patient outcomes, or in legal proceedings, where individual rights are at stake, the ability to provide clear, interpretable explanations for algorithmic decisions is crucial for ensuring fairness, mitigating biases, and upholding ethical standards. Furthermore, explainable AI is essential for fostering user trust and acceptance of machine learning systems. Users are more likely to embrace AI-powered technologies when they can understand and validate the reasoning behind algorithmic decisions. Conversely, opaque or incomprehensible outputs can lead to skepticism, distrust, and even rejection of AI systems, undermining their effectiveness and potential societal benefits. Regulatory and ethical considerations further underscore the need for explainability. Increasingly, policymakers and regulatory bodies recognize the importance of transparency and accountability in algorithmic decision-making, enacting laws and guidelines that mandate the explainability of AI systems, particularly in sensitive domains like healthcare, finance, and criminal justice.

Fortunately, explainable AI is rapidly evolving, with researchers developing various techniques and methodologies to shed light on the inner workings of machine learning models. These approaches

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

range from model-specific interpretability methods, such as feature importance scores and attention mechanisms in neural networks, to model-agnostic techniques like *LIME (Local Interpretable Model-agnostic Explanations)* and *SHAP (SHapley Additive exPlanations)*, which provide post hoc explanations for any black-box model. Moreover, advancements in interdisciplinary research, drawing on insights from computer science, statistics, psychology, and philosophy, enrich our understanding of what constitutes a meaningful explanation in AI. Concepts such as *causal reasoning*, *counterfactual explanations*, and *human-centered design* are gaining traction as researchers seek to bridge the gap between technical accuracy and human comprehension in explaining machine learning models.

The pursuit of explainable AI holds profound implications for future AI development and deployment. As machine learning models become increasingly integrated into society's fabric, the ability to provide transparent, interpretable explanations for their decisions will be essential for fostering trust, ensuring fairness, and maximizing the societal benefits of AI technology. By embracing the principles of explainability, we can navigate the complex interplay between technical innovation, ethical considerations, and human values in pursuing responsible AI.

3. DATASET

The buildings and related technical parameters described in this section are related to masonry (M) buildings. Other buildings with related structural technical parameters describing reinforced concrete (RC) buildings are presented in Appendix A.

3.1 Masonry buildings (M)

The input parameters are divided into three macro categories. In particular, the first one is focused on the material and mechanical parameters characterizing masonry such as *elastic* ($\boldsymbol{E_f}$) and *shear modulus* ($\boldsymbol{G_m}$), *density* ($\boldsymbol{w_m}$), and *average value of compressive* ($\boldsymbol{f_m}$) and *shear strength* ($\boldsymbol{\tau_m}$). Specifically, the material and mechanical parameters used are found in the current/previous *Technical Standards* for the mechanical characterization of stone and clay-brick masonry. The second category focuses on the geometry of the masonry structure, both at the *global level* and at the *local level*. At the *global geometry* level, the parameters are the *number of floors* of which the

Co-funded by the European Union

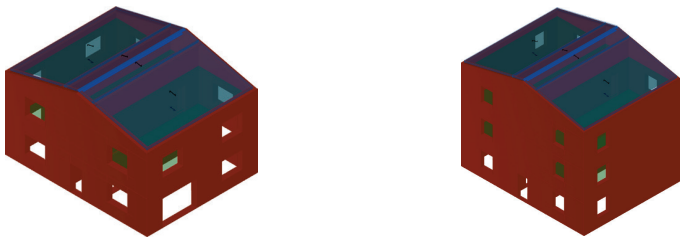
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

masonry structure is composed (n_f), the *average value of the height* of the individual floors (h_f), the ratio of the length of the masonry structure along the two main *X- and Y-directions* (*sides ratio (x-length / y-length)*), the *number of internal shear walls* (evaluated in both main directions), and the *floor area*.

At the *local geometry* level the parameters are evaluated on a floor-by-floor basis. Specifically they are the *average value of the thickness* of the individual (*internal and external*) *shear walls* (t_m), the *area of the openings* (doors and windows) present on each level, and the *the effective shear area*, obtained by subtracting the area of the openings from the cross-sectional area of the walls.

The parameters belonging to the second macro category manage to describe all structures, at the geometric level in a comprehensive and timely manner, even those with non-regular geometries in plan and elevation. Finally, the last macro category is focused on *Seismic Analysis*, which takes into consideration the mass of the individual floor (S_M).

The number of input data for masonry structures depends on the number of floors, since some parameters, such as *openings*, *average shear wall thickness*, etc., must be evaluated individually for each floor. Specifically, the parameters for a two-level masonry structure (called base) are 30, whereas for a three-level masonry structure, they are 42. In other words, 12 parameters are added for each additional floor compared to the base structure. Figure 2 shows the 3D model of masonry structures for 2 and 3 levels. In addition, the number of internal shear walls of 1, 2 (arranged along the two main directions and staggered) was considered. Figure 3 shows two internal shear walls.



(a) (b)

Figure 2: 3D model of masonry structures: (a) 2 levels, and b) 3 levels.

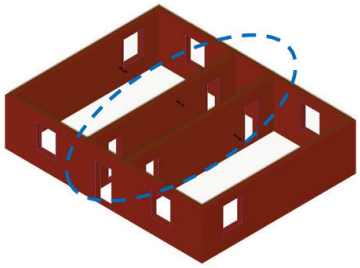


Figure 3: Two internal shear walls.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

In this deliverable, based on considerations by experts in structural engineering, each building is described by 6 global technical parameters (from F_0 to F_5), and each level of a masonry structure is described by 12 technical parameters (from F_6 to F_{17}). Table 1 summarizes the 18 technical parameters considered.

Table 1: Technical parameters for masonry building.

ID	Description
F_0	Number of levels
F_1	Average floor height
F_2	Sides ratio (x-length / y-length)
F_3	Floor area
F_4	Number of internal alignments of masonry wall in X-direction
F_5	Number of internal alignments of masonry wall in Y-direction
F_6	Area of the openings of the external masonry walls
F_7	Area of the openings of the internal masonry walls
F_8	Average thickness of external masonry shear wall
F_9	Average thickness of internal masonry shear wall
F_{10}	Average shear strength of masonry
F_{11}	Average compressive strength of masonry
F_{12}	Masonry gross density
F_{13}	Elastic modulus
F_{14}	Shear modulus
F_{15}	Effective shear area
F_{16}	Seismic floor mass
F_{17}	Ratio seismic floor mass / Effective shear area

3.2 Preprocessing the dataset

Each sample in the dataset describes a building through a set of structural parameters. Each sample is also characterized by the *Peak Ground Acceleration (PGA)* values for which the structure reaches four *limit states* determined by the PGA the building can withstand. These limit states are defined as follows:

- *D1 - Immediate Occupancy (IO):*

  In this state, the structure experiences minimal to no damage during the seismic event. The building remains fully functional and safe for occupancy immediately after the earthquake. The structural response tends to be linearly elastic, meaning the structure behaves within the elastic range of its materials. No immediate repair work is required for the building to resume normal operations.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

- *D2 - Life Safety (LS):*

Structures in this state have sustained moderate damage but are still deemed safe for occupancy. Occupants can evacuate safely during the earthquake, with a low risk of life-threatening collapse. Some repair work may be needed to address the damage, but the structural integrity remains intact.

- *D3 - Damage Control (DC):*

This state indicates significant damage to the structure, rendering it temporarily unusable. Occupants must evacuate the building due to safety concerns, and extensive repair work is required before reoccupation. The damage can be repaired, but it is substantial enough for thorough assessment and repair efforts.

- *D4 - Collapse Prevention (CP):*

In the collapse prevention state, the structure has incurred severe damage that compromises its integrity. There's a high risk of imminent collapse, and the building may need to be evacuated permanently. Repair may not be feasible, and demolition or significant reconstruction is often necessary to prevent catastrophic failure.

These seismic damage limit states provide a framework for assessing and categorizing the level of damage sustained by structures during earthquakes, helping engineers and authorities make informed decisions regarding safety, evacuation, and repair efforts.

The definition of building collapse is not unique. However, *collapse* can be defined as the inability of the structures to guarantee a given performance for a given earthquake. At collapse, a structure loses its capability to carry lateral loads. Collapse can be assessed through either local or global *engineering demand parameters (EDPs)*. Collapse depends on a number of factors, all affected by the cyclic response of the structure to a given earthquake.

As stated in the previous section, the dataset of structures included two-, three- and four-level buildings described by 30, 42, and 54 technical parameters, respectively.

After a preliminary study, we made the number of technical parameters describing a structure independent of the number of its levels. This led to designing and training various machine learning

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

models whose number of inputs is the same. In particular, we considered all the global parameters, and the local parameters relative to the first and last levels of a structure.

Table 2: Global and local technical parameters considered for two-, three- and four-level buildings.

ID	Description
F_0	Number of levels
F_1	Average floor height
F_2	Sides ratio (x-length / y-length)
F_3	Average floor area
F_4	Number of internal alignments of masonry wall in X-direction
F_5	Number of internal alignments of masonry wall in Y-direction
F_6	Area of the openings of the external masonry walls at the ground floor (GF)
F_7	Area of the openings of the external masonry walls at the elevation floor (EF)
F_8	Area of the openings of the internal masonry walls at the ground floor (GF)
F_9	Area of the openings of the internal masonry walls at the elevation floor (EF)
F_{10}	Average thickness of external masonry shear wall (GF)
F_{11}	Average thickness of external masonry shear wall (EF)
F_{12}	Average thickness of internal masonry shear wall (GF)
F_{13}	Average thickness of internal masonry shear wall (EF)
F_{14}	Average shear strength of masonry (GF)
F_{15}	Average shear strength of masonry (EF)
F_{16}	Average compressive strength of masonry (GF)
F_{17}	Average compressive strength of masonry (EF)
F_{18}	Masonry gross density (GF)
F_{19}	Masonry gross density (EF)
F_{20}	Elastic modulus (GF)
F_{21}	Elastic modulus (EF)
F_{22}	Shear modulus (GF)
F_{23}	Shear modulus (EF)
F_{24}	Effective shear area (GF)
F_{25}	Effective shear area (EF)
F_{26}	Seismic floor mass (EF)
F_{27}	Seismic floor mass – roofing plan (RP)
F_{28}	Ratio of Seismic floor mass / Effective shear area (GF)
F_{29}	Ratio of Seismic floor mass / Effective shear area (EF)

Regarding two-level buildings, we considered all 24 local technical parameters (12 technical parameters per level); the 24 local technical parameters related to the first and last levels were considered for three- and four-level buildings. As a result, each building was described by a set of 30 technical parameters as follows:

- 12 local parameters regarding the first level;

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

- 12 local parameters regarding the last level;
- 6 global parameters.

Table 2 summarizes all the parameters used in the experiments. Parameters from F_0 to F_5 are global parameters; local parameters regarding the ground floor and elevation floor (last level) are indicated with (GF) and (EF) in the description, respectively.

4. XAI MODELS

The XAI models developed in this phase of the project are designed to take a specified set of technical parameters provided as input and predict the *Peak Ground Acceleration (PGA)* required for a building to reach the D3 limit state based on the roof displacement EDP. Each machine learning model thus solves a *regression problem* (see Section 2.1). Limit state D3 represents the situation where a building is not immediately usable after the seismic event, and the damage must be repaired before the building can be considered usable again. The D3 limit state is thus the state that distinguishes buildings that need repair after an earthquake from those that can be immediately used as they do not exhibit damage requiring immediate repair. The system architecture is represented in Figure 4. Each model may have different inputs based on optimization techniques to maximize the model performance, i.e., estimating the PGA for the D3 limit state with the lowest error possible. Estimating the PGA of the D3 limit state after an earthquake thus allows us to know which buildings reported a level of damage for which human intervention is required promptly to guarantee people's security.

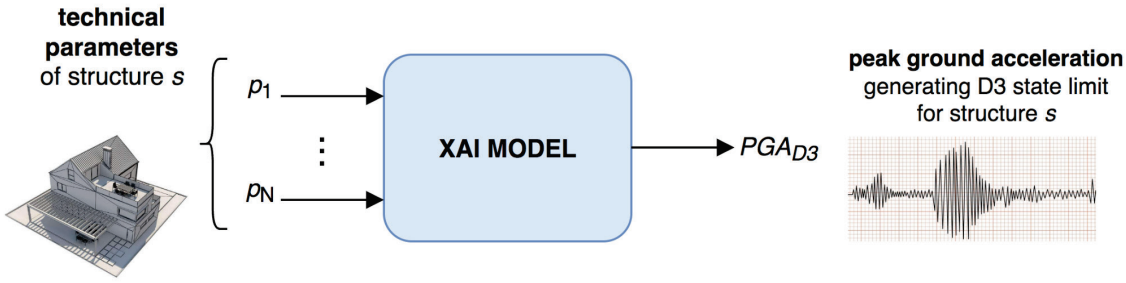


Figure 4: Overview of the XAI model whose inputs are the structure's technical parameters and the output is the peak ground acceleration (PGA) for the D3 state limit.

In order to train the models and guarantee generalization capability, i.e., the independence of all the models from the training data, we divided the dataset into two disjoint sets: the *training set* and

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

the *test set*. The training set was used to select the most relevant technical parameters to estimate the PGA and optimize the various models. Model optimization involved finding the best hyperparameter configuration leading to the highest performance. The *Sequential Feature Selection (SFS)* approach selected the most relevant technical parameters. SFS reduces the dimensionality of the feature space by iteratively selecting a subset of the most informative features. The algorithm starts with an empty set *S* of selected features. At each iteration, the algorithm finds the feature that yields the highest improvement in the model's performance and adds that feature to *S*. The algorithm terminates when a predefined number of features or performance levels are achieved.

The SFS was executed for each regressor. The coefficient of determination, commonly denoted as R-squared (R^2), was considered to select the feature to optimize the hyperparameters of that specific regressor. R^2 is a statistical measure that assesses the proportion of the variance in the dependent variable explained by the independent variables, in a regression model. In other words, it quantifies the model's goodness of fit by indicating the percentage of variability in the response variable that can be accounted for by the predictor variables. Its equation is as follows:

$$R^2 = 1 - \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \,/\, \sum_{i=1}^{N}(y_i - \underline{y_i})^2$$

where y_i represents the observed value of the dependent variable for the i-th observation, $\hat{y}_i$ is the predicted value of the dependent variable for the i-th observation, and $\underline{y_i}$ represents the mean of the observed values of the dependent variable, and N is the number of observations. The numerator is the sum of squares of the residuals (the differences between the observed values and the predicted values), and the denominator is the total sum of squares, which measures the total variance in the dependent variable. The R^2 coefficient ranges from 0 to 1, where 0 indicates that the model does not explain any of the variance in the dependent variable, and 1 indicates that the model explains all the variance. A higher value suggests a better fit of the model to the data, although it should be supplemented with other diagnostic measures to ensure model validity.

Each regressor was trained and tested using 10-fold cross-validation. In particular, the training set was divided into 10 folds (partitions), and the model was trained and evaluated 10 times, using different combinations of folds as training and test sets. SFS was repeated 30 times for each regressor, repopulating the folds after each of the 30 executions. This provides a more robust estimate of model performance.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Mean squared error (MSE) was used to measure the error of each of the 30 regressors on the test folds. The MSE of a regression model is defined as follows:

$$MSE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \; / \; N$$

where y_i represents the observed value of the dependent variable for the i-th observation, $\hat{y}_i$ is the predicted value of the dependent variable for the i-th observation, and N is the number of observations. MSE provides a measure of the average squared deviation between the predicted and actual values. A smaller MSE indicates a better fit of the model to the data. However, the MSE alone might not always provide sufficient information about model performance, which is why it is often used in conjunction with other evaluation metrics and domain knowledge.

In more detail, the loss of a decision tree was used to select the features to optimize the hyperparameters of a decision tree and test it. As illustrated in the next section, this process led to selecting different technical parameters (i.e., features) for each model tested.

Once the most relevant technical parameters were selected, the models were trained in order to search for the best configuration of hyperparameter values, using a grid search. To obtain a robust estimation of model performance, each model was trained and evaluated 10 times using 10-fold cross-validation. For each configuration of hyperparameters, 30 sessions based on 10-fold cross-validation led to training 30 regressors. Folds were repopulated after each of the 30 training sessions. The error of each of the 30 regressors was measured by the MSE on the test folds. The performance, measured as R^2, was higher the lower the average of the 30 mean squared errors on the test folds.

Grid search provides the best hyperparameter configuration that is then used to instantiate and train a new regressor using the entire training set. In fact, as a consequence of applying 10-fold cross-validation in SFS and grid search, only 90% of the data was used to train the models and 10% to test them. Each trained model was then tested using the test set, i.e., those data not used for training. Figures (a) in the next sections show the regression plot obtained on the test data. The two metrics used to evaluate the generalization ability of the models are the R^2 and mean absolute error (MAE), whose values are reported above the regression plot. The MAE of a regression model measures the average of the absolute differences between the actual and predicted values of the dependent variable. Mathematically, it is calculated as follows:

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

$$MAE = \sum_{i=1}^{N} |y_i - \hat{y}_i| \; / \; N$$

where y_i represents the observed value of the dependent variable for the i-th observation, $\hat{y}_i$ is the predicted value of the dependent variable for the i-th observation, and N is the number of observations. MAE provides a measure of the average absolute deviation between the predicted and actual values. Unlike MSE, MAE does not square the differences, which makes it less sensitive to outliers.

4.1 XAI frameworks

Two different frameworks were applied to provide explanations for the predictions of the trained machine learning models: *Local Interpretable Model-agnostic Explanations (LIME)* and *SHAP (SHapley Additive exPlanations)*. Both frameworks are designed to explain the predictions of machine learning models, especially those considered as black-box models.

The primary goal of LIME is to make complex models interpretable on a local level, by approximating their behavior with simpler, interpretable models. The primary goal of LIME is to make complex models interpretable on a local level by approximating their behavior with simpler, interpretable models. Both LIME and SHAP frameworks are model-agnostic as they can be applied to any machine learning model without requiring knowledge of its internal structure. The LIME framework provides local explanations and may not capture the global behavior of the model. In contrast, the SHAP values provide insights into model behavior at both the local and global levels.

4.1.1 Local Interpretable Model-agnostic Explanations (LIME)

LIME starts by selecting a specific instance or data point to interpret the model's prediction. The framework then generates perturbed samples around the selected instance by introducing small changes to the feature values. These samples are used to probe the model's behavior locally. The perturbed samples are then passed through the black-box model to obtain predictions. LIME collects the predictions and the corresponding perturbed samples to train a local, interpretable model, such as a linear model or decision tree, that approximates the behavior of the complex model in the vicinity of the selected instance. The local surrogate model is then analyzed to provide insights into the factors influencing the prediction for the chosen instance. This analysis yields feature

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

importance scores highlighting the contribution of each feature to the model's decision locally.

4.1.2 SHapley Additive exPlanations (SHAP)

SHAP was developed based on cooperative game theory. SHAP values provide a way to fairly distribute the value of a model's prediction among its features. This approach is grounded in the concept of Shapley values, which were originally introduced for fair distribution of payouts in cooperative games. SHAP values are rooted in Shapley values, representing each feature's average contribution to all possible coalitions in a cooperative game. In the context of machine learning, the features are considered *players*, and their contributions are determined based on their impact on the model's predictions. SHAP considers all possible combinations of features (coalitions) and evaluates their impact on the model's output. This involves generating subsets of features and comparing the model's predictions with and without each feature. Shapley values are calculated by averaging the marginal contributions of each feature across all possible coalitions. This process ensures a fair distribution of the prediction value among the individual features. SHAP values provide a clear interpretation of the impact of each feature on a specific prediction. Positive SHAP values indicate a positive contribution to the prediction, while negative values suggest a negative impact. SHAP values adhere to the principles of consistency and linearity. Consistency ensures that the sum of SHAP values for all features equals the difference between the model's output for a specific instance and the average output across all instances. Linearity allows the aggregation of SHAP values across different instances to understand feature importance at a global level. SHAP offers a comprehensive and theoretically grounded approach to interpreting complex model predictions by attributing contributions to individual features.

4.2 Decision Tree

Decision trees are a versatile and intuitive approach to solving classification and regression problems. With their ability to mimic human decision-making processes, decision trees offer a transparent and interpretable framework for analyzing complex data, making them invaluable tools across various domains, from finance and healthcare to marketing and environmental science.

A decision tree is a hierarchical tree-like structure comprising nodes representing decision points and branches denoting possible outcomes or choices. The tree is constructed recursively by

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

partitioning the feature space based on the values of input variables, with each node representing a specific feature and each branch representing a possible decision or outcome. This hierarchical arrangement facilitates a step-by-step decision-making process, where the data is split into increasingly homogeneous subsets until a prediction or decision is reached at the leaf nodes. One of the key strengths of decision trees lies in their simplicity and interpretability.

Unlike complex black-box models such as neural networks, decision trees offer a transparent and easily understandable representation of the decision-making process. Each split in the tree corresponds to a simple decision rule based on a single feature, making it straightforward to interpret and explain the rationale behind the model's predictions. Furthermore, decision trees are highly flexible and capable of handling both categorical and numerical data, as well as multi-class classification and multi-output regression tasks. They can also accommodate missing values and noisy data, making them robust in real-world scenarios where data quality may vary. Additionally, decision trees naturally handle interactions and nonlinear relationships between features without requiring explicit feature engineering or transformation.

Despite their simplicity, decision trees can yield powerful predictive models, especially when combined with ensemble techniques such as random forests and gradient boosting. These ensemble methods leverage the collective wisdom of multiple decision trees to improve predictive accuracy and generalization performance while retaining individual trees' interpretability. However, like any machine learning algorithm, decision trees have limitations. They are prone to overfitting, especially when the tree depth is not properly constrained or when the dataset is noisy or highly imbalanced. Techniques such as pruning, limiting the maximum depth of the tree, or using ensemble methods are often employed to mitigate overfitting. Moreover, decision trees may struggle with capturing complex patterns or interactions in the data, particularly when the decision boundaries are non-linear or when the relationships between features are intricate. More sophisticated models like neural networks or support vector machines may perform better in such cases. In summary, decision trees represent a powerful and interpretable approach to machine learning, offering a balance between simplicity, flexibility, and predictive performance. With their ability to elucidate decision-making processes and handle a wide range of data types and tasks, decision trees continue to be a cornerstone of modern machine learning methodologies, empowering practitioners to extract valuable insights and make informed decisions from complex datasets.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

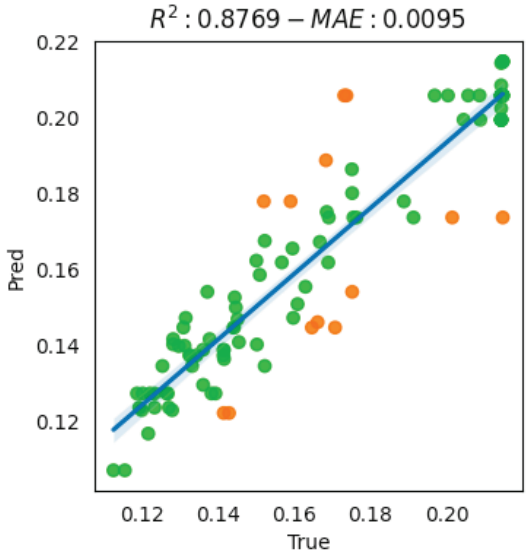


Figure 5: Regression plot for the decision tree model.

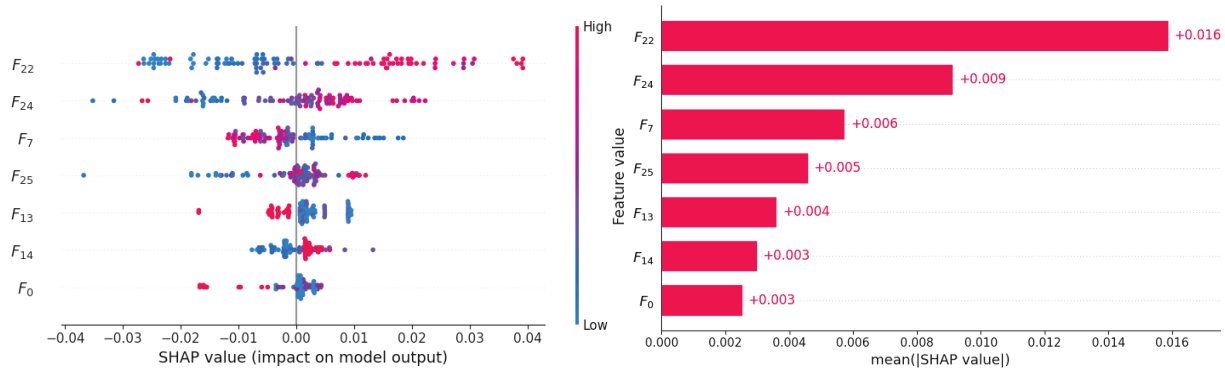


Figure 6: Beeswarm plot (left-hand side) and bar plot (right-hand side) show the distribution of the impact of each feature on the output of the decision tree model.

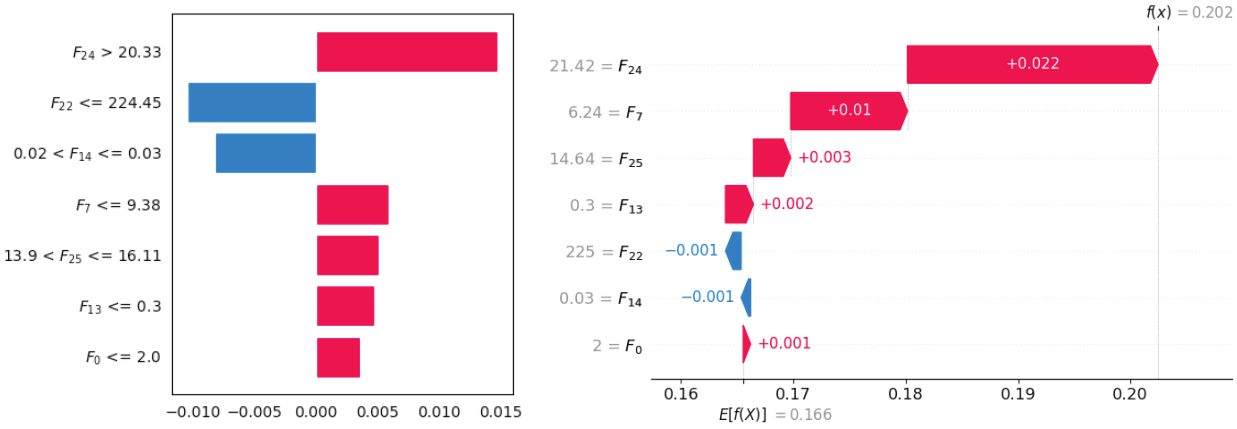


Figure 7: Explanation by LIME (left-hand side) and SHAP (right-hand side) for a representative structure whose PGA value for D3 limit state is calculated by the decision tree model.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

After the SFS, the most selected features were 7: number of level, area of the openings of the external masonry walls at the elevation floor (EF), average thickness of internal masonry shear wall (EF), average shear strength of masonry (GF), shear modulus (GF), effective shear area (GF), effective shear area (EF).

Figure 5 shows the regression plot, which highlights the relationship between two variables, typically, an independent variable (X) and a dependent variable (Y). This type of plot is often used in regression analyses to understand the relationship between variables and visualize how a regression model approximates the data. The regression line represents the best possible fit to the data according to the regression model used. For linear regression, for example, the line represents the line of best fit. The shape of the regression line and its slope indicate the nature of the relationship between the variables. If the line is a straight line, the relationship is linear. If it is a curve, the relationship may be nonlinear. The slope of the line indicates how much the dependent variable changes relative to a unit change in the independent variable. The regression plot may sometimes include a confidence interval around the regression line. This interval represents the expected variability in the model estimate. The wider the interval, the greater the uncertainty in the estimate. Points that deviate significantly from the regression line can be identified as influential points. These points may significantly impact model estimation and require further investigation.

Figure 6 shows the beeswarm and the bar plots generated by SHAP. The beeswarm plot is a visualization that combines the idea of a swarm plot with SHAP information. In a swarm plot, data are distributed along the x-axis so as to avoid overlap. In the context of SHAP, swarm plots are used to show the distribution of SHAP values for each feature. Each point in the swarm plot represents a specific instance of the data, and the position along the x-axis indicates the corresponding SHAP value. This type of visualization is particularly useful for observing how SHAP contributions vary across different data instances. The SHAP plot bar is a graphical representation that shows the relative importance of different features in contributing to model output. Each bar represents a feature, and the length of the bar indicates how much the presence of that feature contributes to the model output. The bars can be oriented ascending or descending depending on the sign of the contribution (positive or negative). This type of visualization is useful for quickly identifying the most influential features in the model and understanding how each contributes to the final output. In

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

either case, these plots provide an intuitive way to interpret the contribution of each feature in the context of the model. They can be used to identify patterns, outliers, and understand how features affect the machine learning model's predictions.

Figure 7 shows the LIME and SHAP plots relative to a representative structure in order to explain how the model came to the PGA_{D3} value. Specifically, throughout the deliverable, the green points in the regression plot correspond to structures whose error in estimating the peak ground acceleration for the D3 limit state is close to the MAE. Conversely, the orange points represent structures for which the model makes an error exceeding twice the MAE when predicting the PGA for the D3 limit state.

The result achieved by this model is quite accurate (see the regression plot) and also makes sense in structural engineering. As the plots in Figure 6 show, the shear modulus (GF), the effective shear area (GF), and the area of the openings of the external masonry walls at the elevation floor (EF) are the technical parameters that most influence the output of the XAI model. For a particular building, the XAI model suggests that the PGA needed to trigger the D3 limit state is highly influenced by two key factors: the effective shear area at the ground floor (GF) and the area of the openings of the external masonry walls at the elevation floor (EF) (LIME plot, left-hand side). This aligns with structural engineering principles. The effective shear area (GF) likely relates to how well structural elements resist lateral loads induced by seismic events. A higher effective shear area (GF) implies better resistance to lateral forces, potentially lowering the likelihood of reaching the D3 limit state during strong ground acceleration. Similarly, the area of the openings of the external masonry walls at the elevation floor (EF) is important. External walls operate as seismic resistance elements, counteracting earthquake-induced lateral forces. However, openings reduce their capacity to resist lateral forces and increase the susceptibility of the structure to reaching the D3 limit state.

Although the explanation aligns with structural engineering, the decision tree model did not perform best compared to the other models, as highlighted by the R^2 coefficient equal to 0.8769. Nevertheless, the MAE was limited to 0.0095, which means that the average error when predicting the PGA is quite low.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

4.3 *K*-Nearest Neighbors

Unlike traditional algorithms, *K*-Nearest Neighbors (*K*-NN) does not involve explicit model training as it relies on stored training data to predict new instances. In *K*-NN, the *K* represents the number of nearest neighbors considered when predicting a new instance's class (for classification) or value (for regression). The algorithm calculates the similarity between instances using a distance metric, often the Euclidean or Manhattan distance. For classification, the majority class among the *K*-nearest neighbors determines the predicted class for the new instance. In regression, the algorithm computes the average (or weighted average) of the target values of the *K*-nearest neighbors, providing the predicted value for the new instance. *K*-NN does not explicitly learn a decision boundary; instead, it adapts to the distribution of training data in the feature space, making it suitable for non-linear decision boundaries. The choice of *K* is critical, with smaller values leading to more sensitive models and larger values resulting in smoother decision boundaries. Feature scaling is crucial in *K*-NN, as it relies on distance calculations, and different feature scales can impact results. The computational cost during prediction may be high, especially for large datasets or high-dimensional feature spaces, as it needs to calculate distances for each query instance against all training instances. While *K*-NN is simple and versatile, its efficiency depends on the size and characteristics of the dataset. It is well-suited for smaller to medium-sized datasets but may face challenges with larger datasets due to computational requirements.

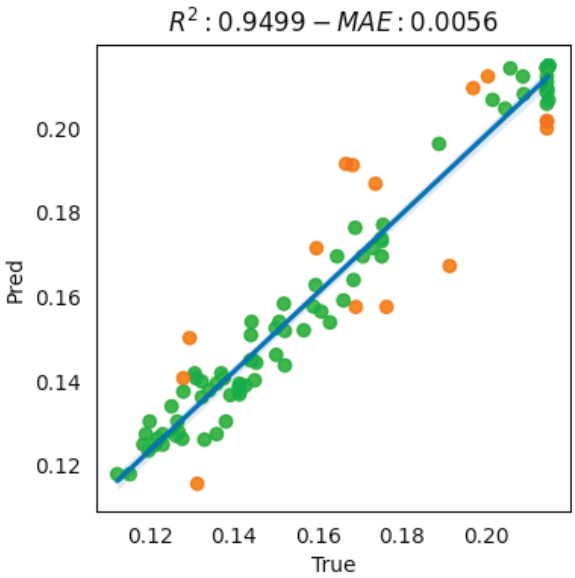


Figure 8: Regression plot for the *K*-Nearest Neighbors model.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

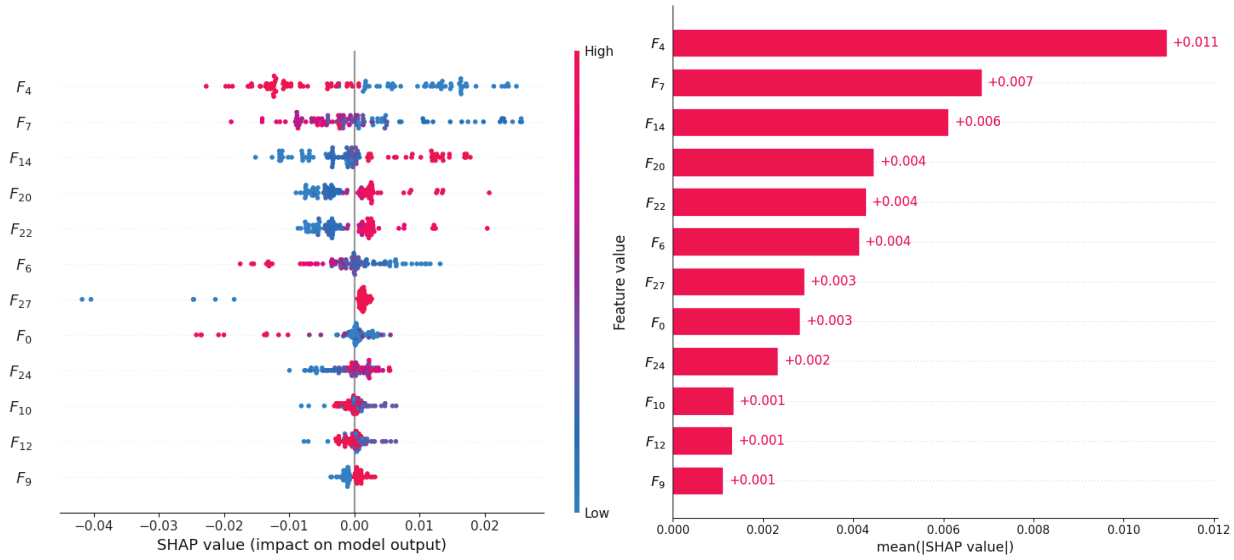


Figure 9: Beeswarm plot (left-hand side) and bar plot (right-hand side) show the distribution of the impact of each feature on the output of the *K*-Nearest Neighbors model.

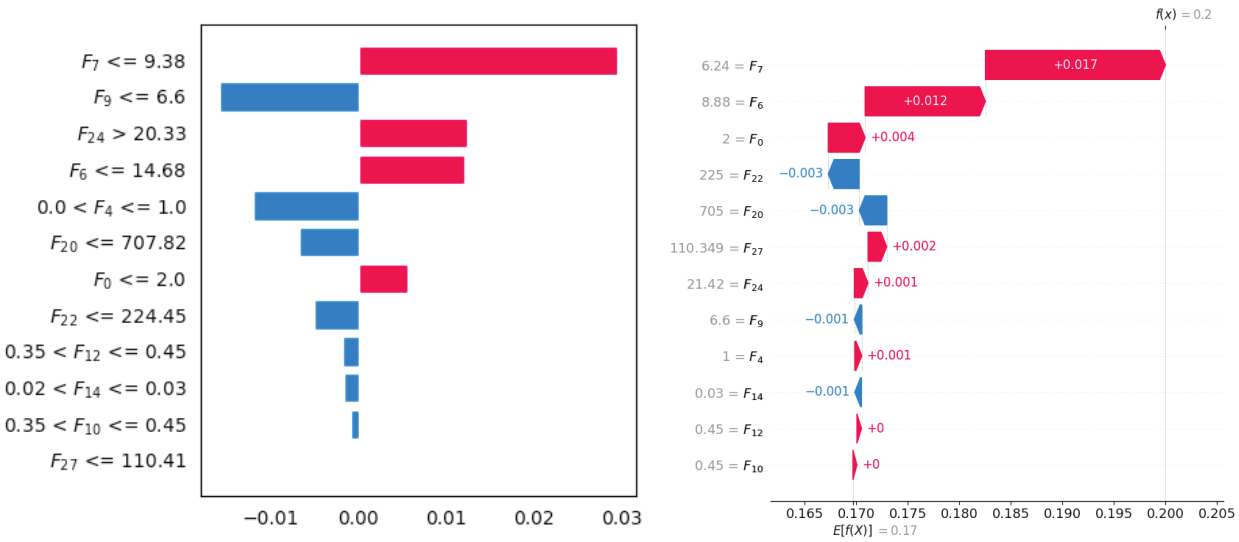


Figure 10: Explanation by LIME (left-hand side) and SHAP (right-hand side) for a representative structure whose PGA value for D3 limit state is calculated by the *K*-Nearest Neighbors model.

After the SFS, the most selected features were 12: number of level, number of internal alignments of the masonry wall, area of the openings of the external masonry walls at the ground floor (GF), area of the openings of the external masonry walls at the elevation floor (EF), area of the openings of the internal masonry walls at the elevation floor (EF), the average thickness of external masonry shear wall (GF), average thickness of internal masonry shear wall (GF), average shear strength of masonry (GF), elastic modulus (GF), shear modulus (GF), effective shear area (GF), seismic floor

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

mass – roofing plan (RP).

Figure 8 shows the regression plot, whereas Figure 9 and Figure 10 show the plots relative to the global explanation and to the local explanation (i.e., explanation relative to a representative structure), respectively. The result achieved by this model is very accurate (see the regression plot) and also makes sense in structural engineering.

Although the *K*-Nearest Neighbors model did not perform best compared to the other models, it achieved a higher R^2 coefficient than that obtained with the decision tree model. In particular, the *K*-Nearest Neighbors model achieves an R^2 coefficient equal to 0.9499 and an MAE equal to 0.0056, which is a very low average error in predicting the PGA.

4.4 Extreme Gradient Boosting

XGBoost, short for eXtreme Gradient Boosting, is an advanced implementation of the gradient boosting framework. It is a powerful and efficient machine learning algorithm that excels in regression and classification tasks. XGBoost is renowned for its speed, accuracy, and versatility, making it a popular choice in various data science competitions and real-world applications. XGBoost builds an ensemble of weak learners, usually decision trees, in a sequential manner. Each tree corrects the errors made by the previous ones, gradually improving the model's predictive performance. The term *gradient boosting* reflects the algorithm's focus on minimizing the gradient of the loss function with respect to the model's predictions.

Key features of XGBoost include its ability to handle missing data, regularization techniques to prevent overfitting, and the flexibility to customize the loss function based on the specific problem requirements. XGBoost leverages parallel and distributed computing, optimizing its computational efficiency and scalability. The algorithm introduces a few innovations, such as tree pruning, which helps prevent overfitting by removing nodes that contribute minimally to the model's improvement. XGBoost also incorporates a regularization term in the objective function, controlling the complexity of individual trees.

XGBoost's training process involves creating a sequence of decision trees, with each subsequent tree aiming to correct the errors of the combined ensemble. The final prediction is obtained by summing the contributions of all the trees, weighted by their respective learning rates. In addition to its high predictive performance, XGBoost provides valuable insights into feature importance. By

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

analyzing the contribution of each feature across the ensemble, users better understand the factors influencing the model's decisions. XGBoost's popularity stems from its efficiency, scalability, and state-of-the-art performance across various machine learning tasks. Its successful integration of boosting techniques with tree-based models has positioned it as a leading algorithm in the machine learning landscape.

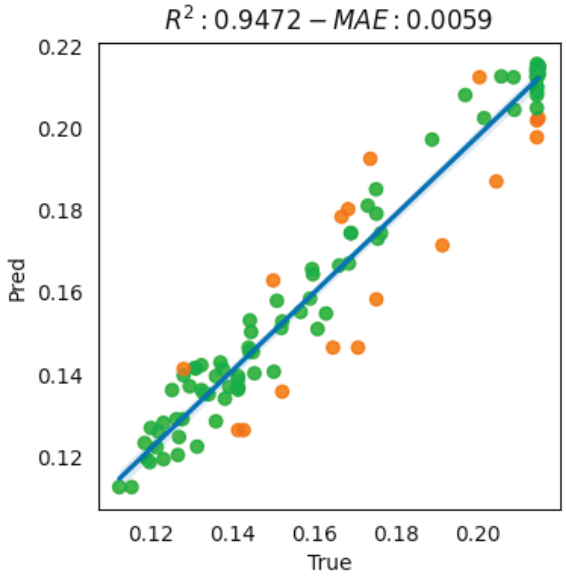


Figure 11: Regression plot for the Extreme Gradient Boosting model.

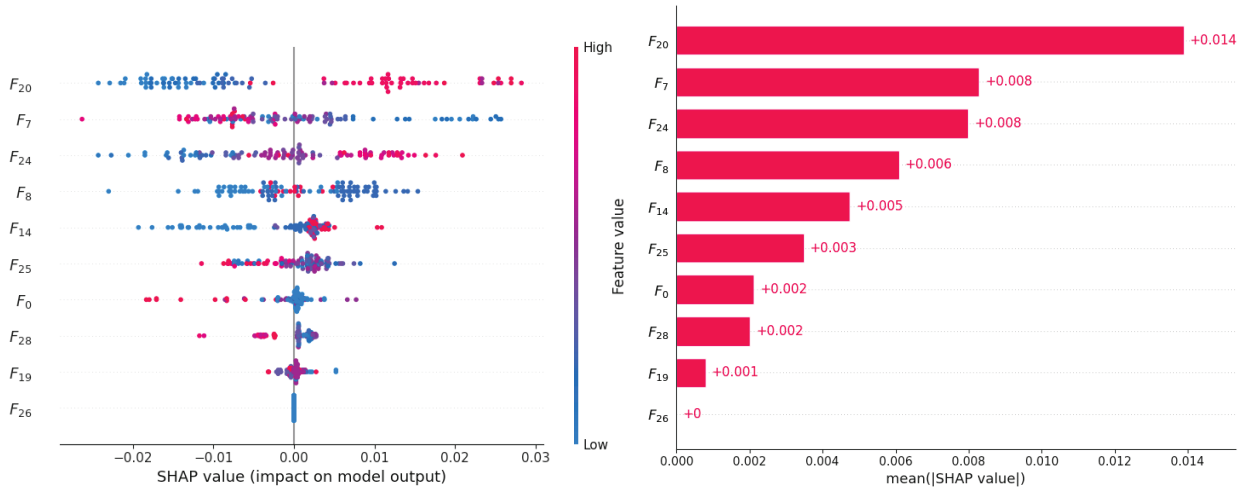


Figure 12: Beeswarm plot (left-hand side) and bar plot (right-hand side) show the distribution of the impact of each feature on the output of the Extreme Gradient Boosting model.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

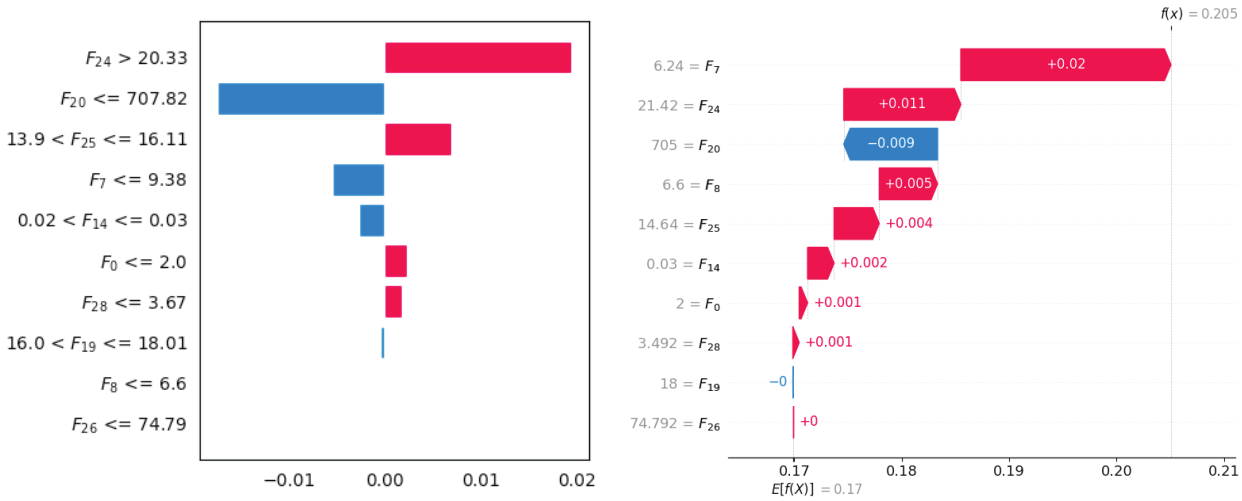


Figure 13: Explanation by LIME (left-hand side) and SHAP (right-hand side) for a representative structure whose PGA value for D3 limit state is calculated by the Extreme Gradient Boosting model.

After the SFS, the most selected features were 10: number of level, area of the openings of the external masonry walls at the elevation floor (EF), area of the openings of the internal masonry walls at the ground floor (GF), average shear strength of masonry (GF), masonry gross density (EF), elastic modulus (GF), effective shear area (GF), effective shear area (EF), seismic floor mass (EF), ratio of seismic floor mass/ effective shear area (GF).

Figure 11 shows the regression plot, whereas Figure 12 and Figure 13 show the plots relative to the global explanation and to the local explanation (i.e., explanation relative to a representative structure), respectively.

The result achieved by this model is very accurate (see the regression plot) and also makes sense in structural engineering. As the beeswarm and the bar plots show, area of the openings of the external masonry walls at the elevation floor (EF), the elastic modulus (GF), and the effective shear area (GF) are the technical parameters that most influence the output of the model. Even if eXtreme Gradient Boosting model uses fewer features, it performs worse than the *K*-Nearest Neighbors model, as highlighted by the R^2coefficient.

4.5 Light Gradient Boosting Machine

LightGBM, short for Light Gradient Boosting Machine, is a high-performance, distributed gradient boosting framework designed to efficiently train large datasets and handle categorical features. Developed by Microsoft, LightGBM is particularly well-suited for problems in machine learning

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

where both speed and accuracy are crucial. LightGBM employs a histogram-based learning approach to construct decision trees during the boosting process. Instead of using traditional depth-first or level-wise tree growth strategies, LightGBM uses a leaf-wise approach. This allows the algorithm to quickly grow trees and prioritize splitting nodes, leading to greater loss function reductions. One notable feature of LightGBM is its ability to handle categorical features directly without needing one-hot encoding. This reduces memory usage and computational costs associated with encoding categorical variables, making LightGBM especially efficient for datasets with a large number of categorical features. LightGBM incorporates regularization techniques such as L1 and L2 regularization on leaf values and feature-wise binomial or Gaussian likelihood to address overfitting concerns. Additionally, it implements a technique called Gradient-based One-Side Sampling (GOSS) to efficiently select a subset of data points for training, reducing computation while retaining the overall gradient information. LightGBM supports parallel and distributed computing, making it suitable for large-scale datasets. It also offers GPU acceleration, further enhancing its training speed. The flexible framework allows users to customize various hyperparameters to fine-tune model performance. Due to its efficiency and scalability, LightGBM has gained popularity in both research and industry settings. It has proven effective in classification, regression, and ranking tasks. It is widely used in machine learning competitions and real-world applications where rapid model training and high predictive accuracy are paramount.

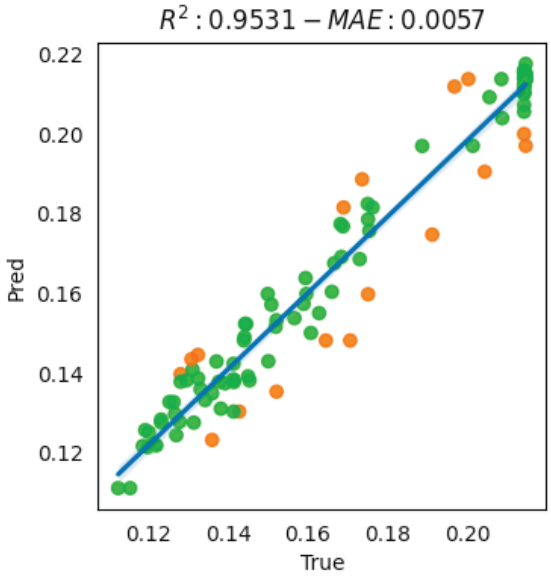


Figure 14: Regression plot for the Light Gradient Boosting Machine model.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

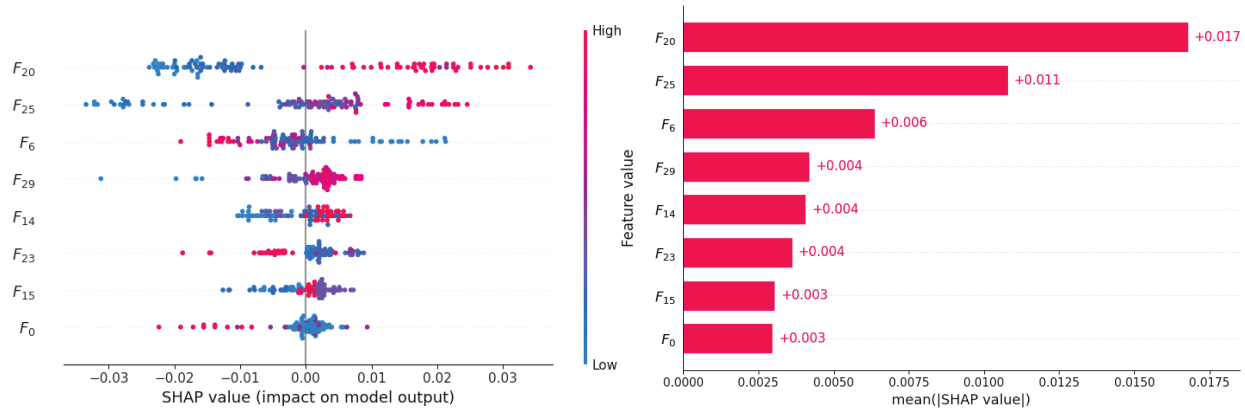


Figure 15: Beeswarm plot (left-hand side) and bar plot (right-hand side) showing the distribution of the impact of each feature on the output of the Light Gradient Boosting Machine model.

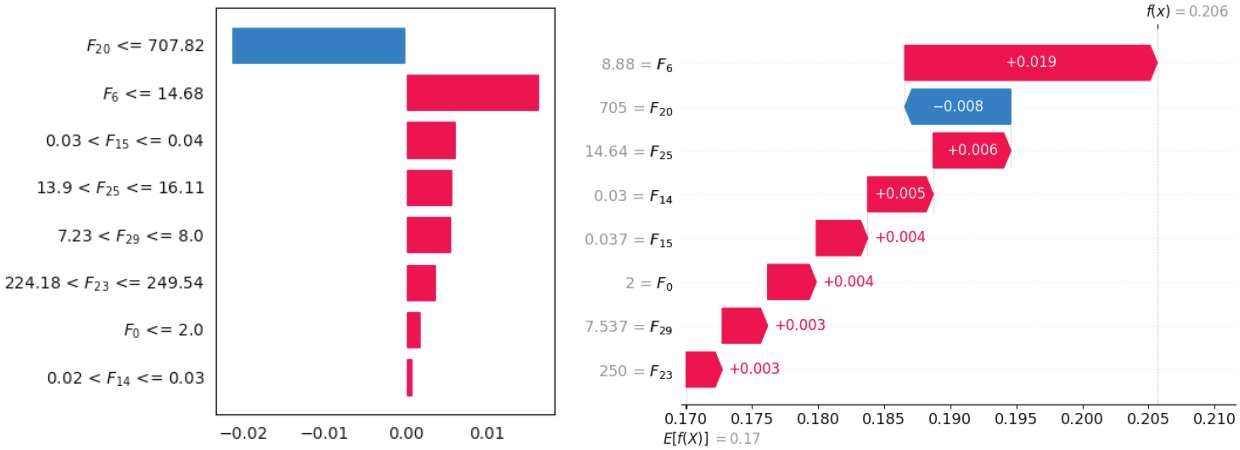


Figure 16: Explanation by LIME (left-hand side) and SHAP (right-hand side) for a representative structure whose PGA value for D3 limit state is calculated by Light Gradient Boosting Machine.

After the SFS, the most selected features were 10: number of level, area of the openings of the external masonry walls at the ground floor (GF), average shear strength of masonry (GF), average shear strength of masonry (EF), elastic modulus (GF), shear modulus (EF), effective shear area (EF), ratio of Seismic floor mass/ effective shear area (EF).

Figure 14 shows the regression plot, whereas Figure 15 and Figure 16 show the plots relative to the global explanation and to the local explanation (i.e., explanation relative to a representative structure), respectively.

The result achieved by this model is the most accurate (see the regression plot) and also makes sense in structural engineering. As the beeswarm and the bar plots show, the area of the openings of the external masonry walls at the ground floor (GF) and the effective shear area (EF) are the

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

technical parameters that most influence the output of the model. The Light Gradient Boosting Machine achieved an R^2coefficient equal to 0.9531. Also, the MAE was limited to 0.0056, which means that the average error when predicting the PGA is very low.

4.6 Random Forest

Among the various machine learning methodologies, the random forest algorithm combines predictive performance and interpretability. This unique blend of accuracy and transparency has propelled random forest to the forefront of data-driven decision-making across various domains. Random forest uses the power of ensemble learning, where multiple decision trees collaborate to form a predictive model. Each decision tree is trained on a random subset of the training data and features, fostering diversity within the ensemble. During inference, the collective decisions of these individual trees are aggregated, resulting in a robust and reliable prediction that reflects the wisdom of the forest. While the primary objective of random forest is undoubtedly predictive accuracy, its inherent structure also offers a gateway to understanding and explaining the complex dynamics of the underlying data. Unlike opaque black-box models, random forest provides a clear and intuitive representation of the decision-making process, allowing practitioners to dissect the model's inner workings easily. By delving into the constituent decision trees within the forest, experts can uncover the rationale behind each prediction and identify the key features driving model behavior. This analysis highlights the factors that influence individual predictions and sheds light on the broader patterns and relationships inherent in the dataset. In addition, random forest offers a built-in mechanism for assessing feature importance, enabling practitioners to gauge the relative contribution of each input variable to the model's predictions. This feature importance analysis not only aids in feature selection and dimensionality reduction but also provides valuable insights into the underlying mechanisms shaping the data. Random forest's resilience to noisy or imbalanced datasets also enhances its interpretability capabilities, ensuring the model's explanations remain robust and trustworthy in real-world complexities. By incorporating randomness into the training process and aggregating the predictions of multiple trees, random forest mitigates the risk of overfitting. It generalizes effectively to unseen data, bolstering the reliability of its explanations.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

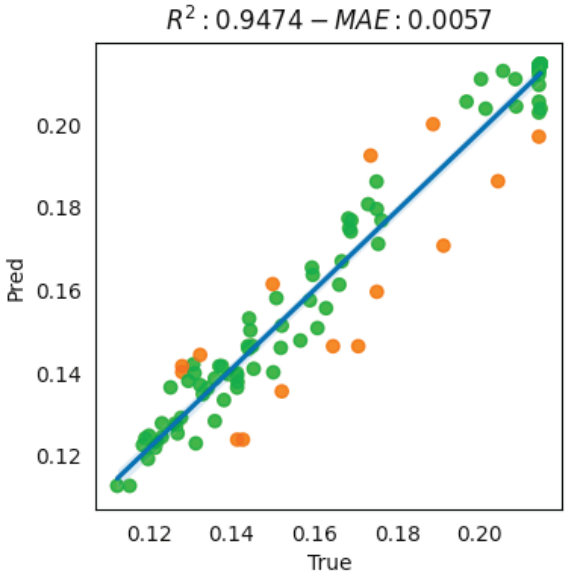


Figure 17: Regression plot for the Random Forest model.

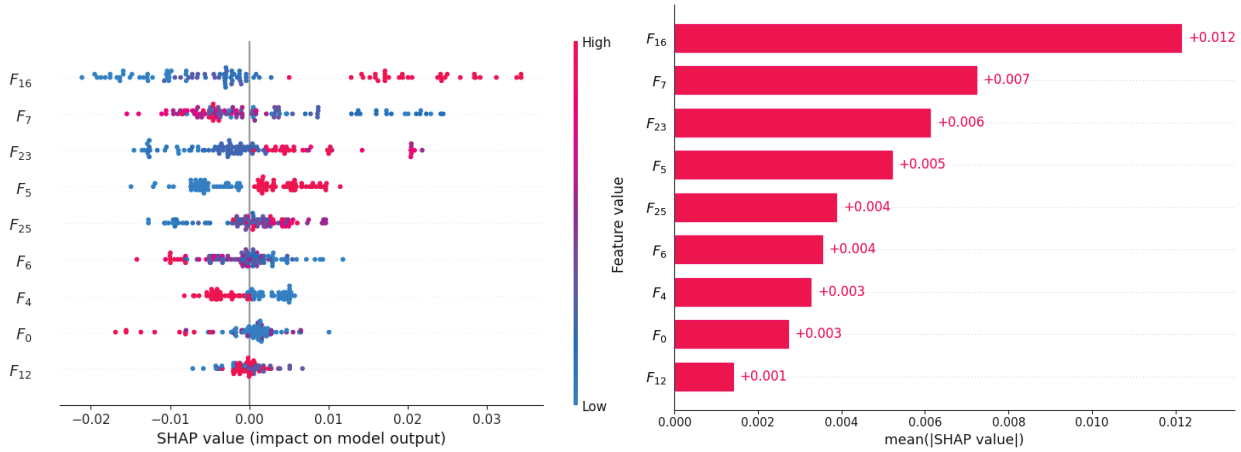


Figure 18: Beeswarm plot (left-hand side) and bar plot (right-hand side) show the distribution of the impact of each feature on the output of the Random Forest model.

As a consequence of the SFS, the most selected features were 9: number of level, number of internal alignments of masonry wall, area of the openings of the external masonry walls at the ground floor (GF), area of the openings of the external masonry walls at the elevation floor (EF), average thickness of internal masonry shear wall (GF), average compressive strength of masonry (GF), shear modulus (EF), effective shear area (EF).

Figure 17 shows the regression plot, whereas Figure 18 and Figure 19 show the plots relative to the global explanation and to the local explanation (i.e., explanation relative to a representative structure), respectively.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

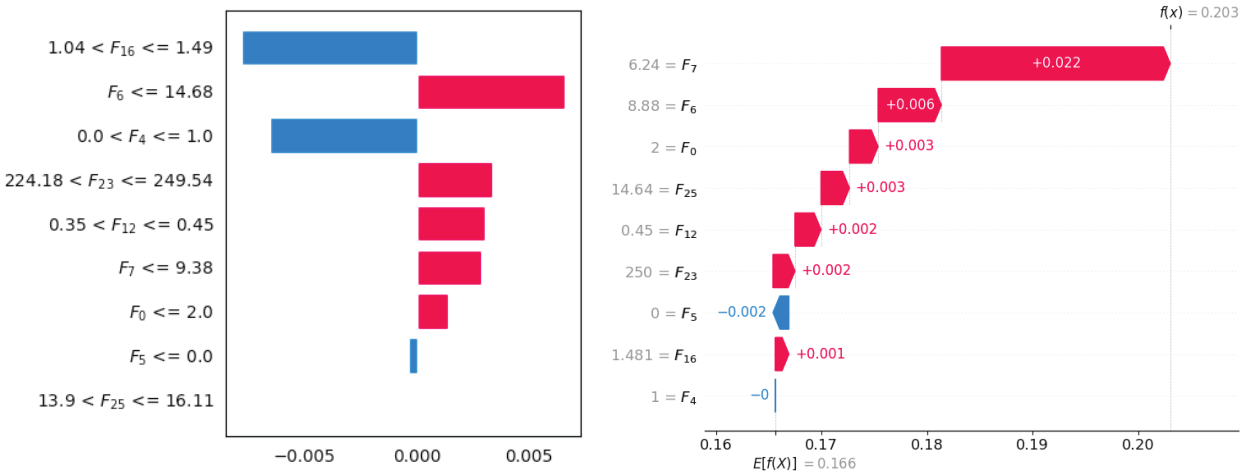


Figure 19: Explanation by LIME (left-hand side) and SHAP (right-hand side) for a representative structure whose PGA value for D3 limit state is calculated by the Random Forest model.

The result achieved by this model is accurate (see the regression plot) and also makes sense in structural engineering. As the beeswarm and the bar plots show, the average compressive strength of masonry (GF) and the area of the openings of the external masonry walls at the ground floor (GF) are the technical parameters that most influence the output of the model. The XAI model suggests that these two technical parameters influence the prediction of the PGA needed to trigger the D3 limit state for a particular building. Although the random forest model uses only 9 features, it achieves an R^2 coefficient equal to 0.9474 and an MAE limited to 0.0057.

4.7 Gradient Boosting

Gradient Boosting is a machine learning ensemble technique that builds a predictive model in a sequential manner by combining the predictions of weak learners, usually decision trees. The primary goal of Gradient Boosting is to improve predictive accuracy by focusing on the mistakes made by the previous models in the ensemble. The process begins with an initial weak learner, typically a shallow decision tree, which makes predictions on the training data. Subsequent weak learners are then added sequentially to the ensemble, with each new learner aiming to correct the errors of the combined ensemble so far. The key innovation of Gradient Boosting lies in its focus on the gradient of the loss function with respect to the model's predictions. In each iteration, the algorithm fits a new weak learner to the negative gradient of the loss function, effectively moving

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

the model towards the direction of the steepest decrease in the loss. The learning rate parameter controls the contribution of each weak learner to the ensemble. A smaller learning rate makes the process more conservative by reducing the impact of each new learner, while a larger learning rate allows the model to learn more quickly but might lead to overfitting. Gradient Boosting can handle various loss functions, making it versatile for regression, classification, and ranking problems. Common loss functions include mean squared error for regression and logistic loss for binary classification. While Gradient Boosting is a powerful technique with high predictive accuracy, it can be computationally expensive and prone to overfitting, especially if the model is allowed to become too complex. Regularization techniques, early stopping, and tuning hyperparameters are common strategies to address these challenges and achieve a well-balanced model.

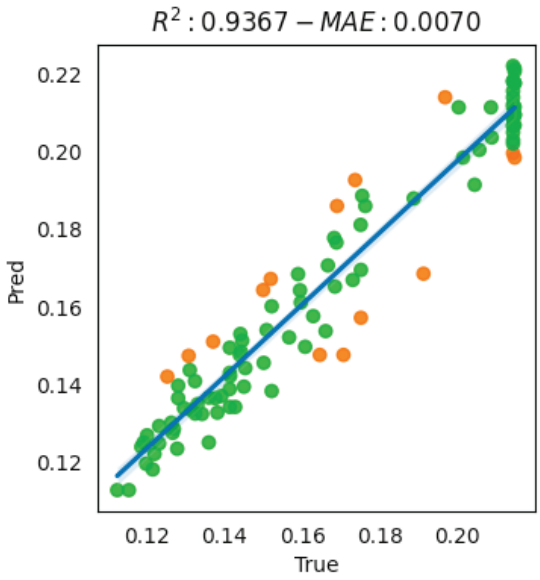


Figure 20: Regression plot for the Gradient Boosting model.

After the SFS, the most selected features were 15: Number of level, Number of internal alignments of masonry wall in the X-direction, Number of internal alignments of masonry wall in the Y-direction, Area of the openings of the external masonry walls at the ground floor (GF), Area of the openings of the external masonry walls at the elevation floor (EF), Area of the openings of the internal masonry walls at the ground floor (GF), Average thickness of external masonry shear wall (GF), Average thickness of external masonry shear wall (EF), Average shear strength of masonry (GF), Masonry gross density (EF), Shear Modulus (GF), Shear Modulus (EF), effective shear area (EF), Ratio of Seismic floor mass/ effective shear area (GF), Ratio of Seismic floor mass to effective shear area.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

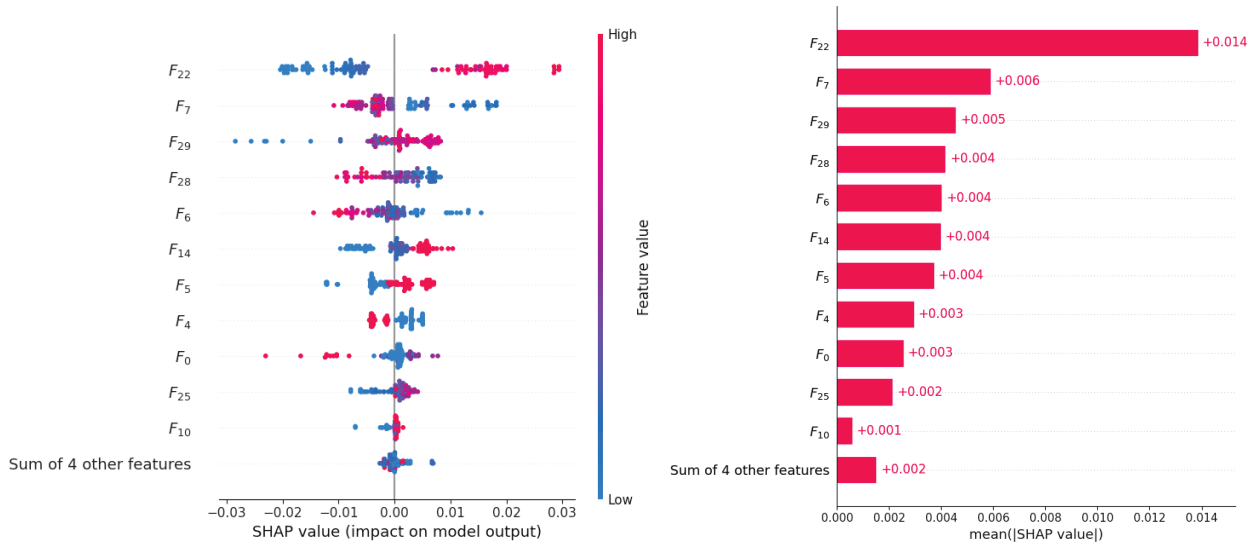


Figure 21: Beeswarm plot (left-hand side) and bar plot (right-hand side) show the distribution of the impact of each feature on the output of the Gradient Boosting model.

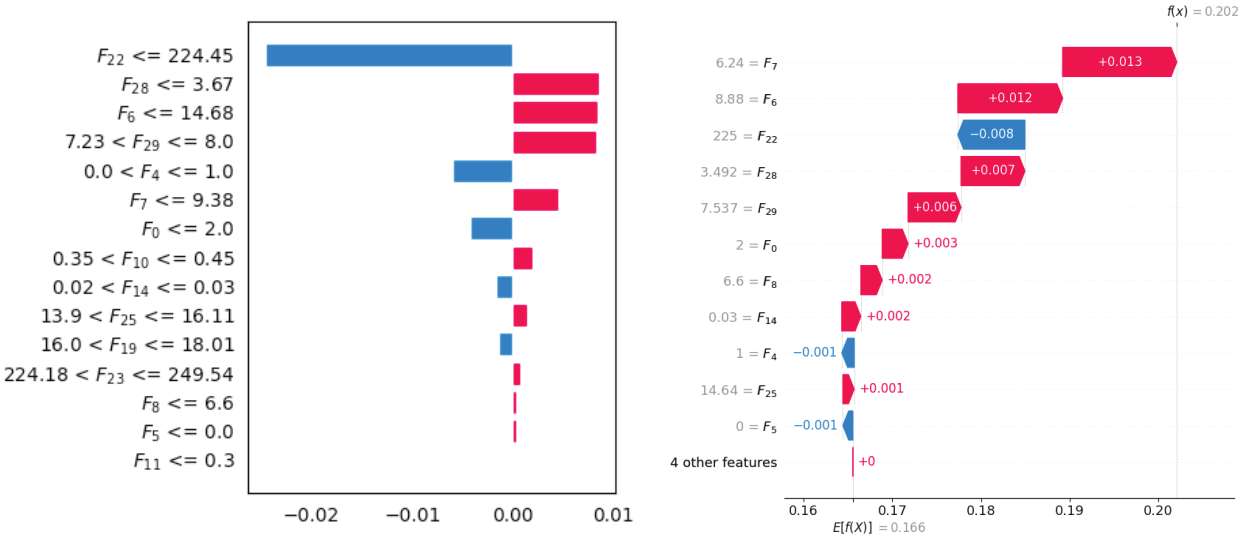


Figure 22: Explanation by LIME (left-hand side) and SHAP (right-hand side) for a representative structure whose PGA value for D3 limit state is calculated by the Gradient Boosting model.

Figure 20 shows the regression plot, whereas Figure 21 and Figure 22 show the plots relative to the global explanation and to the local explanation (i.e., explanation relative to a representative structure), respectively.

The result achieved by this model is quite accurate (see the regression plot) and also makes sense in structural engineering. As the beeswarm and the bar plots show, the shear modulus (GF) and the area of the openings of the external masonry walls at the elevation floor (EF) are the technical

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

parameters that most influence the output of the model. Although the gradient boosting model uses more technical parameters than other models, it did not achieve the best R^2 coefficient.

4.8 Adaptive Boosting

AdaBoost, short for Adaptive Boosting, is a machine learning ensemble method that combines weak learners' predictions to create a strong and accurate predictive model. The key idea behind AdaBoost is to give more weight to the instances that are misclassified by the current weak learners in each iteration, allowing subsequent weak learners to focus on the harder-to-predict instances. The process begins with an initial weak learner, which could be a simple model like a decision stump (a one-level decision tree). After the first learner is trained, AdaBoost increases the weights of misclassified instances and decreases the weights of correctly classified instances. This adjusted dataset is then used to train the next weak learner, and the process repeats. Each weak learner is assigned a weight based on its accuracy in predicting the instances. The final prediction of the AdaBoost model is a weighted sum of the individual weak learners, with higher weights given to more accurate learners. AdaBoost's adaptability lies in its ability to sequentially emphasize misclassified instances, guiding subsequent learners to improve in areas where the model struggles. This iterative process continues until a predefined number of weak learners have been trained or a specified level of accuracy is achieved. AdaBoost is particularly effective in improving the performance of weak learners, and it is less prone to overfitting than training a single, more complex model. It is versatile and can be applied to binary classification and regression problems.

One notable aspect of AdaBoost is that it is sensitive to noisy data and outliers. Outliers might significantly impact the learning process as the algorithm adapts to misclassified instances. To mitigate this, data preprocessing techniques and robust weak learners can be employed. Overall, AdaBoost is a powerful algorithm for boosting the performance of weak learners, providing a robust and accurate predictive model across a range of machine learning tasks.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

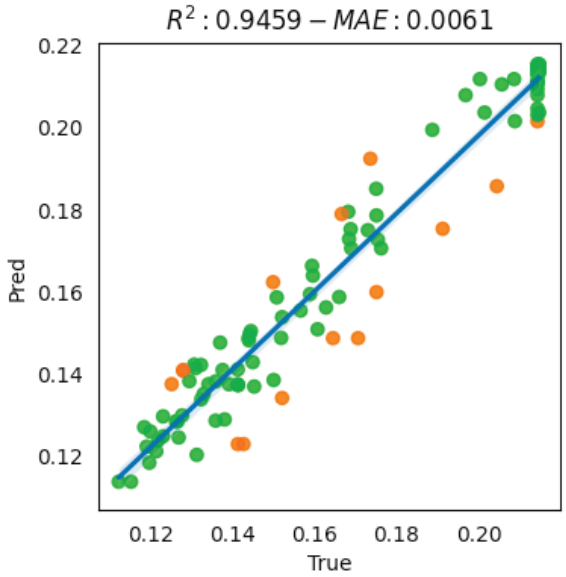


Figure 23: Regression plot for the Adaptive Boosting model.

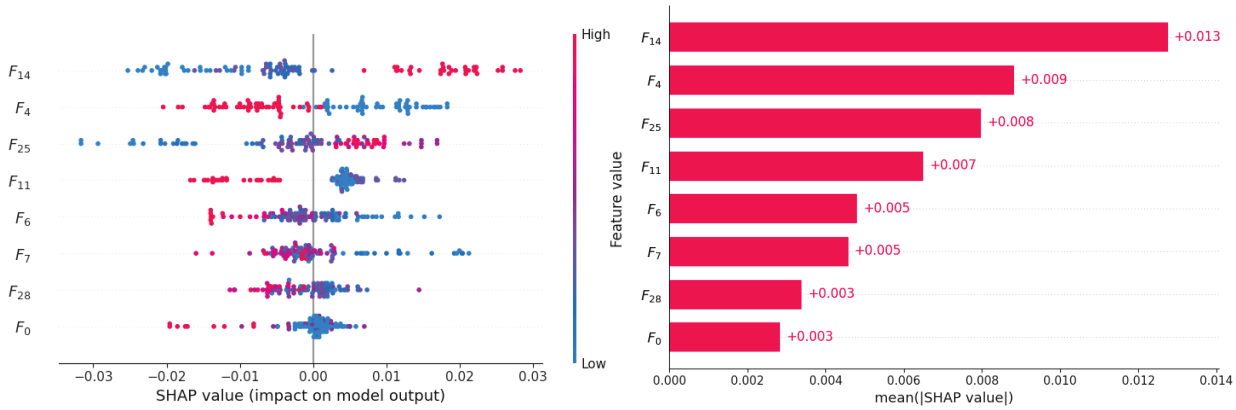


Figure 24: Beeswarm plot (left-hand side) and bar plot (right-hand side) show the distribution of the impact of each feature on the output of the Adaptive Boosting model.

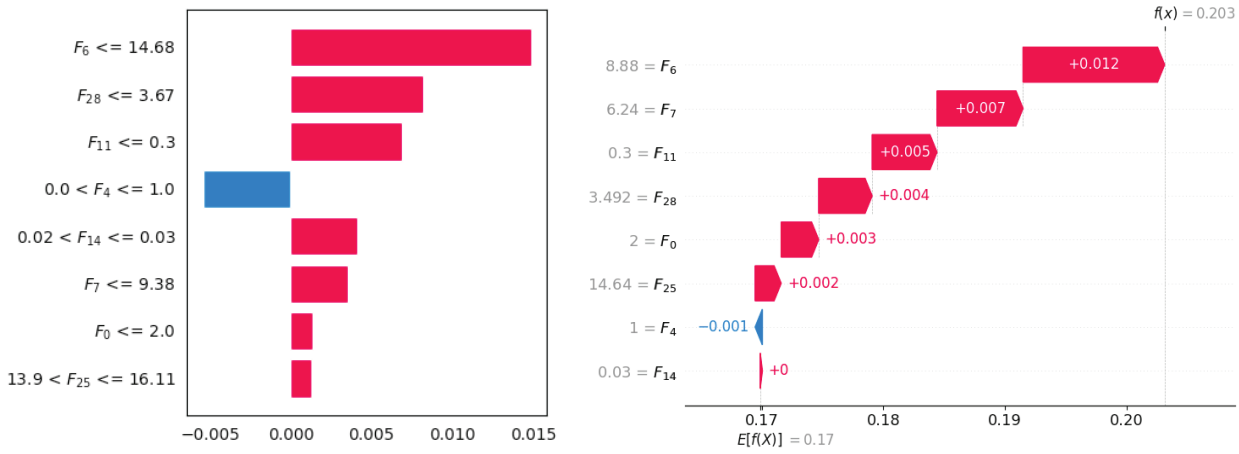


Figure 25: Explanation by LIME (left-hand side) and SHAP (right-hand side) for a representative structure whose PGA value for D3 limit state is calculated by the Adaptive Boosting model.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

After the SFS, the most selected features were 8: Number of level, Number of internal alignments of masonry wall in X-direction, Area of the openings of the external masonry walls at the ground floor (GF), Area of the openings of the external masonry walls at the elevation floor (EF), Average thickness of external masonry shear wall (EF), Average shear strength of masonry (GF), Effective shear area (EF), Ratio of Seismic floor mass/ effective shear area (GF).

Figure 23 shows the regression plot, whereas Figure 24 and Figure 25 show the plots relative to the global explanation and to the local explanation (i.e., explanation relative to a representative structure), respectively.

The result achieved by this model is accurate (see the regression plot) and also makes sense in structural engineering. As the beeswarm and the bar plots show, the average shear strength of masonry (GF), the number of internal alignments of masonry wall, and the effective shear area (EF) are the technical parameters that most influence the output of the model. IAdaptive boosting uses only 8 technical parameters to predict the PGA. However, it achieves an R^2 coefficient equal to 0.9459 with an MAE limited to 0.0061, which are worse than those achieved by the Bootstrap Aggregating model.

4.9 Bootstrap Aggregating

Bagging, short for Bootstrap Aggregating, is an ensemble learning technique used to improve the accuracy and stability of machine learning models. The fundamental idea behind bagging is to create multiple independent instances of a base learning algorithm by training each instance on a different subset of the original training data. The process begins by randomly selecting subsets, with replacements, from the training data. Since sampling is done with replacement, some instances may appear more than once in a subset, while others may not be included at all. These subsets are used to train individual base models, often referred to as weak learners or base learners. After training, predictions from each base model are combined through a process of averaging (for regression tasks) or voting (for classification tasks). This ensemble approach helps reduce the variance of the model, improving its generalization performance on unseen data. One of the key advantages of bagging is that it provides a way to address overfitting, especially in the context of high-variance models. By creating diverse subsets of the data for each base learner, bagging helps ensure that the

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

ensemble is less sensitive to noise and outliers in the training data. One popular algorithm that utilizes bagging is the Random Forest, where the base learners are decision trees. Random Forest builds multiple trees by bootstrapping the data and introducing additional randomness during tree-building. The final prediction is then obtained by aggregating the predictions of all individual trees. Bagging is a versatile technique applicable to various machine learning algorithms, and it is particularly effective when dealing with complex models prone to overfitting. It provides a straightforward and effective approach to enhance model robustness and improve predictive performance.

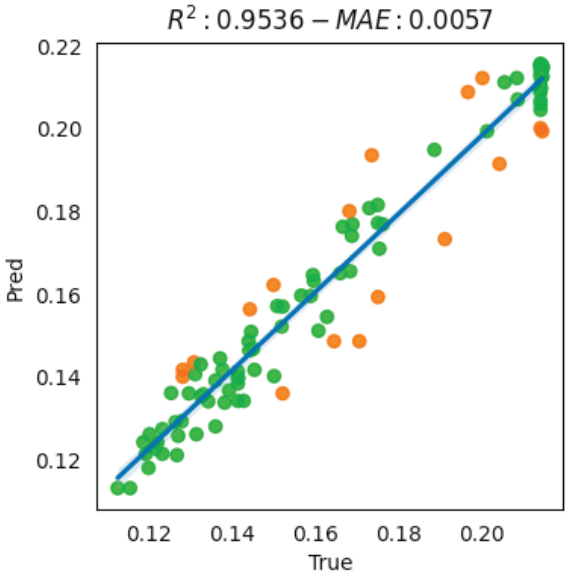


Figure 26: Regression plot for the Bootstrap Aggregating model.

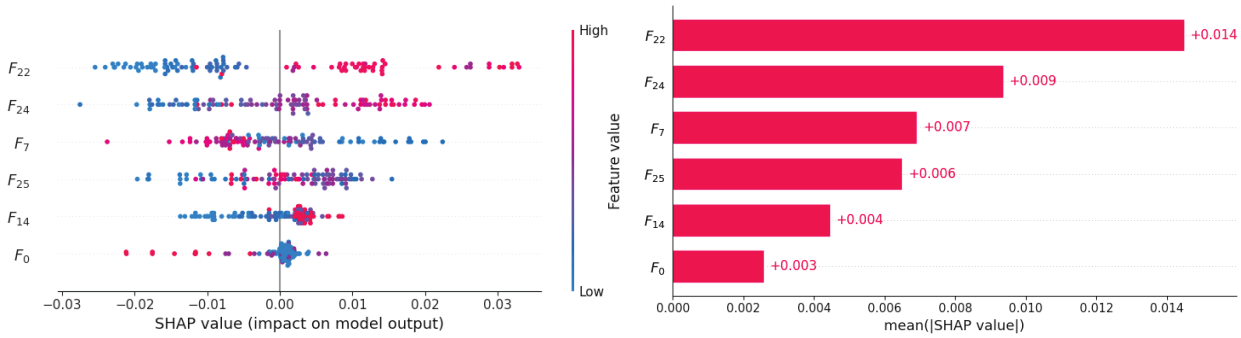


Figure 27: Beeswarm plot (left-hand side) and bar plot (right-hand side) show the distribution of the impact of each feature on the output of the Bootstrap Aggregating model.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

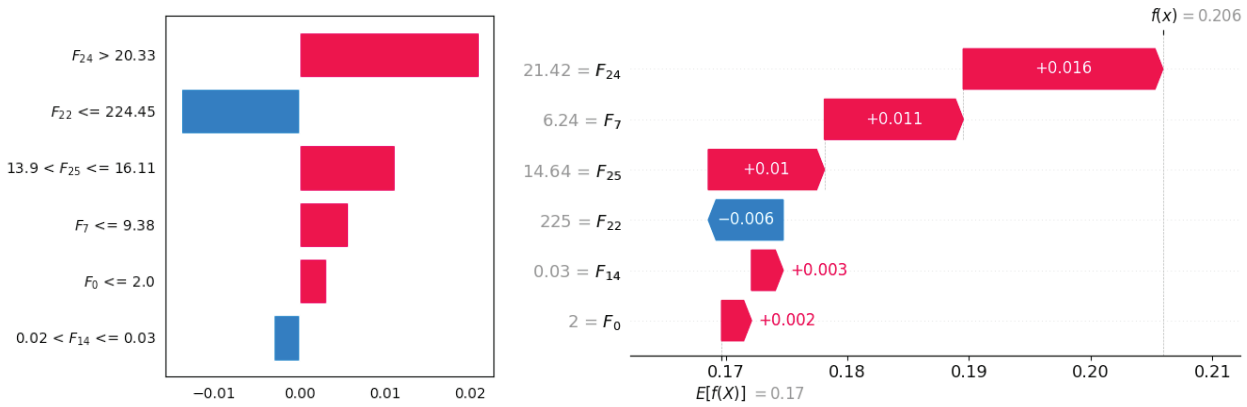


Figure 28: Explanation by LIME (left-hand side) and SHAP (right-hand side) for a representative structure whose PGA value for D3 limit state is calculated by the Bootstrap Aggregating model.

After the SFS the most selected features were 6: Number of level, Area of the openings of the external masonry walls at the elevation floor (EF), Average shear strength of masonry (GF), Shear Modulus (GF), Effective shear area (GF), Effective shear area (EF).

Figure 26 shows the regression plot, whereas Figure 27 and Figure 28 show the plots relative to the global explanation and to the local explanation (i.e., explanation relative to a representative structure), respectively.

The result achieved by this model is quite accurate (see the regression plot) and also aligns with structural engineering. As the beeswarm and the bar plots show, the shear modulus (GF), the effective shear area (GF), and the area of the openings of the external masonry walls at the elevation floor (EF) are the technical parameters that most influence the output of the model.

Bootstrap Aggregating achieves the best R^2 coefficient and the best MAE, which are equal to 0.9536 and 0.0057, respectively. Although the MAE is equal to that achieved by the Light Gradient Boosting Machines, the Bootstrap Aggregating model needs fewer technical parameters as input to predict the PGA.

4.10 Discussion on the influence of technical parameters on the PGA_{D3}

According to the XAI models presented, the technical parameters of the masonry buildings that most influence the PGA_{D3} are the shear modulus at GF (F22), the effective shear area at GF (F24), and the area of the openings of the external masonry walls at GF (F7). This is consistent with the mechanics of masonry structures.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

The attainment of severe damage to a masonry structure as a consequence of seismic action is often associated with the shear failure of its masonry panels. The shear failure of a masonry panel occurs when the shear force acting on the panel attains its shear strength, which is proportional to the product of the cross-sectional area of the panel by the shear strength of the masonry. The overall capacity of a masonry structure to resist lateral forces produced by seismic actions is thus affected by the product of the effective shear area (which is in turn affected by the number of openings) by the shear strength of the masonry.

It should be noted that the openings in the external walls affect the seismic capacity more than those in the internal walls due to torsional effects which occur when the center of rigidity does not coincide with the center of mass. The influence of the shear modulus is also consistent with the mechanics of masonry structures because the shear modulus influences the lateral stiffness of the structure and, as a consequence, its natural period, which, in turn, affects the lateral force applied to the structure for a given intensity of the seismic action.

Some XAI models (e.g., the Random Forest) identify the masonry compressive strength (F_{16} and F_{17}) as an essential parameter affecting the seismic performance of the masonry buildings. From the structural point of view, this is related to attaining the flexural capacity of masonry panels when the structure is subject to seismic actions.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

5. CONCLUSIONS

This deliverable has presented the XAI models designed and developed to predict the damage to structures after a seismic event using benchmark structures similar to those in the pilot areas, which will be the focus of the case study. The Bootstrap Aggregating model achieved the best performance, with an R^2 coefficient of 0.9536 and a mean absolute error of 0.0057.

Also, the explanations provided by the models were consistent with the foundational aspects of structural engineering. This means that the developed models can provide explanations that help experts understand the reasoning behind each model prediction. Therefore, we can affirm that the designed and developed artificial intelligence models resulted precisely in line with the initial requirements and fully met the predetermined expectations.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

APPENDIX

This appendix presents new case studies involving reinforced concrete structures instead of masonry structures. We started to generate the technical parameters for this new building type. The resulting analysis will be carried out in a similar manner to that presented in this deliverable that was carried out for masonry buildings.

Reinforced Concrete building (RC)

The input parameters are divided into three macro categories. Specifically, the first is focused on the *geometric* parameters characterizing the structure at the global level. At the global geometry level, the parameters are the number of floors underground (*number of underground levels*) and the number of floors in elevation (*number of floors*) of which the concrete structure is composed. In addition to defining the floors, the structure's total height is identified after identifying the seismic zero (*Estimate building height*). In the *geometry* of the concrete structure, the total length of the building in the X- and Y-direction (*total length X-side and total length Y-side*) is also reported.

The second macro category focuses on the *mechanical* parameters characterizing reinforced concrete structures, such as the normal stress acting in the ground floor column, three categories of columns belonging to 3 different zones within the structure are considered (*Average stress in the columns in the corner at the ground floor, Average stress in the external column at the ground floor and Average stress in the internal columns at the ground floor*). This category also includes the ratio between maximum floor eccentricity and minimum floor eccentricity (*Ratio between maximum/minimum eccentricity*). Additionally, the frame stiffness is determined, then the ratio of the column stiffness to the beam stiffness of all frames in the *X-direction* and *Y-direction* is made, then for each direction, the maximum and minimum values are compared (*Ratio between maximum/minimum stiffness X direction, Ratio between maximum/minimum stiffness Y direction*). The parameters belonging to the second macro category manage to describe in depth, and to the point, the mechanical behavior of structural elements (*columns and beams*). Thus, in the second category, all *mechanical* parameters of concrete and steel are indirectly considered.

Finally, the last macro category is focused on the *Seismic Analysis*, taking into consideration, the mass of the individual floor the floor with the highest and the one with the lowest mass is identified

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

and then the comparison is made (*Seismic floor mass MAX / Seismic floor mass MIN*), then the ratio between the maximum and the minimum value of the ratio of the seismic mass of a floor with the total area of resistant concrete for that floor is identified (*Parameter between the max and the min (Ratio of Seismic floor mass / Concrete area)*). In this category, the period of the structure (*Period of the structure*) is evaluated through a simplified method and the shear force at the base (*Shear force at the base*).

The number of input data for reinforced concrete structures does not depend on the number of floors since the ratio of the maximum value to the minimum value of the parameter under consideration is used. Therefore, in such structures, there are 15 global parameters.

Based on considerations by structural engineering experts, this deliverable describes each building by 15 global technical parameters. After a preliminary study, we made the number of technical parameters describing a structure independent of the number of its floors. Table 3 summarizes all the parameters used in the experiments.

Table 3: Technical parameters for reinforced concrete buildings.

ID	Description
F_0	Number of underground levels
F_1	Number of floors
F_2	Estimate building height
F_3	Total length X-side
F_4	Total length Y-side
F_5	Seismic floor mass MAX / Seismic floor mass MIN
F_6	Parameter between the max and the min (Ratio of Seismic floor mass/ Concrete area)
F_7	Average stress in the columns in the corner at the ground floor
F_8	Average stress in the external column at the ground floor
F_9	Average stress in the internal columns at the ground floor
F_{10}	Ratio between maximum/minimum eccentricity
F_{11}	Ratio between maximum/minimum stiffness X direction
F_{12}	Ratio between maximum/minimum stiffness Y direction
F_{13}	Period of the structure
F_{14}	Shear force at the base

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.