Building typological classification in Switzerland using deep learning methods for seismic assessment

Available online at www.sciencedirect.com

ScienceDirect

Procedia Structural Integrity 44 (2023) 1522–1529

Structural Integrity Procedia

www.elsevier.com/locate/procedia

XIX ANIDIS Conference, Seismic Engineering in Italy

Building typological classification in Switzerland using deep learning methods for seismic assessment

A. Casciato[a,b]*, A. Khodaverdian[a], G. Coletta[b], L. Scussolini[b], P. Lestuzzi[a], R. Ceravolo[b]

[a]École Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Earthquake Engineering and Structural Dynamics Laboratory (EESD), 1015 Lausanne, Switzerland.
[b]Politecnico di Torino, Department of Structural, Building and Geotechnical Engineering (DISEG), 10129 Turin, Italy.

Abstract

Natural disasters, such as earthquakes, have always represented a danger to human life. Seismic risk assessment consists of the evaluation of existing buildings and their expected response in case of an earthquake; the exposure model of buildings plays a key role in risk calculations. With this respect, in recent years, advanced techniques have been developed to speed up and automatize the processes of data acquisition to data interpretation, although it is worth mentioning that the visual survey is essential to train and validate Machine Learning (ML) methods. In the present study, the identification of building types is conducted by exploiting the traditional visual survey to implement a Deep Learning (DL) classification model. As a first step, city mapping schemes are obtained by classifying buildings according to the main features (i.e., construction period and height classes). Then, Random Forest (RF), a supervised learning algorithm, is applied to classify different building types by exploiting all their attributes. The RF model is trained and tested on the cities of Neuchatel and Yverdon-Les-Bains. The decent accuracy of the results encourages the application of the method to different cities, with proper adjustments in datasets, features and algorithms.

Keywords: SHM; Visual survey; Machine Learning; Random Forest; Seismic Assessment.

1. Introduction

The last 20 years have seen significant losses due to natural catastrophes (Silva, 2018). Earthquakes contribute to a significant amount of annual losses, reaching 60% for some years (MunichRe, 2012). A recent example in Europe is the severe earthquake happened in L'Aquila (IT) with magnitude of 5.9 ML on April 6th, 2009 which caused more than 300 victims, 1,600 injured people and financial losses of ~ 10 billion Euros (Greco et al., 2018). Moreover,

10.1016/j.prostr.2023.01.195

environmental factors such as global urbanization processes and increasing spatial concentration of exposed elements (e.g., people, buildings, infrastructure, etc.) in earthquake-prone area lead to an increased seismic risk (Geiß et al., 2016). To implement risk mitigation strategies effectively, it is essential to fully understand the potential losses to humans and the economy in a given urban area.

The assessment of earthquake damages requires three components: (1) a hazard model which describes possible earthquakes and consequent ground motion, (2) an exposure model with a list of assets. (3) a fragility model which links the probability of damage to the level of ground motion. For instance, concerning the exposure model, Lestuzzi et. al. (Lestuzzi et. al., 2016) inspected more than 2500 buildings in two Swiss cities and, using blueprints, prepared a detailed inventory. However, even though in situ field surveys are the most comprehensive and accurate way to acquire inventories, they are rather expensive, especially when the number of buildings is large. That could be even more challenging without detailed data of the existing structure stock, which is almost expected in different areas in either developed or developing countries. Over the last few decades, several methodologies have been implemented to obtain information about exposed built environment in terms of its vulnerability. For instance, Remote Sensing (RS) techniques are effectively used to create an inventory of assets exploiting a large variety of mid-resolution (i.e., satellite/ aerial imagery) and/or high-resolution (i.e., LiDAR) datasets available nowadays (Pittore, Wieland, and Fleming 2017). Besides providing an overview of building stock distribution (Geiß et al., 2016), remote sensing can also extract its envelope characteristics (e.g., height, roof shape, and material). An exposure dataset was recently produced by Torres et al., (Torres et al., 2019), integrating aerial LiDAR points, orthophotos, and satellite images to characterize the building stocks in Lorca, Spain. It is shown that the proposed procedure for data integration is fast and easy to deploy in other cities.

Seismic damage assessment is commonly based on the type of structure, which represents the lateral load-reassuring system. Datasets often lack this characteristic. Data mining methodologies (Campostrini et al. 2018; Guettiche, Guéguen, and Mimoune 2017; Riedel et al. 2015) have been used to develop proxies, that link the building features, which are available from cadastre/census databases or remote sensing, to structural vulnerability. Recent studies demonstrated that Machine Learning (ML) models, fine-tuned on the basis of an adequate ground-truth dataset, can perform well in earthquake risk assessment and reduce the cost of a large-scale survey. For instance, Riedel et al., (Riedel et al., 2014) proposed the Association Rule Learning (ARL) method for discovering relationships among features in large building databases and implemented it using the elementary attributes in Grenoble, France. Riedel et al., (Riedel et al., 2015) also applied Support Vector Machine (SVM) (C. Cortes, 1995, Noble, 2006), a classification/regression algorithm, to develop two vulnerability agents based on ARL and SVM methods.

The goal of the present study is to implement a Deep Learning (DL) model combined with traditional visual survey to identify building types. The case study is represented by a dataset of 3537 buildings of Neuchatel and 2808 of Yverdon-Les-Bains. In section 2, the outcomes of the visual survey are presented exploiting the building taxonomy proposed by Lagomarsino et al. (Lagomarsino et al., 2006) and obtaining city mapping schemes. In section 3, the Random Forest (RF) a supervised learning algorithm is presented and applied on the available dataset. The accuracy resulting from the classification of three data sets, concerning the two cities individually and in a combined way, is calculated and then compared. Conclusions are finally drawn in section 4.

2. Survey and building data analyses

Although innovative methods (e.g., RS and DL techniques) strongly simplify the process of classifying building types, it is worth mentioning that visual surveys remain crucial in providing ground-truth labeled datasets used to evaluate the accuracy of the method. Construction practice, which varies among countries, is also a significant factor, making surveys necessary. In this section, the visual survey and the analysis of the results are presented.

2.1. Building datasets

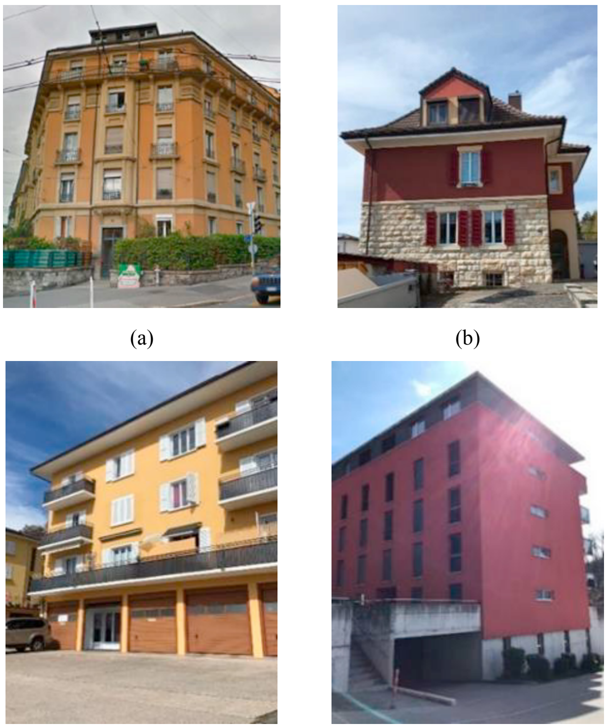
Thanks to the building dataset from the federal office for the environment (BAFU), several detailed features of buildings located in Switzerland are available. The dataset includes a wide range of features, ranging from building location (canton’s name, ZIP-code, coordinates of building location), to construction characteristics (period of construction, footprint, number of storeys, roof type) details of housing units (number of housing units, cumulative number of rooms in housing units), usage (number of inhabitants or equivalent full-time employees) and financial value (replacement value in CHF). Among these, 13 features are numerical and 4 of them are categorical. However, in order to facilitate visual surveys, a second dataset, containing street and number of each building, is considered and merged with the BAFU dataset according to EGID (Federal building identification number).

2.2. Visual surveys

In exposure modeling, the building taxonomy (Porter et al., 2001), a list of building types, classified based on method and/or materials used for construction, need to be defined. Here, the building types proposed by Lagomarsino et al. (Lagomarsino et al., 2006) are used. This taxonomy has been used in the Risk-UE project and can be described as an updated version of the taxonomy proposed in EMS-98. Building types and their definitions are presented in Table 1 and images of the building types are shown in Figure 1.

Table 1. Building types.

Building type	Description
M3	Masonry buildings with simple stone
M4	Masonry buildings with massive stone
M5	Unreinforced masonry (bricks) with flexible floors
M6	Unreinforced masonry - RC floors (rigid floors)
RCW	Reinforced concrete buildings with shear walls
RCF	Reinforced concrete buildings
W	Wood structures
S	Steel structures



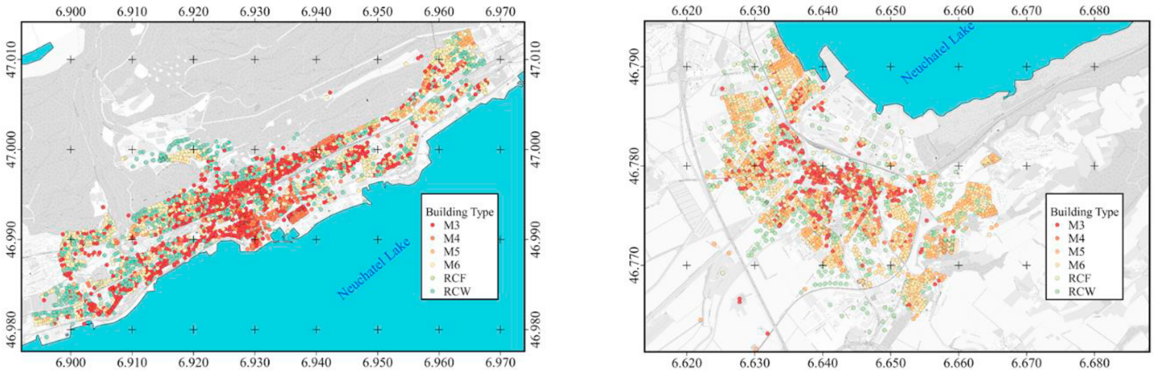
(a) (b) (c) (d)

Fig. 1. Building types: (a) M3; (b) M4; (c) M6; (d) RCW.

In order to have a ground-truth dataset, all buildings in the cities of Neuchatel and Yverdon-Les-Bains have been surveyed and the structural type for all buildings has been determined based on the taxonomy presented in Table 1.

2.3. Building feature analyses

The maps of the distribution of the building types in Neuchatel and Yverdon-Les-Bains are presented in Fig. 2.



(a) (b)

Fig. 2. Building type distribution in (a) Neuchatel and (b) Yverdon-Les-Bains.

As expected, stone masonry buildings (i.e., M3 and M4) are highly concentrated in the city center while the unreinforced masonry buildings (i.e., M5 and M6) and the reinforced concrete buildings (i.e., RCW and RCF) are mainly located in the suburban area of the cities. The comparison Fig. 2 (a) and (b) shows that the main difference between the two cities is the contribution of M5 building types: while a large density of M5 buildings can be found in Yverdon-Les-Bains, their contribution in Neuchatel is considerably lower. This can be explained by the presence of more recently-built residential M5 buildings in Yverdon-Les-Bains.

Although very detailed information (e.g., year of construction, number of storeys) are available in the building datasets and will be exploited in the classification (see section 3), some features are categorized here for the sake of simplicity in the interpretation of surveys' result. The year of construction has been divided into six categories (Table 2) while the number of storeys has been splitted into three categories (Table 3).

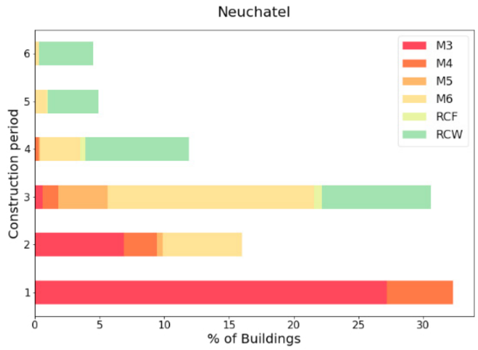
Table 2. Classification of the construction periods.

Construction period	Code
< 1919	1
1919 - 1945	2
1946 - 1970	3
1971 - 1990	4
1991 - 2003	5
> 2003	6

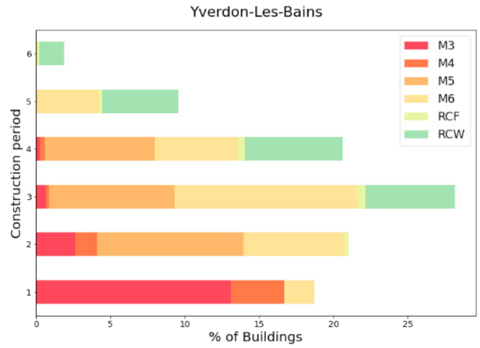
Table 3. Classification of the number of stories.

Number of stories	Height class
1-3	L
4-6	M
>7	H

Among the building features, the period of construction is a crucial characteristic to define the building type, considering that it gives information regarding materials and/or construction techniques that were available in different periods. Fig. 3 shows the distribution of building types in different construction periods, introduced in Table 2, for two cities. It is clear that buildings constructed in period 1 (i.e., <1919) are mainly made of stones (i.e., M3, M4) while the M5 and M6 buildings belong to periods of 2-4. Construction of RC structures (i.e. RCF, RCW) started in period 3 and spread in periods of 5 and 6.



(a)



(b)

Fig. 3. Distribution of building per period of construction: (a) Neuchatel; (b) Yverdon-Les-Bains.

3. Building typological classification

Assigning typological classes to the building stocks is an essential process for seismic assessment and can hardly be done using the traditional (i.e., visual) survey when dealing with seismic vulnerability of existing buildings on a large scale. A supervised learning algorithm, RF, is briefly introduced in this section and it is applied to predict building types based on their features. The method can be used for both classification and regression.

3.1. Random Forest classifier method

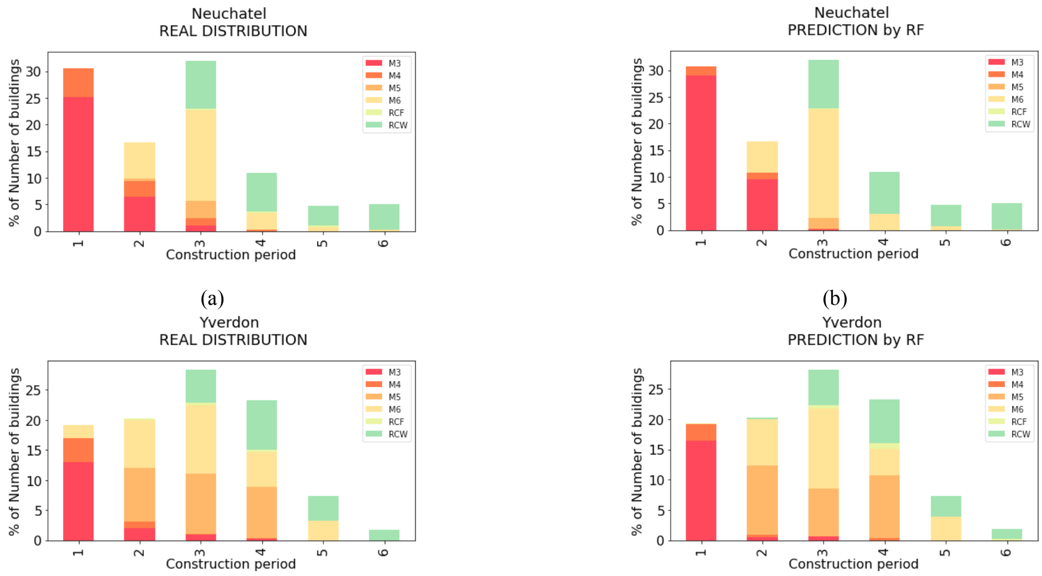
RF method is an ensemble of random decision tree classifiers (Ho, 1995), which here is used to discriminate between different classes based on building features. The final prediction is made by combining the predictions of individual trees that form the decision forests. In other words, a decision forest includes a set of expert tree classifiers and all of these would vote for the most probable class of an input vector of features (Ho, 1998). RF has been widely used in areas of geography, economics, medicine, and engineering (Navlani, 2018). Fan et al. (Fan et al., 2013) extracted building geometrical features from LiDAR point clouds using the RF method. Bosch et al. (Bosch et al., 2007) explored the problem of classifying images by the object categories by combining RF classification and multi-way SVM. Based on literature results, RF can be considered among the most performing predictive models (Vens, 2013).

The Scikit-learn, an open-source Python module for ML, is used to implement the RF classifier (Pedregosa et al., 2012). The dataset of labelled samples is randomly divided into two sets: a training set, involving the 80% of samples and a test set for validation including the rest of the data. This division resulted from several simulations carried out in order to find the optimal balance leading to the best accuracy.

Among the building features that have been mentioned in 2.1, only the most significant ones have been used in the application of RF. In particular, only those features that are strongly linked to the building type were selected (i.e., Roof type), while the others were discarded (i.e., Federal building identification number) in order not to provide confusing or useless information to the algorithm.

3.2. Results of RF method and comparison with visual survey

Considering the test dataset, a comparison between the real and the predicted distribution of building types is shown in Fig. 4(a-b) and in Fig. 4(c-d). As observable, the models provide a good prediction of the building types in general. The biggest confusion is found between M3 and M4 in period of construction of 1 and 2. This could be explained by the low population of M4 building type.



(a) (b) (c) (d)

Fig. 4. Real and predicted distributions of building types in (a-b) Neuchatel and (c-d) Yverdon-Les-Bains.

3.3. Accuracy of RF model

To evaluate the performance of the RF, the Confusion Matrix (CM), also known as error matrix, is obtained. The CM is a $N \times N$ matrix; where N is the number of building types. A CM cell indicates the number of test samples for each combination of ground-truth building types (N) and assigned building types (N). The diagonal elements show the number of buildings that have been correctly predicted by the RF model. On the contrary, the off-diagonal elements show the number of buildings that have been incorrectly predicted by the RF model. It is important to mention that, the fewer building number for each building type, the harder it will be for the model to predict that building type, due to the scarcity of training examples. For this reason, more errors are expected for the M4 and RCF building types, which have the lowest contributions in the dataset. Fig. 5 shows the CMs obtained from the RF models trained and tested on the databases of (a) Neuchatel, (b) Yverdon-Les-Bains, and (c) Neuchatel and Yverdon-Les-Bains together.

The results of the RF models are evaluated through the accuracy measure AM1, that is the overall accuracy of building types, which is based on the confusion matrix. AM1 is calculated as the ratio between the number of correctly classified buildings over the total number of buildings, as given by (1).

$$AM1 = \frac{\sum_i A_{ii}}{\sum_i \sum_j A_{ij}} \tag{1}$$

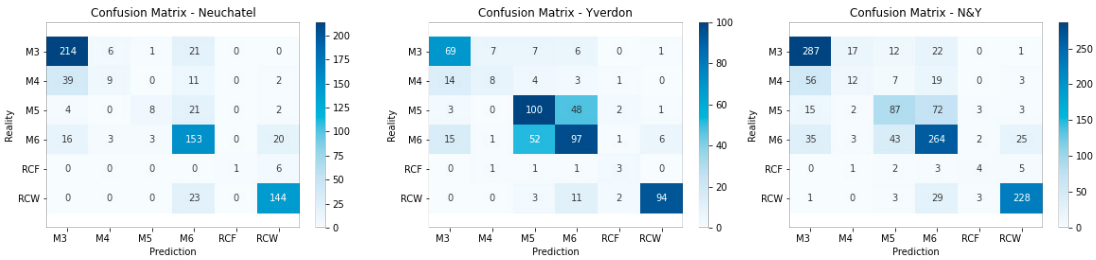


Fig. 5. Confusion matrices for (a) Neuchatel; (b) Yverdon-Les-Bains; (c) concatenated dataset of Neuchatel and Yverdon-Les-Bains.

In **Errore. L'origine riferimento non è stata trovata.** a comparison is carried out between the values of accuracy of the RF models which are separately trained based on the datasets of Neuchatel, Yverdon-Les-Bains and on the concatenated datasets of the two cities (RF N&Y). The values obtained show that RF N&Y model generalize the test set with a slightly higher accuracy than the average of the RF Neuchatel and the RF Yverdon-Les-Bains models.

Table 4. Values of AM1 accuracy for RF models trained and tested on datasets of (a) Neuchatel; (b) Yverdon-Les-Bains; (c) Neuchatel and Yverdon-Les-Bains.

	Neuchatel	Yverdon-Les-Bains	N&Y
AM1	0.73	0.65	0.70

4. Conclusion and future works

The prevention of the tragic consequences of natural disasters is a topic of growing interest. To this end, it is essential to assess the risk to which cities are exposed and to plan, if necessary, retrofitting measures. In the proposed research, a method based on the combination of traditional visual survey and modern DL algorithm is suggested for the classification of building types. First, a large number of buildings in Neuchatel and Yverdon-Les-Bains were investigated with a visual survey. The main characteristics of the buildings were considered, such as the roof shape, the façade aspect, the presence of balconies, and each building has been assigned to a building type. This very demanding step (both in terms of time and energy) was carried out according to expert engineering judgment. The taxonomy used refers to Lagomarsino et al. (Lagomarsino et al. 2006). Then, using the information from the Federal Office of Buildings and Logistic of Switzerland, the original dataset was enriched with more than 18 building characteristics.

After that, RF, a supervised learning algorithm used for both classification and regression, was applied. This algorithm was exploited to get the building type for each sample, by using the attributes and applying the RF and implementing a classification. The method was separately trained and tested on the datasets of the cities of Neuchatel and Yverdon-Les-Bains. The results of the RF models are evaluated with the accuracy measure AM1, that is the overall accuracy of building types, which is based on the confusion matrix (also known as the error matrix). Discrete accuracies have been reported for both cities. Then, a new RF model was created by training and testing on the concatenated datasets of Neuchatel and Yverdon-Les-Bains. For this model, a better performance was observed, compared to the models trained and tested on the individual cities, as a higher number of data provides an improvement in accuracy.

The decent accuracy of the results showed the robustness of the method in the classification of building types, paving the path for a wider application. In fact, the strength of a DL model is that by expanding the training dataset by including different building types from different cities, its accuracy and efficiency can be increased.

As a future development of this study, the already trained RF models may be applied to other Swiss cities. Indeed, to cover the possible spatial diversity of building type distribution in Switzerland, some additional districts of cities (e.g. big cities, rural areas) may be surveyed and will be serving as validation datasets. Moreover, different classification algorithm, such as Neural Networks, SVM could be tested to evaluate possible improvement.

References

Bosch, A., Zisserman, A., & Muñoz, X. (2007). Image Classification using Random Forests and Ferns. Undefined. https://doi.org/10.1109/ICCV.2007.4409066

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

Fan, S., Zhang, A., Hu, S., & Sun, W. (2013). A method of classification for airborne full waveform LiDAR data based on random forest. Zhongguo Jiguang/Chinese Journal of Lasers, 40(9). https://doi.org/10.3788/CJL201340.0914001

Geiß, C., Jilge, M., Lakes, T., & Taubenböck, H. (2016). Estimation of Seismic Vulnerability Levels of Urban Structures With Multisensor Remote Sensing. 9(5), 1913–1936.

Greco, A., Pluchino, A., Barbarossa, L., Caliò, I., Martinico, F., & Rapisarda, A. (2018). Seismic and Energy Renovation for Sustainable Cities A simplified model based on self-organized criticality framework for the seismic assessment of urban areas.

Ho, T. K. (1995). Random Decision Forests. 47, 4–8.

Ho, T. K. (1998). The Random Subspace Method for Constructing Decision Forests. 20(8), 832–844.

Lagomarsino, S., & Giovinazzi, S. (2006). Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings. 415–443. https://doi.org/10.1007/s10518-006-9024-z

MunichRe (2012). Topics Geo: Natural Catastrophes 2011 Analyses, Assessments, Positions, Munich, Germany

Navlani, A. (2018). Sklearn Random Forest Classifiers in Python - DataCamp. https://www.datacamp.com/community/tutorials/random-forests-classifier-python

Noble, W. S. (2006). What is a support vector machine? NATURE BIOTECHNOLOGY, 24. http://www.nature.com/naturebiotechnology

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://arxiv.org/abs/1201.0490v4

Porter, K. A., Kiremidjian, A. S., & LeGrue, J. S. (2001). Assembly-Based Vulnerability of Buildings and Its Use in Performance Evaluation: Https://Doi.Org/10.1193/1.1586176, 17(2), 291–312. https://doi.org/10.1193/1.1586176

Riedel, I., Guéguen, P., Dalla Mura, M., Pathier, E., Leduc, T., & Chanussot, J. (2015). Seismic vulnerability assessment of urban environments in moderate-to-low seismic hazard regions using association rule learning and support vector machine. 1111–1141. https://doi.org/10.1007/s11069-014-1538-0

Riedel, I., Gueguen, P., Dunand, F., Cottaz, S., Riedel, I., Gueguen, P., Dunand, F., & Macro-scale, S. C. (2014). Macro-scale vulnerability assessment of cities using Association Rule Learning To cite this version : HAL Id : hal-01015723.

Silva V. (2018). Critical Issues on Probabilistic Earthquake Loss Assessment,
Journal of Earthquake Engineering, 22:9, 1683-1709, DOI: 10.1080/13632469.2017.1297264

Torres, Y., Juan, J., Gaspar-escribano, J. M., Haghi, A., Martínez-cuevas, S., Benito, B., & Carlos, J. (2019). Int J Appl Earth Obs Geoinformation Integration of LiDAR and multispectral images for rapid exposure and earthquake vulnerability estimation . Application in Lorca , Spain. Int J Appl Earth Obs Geoinformation, 81(May), 161–175. https://doi.org/10.1016/j.jag.2019.05.015

Vens. (2013). Random Forest. https://scholar.google.it/scholar?hl=it&as_sdt=0%2C5&q=random+forest+vens+2013&oq=vens

Wu, H., Cheng, Z., Shi, W., & Miao, Z. (2014). An object-based image analysis for building seismic vulnerability assessment using high-resolution remote sensing imagery. 151–174. https://doi.org/10.1007/s11069-013-0905-6