14th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP14
Dublin, Ireland, July 9-13, 2023

Exposure Modelling through Machine Learning for Multi-Hazard Risk Assessment

Vitor Silva
Principal Investigator, Dept. of Civil Engineering, University of Aveiro, Aveiro, Portugal

Jorge Lopes
Graduate Student, Faculty of Science and Technology, University Fernando Pessoa, Porto, Portugal

Feliz Gouveia
Full Professor, Faculty of Science and Technology, University Fernando Pessoa, Porto, Portugal

Romain Sousa
Researcher, Polytechnic of Leiria, Leiria, Portugal

ABSTRACT: In the last decades, most efforts to catalogue and characterize the built environment for multi-hazard risk assessment have focused on the exploration of housing census data, cadastral datasets, and local surveys. Some of these datasets are only updated every 10 years, do not provide building locations, and might be cost-prohibitive for large-scale risk studies. It is thus clear that methods to characterize the built environment for large-scale risk analysis at the asset level are currently missing, which hampers the assessment of the impact of natural hazards. Some recent efforts have demonstrated how machine learning algorithms can be trained to recognize specific architectural and structural features of buildings based on their façades, and probabilistically propose one or multiple building classes. In this study, we demonstrate how such algorithms can be combined with street level images to develop exposure models for multi-hazard risk analysis. We created a training dataset with 5276 images of buildings from the parish of Alvalade (in Lisbon) and used it to calibrate an algorithm to develop exposure models. The resulting model was used to estimate the impact of an earthquake scenario, and the results were compared with the impact calculated using the ground truth data. We discuss how the uncertainty from such models can be propagated into the risk results, and how this approach can be used for the assessment losses due to hydrological and meteorological hazards.

The assessment of the impact caused by natural hazards requires detailed exposure models, characterizing the location, value, occupants, and construction attributes of the elements exposed to the hazards. A high resolution in exposure models is particularly important when assessing the impact of localized hazards such as floods, landslides, or lahars. Yet, most studies tend to use cadastral datasets (e.g., Tyagunov et al. 2014; Acevedo et al. 2017), data from the national housing census (e.g., Silva et al. 2014a), or approximated exposure datasets often developed using proxy data such as population or socio-economic data. Despite the value of these sources of information, they do not contain the exact location of each building, and tend to classify the building stock according to general indicators. Alternative approaches for the characterization of the building stock at the urban level have been proposed by several studies, spanning from the organization of field surveys (e.g., Vicente et al. 2010), classification of the building stock through virtual visits (e.g., Santa-Maria et al. 2017), or the use of volunteered geographic information (VGI), such as OpenStreetMap (Cerri et al. 2021). These approaches have their own strengths and limitations, but tend to cover also specific areas of a given region or country due to the associated high costs.

It is thus clear that it is fundamental to explore new sources of data and innovative mathematical formulations to facilitate the development of exposure models with a high resolution and detail in the structural attributes. The recent availability of building imagery has led researchers to explore techniques for the development of building datasets at the urban level, for the purposes of natural hazards risk assessment. These approaches usually involve the combination of street-level building imagery with machine learning algorithms to automatically identify specific building attributes (e.g., Wang et al. 2019; Gonzales et al. 2020; Pelizari et al. 2021). These features typically include the main material of construction, height of the buildings, epoch of construction, type of use, irregularities, and level of ductility (which is usually a proxy for the seismic performance). This information is fundamental to assign a vulnerability class to each building, which can then be combined to one or multiple fragility for vulnerability functions (e.g., Martins and Silva 2021). This is an important step in the process of assessing the potential impact of natural hazards.

Some examples of the application of this technique can be found in the literature (e.g., San Francisco, United States - Wang et al. 2019; Medellin, Colombia - Gonzales et al. 2020; Santiago, Chile - Pelizari et al. 2021; Oslo, Norway - Guione et al. 2022). However, with the exception of the research carried within the SimCenter (https://simcenter.designsafe-ci.org/), it does not seem like the data used for the training of the algorithms is publicly available. In this study, we use an open database of 5276 building images for Alvalade, a parish in the district of Lisbon, to train a number of machine learning algorithm to automatically classify buildings according to a number of attributes. Then, to test how misclassifications by these algorithms can affect risk estimates, we compared the impact caused by an earthquake scenario using the real exposure data (i.e., ground truth) and an exposure model generated by the best-performing machine learning algorithm.

1. THE PARISH OF ALVALADE, LISBON

We tested various machine learning algorithms using the parish of Alvalade as a case study. The construction of Alvalade took place between the decade of 1940s and 1970s, and aimed at reducing the residential needs of Lisbon (Ferreira, 2014). Alvalade includes social equipment and infrastructures, namely the main Campus of the University of Lisbon ("Cidade Universitaria"), the largest hospital in Portugal (Santa Maria), the National Laboratory for Civil Engineering (LNEC), undergraduate schools and sports facilities.

The period of construction of this parish marks the transition in Portugal from predominantly masonry to concrete building construction. Reinforced concrete started being used to construct the floors of masonry buildings, and later on also to incorporate beams and columns in the structures (Bernardo et al, 2021), though the structural behavior might still be similar to an unreinforced masonry due to the small dimension of the concrete elements and reduced area of reinforcement bars. According to studies carried out by Lamego (2014), Ferrito et at (2016) and Milosevic et al., (2020), most buildings in this area may not present an adequate seismic performance for the hazard level expected for Lisbon. Moreover, most of the existing building stock was built before the enforcement of the first seismic code in 1958, and certainly before the first modern seismic code (1983). A satellite image of the parish of Alvalade is presented in Figure 1.

Figure 1. Satellite image of the parish of Alvalade, within the district of Lisbon, Portugal (source: Google Satellite).

The database of building imagery used in this study comprises 5,276 images covering 2,457 buildings within the parish of Alvalade, all classified according to the main structural characteristics. The database is publicly available through an open Github repository at https://github.com/vsilva028/ML, and users can download the data, or contribute with additional information.

2. TRAINING OF MACHINE LEARNING ALGORITHMS

We selected the following 6 machine learning algorithms from the ImageNet (ILSVRC) database (Russakovsky et al., 2015): ResNet50V2, InceptionResNetV2, NASNetLarge, Xception, InceptionV3 and DenseNet201. We ran all the algorithms in a i7 machine, 64 GB RAM, using a recent graphics card (Zotac Gaming GeForce RTX 3090 Trinity OC 24GB GDDR6X).

We cleaned the dataset using the VGG16 model trained on the Places365 dataset, and using a set of 24 classes to detect a façade: 'apartment_building/outdoor', 'beach_house', 'building_facade', 'chalet', 'church/outdoor', 'cottage', 'courthouse', 'embassy', 'fire_station', 'hangar/outdoor', 'hospital', 'hotel/outdoor', 'house', 'hunting_lodge/outdoor', 'mansion', 'manufactured_home', 'motel', 'office_building', 'palace', 'schoolhouse', 'shed', 'skyscraper', 'synagogue/outdoor', and 'tower'. The rule to detect a façade was taken from Pelizari et al., (2021), with L_i being the label assigned to an image *i* and the C_i the 4 most likely predicted classes:

$$L_i = \text{“façade” if } |\, C_i \cap S \,| \geq 2, \text{ “other” otherwise}$$

Images not having a “façade” label were discarded. After cleaning, the images were assigned one of the following classes, combing the two attributes that influence the most the vulnerability of the assets to earthquake risk: construction material and number of floors. Based on these attributes, the following classes were defined:

- Masonry 1-3
- Masonry 4+
- Concrete 1-3
- Concrete 4-6
- Concrete 7+

In another ongoing study, additional attributes (e.g., age of construction, type of roof) more relevant to other natural hazards are also being incorporated in the list of vulnerability classes. However, the consideration of several classes may lead to a low number of images per class, which can reduce the accuracy of the models.

We performed data augmentation using rotation, vertical and horizontal shifting, random zoom, horizontal flipping and brightness change. The dataset was divided into train (80%, 2713 images) and test (20%, 678 images) subsets. From the training subset, a validation subset was created (20%, resulting in 848 images). We used transfer-learning, from the Places365 dataset, and fine-tuning procedures to get the best performing machine learning model for the given problem context with the aim of predicting the building material and the number of floors. The models were trained with 50 epochs, and a batch size of 32.

Without fine-tuning, the best performing models were ResNet50V2 and DenseNet201, both with an accuracy of 0.77. Using fine-tuning, the best accuracy, 83%, was obtained with Xception. The following figures show the loss and the accuracy of these models.

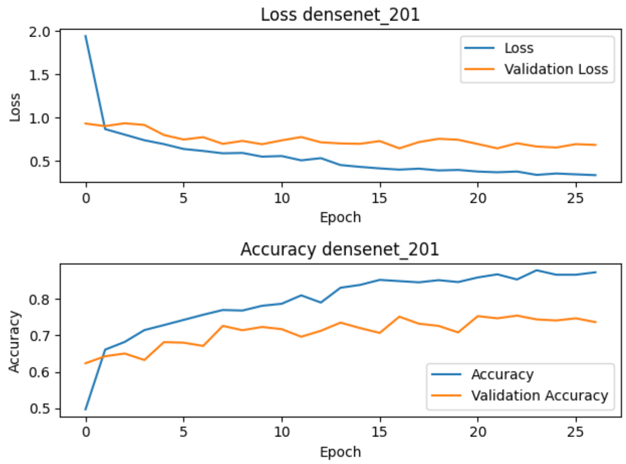

Figure 2: Results for the DenseNet201 algorithms.

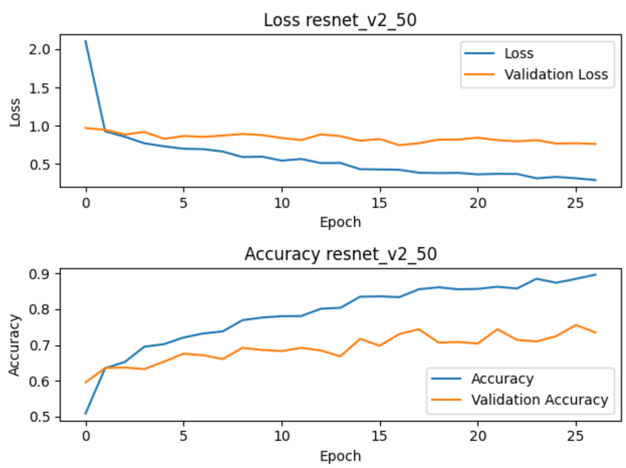

Figure 3: Results for the ResNet50V2 algorithm.

With fine-tuning the best model was Xception, with an accuracy of 0.83 in the test dataset.

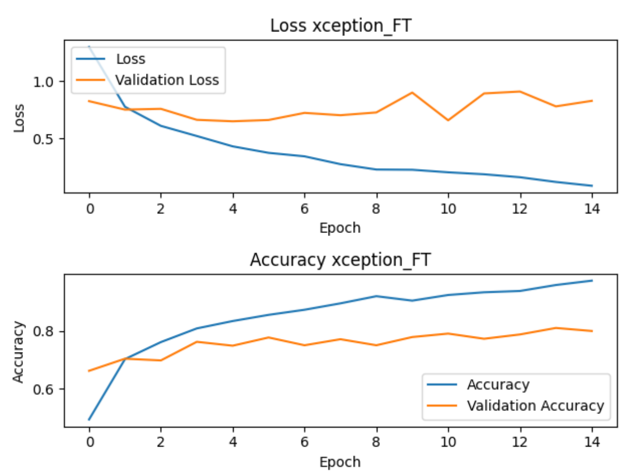

Figure 4: Results with the Xception algorithm.

Using the Xception algorithm, we developed an exposure model for Alvalade, and employed this dataset in the assessment of casualties, as described in the following section.

3. APPLICATION TO AN EARTHQUAKE SCENARIO

Earthquake scenarios are a fundamental tool for the development of post-disaster response plans or to raise risk awareness amongst the population. For example, the results from this type of analysis are frequently used to design earthquake drills to train the population on how to react to destructive events.

In this study, we considered an onshore event with a moment magnitude of 6.0, located northeast of Lisbon. The location and magnitude of this event is identical to the 1909 M6.0 Benavente earthquake, which heavily damaged a few villages in the district of Santarém. We defined the location of this event slightly closer to Lisbon to increase the impact of the scenario, considering the location of faults described in Carvalho et al. (2008). We used the OpenQuake-engine (Pagani et al. 2014) and the ground motion model from Atkinson and Boore (2006) for the calculation of the ground shaking in the affected region. We simulated 1000 ground motion fields considering the spatial correlation in the ground motion residuals. The spatial distribution of peak ground acceleration (PGA) for Portugal is depicted in Figure 5. For this event, the PGA in Alvalade is approximately 0.15g.

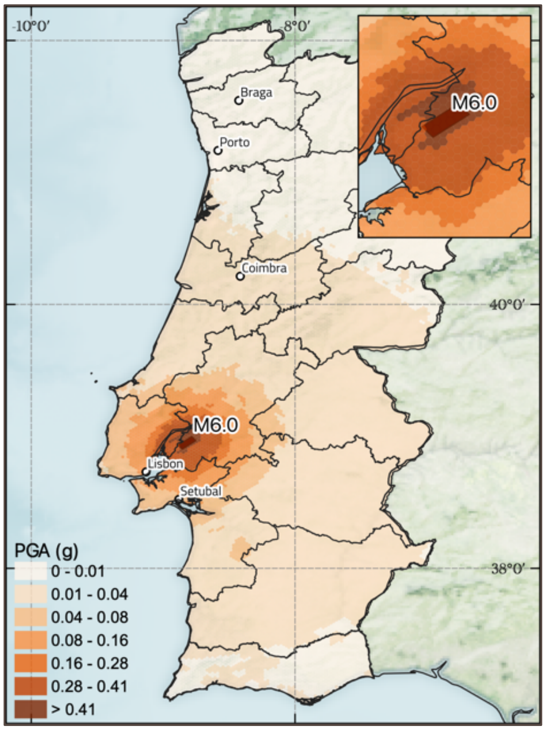

Figure 5: Spatial distribution of peak ground acceleration for the M6.0 earthquake scenario.

For the calculation of casualties due to this seismic event, we used the fragility and vulnerability functions proposed by Martins and Silva (2020) for reinforced concrete and masonry, and used the scenario damage and risk calculator of the OpenQuake-engine (Silva et al. 2014b). The purpose of these analyses was not only to demonstrate the application of the previously described machine learning model, but also to evaluate the potential error introduced due to misclassifications. To this end, we performed two earthquake scenarios: in the first assessment we directly used an exposure model using the data collected in the field (i.e., "true" exposure). In the second analysis, we used the exposure model that was created automatically by the machine learning model. We note that in this process, since we are considering the entire building stock of the parish of Alvalade, all the images have been used by the machine learning algorithm. This means that images that were used to train the algorithm, were also used to generate the exposure model. This led to an accuracy (88%) above the one presented previously for the test dataset (in which only images unseen by the algorithm had been used). To mitigate this bias, a different region should be used for this algorithm, but such data was not available in this study.

In both cases, the same ground shaking input and set of fragility and vulnerability functions were used. Figure 6 presents the casualties (fatalities, seriously injured and critically injured) for these two analyses, and the difference between the two (i.e., error).

The results from the two analyses indicate practically the same impact. Even though 12% of the buildings in Alvalade have been misclassified, the error in the results is practically insignificant. We evaluated further the structures that were misclassified to understand the reason behind these minor differences. Two situations were observed: 1) some buildings were (incorrectly) classified into classes with a higher vulnerability (i.e., masonry instead of concrete), but for some of the simulations the ground shaking was relatively low, which would result in minor damage regardless of the building class; 2) some buildings were misclassified into classes with a similar vulnerability (i.e., reinforced concrete with 4-6 or 7+ storeys), and thus no significant discrepancies would be expected in any case. Finally, we also noted that due to the fact that the region has more than 2000 buildings, we also have the averaging effect, in which some overestimations of casualties "compensate" for the underestimations. Additional analysis considering more refined building classes (for example, considering the period of construction) and other earthquake scenarios are the topic of another ongoing study by the authors.

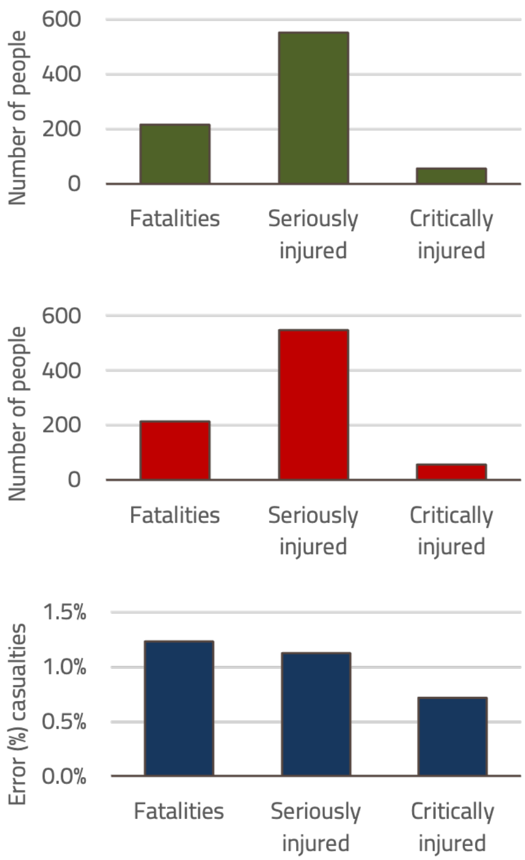

Figure 6: Casualties caused by the M6.0 earthquake scenario using the "true" exposure model (top), derived exposure model using the machine learning model (middle), and difference (i.e., error) between the two results (bottom).

4. CONCLUSIONS

In this study, we presented the calibration and testing of 6 machine learning algorithms (i.e., convolutional neural networks) for the purposes of automatically identifying specific structural and architectural features and defining a vulnerability class. The results using the Xception algorithm indicated an accuracy of 83% using the test dataset, in line with similar studies performed for Medellin, Chile, and Oslo. The algorithms were trained using an open database of building imagery, available at https://github.com/vsilva028/ML.

The findings from this study suggest that that machine learning algorithms are reliable and suitable for the classification of building stocks for the purposes of performing natural hazards risk analysis. Granted that street level imagery is available for a given region, it is possible to employ this technique to build a high-resolution exposure model, containing not just the exact location of each building, but also a set of vulnerability features. This level of detail in exposure models is not achievable with current methodologies that rely on cadastral datasets or housing census surveys. As previously mentioned, this capacity is particularly important for the assessment of risk due to localized hazards, such as hydrological and volcanic (i.e., lava flow, lahars) hazards. However, we note that services such as Google Street View are not available globally, and even in the western world rural areas are often not well covered. Nonetheless, the increase in the coverage offered by such services and the ever-improving accuracy of machine learning algorithms have the potential to drastically change the field of exposure modelling.

5. ACKNOWLEDGEMENTS

The authors acknowledge the support of Portuguese Science and Technology Foundation (FCT), through grant EXPL/ECI-EGC/1555/2021, project AI4DRR.

6. REFERENCES

Acevedo AB, Yepes C, Gonzalez D, Silva V, Mora M, Posada G, Arcila M, Rosales C (2019). "Seismic risk assessment for the residential buildings of the major three cities in Colombia: Bogotá, Medellín and Cali". *Earthquake Spectra*, Vol 36, Issue 1_suppl, https://doi.org/10.1177/8755293020942537.

Atkinson G, Boore D (2006). "Earthquake ground-motion prediction equations for eastern North America". *Bulletin of the Seismological Society of America*, 96(6), 2181-2205.

Bernardo, V., Sousa, R., Candeias, P., Costa A., Campos Costa A. (2021). "Historic Appraisal Review and Geometric Characterization of Old Masonry Buildings in Lisbon for Seismic Risk Assessment". *International Journal of*

Architectural Heritage, DOI: 10.1080/15583058.2021.1918287

Carvalho A, Zonno G, Franceschina G, Bilé Serra J, Campos Costa A (2008). "Earthquake shaking scenarios for the Metropolitan Area of Lisbon". *Soil Dynamics and Earthquake Engineering*, 28(5):347–364.

Cerri M, Steinhausen M, Kreibich H, Schröter K (2021). "Are OpenStreetMap building data useful for flood vulnerability modelling?" *Natural Hazards Earth Systems Sciences*, 21:643–662.

Ferreira J (2014). "Levantamento de Informação sobre Edifícios do Tipo "Placa" na Cidade de Lisboa. Análise Sísmica de um Caso Específico". *Master Thesis*. Instituto Superior Técnico, Universidade de Lisboa, Portugal. (in portguese)

Ferrito T, Milosevic J, Bento R (2016). "Seismic vulnerability assessment of a mixed masonry–RC building aggregate by linear and nonlinear analyses". *Bulletin of Earthquake Engineering*, 14:2299–2327, 10.1007/s10518-016-9900-0

Ghione, F., Mæland, S., Meslem, A., & Oye, V. (2022). "Building Stock Classification Using Machine Learning: A Case Study for Oslo, Norway". *Frontiers in Earth Science*, 10. https://doi.org/10.3389/feart.2022.886145

Gonzalez D, Rueda-Plata D, Acevedo A, Duque J, Ramos-Pollán R, Betancourt A, García S (2020). "Automatic detection of building typology using deep learning methods on street level images". *Building and Environment*, 177(15):106805.

Lamego P (2014). "Reforço sísmico de edifícios de habitação. Viabilidade da mitigação do risco." *PhD Thesis*. University of Minho, Portugal. (in Portuguese)

Martins L, Silva V (2020). "Development of a Fragility and Vulnerability Model for Global Seismic Risk Assessment". *Bulletin of Earthquake Engineering*, https://doi.org/10.1007/s10518-020-00885-1.

Milosevic J, Cattari S, Bento R (2020). "Definition of fragility curves through nonlinear static analyses: procedure and application to a mixed masonry-RC building stock". *Bulletin of Earthquake Engineering*, 18:513–545, 10.1007/s10518-019-00694-1.

Pagani M, Monelli D, Weatherill G, Danciu L, Crowley H, Silva V, Henshaw P, Butler L, Nastasi M, Panzeri L, Simionato M, Vigano D (2014). "OpenQuake Engine: An open hazard (and risk) software for the Global Earthquake Model". *Seismological Research Letters*, (85)3:692-702. doi.org/10.1785/0220130087

Pelizari, P., Geiß, C., Aguirre, P., Santa María, H., Merino Peña, Y., & Taubenböck, H. (2021). "Automated building characterization for seismic risk assessment using street-level imagery and deep learning." *ISPRS Journal of Photogrammetry and Remote Sensing*, 180. https://doi.org/10.1016/j.isprsjprs.2021.07.004

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). "ImageNet Large Scale Visual Recognition Challenge." *International Journal of Computer Vision*, 115(3). https://doi.org/10.1007/s11263-015-0816-y

Santa María, H., Hube, M., Rivera, F., Yepes-Estrada, C., Valcárcel, J. (2017). "Development of national and local exposure models of residential structures in Chile." *Natural Hazards* 86, 55–79 (2017). doi: 10.1007/s11069-016-2518-3.

Silva V, Crowley H, Pinho R, Varum H (2014a). "Seismic Risk Assessment for mainland Portugal." *Bulletin of Earthquake Engineering*, 13(2):429-457.

Silva V, Crowley H, Pagani M, Monelli D, Pinho R (2014b). "Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment." *Natural Hazards*, 72(3):1409-1427. doi.org/10.1007/s11069-013-0618-x

Tyagunov S, Pittore M, Wieland M, Parolai S, Bindi D, Fleming K, Zschau J (2014). "Uncertainty and sensitivity analyses in seismic risk assessments on the example of Cologne, Germany." *Natural Hazards and Earth System Sciences*, 14(6):1625:1640.

Vicente, R., Parodi, S., Lagomarsino, S., Varum, H., Mendes Silva, J. (2010). "Seismic vulnerability and risk assessment: case study of the historic city centre of Coimbra, Portugal". *Bulletin of Earthquake Engineering*, 9:4, 1067-1096.

Wang C, Yu Q, McKenna F, Cetiner B, Yu S, Taciroglu E, Law K. (2019). "Building Recognition using AI at Large-Scale NHERI-SimCenter/BRAILS: v1.0.1 (Version v1.0.1)."