Accepted version
doi: 10.1016/j.envsoft.2022.105604

Leveraging data driven approaches for enhanced tsunami damage modelling: insights from the 2011 Great East Japan event

Mario Di Bacco[a], Pierfrancesco Rotello[a], Anawat Suppasri[b] and Anna Rita Scorzini[a,*]

a Department of Civil, Environmental and Architectural Engineering, University of L'Aquila, L'Aquila, 67100, Italy
(M.D.B.: mario.dibacco@univaq.it ; P.R.: pierfrancescorotello@libero.it ; A.R.S.* (*corresponding author*: annarita.scorzini@univaq.it ; tel: +39 0862 434112)

b International Research Institute of Disaster Science, Tōhoku University, Sendai 980-8572, Japan
suppasri@irides.tohoku.ac.jp

Abstract

This study aims at developing an empirical, multi-variable tsunami damage model for buildings, based on machine-learning algorithms which leverage about 250.000 ex-post data surveyed by the Japanese Ministry of Land, Infrastructure and Transportation after the 2011 Great East Japan event in the Tōhoku region. By implementing simple geospatial tools, the dataset is integrated with additional explanatory variables, including, among others, factors accounting for the mutual interaction between the inundated structures. Tests on models' sensitivity to the number and type of input features used for model development reveal the importance, on the predictive performance, of considering usually neglected mechanisms like the shielding effect and the debris impact generation. The analysis for the potential spatial transferability indicates a reduction in the accuracy, thus suggesting a better suitability of empirical models for descriptive purposes, limiting their predictive ability only to region-specific cases.

Keywords

Building damage; tsunami; machine-learning; feature importance; spatial transferability; Japan

Software and data availability

Model development relies on the Python programming language (version 3.8.5) and open-source libraries including NumPy (Harris et al.; 2020), Pandas (McKinney, 2010), scikit-learn (Pedregosa et al., 2011), TensorFlow (Abadi et al., 2016) and XGBoost (Chen et al., 2016). The geospatial computations are performed with the open-source software QGis 3.16 with the modules SAGA 7.8.2 and GRASS GIS 7.8.6.
The original MLIT dataset is publicly available from the website of the Ministry of Land, Infrastructure, and Transportation of Japan http://www.mlit.go.jp/toshi/toshi-hukkou-arkaibu.html (and related webGIS (doi: 10.5638/thagis.21.87): http://fukkou.csis.u-tokyo.ac.jp/, for registered users). The additional explanatory variables computed in this study are available at the following link

https://drive.google.com/file/d/1hGPY1qCWD8F8N1hJV08jZTnzflkdjipI/view?usp=sharing (during the review process) and they will be stored in Mendeley Data upon acceptance.

1. Introduction

The development of damage models, enabling the estimation of the impacts generated by the occurrence of extreme natural events, plays a key role in supporting decision making for the adoption of effective and efficient disaster management strategies.

The conventional tool used in tsunami risk modelling is represented by fragility curves, which describe the probability of reaching or exceeding a certain damage state as a function of (one or more) event intensity parameters (usually inundation depth) and vulnerability characteristics of the exposed object. These functions can be derived empirically, based on post-event observations, or using synthetic expert-based approaches (Tarbotton et al., 2015; Charvet et al., 2017). The ability of a building to withstand the impact of a tsunami depends on several factors, including the structural features and shape of the building, its construction material, the foundation type, as well as the characteristics of the surrounding environment (Reese et al., 2007, 2011; Kappes et al., 2012; Suppasri et al., 2013, 2015; Charvet et al., 2014; Leelawat et al., 2014; Tarbotton et al., 2015; Dall'Osso et al., 2016; Charvet et al., 2017). Regarding the latter factors, coastal topography is expected to have a significant influence on tsunami damage, due to possible amplifying effects occurring in ria coasts, which would induce, for a same value of the inundation depth, higher damage levels than in plain areas (Suppasri et al., 2013, 2015; Leelawat et al., 2014; De Risi et al., 2017). In addition, as observed in past tsunami events, the mutual interaction between neighboring buildings can result in a "shielding effect" (i.e., robust buildings could act as a shield for other ones (Leone et al., 2011; Reese et al., 2011; Tomiczek et al., 2016; Winter et al., 2020; Moris et al., 2021)) or, conversely, in a negative source of debris, in the case of collapsed or washed away buildings (Charvet et al., 2015; Nistor et al., 2017).

In the complexity and heterogeneity of urban environments, these factors can combine in different ways, thus making the modelling of building vulnerability to tsunamis a rather arduous exercise, exacerbated by the general limited availability of empirical data, which are the basis for model development and/or validation. Therefore, complexity of involved damage phenomena and data availability are the main reasons behind the limited ability of most existing models in describing tsunami damage comprehensively, with obvious potential repercussions on the results of risk analyses.

Machine learning-based methods, although adopted in many fields of the scientific research (Gibert et al., 2018; Razavi, 2021), have still found, to date, limited application for the development of empirical tsunami damage models. Indeed, even where large empirical damage datasets are available, these have been traditionally employed in simple regression models (e.g., Koshimura et al. (2009) and Suppasri et al. (2011) for the 2004 Indian Ocean tsunami or Suppasri et al. (2013, 2015) and Leelawat et al. (2014) for the 2011 Great East Japan tsunami) that, however, as also pointed out by Charvet et al. (2015), may bring considerable uncertainty in damage estimation. While data driven approaches for flood damage modelling have gained increasing traction in the last decade (Merz et al., 2013; Wagenaar et al., 2017; Schröter et al., 2018; Amadio et al., 2019), only recently they started to receive attention for the case of tsunamis, as shown by Saengtabtim et al. (2021), who analyzed damage data for 18.000 buildings hit by the 2011 tsunami in Ishinomaki City, by considering only inundation depth and flow velocity as the main variables for developing decision tree related algorithms.

In this context, the present study aims at further exploiting data-driven approaches for developing an empirical, multi-variable tsunami damage model for civil buildings, based on the full dataset, prepared by the Japanese Ministry of Land, Infrastructure and Transportation (MLIT, 2012) after the 2011 Great East Japan tsunami, containing micro-scale information on the damage level and other explanatory parameters for about 250.000 affected buildings in the Tōhoku area (Leelawat et al., 2014; Suppasri et al., 2013, 2015). The considered machine learning techniques, characterized by the ability to capture nonlinear interactions among the descriptive features of the phenomenon, will allow to analyze the relative importance of the different parameters at stake on modelling accuracy. In particular, the main novelty of this study is to evaluate, besides traditional building vulnerability parameters, the effect of other explanatory factors often described in the literature as potentially important, but actually neglected in the development of existing damage models. To this aim, the original MLIT database will be enhanced by implementing geospatial algorithms that will enable to calculate other geometric building-related features (i.e., shape factors) or site-related parameters, as a function of building position and orientation with respect to the coastline (as minimum distance, direction of wave origin), which will, in turn, be classified according to the coastal morphology (ria or plain coast). In addition, an approach will be proposed to account for the effect of the reciprocal interaction between buildings and/or singular structures, with reference to the following aspects: (i) the potential "shielding effect" exerted between interacting buildings, which may reduce the impact of water, (ii) the presence of collapsed buildings that may themselves be a source of debris, and (iii) the presence of protective barriers, as seawalls. Criteria for the selection of potentially interacting elements will be developed and analyzed, by evaluating their influence on models' performance. Finally, this study will compare the predictive power of several models trained by considering different combinations of the input parameters, classified in terms of their complexity for data retrieval or pre-processing, thus providing useful insights to address an efficient characterization of buildings' vulnerability for tsunami damage modelling: indeed, since the characterization of the various input data may not always be possible in all contexts or cases (e.g., due to the lack or difficulties in retrieving or pre-processing specific local information), from a practical perspective, it is key to find an efficient trade-off between predictive accuracy and complexity of input data collection (i.e., "input data cost").

2. Materials and Methods

2.1. Extended MLIT dataset

The main data source for this study was the database prepared by the Japanese Ministry of Land, Infrastructure and Transportation (MLIT, 2012) after conducting fields surveys in the aftermath of the 2011 tsunami in the Tōhoku Region. The dataset consists of a polygon shapefile containing micro-scale (i.e. building scale) information on the degree of observed damage (distinguished in 7 classes: (1) no damage; (2) minor damage, (3) moderate damage, (4) major damage, (5) complete damage, (6) collapse and (7) washed away) and other explanatory variables for about 250.000 affected buildings. The original data for each building include, among the descriptive variables, information on both hazard and vulnerability characteristics (Table 1): the former is represented by the inundation depth at the building location (h) and any other concurrent hazard factors (OthHaz) (e.g., earthquake, fire, liquefaction (Hokugo et al., 2011; Yamaguchi et al., 2012; Nishino et al., 2015)), while the latter include the structural type (BS) (i.e., construction material) and number of floors (NF) of the building, its intended use (Use), and the year of construction (Year). Detailed descriptions of each damage class and explanatory features of the MLIT can be found in previously published literature (e.g., Suppasri et al., 2013, 2015; Charvet et al., 2014; Leelawat et al., 2014).

Table 1. Variables included in the extended MLIT dataset used in this study

Variable*	Description	Data source type
h	Surveyed inundation depth at building location [m]	Original MLIT dataset
FA	Footprint area of the building [m^2]	Geospatial analysis
BS	Building structure, i.e., material of construction of the building [1: reinforced concrete; 2: steel; 3: wood; 4: masonry; 5: other]	Original MLIT dataset
NF	Number of floors of the building	Original MLIT dataset
Use	Building use classes, as defined by the MLIT	Original MLIT dataset
Year	Year of construction of the building	Original MLIT dataset
OthHaz	Other possible hazards at site location [0: no; 1: yes]	Original MLIT dataset (converted to a binary variable)
DegComp	Compactness of the building shape, i.e., relative proximity of internal points, normalized to a circle [0-1]	Geospatial analysis
Distance	Minimum distance between the building centroid and the coastline [m]	Geospatial analysis
CoastType	Indicator for coast morphology [0: plain coast; 1: ria coast]	Surveyed
WDir	Probable direction of tsunami wave attack, calculated as the direction of the line of minimum distance between the building centroid and the coastline	Geospatial analysis
BOrient	Building orientation, calculated as the rotation of the reference system which minimizes the ratio between the projections of the building on a cartesian coordinate system, computed with a 5° resolution	Geospatial analysis
BSideRatio	Proxy for building shape elongation, calculated as the minimum ratio between the lengths of building sides [0 (elongated shape) - 1 (squared shape)]	Geospatial analysis
NDIArea NDIVol	Proxies for the debris impact (DI) effect calculated on a narrow-type (N) buffer created around building's centroid [Eq. (1, 2)]	Geospatial analysis
LDIArea LDIVol	Proxies for the debris impact (DI) effect calculated on a large-type (L) buffer created around building's centroid [Eq. (1, 2)]	Geospatial analysis
NShArea NShVol	Proxies for the shielding effect (Sh) calculated on a narrow-type (N) buffer created around building's centroid [Eq. (3, 4)]	Geospatial analysis
LShArea LShAvol	Proxies for the shielding effect (Sh) calculated on a large-type buffer (L) created around building's centroid [Eq. (3, 4)]	Geospatial analysis

NSW LSW	Proxies for the shielding effect provided by the presence of seawalls (SW), calculated on a narrow- (N) or large-type (L) buffer created around building's centroid [Eq. (5)]	Geospatial analysis

* The original MLIT database is in Japanese; the names of the corresponding variables reported here have been assigned for enhancing their interpretation.

The original MLIT database was then enriched by including other possible damage explanatory variables (Table 1), as described hereinafter:

- Geometric features of the buildings were determined based on straightforward geospatial operations on building polygons contained in the MLIT database. These included the attribution of the footprint area (FA) and the calculation of the following shape factors: degree of compactness (DegComp = $4\pi A/P^2$, where A and P are, respectively, the area and perimeter of the building polygon); building orientation (BOrient), defined as the rotation of the reference system which minimizes the ratio between the projections of the building on a cartesian coordinate system, computed with a 5° resolution (Figure 1); and a proxy for the shape elongation (BSideRatio = dy/dx).

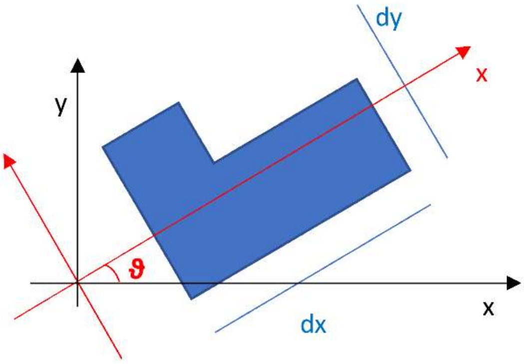

Figure 1. Identification of the parameters for calculating BOrient and BSideRatio. [*1-column fitting image*]

- Coast-related parameters were calculated by combining the MLIT dataset with the polyline of the Japanese shoreline. The simplest one involved the assessment of the minimum distance between the centroid of each building and the coastline (Distance). Then, the direction of the line of minimum distance was also considered as a proxy for the probable direction of tsunami wake attack (WDir). Furthermore, a Coast Type binary label was associated to each building based on its location and on the morphological characteristics of the coast (0 for plain and 1 for ria coast): in particular, the Tōhoku region is characterized to have a ria coast in Sanriku area in Iwate prefecture and North of Miyagi prefecture, and a plain coast in the South of Miyagi prefecture and Fukushima prefecture (Suppasri et al., 2013, 2015).
- The influence of the mutual interaction between the individual buildings and the neighboring structures in terms of "shielding effect" and debris impact was quantitatively taken into account by defining a criterium for the selection of the potentially interacting elements. Based on the location of dispersed vessels, Naito et al. (2014) assumed that debris propagates from its origin towards a direction perpendicular to the coast, with a 45° spread angle, which can also be larger in the areas close to the coast

due to drawdown. Moreover, it seems reasonable to assume that the area spanning from the considered building to the shoreline is the more interesting for both mechanisms. As a consequence, we defined and tested two different buffer geometries (denoted as "narrow" (N) and "large" (L) hereinafter) for the selection of the interacting elements, with the extreme points represented by the centroid of the examined building and the nearest point on the shoreline. Figure 2 shows an example of the shapes of the different buffer types generated for a target building, while analytical details on the construction of the two geometries is provided in the Supplementary Material.

The intersections between each building's buffer polygon (N or L) and the MLIT shapefile was then used to compute the following synthetic parameters describing the mentioned mechanisms. In particular, coherently with Charvet et al. (2015), only washed away buildings in the intersections were considered as a potential source of debris to compute the following indices:

$$LDIArea, NDIArea = \frac{\sum_{i=1}^{N_{wa}} A_i}{A_b} \tag{1}$$

$$LDIVol, NDIVol = \frac{\sum_{i=1}^{N_{wa}} A_i\, NF_i}{A_b} \tag{2}$$

Figure 2. Example of tested buffer geometries for a target building (in the blue circle). On the left, narrow (N) buffer; on the right, large (L) buffer. [*2-column fitting image*]

where N_{wa}, A_i and NF_i are, respectively, the number of washed away buildings in the buffer, the footprint area and the number of floors of the i-th building in the buffer, while A_b is the area of the buffer polygon. DIArea (calculated on L or N buffers) is then an indicator of the areal density of washed away elements in the surrounding of the building of interest, while Eq. (2) integrates this formulation by considering also the height of the buildings, by means of NFi, as a proxy for their volumetric density.

A similar approach was used also for the "shielding effect", with the only difference that, in this case, the total number of buildings in the buffer, N_{tot}, was considered:

$$LShArea, NShArea = \frac{\sum_{i=1}^{N_{tot}} A_i}{A_b} \quad (3)$$

$$LShVol, NShVol = \frac{\sum_{i=1}^{N_{tot}} A_i\, NF_i}{A_b} \quad (4)$$

When accounting for the shielding effect, a question could arise whether all buildings in the buffers should be considered or the damaged ones (of any damage level) should be neglected. From a physical perspective, it is reasonable to assume that all buildings subtract momentum to the incoming floodwater, but it is also true that collapsed ones may exert less resistance. However, since $DIArea$ is also accounted for, from a modelling point of view, there would be no difference among the two options, because the footprint area of the less damaged buildings is exactly the difference between ShArea and DIArea.

Finally, the shielding action produced by seawalls was also taken into account by calculating the following index:

$$LSW, NSW = \frac{\sum_{i=1}^{N_{sw}} L_i h_i}{A_b} \quad (5)$$

where N_{SW} is the total number of seawall elements within the buffer, while Li and hi are, respectively, the length and the heigth of the i-th element in the buffer. To this aim, Li and hi were estimated by creating a digital file with the linear extension and corresponding height of seawalls and barriers, based on information from Ranghieri and Ishiwatari (2014), integrated with interpretation of satellite images and virtual surveys by means of Google Street View.

2.2. Data handling

About 1% of the data were kept out from the training procedure in order to test the spatial transferability of the developed models. To this aim, a “Spatial Transferability Test Set” was generated by extracting the data from a few municipalities, characterized to be heterogeneously distributed across the impacted region and by having different hazard and vulnerability features: the resulting subset (for a total of 2.762 entries) included the data for the municipalities of Hashikami, Iwaizumi, Matsushima, Rifu and Shinchi.

From the remaining (237.563) entries, the 5% was holdout as a validation set for hyperparameter tuning (Section 2.3) and then, again, the remaining entries were split in training (95%) and test set (5%). In total, about 11% of the data were excluded from the training set for validation and testing purposes. All splittings were performed by stratifying the dataset according to the damage level to ensure similar distributions of the target feature.

Furthermore, as a final step, the accuracy of trained models was tested on the “Spatial Transferability Test Set” in order to obtain more insights on their generalization ability, which is recognized to be one of the main problems that empirical damage models face in their practical application (Amadio et al., 2019; Cerri et al., 2021; Lüdtke et al., 2019; Wagenaar et al., 2021).

2.3. Implemented machine learning algorithms

Four different machine learning algorithms were selected to analyze the data, with the first being a Multi-Layer Perceptron (MLP), and the others being based on Decision Trees with ensemble methods: Random Forest (RF), Extra Trees (XT, also known as Extremely Randomized Trees) and Extreme Gradient Boosting (XGBoost (XGB)).

The Multi-Layer Perceptron (Rumelhart et al., 1986), here implemented by using the Python library TensorFlow (Abadi et al., 2016), is an evolution of the classic linear perceptron, which can tackle non-linear problems by virtue of the presence of additional intermediate hidden layers.

All tree-based algorithms were instead developed through the Python libraries scikit-learn (Pedregosa et al., 2011) and XGBoost (Chen et al., 2016). Both Random Forest (Ho, 1995) and Extra Trees (Gert et al., 2006) apply a bagging algorithm as ensemble method, obtaining the classification result by taking into account the predictions of every tree, by majority vote, with slight differences in the growing tree procedure. Random forest selects the instances for each splitting using bootstrap replicas of the dataset and it defines the cut points in order to obtain the optimal split for each feature. Extra Trees, conversely, uses the original entries of the dataset for each splitting node and randomly selects the cut points, with lower computational costs under the same conditions. After selecting the split points for the whole subset of features, both algorithms select the optimal one.

XGBoost (Chen et al., 2016) applies a Gradient Boosting algorithm as ensemble method: when selecting the dataset instances for the splitting nodes for a new tree, instead of a random selection, only a subset from the residuals (i.e., the misclassified entries) of the previous tree is considered for the optimization. Also, the prediction of each tree has its own weight, and the optimal ones are determined with a gradient optimization process to minimize a loss function.

For each model, a preliminary Random Search was performed to properly tune the hyperparameters and assess the variability in the validation accuracy. For the MLP, the tuning mainly focused on the number of layers and neurons per layer, and the best results were achieved by using 3 hidden layers of 80 nodes, with *tanh* as activation function, and a last layer with 7 *softmax* nodes. The maximum number of epochs for the training process was set at 300, with an early stopping after 10 epochs without improvements. Similarly, an early stopping procedure was established to determine the number of trees for the XGBoost model, with the growth arrested after adding 10 trees without improvements, with the maximum set at 300 and the depth of each tree limited to 9.

The Random Forest was built with 150 fully grown trees, using the Shannon entropy as splitting evaluation criterium and 90% of the training entries to build the bootstrap replicas at each splitting. For the Extra Trees classifier, the number of trees was fixed at 130 and the splitting process was performed by evaluating the decrease in Gini impurity on random subsets with 60% of the training entries. The adoption of methods to contain the tree growth (e.g., maximum depth, minimum number of samples per leaf, minimum required impurity decrease for splitting) affected the validation accuracy for the last two models.

The metric chosen for the evaluation of models' performance was the relative hit rate (HR), i.e., the ratio of the number of correct predictions to the total number of predictions. Moreover, a normalized confusion matrix was computed for each model to have a more comprehensive picture of the error patterns (i.e., to understand whether some damage classes were more commonly misclassified).

2.4. Assessment of the feature importance

Different approaches are available in the literature to obtain insights on the feature importance, which consists in measuring the drop in the performance of the model when that feature is not considered in the training. The more complete approach would be to train the model with all the available N features and then train it again with all possible combinations of N-1 features. This can be time-consuming when the number of features is large, but examples exist with a limited number of input variables (Azmathullah et al., 2005; Azimi et al., 2018; Di Bacco and Scorzini, 2019).

For tree-based ensemble methods, a first simple variable importance measure is the number of times each feature is selected by the trees. A more elaborate measure is obtained by computing a weighted average improvement caused by each variable in the splitting criterion of the trees (Friedman, 2001); for classification ensembles the improvement is evaluated as Gini impurity or entropy reduction, while the variance reduction is considered for regression problems. However, besides being suitable only for tree-based ensembles, these methods tend to overestimate the importance of the features with a higher number of classes (or higher number of different values), when feature selection is based on impurity reduction (Strobl et al, 2007). In this study, we instead selected a more robust estimation of the feature importance, based on the relative mean decrease accuracy (mda), which was computed by randomly shuffling each feature in turn (Breiman, 2001) and then measuring the consequent variation in the accuracy (acc_{shuff}) with respect to the original one (acc_0):

$$mda = \frac{acc_0 - acc_{shuff}}{acc_0} \tag{7}$$

In addition to the analysis of the feature importance, it would be also interesting to investigate the sensitivity of models' accuracy to the comprehensiveness and complexity of the input data considered in model development. Indeed, in practical applications, information for the characterization of the damage explanatory variables may not be always available in any regional context, due to possible lack of appropriate data or considerable retrieval and/or pre-processing costs. To this aim, the performances for each of the four machine-learning algorithms were tested by considering, for the training process, different combinations of the input parameters with increasing degree of data type complexity. In detail, we analyzed a total of 29 possible combinations, which were classified into the following three macro-groups (Figure 3):

RUN / VAR.	A.					B.					C.																		
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
h	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
FA	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
BS		■		■	■			■						■	■	■				■	■	■	■	■	■	■	■	■	■
NF		■		■	■			■							■	■					■	■	■	■	■	■	■	■	■
Use		■		■	■			■								■						■	■	■	■	■	■	■	■
Year		■		■	■			■															■	■	■	■	■	■	■
OthHaz				■	■			■															■	■	■	■		■	■
DegComp			■			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Distance						■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
CoastType							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
WDir								■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
BOrient								■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
BSideRatio								■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
NShVol																	■	■	■	■	■	■	■		■			■	
LShVol											■	■	■	■	■	■								■		■	■		■
NShArea																									■				
LShArea																										■			
NSW																		■	■	■	■	■			■			■	
LSW												■	■	■	■	■										■	■		■
NDIVol																			■	■	■	■			■			■	
LDIVol													■	■	■	■										■	■		■
NDIArea																									■				
LDIArea																										■			

Figure 3. Feature combinations for testing the sensitivity of models' accuracy to the comprehensiveness and complexity of the input data considered in model development. [*2-column fitting image*]

- "Group A" (indicated in green in Figure 3): in these runs, only original MLIT data and simple geometrical building features were used as input for the algorithms;
- "Group B" (indicated in yellow in Figure 3): other geometrical building features and simple coast-related indicators were added to "Group B" input parameters;
- "Group C" (indicated in red in Figure 3): these combinations included also buffer-related indices (calculated on N or L geometries), accounting for shielding mechanisms and debris impact.

3. Results and discussion

3.1. Models' accuracy and feature importance

Figure 4 summarizes the achieved accuracies, in terms of relative hit rate (HR), for the four tested models trained with all the different feature combinations listed in Figure 3. Lower accuracies are obviously evident for the "Group A" (green) runs, with MLP and XGB performing slightly better (HR~0.60) than RF and XT (HR~0.51) for the simplest combination considering water depth and building footprint area as the only input parameters. For all the tested algorithms, the accuracy is observed to increase with models' complexity (i.e., with the number of considered variables), although at different rates, with XGB providing the best performances for all the "Group A" combinations. When coast-related parameters (in particular: Distance, CoastType and WDir – "Group B" runs) are included in the models, the HR increases up to 0.70 for MLP (which systematically reports the worst performances) and to 0.74 for XT, RF and XGB.

A further significant improvement (HR>0.80) is achieved when considering more complex input feature combinations and especially those including buffer related parameters accounting for the debris impact and "shielding effect", then supporting empirical observations on the importance of these mechanisms on tsunami damage (Leone et al., 2011; Reese et al., 2011; Charvet et al., 2015; Moris et al., 2021).

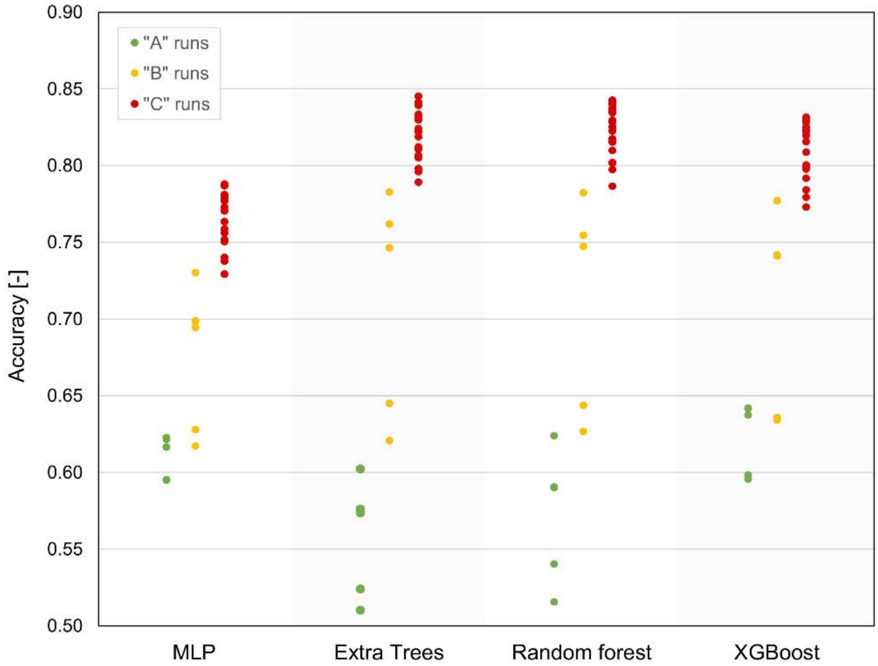

Figure 4. Models' accuracies (in terms of relative hit rate) for the four tested algorithms under the different feature combinations shown in Figure 3. [*1.5-column fitting image*]

In addition to overall results, confusion matrices, by depicting the hit and misclassification rates among the different damage classes, can provide a more detailed description of models' accuracy.
However, before discussing these results in detail, it would be useful to analyze the correlation matrix of the features considered in the extended MLIT dataset (Figure 5, in case of buffer related indicators calculated on L (left) or N (right) geometries), given that any strong correlation among the explanatory variables should be reminded when interpreting accuracy and feature importance results.

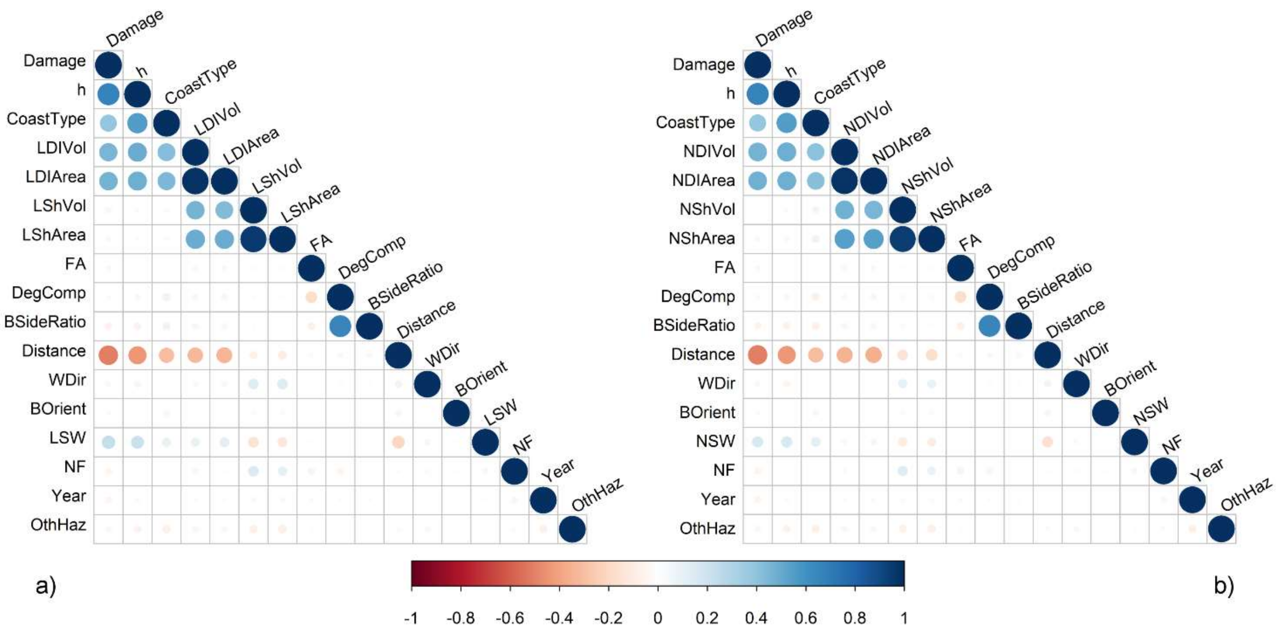

Figure 5. Correlation matrix for the variables included in the extended MLIT dataset: cases of buffer related indicators calculated on L (a) and N (b) geometries (Note: only ordinal and continuous variables are considered in these plots). [*2-column fitting image*]

As expected, the damage class has a strong positive correlation with inundation depth (0.66) and a negative correlation with the distance from the shoreline (0.51); it also shows a positive correlation (0.39) with CoastType, coherently with the discussed amplifying effects occurring in ria coasts (Suppasri et al., 2013, 2015; Leelawat et al., 2014; De Risi et al., 2017). Debris Impact parameters also appear to have a strong positive correlation (≥ 0.45) with the damage, revealing a possible relevant impact of this mechanism on damage occurrence; however, it should be reminded that these indicators were computed based on the presence of interacting washed away buildings, resulting in a stronger correlation with h and Damage, differently from "shielding effect" indicators, which instead exhibited very weak linear correlation with the two mentioned parameters. The (weak) negative correlation between OthHaz and Damage is explained by the fact that occurrences in the dataset with OthHaz=1 were mainly associated to buildings which experienced relatively shallow water depths and then minor damages.

Figure 6 reports an example of the obtained confusion matrices for feature combinations C.19 and C.16 (which are two of the most complete, calculated on L buffer geometries) with the four selected algorithms. Interestingly, Figure 6 indicates that damage classes 3, 4 and 7 are the more easily recognized, while class 5 is not only the one with the higher misclassification rates, but also the less common guess for all models, probably because it shares the lowest number of occurrences in the dataset, along with class 1, as visible in Figure 7.

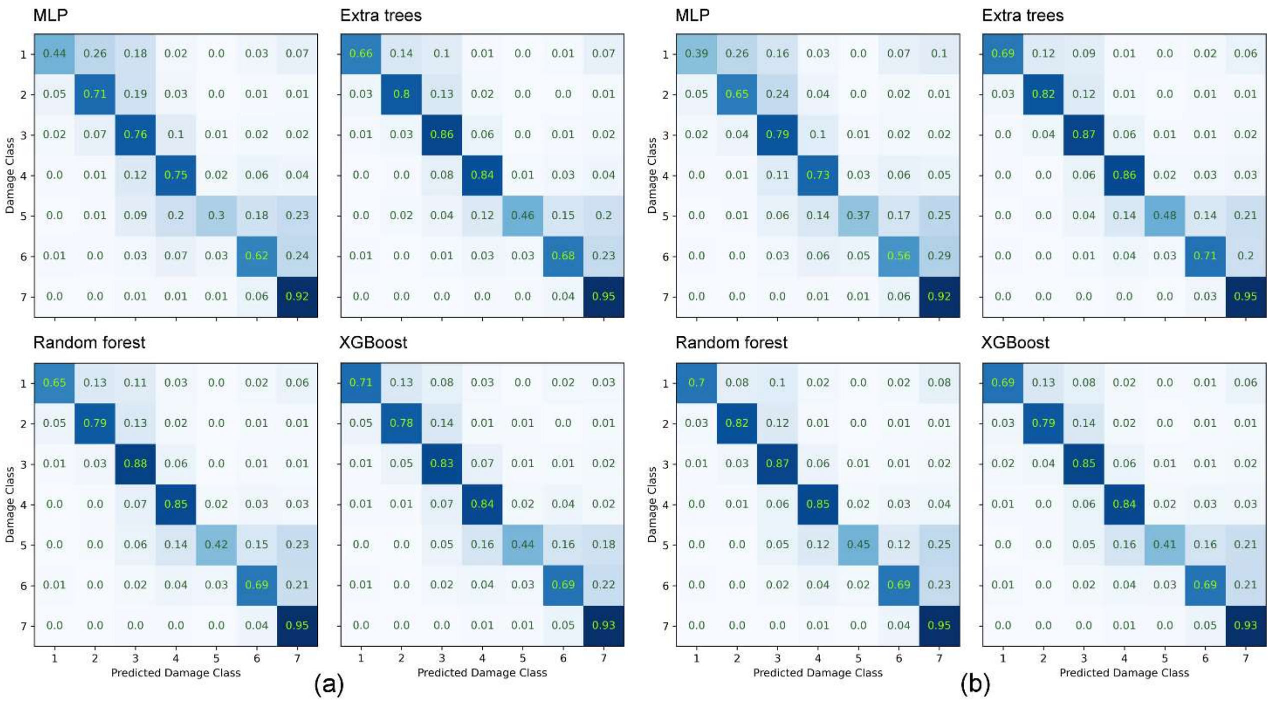

Figure 6. Normalized confusion matrices for the four tested models: a) feature combination C.19; b) feature combination C.16. The numbers in the cells indicate the relative hit rate. [*2-column fitting image*]

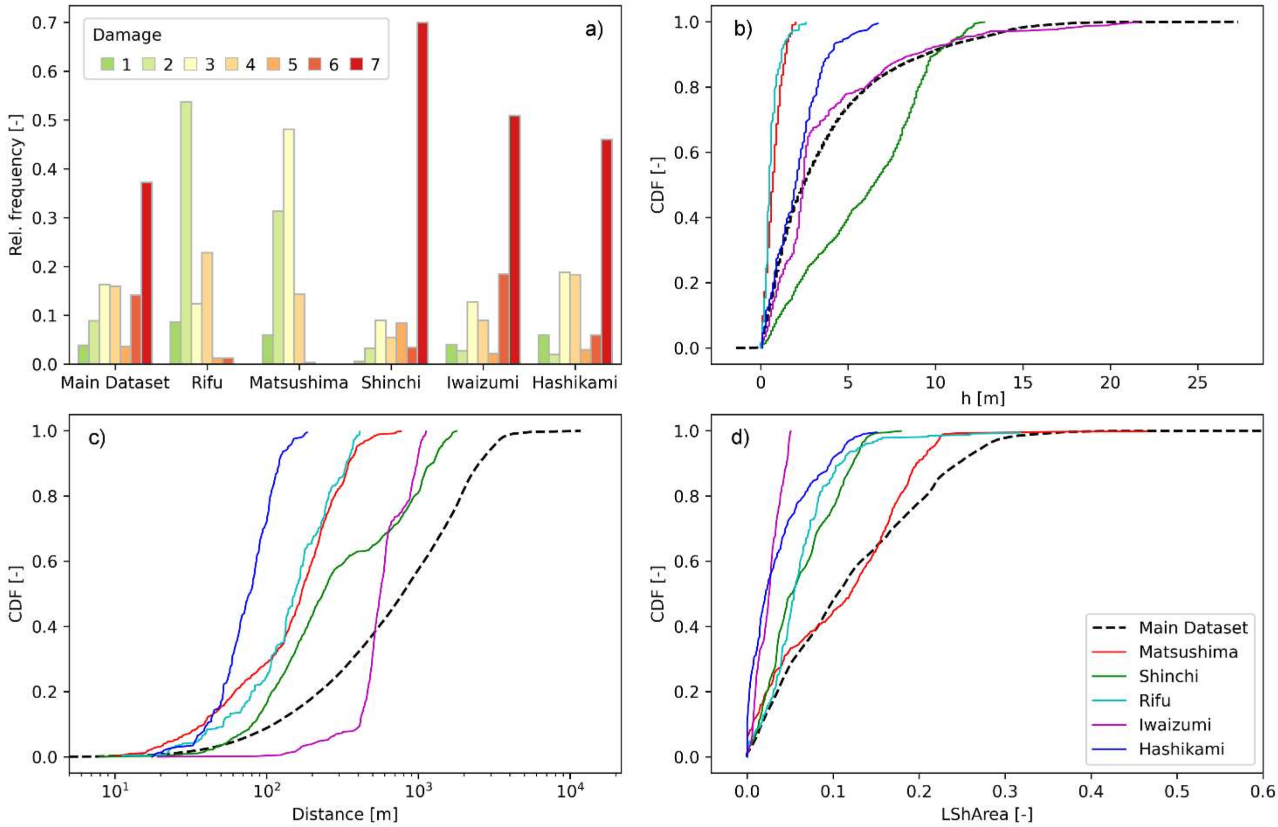

Figure 7. Empirical distributions for some of the key variables for the total extended MLIT dataset and the cities of the "Spatial Transferability test set": a) Damage; b) Inundation depth; c) Distance from the shoreline; d) Areal shielding effect indicator calculated on large buffer geometry. [*2-column fitting image*]

Figure 6 also highlights that most of the misclassified buildings are still assigned to the two nearest classes, thereby indicating reasonable accuracy even when the prediction is not exact (the only exception is represented by buildings in class 5, which are observed to be more frequently assigned to class 7 rather than 4 or 6). In general, the error distributions have been found to be model-independent, being qualitatively similar for all the considered algorithms: this may suggest that identified patterns could have be driven also by uncertainties in the attribution of individual buildings to the specific damage classes in the post-event field surveying phase, with central classes being prone to larger uncertainties (Lin et al., 2018; Lorenzo and Reuland, 2019).

The confusion matrices for feature combinations C.18 and C.15 (similar to C.19 and C.16, but with buffer related indices calculated on N geometries), are reported in the Supplement (Figure S1): the comparison of Figure 6 and Figure S1 reveals very similar performances, then indicating that, while the consideration of debris impact and shielding mechanisms are essential for model accuracy, the choice of the buffer shape for calculating the related indicators (between the two proposed in this study), has only a slight influence.

An example of the results regarding the feature importance analysis is shown in Figure 8, which describes the mean relative decrease in hit rate caused by the shuffling of each input feature for the most complete run C.16, averaged on 10 different shufflings. The corresponding plots for the feature combination C.19 (Figure

S2), together with those calculated for N geometries (runs C.18 and C.15), are instead reported in the Supplement (Figures S3-S4).

The indications provided by Figure 8 roughly confirm the overall results displayed in Figure 4. As expected, according to the relative mean decrease accuracy, inundation depth is always, by far, the most significant explanatory variable, together with other related parameters, such as the distance from the shoreline and the debris impact indicators. This highlights, one more time, the importance of relying on accurate representations of inundation scenarios when performing ex-ante or ex-post damage assessments (Scorzini et al., 2022).

Interestingly, although with almost no linear correlation with building damage or water depth (Figure 5), the proxies accounting for the "shielding effect" are identified among the key features, thus confirming the importance of considering this mechanism when modelling tsunami damage. Other critical features are the proxy variable for the wave direction (WDir), the building structure (BS) and the indicator for the presence of seawalls (LSW). CoastType results to be one of the variables with higher mean decrease accuracy for MLP and Extra Trees, but it appears to be less influent for RF and XGB. This different behavior could be ascribed to the selection of the cut points during the splitting procedure, which is completely random for XT and optimized in the other models, with no difference for binary variables: as a consequence, XT should be relatively more prone to select binary features. Among the typical building vulnerability parameters, Use and Year, although sometimes indicated in the literature as descriptors for building resistance (Suppasri et al., 2015), are here observed to be among the less important ones, together with other shape parameters of the buildings. OthHaz also appears to be ranked in the last positions, but probably due to the characteristics of the considered tsunami event, where the inundation depth was the main driver for damage occurrence, as highlighted in the correlation matrix in Figure 5.

It can be noted that the results for the most complete runs C.16 and C.19 (with the latter including only the volumetric version of the buffer related indices) are practically identical in terms of accuracy (Figure 6), given that, due to the collinearity of (LDIArea, LShArea) and (LDIVol, LShVol), the exclusion of one of them has little effect on the models' performance, because they can get the same information from the correlated feature. This is also the reason explaining why the two added areal variables are still among the most important ones for combination C.16, even though they share their importance with the volumetric versions of the indicators (Figure 8 and S2). Moreover, the comparable feature importance depicted by Figures 8 and S2 and by the corresponding Figures S3 and S4, respectively for large and narrow buffer-related indices, confirms the limited influence of buffer geometry on the overall results.

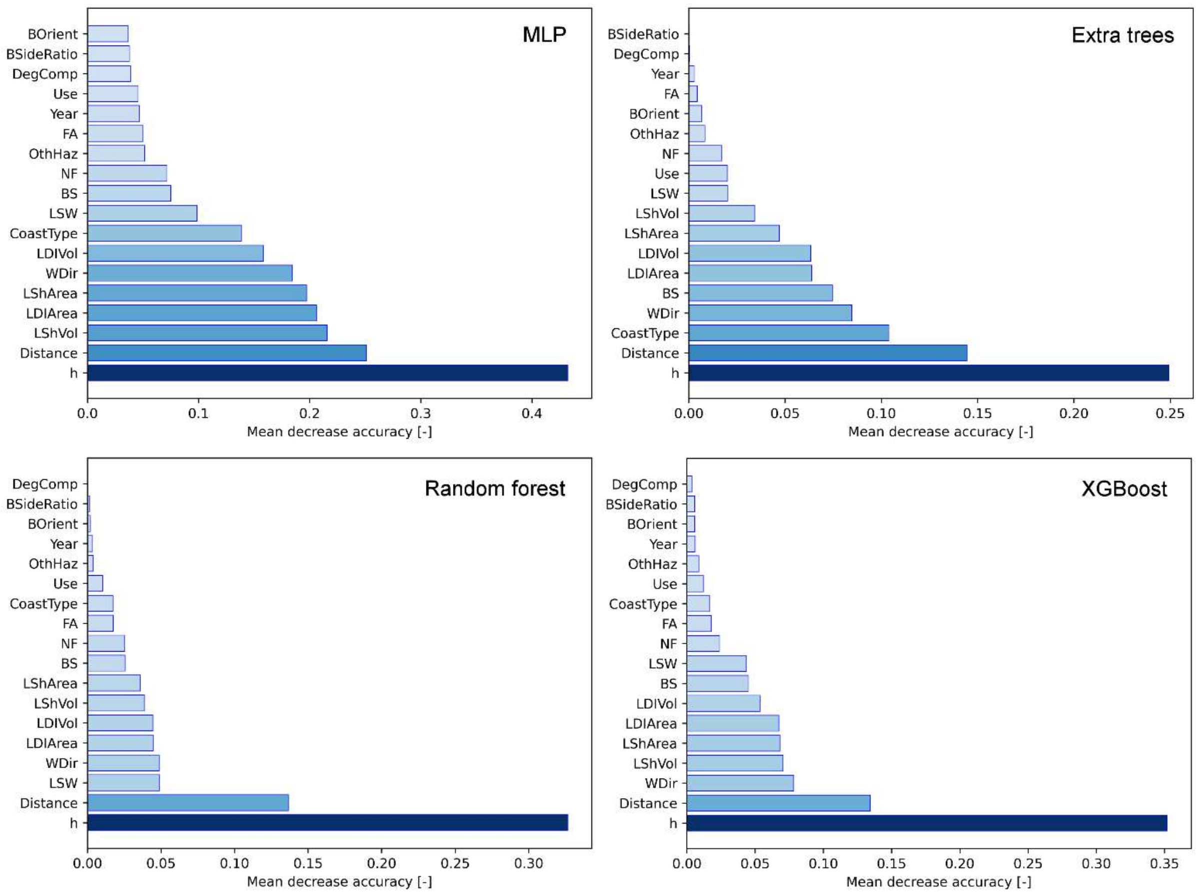

Figure 8. Results of the feature importance analysis for the four tested models (under feature combination C.16). [*2-column fitting image*]

3.2. Spatial transferability test

Table 2 summarizes the results of the spatial transferability analysis, which consisted in testing the most complete and accurate RF, XGB and XT models (i.e., under feature combination C.16) on the data from the municipalities of Hashikami, Iwaizumi, Matsushima, Rifu and Shinchi that were excluded from the training procedure.

Table 2 indicates a significant worsening of models' performances on the five cities, with the exception of Shinchi, where the hit rate was still reasonably high (0.73-0.76). The main reason for these heterogeneous results may be probably attributed to the different distributions of some key features of the datasets, as shown in Figure 7 for Damage, h and Distance. In particular, the very low accuracies found for Rifu and Matsushima (~0.3-0.5) can be clearly explained by the reported larger shares of low damages and practically no washed away buildings resulting from shallower inundation depths registered at these locations, compared to those depicted in the whole dataset. Low accuracies are also evident for Iwaizumi, although being characterized by empirical cumulative distribution functions for h and Distance similar to those of the main dataset; however, in this case, the reason is to be found in the less dense urban fabric of the city, as highlighted in Figure 7d for LShArea (and, similarly, in Figure S5 for NShArea), which is obviously a function of building density. Therefore, this analysis confirms that consistency in input data is a critical requirement for the spatial transfer of damage models among regions (Amadio et al., 2019; Lüdtke et al., 2019; Cerri et al., 2021; Wagenaar et al., 2021), which, otherwise, could negatively affect the predictive performance.

Table 2. Hit rates of the models (under feature combination C.16) on the spatial transferability test.

	Accuracy (relative hit rate) [-]								
Municipality	*Extra Trees*			*Random forest*			*XGBoost*		
	Training	Valid.	Test	Training	Valid.	Test	Training	Valid.	Test
Hashikami	1.00	0.84	0.52	1.00	0.84	0.55	0.95	0.84	0.54
Iwaizumi	1.00	0.84	0.59	1.00	0.84	0.60	0.95	0.84	0.60
Matsushima	1.00	0.84	0.53	1.00	0.84	0.56	0.95	0.84	0.49
Rifu	1.00	0.84	0.33	1.00	0.84	0.33	0.95	0.84	0.36
Shinchi	1.00	0.84	0.75	1.00	0.84	0.74	0.95	0.84	0.76

Such results then suggest that data-driven models, to be applied in a more effective way, should leverage on (very) numerous ex-post building data, capable of resembling the individual local hazard and vulnerability features of the different sites of model implementation; otherwise, they should be considered more suitable as descriptive models rather than for the predictive purposes.

However, for homogeneous areas characterized by the availability of a sufficient number of empirical event data (to avoid overfitting), a feasible approach could be to train the models on subsets of these local data, in order to have a faster calibration and higher similarity between the feature distributions of the training and test sets. The suitability of this operation is clearly confirmed by the results reported in Table 3, which shows, as an example, the performances of RF, XGB and XT models (under feature combination C.16) trained and tested (using the same splitting procedure and tuning as performed with the whole dataset) on the subsets of the extended MLIT dataset based only on the city of Ishinomaki (for a total of 68.596 data): the values of the global hit rate (~0.89), as well as of its distribution among the damage classes (confusion matrices in Figure S6) are indeed found to be comparable or even higher than the ones of the main models, but with lower computational time and no significant difference between validation and test accuracy.

Table 3. Hit rates of the local models for Ishinomaki (under feature combination C.16).

	Accuracy (relative hit rate [-])		
Model	Training	Valid.	Test
Extra Trees	1.00	0.89	0.89
Random forest	1.00	0.89	0.89
XGBoost	0.99	0.89	0.89

4. Conclusions

The present study focused on the implementation of machine-learning approaches for tsunami damage modelling based on an enhanced version of building damage data from the 2011 Great East Japan event, which involved the assessment and integration of usually neglected predictors accounting for building- and site-related features, as well as mechanisms of building interaction, like the "shielding effect" or debris impact from collapsed structures, here modelled by proposing a new buffer-based method.

Compared to simpler and commonly used regression methods, machine-learning approaches can capture complex input-output relationships without being constrained by a functional shape or by any a priori assumption on the statistical distribution of the data. These features make them suitable tools for improving knowledge on physical mechanisms involved in tsunami damage and then for shedding light on the importance of the different variables at stake on the modelled phenomenon. Four algorithms have been tested and compared, starting from simple multi-layer perceptron neural networks (MLP) to decision tree-based algorithms (Random Forest (RF) and Extra Trees (XT)) and boosting methods (XGBoost (XGB)). In particular, models' accuracy in damage classification has been analyzed for different combinations of the explanatory variables, which were distinguished based on their complexity for retrieval and/or computation ("input data cost"), thus providing useful insights for the identification of key and ancillary variables to be used in tsunami damage modelling. The results revealed that, independently of the models and/or input feature combinations (especially for RF, XT and XGB), a critical role, even more than typical vulnerability parameters included in standard fragility functions, is played by the proposed indicators describing the mutual interaction between the inundated buildings, which, then, should not be neglected when modelling tsunami damage. Furthermore, while providing such new insights, the developed models can be also considered as predictive tools to be used for damage assessment under potential simulated tsunami scenarios. Finally, in order to assess the potential spatial transferability (i.e., generalizability) of the developed models, their predictive ability on data kept out from the training dataset has been analyzed ("Spatial transferability test set"). The results, in line with similar studies carried out on flood damage modelling (Amadio et al., 2019; Lüdtke et al., 2019; Cerri et al., 2021; Wagenaar et al., 2021), confirmed that a model transfer from one region to another often comes with a reduced predictive performance, thus suggesting the need to rely on more specific local models (as presented for the Ishinomaki case), able to give an adequate representation of the typical characteristics and damage processes occurring in the different areas.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., [...] Zheng, X. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467. Software available from tensorflow.org.

Amadio, M., Scorzini, A.R., Carisi, F., Essenfelder, A.H., Domeneghetti, A., Mysiak, J., Castellarin, A. (2019). Testing empirical and synthetic flood damage models: the case of Italy. Nat. Hazards Earth Syst. Sci. 19(3), 661-678. doi: 10.5194/nhess-19-661-2019.

Azimi, H., Bonakdari, H., Ebtehaj, I., Gharabaghi, B., Khoshbin, F. (2018). Evolutionary design of generalized group method of data handling-type neural network for estimating the hydraulic jump roller length. Acta Mech. 229(3), 1197-1214. doi: 10.1007/s00707-017-2043-9.

Azmathullah, H.M., Deo, M.C., Deolalikar, P.B. (2005). Neural networks for estimation of scour downstream of a ski-jump bucket. J. Hydraul. Eng. 131(10), 898-908. doi: 10.1061/(ASCE)0733-9429(2005)131:10(898).

Breiman, L. (2001). Random forests. Mach. Learn. 45(1), 5-32. doi: 10.1023/A:1010933404324.

Charvet, I., Suppasri, A., Imamura, F. (2014). Empirical fragility analysis of building damage caused by the 2011 Great East Japan tsunami in Ishinomaki city using ordinal regression, and influence of key geographical features. Stochastic Environ. Res. Risk Assess. 28(7), 1853-1867. doi: 10.1007/s00477-014-0850-2.

Charvet, I., Suppasri, A., Kimura, H., Sugawara, D., Imamura, F. (2015). A multivariate generalized linear tsunami fragility model for Kesennuma City based on maximum flow depths, velocities and debris impact, with evaluation of predictive accuracy. Nat. Hazard. 79(3), 2073-2099. doi: 10.1007/s11069-015-1947-8.

Cerri, M., Steinhausen, M., Kreibich, H., Schröter, K. (2021). Are OpenStreetMap building data useful for flood vulnerability modelling? Nat. Hazards Earth Syst. Sci. 21(2), 643-662. doi: 10.5194/nhess-21-643-2021.

Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD'16. New York, NY, USA, 785–794. doi: 10.1145/2939672.2939785.

Dall'Osso, F., Dominey-Howes, D., Tarbotton, C., Summerhayes, S., Withycombe, G. (2016). Revision and improvement of the PTVA-3 model for assessing tsunami building vulnerability using "international expert judgment": introducing the PTVA-4 model. Nat. Hazard. 83(2), 1229-1256. doi: 10.1007/s11069-016-2387-9.

De Risi, R., Goda, K., Yasuda, T., Mori, N. (2017). Is flow velocity important in tsunami empirical fragility modeling? Earth Sci. Rev. 166, 64-82. doi: 10.1016/j.earscirev.2016.12.015.

Di Bacco, M., Scorzini, A.R. (2019). Are We Correctly Using Discharge Coefficients for Side Weirs? Insights from a Numerical Investigation. Water. 11(12), 2585. doi: 10.3390/w11122585.

Friedman, J. (2001) Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 29(5), 1189-1232.

Geurts, P., Ernst, D., Wehenkel, L. (2006). Extremely randomized trees. Mach. Learn. 63, 3-42. doi: 10.1007/s10994-006-6226-1.

Gibert, K., Izquierdo, J., Sànchez-Marrè, M., Hamilton, S. H., Rodríguez-Roda, I., & Holmes, G. (2018). Which method to use? An assessment of data mining methods in Environmental Data Science. Environ. Modell. Software. 110, 3-27. doi: 10.1016/j.envsoft.2018.09.021.

Harris, C.R., Millman, K.J., van der Walt, S.J. et al. (2020) Array programming with NumPy. Nature. 585, 357-362. doi: 10.1038/s41586-020-2649-2.

Ho, T.K. (1995). Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition, 1, 278-282. doi: 10.1109/ICDAR.1995.598994.

Hokugo, A., Nishino, T., Inada, T. (2011). Damage and effects caused by tsunami fires: fire spread, fire fighting and evacuation. Fire Science and Technology. 30(4), 117-137.

Kappes, M.S., Papathoma-Koehle, M., Keiler, M. (2012). Assessing physical vulnerability for multi-hazards using an indicator-based methodology. Appl. Geogr. 32(2), 577-590. doi: 10.1016/j.apgeog.2011.07.002.

Koshimura, S., Oie, T., Yanagisawa, H., Imamura, F. (2009). Developing fragility functions for tsunami damage estimation using numerical model and post-tsunami data from Banda Aceh, Indonesia. Coastal Eng. J. 51(3), 243-273.

Leelawat, N., Suppasri, A., Charvet, I., Imamura, F. (2014). Building damage from the 2011 Great East Japan tsunami: quantitative assessment of influential factors. Nat. Hazard. 73(2), 449-471. doi: 10.1007/s11069-014-1081-z.

Leone, F., Lavigne, F., Paris, R., Denain, J.C., Vinet, F. (2011). A spatial analysis of the December 26th, 2004 tsunami-induced damages: Lessons learned for a better risk assessment integrating buildings vulnerability. Appl. Geogr. 31(1), 363-375. doi: 10.1016/j.apgeog.2010.07.009.

Lin, S.L., Uma, S.R., King, A. (2018). Empirical fragility curves for non-residential buildings from the 2010–2011 Canterbury earthquake sequence. J. Earthquake Eng. 22(5), 749-777. doi: 10.1080/13632469.2016.1264322.

Lorenzo, D., Reuland, Y. (2019). Assessing the residual capacity of buildings for post-earthquake asset management at urban scale. Valori e Valutazioni. 22, 85-107.

Lüdtke, S., Schröter, K., Steinhausen, M., Weise, L., Figueiredo, R., Kreibich, H. (2019). A consistent approach for probabilistic residential flood loss modeling in Europe. Water Resour. Res. 55(12), 10616-10635. doi: 10.1029/2019WR026213.

McKinney, W. (2010). Data structures for statistical computing in Python. Proceedings of the 9th Python in Science Conference. 445 (1), 51-56. doi: 10.25080/Majora-92bf1922-00a.

Ministry of Land, Infrastructure, and Transportation of Japan (MLIT). (2012). Survey of Tsunami Damage condition, http://www.mlit.go.jp/toshi/toshi-hukkou-arkaibu.html (last accessed 20 June 2022).

Merz, B., Kreibich, H., Lall, U. (2013). Multi-variate flood damage assessment: a tree-based data-mining approach. Nat. Hazards Earth Syst. Sci. 13(1), 53-64. doi: 10.5194/nhess-13-53-2013.

Moris, J.P., Kennedy, A.B., Westerink, J.J. (2021). Tsunami wave run-up load reduction inside a building array. Coastal Eng. 169, 103910. doi: 10.1016/j.coastaleng.2021.103910.

Naito, C., Cercone, C., Riggs, H.R., Cox, D. (2014). Procedure for site assessment of the potential for tsunami debris impact. J. Waterway, Port, Coastal, Ocean Eng. 140(2), 223-232. doi: 10.1061/(ASCE)WW.1943-5460.0000222.

Nishino, T., Imazu, Y., Hokugo, A., Notake, H. (2015). Spatio-temporal distribution modeling of tsunami-driven combustible objects released from destroyed buildings towards tsunami fire simulation. J. Environ. Eng. 80 (716), 857-866. doi: 10.3130/aije.80.857. (in japanese)

Nistor, I., Goseberg, N., Stolle, J. (2017). Tsunami-driven debris motion and loads: A critical review. Front. Built Environ. 3, 2, 1-11. doi: 10.3389/fbuil.2017.00002.

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. (2011) Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

Ranghieri, F., Ishiwatari, M. (Eds.) (2014). Learning from megadisasters: lessons from the Great East Japan Earthquake. World Bank Publications, Washington, DC, USA. doi:10.1596/978-1-4648-0153-2.

Razavi, S. (2021). Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling. Environ. Modell. Software. 144, 105159. doi: 10.1016/j.envsoft.2021.105159.

Reese, S., Cousins, W.J., Power, W.L., Palmer, N.G., Tejakusuma, I.G., Nugrahadi, S. (2007). Tsunami vulnerability of buildings and people in South Java-field observations after the July 2006 Java tsunami. Nat. Hazards Earth Syst. Sci. 7(5), 573-589. doi: 10.5194/nhess-7-573-2007.

Reese, S., Bradley, B.A., Bind, J., Smart, G., Power, W., Sturman, J. (2011). Empirical building fragilities from observed damage in the 2009 South Pacific tsunami. Earth Sci. Rev. 107(1-2), 156-173. doi: 10.1016/j.earscirev.2011.01.009.

Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986). Learning representations by back-propagating errors. Nature. 323(6088), 533-536. doi: 10.1038/323533a0.

Saengtabtim, K., Leelawat, N., Tang, J., Treeranurat, W., Wisittiwong, N., Suppasri, A., Pakoksung, K., Imamura, F., Takahashi, N., Charvet, I. (2021). Predictive Analysis of the Building Damage From the 2011 Great East Japan Tsunami Using Decision Tree Classification Related Algorithms. IEEE Access, 9, 31065-31077. doi: 10.1109/ACCESS.2021.3060114.

Schröter, K., Lüdtke, S., Redweik, R., Meier, J., Bochow, M., Ross, L., Kreibich, H. (2018). Flood loss estimation using 3D city models and remote sensing data. Environ. Modell. Software. 105, 118-131. doi: 10.1016/j.envsoft.2018.03.032.

Scorzini, A. R., Dewals, B., Rodriguez Castro, D., Archambeau, P., Molinari, D. (2022). INSYDE-BE: adaptation of the INSYDE model to the Walloon region (Belgium). Nat. Hazards Earth Syst. Sci. 22(5), 1743-1761. doi: 10.5194/nhess-22-1743-2022.

Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinf. 8(1), 1-21. doi:10.1186/1471-2105-8-25.

Suppasri, A., Koshimura, S., Imamura, F. (2011). Developing tsunami fragility curves based on the satellite remote sensing and the numerical modeling of the 2004 Indian Ocean tsunami in Thailand. Nat. Hazards Earth Syst. Sci. 11(1), 173-189. doi:10.5194/nhess-11-173-2011.

Suppasri, A., Mas, E., Charvet, I., Gunasekera, R., Imai, K., Fukutani, Y., Abe, Y., Imamura, F. (2013). Building damage characteristics based on surveyed data and fragility curves of the 2011 Great East Japan tsunami. Nat. Hazard. 66(2), 319-341. doi: 10.1007/s11069-012-0487-8.

Suppasri, A., Charvet, I., Imai, K., Imamura, F. (2015). Fragility curves based on data from the 2011 Tohoku-Oki tsunami in Ishinomaki city, with discussion of parameters influencing building damage. Earthquake Spectra. 31(2), 841-868. doi: 10.1193/053013EQS138M.

Tarbotton, C., Dall'Osso, F., Dominey-Howes, D., Goff, J. (2015). The use of empirical vulnerability functions to assess the response of buildings to tsunami impact: comparative review and summary of best practice. Earth Sci. Rev. 142, 120-134. doi: 10.1016/j.earscirev.2015.01.002.

Tomiczek, T., Prasetyo, A., Mori, N., Yasuda, T., Kennedy, A. (2016). Physical modelling of tsunami onshore propagation, peak pressures, and shielding effects in an urban building array. Coastal Eng. 117, 97-112. doi: 10.1016/j.coastaleng.2016.07.003.

Wagenaar, D., de Jong, J., Bouwer, L. M. (2017). Multi-variable flood damage modelling with limited data using supervised learning approaches. Nat. Hazards Earth Syst. Sci. 17(9), 1683-1696. doi: 10.5194/nhess-17-1683-2017.

Wagenaar, D., Hermawan, T., van den Homberg, M. J., Aerts, J. C., Kreibich, H., de Moel, H., Bouwer, L.M. (2021). Improved Transferability of Data-Driven Damage Models Through Sample Selection Bias Correction. Risk Anal. 41(1), 37-55. doi: 10.1111/risa.13575.

Winter, A.O., Alam, M.S., Shekhar, K., Motley, M.R., Eberhard, M.O., Barbosa, A. R., Lomonaco, P., Arduino, P., Cox, D.T. (2020). Tsunami-like wave forces on an elevated coastal structure: effects of flow shielding and channeling. J. Waterw. Port, Coast. Ocean Eng. 146(4), 04020021. doi: 10.1061/(ASCE)WW.1943-5460.0000581.

Yamaguchi, A., Mori, T., Kazama, M., Yoshida, N. (2012). Liquefaction in Tohoku district during the 2011 off the Pacific Coast of Tohoku Earthquake. Soils Found. 52(5), 811-829. doi: 10.1016/j.sandf.2012.11.005.