Sustainable Cities and Society 88 (2023) 104251

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

# Simplified automatic prediction of the level of damage to similar buildings affected by river flood in a specific area

David Marín-García [a,*], Juan Rubio-Gómez-Torga [a], Manuel Duarte-Pinheiro [b], Juan Moyano [a]

[a] *Department of Graphical Expression and Building Engineering, Higher Technical School of Building Engineering, University de Seville, 4A Reina Mercedes Avenue, Seville 41012, Spain*

[b] *Department of Civil Engineering, Architecture and Georessources, Technical Institute (IST), Lisbon University, CERIS– Civil Engineering Research and Innovation for Sustainability, Universidade de Lisboa, Avenida Rovisco Pais 1, 1049-001, Lisboa, Portugal*

ARTICLE INFO

*keywords:*
River floods
Buildings
Vulnerability
Damage
Machine learning

ABSTRACT

Flooding due to overflowing rivers affects the construction elements of many buildings. Although significant progress has been made in predicting this damage, there is still a need to continue studying this issue. For this reason, the main **goal** of this research focuses on finding out if, based on a small dataset of cases of a given area, it is possible to predict at least three degrees of affectation in buildings, considering only three environmental factors (minimum distance from the river, unevenness and possible water communication). To meet this goal, the **methodological** approach followed considers scientific literature review and collection and analysis of a small dataset from 101 buildings that have been affected by floods in the Guadalquivir River basin (Andalusia. Spain). After analyzing this data, algorithms based on machine learning (ML) are applied to predict the degree of affection. The **results, analysis and conclusions** indicate that, if the study focuses on a specific area and similar buildings, using a correlation matrix and ML algorithms such as the "Decision Tree" with cross-validation, around 90% can be achieved in the "Recall" and "Precision" of "High-Level-Affection" class, and an "Accuracy" around 80% in general.

## 1. Introduction

Floods in general (Adhikari et al., 2010) and in particular those caused by overflowing rivers are often responsible for varying damage to buildings. Therefore, it is very important to be able to foresee these floods and the damage they cause, thus being able to quantify the consequences of these events and develop regulations to avoid or mitigate them (Piyumi et al., 2021), especially in households with low-income, since these are usually the most vulnerable (Chen et al., 2021; Deria et al., 2020).

On the other hand, most of the studies detected on this issue focus on the calculation of the risk of flooding, being less frequent studies on damage prediction, especially when it comes to buildings with similar construction characteristics, which are located in an area or specific region, with similar weather and a proven history of flooding. This study focuses precisely on the latter issue. However, a brief introduction to the state of the art regarding the calculation of the risk of flooding and damage is previously exposed.

### 1.1. State of the art in terms of calculating the risk of flooding

At the present, to calculate the risk of flooding (Merz et al., 2010; Osés-Eraso & Foudi, 2020) that can affect a building located in certain geographical coordinates, it is usual to consult the historical and cartographic data (Wan Mohtar et al., 2020) and apply various analysis and prediction techniques and technologies in the form of models mathematicians (Dutta et al., 2003). The results can later be captured graphically using technologies that have to do with GIS geographic information systems (Deckers et al., 2009; Meyer et al., 2009), in which flood risk maps are reflected that, as indicated, are based on environmental data, statistical and mathematical studies, machine learning (Armenakis et al., 2017; Li et al., 2019), etc., and that they can be supported, among others, by LiDAR and UAV point clouds for the creation of digital elevation model (DEM) (Jakovljevic et al., 2019).

In this context, as indicated, geospatial analysis is usually based on statistical methods and historical data processed in a GIS platform, but also on hydrological modeling based on equations and multi-criteria decision analysis (MCDA) (Arabameri et al., 2019; Doorga et al., 2021;

* Corresponding author.
*E-mail address:* damar@us.es (D. Marín-García).

https://doi.org/10.1016/j.scs.2022.104251
Received 7 August 2022; Received in revised form 20 September 2022; Accepted 10 October 2022
Available online 15 October 2022

Rahman et al., 2019). Thus, the so-called analytical hierarchy process (AHP) is usually used to determine the weights of influential factors. For the combination of weighted factors, the so-called weighted linear combination (WLC) is applied, thus being able to generate a risk map of floods. In this sense, unlike approaches based on determining the economic and physical risks (Merz et al., 2010), the MCDA may also consider, among other aspects, the environmental, social and cultural aspects.

### 1.2. State of the art in terms of damage to buildings and goal of this research

As has been shown, there is currently abundant information available on the calculation of flood risk, especially in developed regions, as is the case in European countries (De Moel, Van Alphen and Aerts, 2009; European Commission, 2007; Feyen et al., 2012; Vanneuville, Wolters, Scholz, & Uhel, 2016), such as Spain (Perles Roselló, Olcina, & Mérida Rodríguez, 2018; CNIH 2014; Junta de Andalucía REDIAM, 2022; Government of Spain. Sistema Nacional de Cartografía de Zonas Inundables/National Flood Zone, 2010 ),and there continues to be concern about the damage that these floods can cause, especially as to whether these the damages are foreseeable, and what factors influence their greater or lesser magnitude.

In this sense, in addition to studying damage to people, infrastructure, farms and industry, among others, various studies have been carried out on the evaluation of vulnerability to buildings (Galasso et al., 2021), recommendations for the construction and rehabilitation of buildings in flood zones (Núñez Collado et al., 2019), technical assessment of the deterioration (Domínguez Gutiérrez & González Pajaro, 2015), damage to buildings (Galasso et al., 2021;García-Prieto Rui et al., 2009), and the determination of Process Analysis Hierarchy (PAH) (Dall'Osso et al., 2009).

However, sometimes the type and amount of data available are limited, which initially makes it difficult to predict what level of damage buildings in a specific area will have when they are affected by a flood due to an overflow of the river. For this reason, the main goal of this research focuses on finding out if, based on a small set of case data from a certain area, it is possible to predict at least three degrees of damage to buildings, considering only three environmental factors (distance to the river, unevenness and possible communication of the water). It must be taken into account that the evaluation of the vulnerability of the building, that is, the estimation of the probable maximum losses (PML), is an essential issue to determine the strategies and policies of preparation, response mitigation and implementation of regulations and appropriate technical standards.

For this reason, it is necessary to look for models that obtain the relationship between the flood and its effect on the buildings it reaches and help carry out evaluations that allow the establishment of adequate strategies, narrowing the knowledge gap on this matter.

Thus, machine learning (ML) algorithms based on supervised learning offer a very interesting option to achieve the objective sought to estimate the damage that a building could suffer previously.

In this sense, during the development of this work the question of why use ML and not statistical methods or traditional programming was raised. Although it is possible to carry out the task by said methods, ML has important advantages, among which two stand out: i) Traditional programming requires formulating or coding rules (logic) manually, however in ML, the algorithm automatically formulates the rules (logic) of the data; ii) traditional statistical methods make a priori assumptions, while ML can constantly learn from the data and thus build and refine the prediction model, which can be trained over and over again to further improve the results or for other cases.

On the other hand, until a few years ago, ML algorithms were usually applied to predict the risk of flooding (Wang et al., 2015), susceptibility mapping and assessment (Ahmadlou et al., 2021) or buildings' flood exposure (Pham et al., 2022), but not so much to predict the damage that buildings could suffer. However, there are some works (Amadio et al., 2019) in which they have carried out very interesting checks as well as some empirical and synthetic models to predict and evaluate flood damage. Despite this, these jobs are often based on data that is sometimes difficult to obtain. That is why this work will focus on the premise that the values to be introduced in the algorithms, for their training and subsequent use, are few and in principle easy to obtain directly at any time, in such a way that the interested party, through a simple and quick operation of collecting this data can know how a hypothetical flood could affect a given building.

It is true that data of the selected factors are easier to obtain in developed regions, while for underdeveloped and developing regions they may not be available. However, it is understood that in any case the factors used in this study, can be collected both by sophisticated means (satellite data, digital applications, etc.) and by simple observation and direct measurement, simply resulting in the latter case a slightly more laborious operation.

## 2. Methods and materials

The methodology followed is based in the first place on the detection of studies carried out to date, related to the factors that influence in one way or another the degree of affectation of floods in buildings, focusing attention on the factors that have the most interesting results (Merz et al., 2004) and the greater or lesser difficulty involved in obtaining data from them. Thus, the easiest data to obtain is selected. The data collection areas (areas that have suffered flooding) and the corresponding buildings are selected. The data of these buildings and the selected factors are collected. The appropriate ML algorithms are selected for the class of data and type of classification-prediction that is intended and the data is processed to apply them to said ML algorithms. An analysis of a data set of 101 buildings that have been affected by floods in the Guadalquivir river basin (Andalusia, Spain) is carried out. Once the data is collected, the feasibility of establishing patterns is analyzed through the use of different machine learning algorithms, which allow us to reliably predict the level of severity of the damage.

In addition to these steps, it should be noted that studies carried out in the area are analyzed (Bohorquez & del Moral-Erencia, 2017; Vallejo Villalta, 2000).

To make this methodology as clear as possible, Fig. 1 shows the steps followed.

### 2.1. Factors

One of the first steps of the work focuses on identifying the factors that influence the degree of affectation of the buildings, to later classify these factors according to whether or not they can be easily obtained in most cases that may arise. The list of factors is obtained based on the existing literature (Dall'Osso et al., 2009; Galasso et al., 2021; Núñez Collado et al., 2019). The degree of difficulty in obtaining data has been divided into two levels: i) Difficult (it is not always possible to obtain the data); ii) Easy (virtually always are possible to achieve easily). In addition, an intermediate level (Easy-Difficult) could occur depending on the circumstances. For the assignment of these levels to the different factors, the experience in this type of work of one of the authors in the field of data review for insurance companies. However, in order for the classification to be as scientific as possible, it has been taken into account whether it is possible (easy) to collect the data: (1) at any time, regardless of whether the flood has occurred or it has not occurred yet; (2) with few means (observation of the environment and basic measurement); (3) without the need for knowledge or complex calculations or to collect other data not always available. As can be seen, these criteria buffer the reality that data that is readily available in developed regions may not be available in underdeveloped and developing regions. Thus, the "difficult" and "easy" standards have been classified in turn into those that meet the three conditions, those that meet two and those that

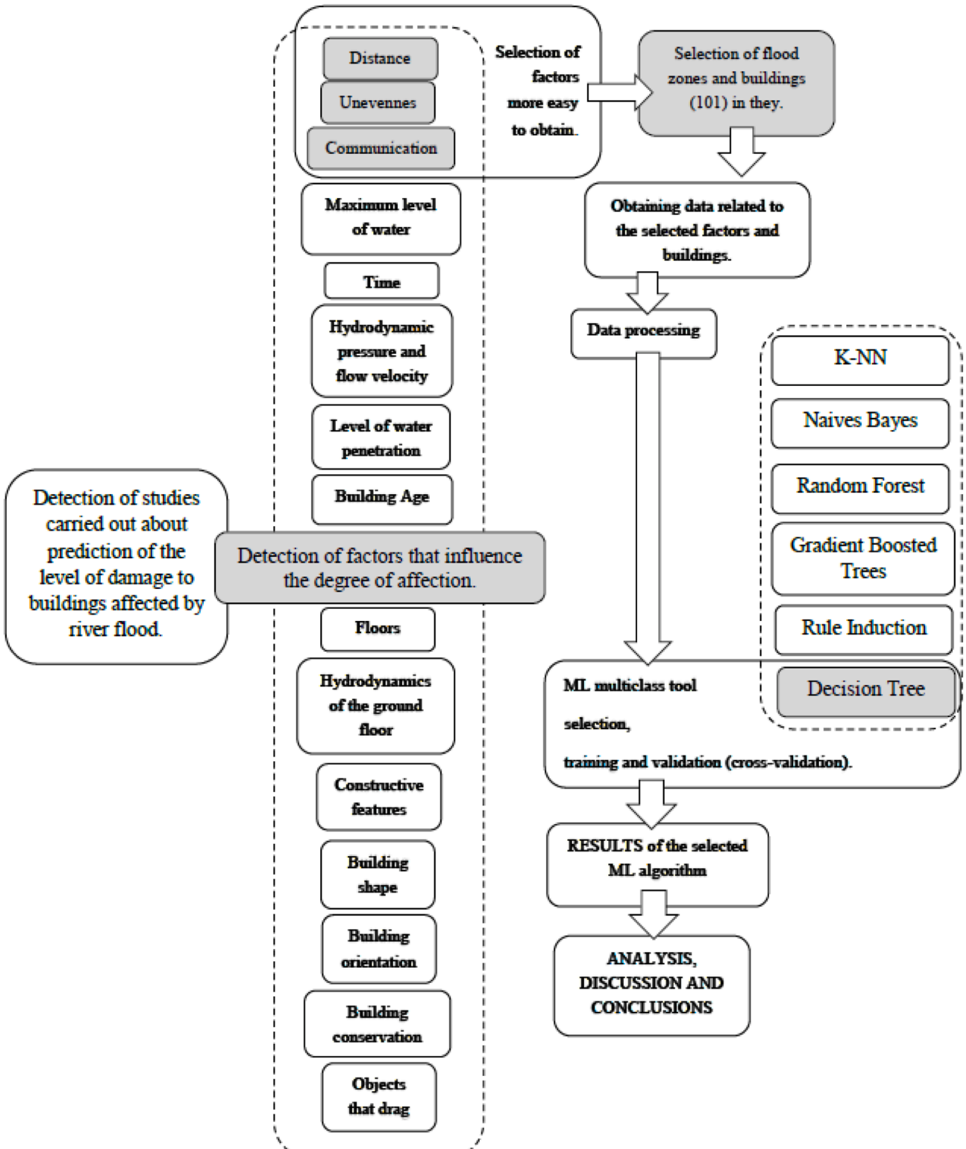



**Fig. 1.** Conceptual flowchart of the methodology.

meet only one of them. The initial relation between the influencing factors and said degree of difficulty is reflected in Table 1.

As can be seen, the value of certain factors can be obtained more easily, as is the case of the minimum distance from the building to the river, the difference between the level of the ground floor of the building and the level of maximum ordinary of the river and the existence or not of possible communication between the building and the river in case of flooding. It is also easy to obtain the shape of the building, but since these are cases of buildings, all of which have rectangular floor plans, this factor is not the object of study.

On the other hand, it is common to be able to obtain this data online, consulting freely accessible sources. Even in some countries, such as Spain, other factors can be consulted, such as the age, the heights, and the location of the building, among others, for example, through the electronic headquarters of the cadastre (Government of Spain, 2022), without the need carry out visits to the place and consultations to owner or users, as well as to the plans of the architectural project.

A recurring factor in the existing research already mentioned is the maximum level of the water surface above ground (ml) reached in contact with the building during the flood (Manrique Ruíz et al., 2017; Sánchez et al., n.d.). However, although there are studies that propose certain methodologies for its determination, as is the case of the statistical method (Galasso & Senarath, 2014) and numerical-computational simulations (Echeverribar et al., 2017), its approximate calculation is carried out resorting to complex technologies such as the satellite sensors used by Aguirre and Caro (Aguirre & Caro Gómez, 2014) or Vales et al. (Vales et al., 2010) that carried out a study collecting data on the extension of the flood and capturing them on maps of affected areas in Andalusia (Spain). Thus, the extent of the sheet of water produced is detected. Subsequently, cartography and characterization are obtained from direct observation of the territory during the development of the phenomenon and not based on simulation models, using Radar technology (for obtaining photographs).

On the other hand, the CNIH offers historical data on homes affected by floods in Spain. In the case of Andalusia, the years 1963 and 2010 (Fig. 2) are the periods in which most of them are found, so These years and the most affected areas will be the scope of this study, to be able to take advantage of precisely these existing data and relate it to new ones for the objective pursued.

Regarding the importance or weight of these factors (Kreibich et al., 2009), accurately determining some of them is sometimes a difficult matter, however, with the available data, in this study, a matrix of

**Table 1**
Factors influencing the degree of affectation of buildings (Dall'Osso et al., 2009; Galasso et al., 2021; Núñez Collado et al., 2019) and the degree of difficulty in obtaining data.

| Outstanding factors related to the degree of affectation of a flood in buildings | Compliance with ease conditions to collect data | Final classification |
|---|---|---|
| Minimum distance from the building to the bank of the riverbed when it has its maximum ordinary level. | 1,2,3 | Easy |
| Unevenness between the level of the ground floor of the building and the maximum ordinary level of the river. (Measured considering the point of the minimum distance between the building and the river). | 1,2,3 | Easy |
| Existence of possible water communication between the building and the river in case of flooding regardless of the existence or not of defense elements. | 1,2,3 | Easy |
| The maximum level of the water surface above ground (ml). | 2,3 | Easy-Difficult |
| The time that the maximum level of the water sheet remains. | 2,3 | Easy-Difficult |
| Hydrodynamic pressure and flow velocity. | – | Difficult |
| Level of water penetration inside the house. | 2,3 | Easy-Difficult |
| Age in years of the building. | 1,2 | Easy-Difficult |
| Number of floors in height. | 1,2,3 | Easy |
| Hydrodynamics of the ground floor. | – | Difficult |
| Constructive features. | 1,2 | Easy-Difficult |
| Building shape (circular, rectangular, polygonal, etc.). | 1,2,3 | Easy |
| Building orientation (perpendicular, parallel or at an angle to the water flow). | 2,3 | Easy-Difficult |
| State of conservation. | 1,2 | Easy-Difficult |
| Objects that drag the sheet of water. | 2,3 | Easy-Difficult |
| Others. | – | |

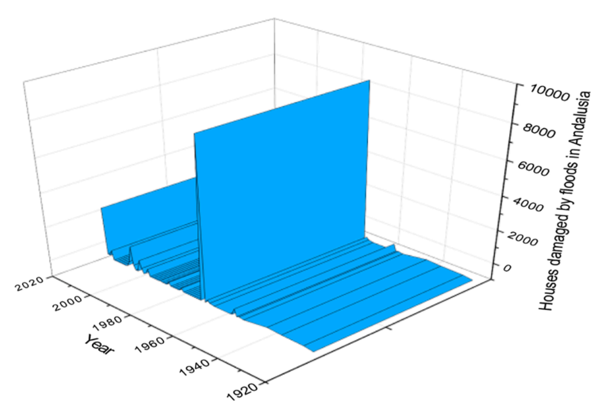



**Fig. 2.** Number of houses affected by floods (1920 to 2010). Andalusia (Spain) (CNIH 2014)

correlation is made to said effects.

### 2.2. Study area and characteristics of the houses analyzed

For the case study, the dwellings located in Spain have been taken, specifically in the andalusian municipalities of Peñaflor, Lora Del Río, Alcolea Del Río, Villanueva del Río, Tocina, Los Rosales, Cantillana, Villaverde del Río, and Écija.

Regarding this last locality, only some properties that were flooded by the overflow of the Gentil River have been studied.

These locations in Andalusia (Spain) are towns with between 2000 and 40,000 inhabitants, most of them close to the Guadalquivir river as it passes through the province of Seville.

Fig. 3 shows the situation map, the location of said populations and the levels of risk detected based on the Prevention Plan for floods and floods in Andalusian urban channels. This study area is selected for this model due to the existence of said risk and because floods have historically occurred due to river overflow.

These areas have a continental Mediterranean climate. Advancing inland, the continentally increases. The average rainfall ranges between 500 and 700 mm, with between 75 and 100 days of rainfall per year and an average annual temperature of 17–19 °C. Winters are usually mild with irregular rainfall, and dry, hot and sunny summers. These last characteristics are accentuated as move from the coast to the interior of the region. The month with the least rainfall is usually in July with only an average of 0.020 mm. The month of November is usually the rainiest. In the analyzed areas, the height above sea level is between 20 and 100 m.

Regarding ground-level elevations, accurate representation is a fundamental objective in geodetic surveying and is essential for flood modelling studies (Sampson et al., 2016; Yamazaki et al., 2014). In this sense, the accuracy of Digital Elevation Models (DEMs) has improved in recent years due to advances in remote sensing techniques (Yamazaki et al., 2017). There are various databases and maps available online. However, for adequate accuracy, the recently built DEM (MERIT DEM) is highly accurate and is known (Topographic-map, 2022).

Regarding the characteristics of the houses, they are two-story, with a reinforced concrete structure with a brick factory enclosure and a flat walkable roof, or a structure with load-bearing walls and a sloping roof with curved tiles. The exterior finishes are cement mortar coating and plastic paint. On the ground floor, there is a façade clad with natural or artificial stone at the height of 1.20 m. Regarding conservation and maintenance, the houses are in a good state of preservation.

The entrances to the buildings are generally: i) Pedestrian, saving a height of 30 cm; ii) access above ground from the roadway and finished in non-slip flooring.

The houses analyzed are residences, and the water has penetrated the property to practically the same height as the level of water outside.

### 2.3. Affectation levels

To define the levels of damage, the lesions observed in the buildings were analyzed following the methodology mentioned by Domínguez Gutiérrez, J. & González Pájaro, Abel. (Domínguez Gutiérrez & González Pajaro, 2015) and what is specified in the "Guide for the inspection and evaluation of damage to buildings due to flooding" edited by the Valencian Institute of Building (García-Prieto Rui et al., 2009). This information was compiled orderly, and a state of conservation was defined based on the levels of damage detected. The states of degradation were rated employing a numerical scale. This numerical gradation was defined, in a general way, by the following characteristics: i) **Level 1 (Low).** It was generally assigned to cleaning operations and generalized maintenance or light and punctual repairs. Light damage; ii) **Level 2 (Medium).** Corresponds to major repairs, up to 50% in the extension of the element. Serious damage; iii) **Level 3 (High).** The element that requires very important repairs (affecting more than 50% of its extension) or requires its total replacement. Very serious damage.

### 2.4. Data processing

After setting the factors whose data they want to get, they are collected. Once obtained, they are pre-processed before using them in modeling. Said pre-processing includes: i) Integration; ii) Cleaning and organization; iii) checking if some data are outliers or missing; iv) carrying out training, testing and validation of data sets where appropriate.

For the identification of outliers, the Inter Quantile Range (IQR) is

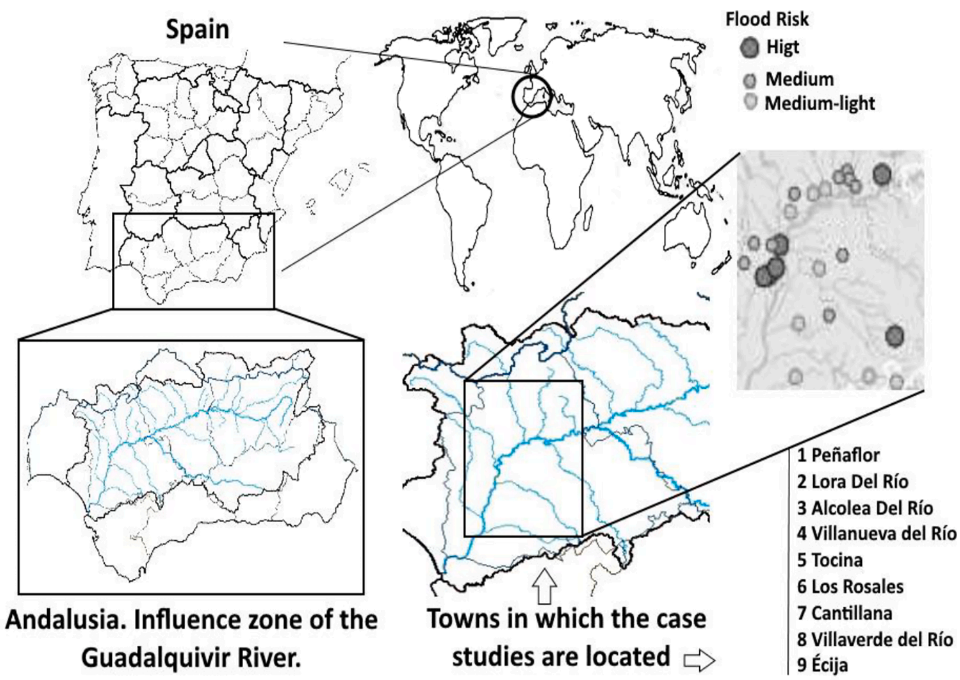



**Fig. 3.** Situation map, location of the study populations and levels of risk detected based on the Prevention Plan for floods in Andalusian urban channels.(Junta de Andalucía, 2002).

used, and a scaling method. However, in order not to eliminate values that could be considered atypical but were real, some of them were not eliminated.

Correlation analysis of the data was carried out to understand the relation between the input and the output parameters.

### 2.5. Machine learning algorithms

As for machine learning models, there is currently a great variety and in turn, they can be divided into two types: univariate and multivariate. In the case of the study, since there are more than two input and output variables, the multivariate ones will be used. On the other hand, depending on the way in which they carry out learning, these models can be of the supervised, unsupervised and semi-supervised types. In this work, those of the supervised type are used. At this point, it should be noted that to obtain results a validation process is necessary. The simplest way to perform the validation process is to use the holdout technique. This technique simply consists of dividing the original data set into two sets, one for training and the other for testing (generally 70% of the data is reserved for the first and 30% for the second), trying to maintain the same distributions in the data two sets.

The main problem with this type of validation is that the test set can be too small, so the variance of the estimate is high and since there is a disproportion with respect to the training data, the final estimate may tend to be pessimistic. On the other hand, there is also the problem that overfitting may occur during the process.

To evited these problems, some authors (Kohavi, 1995) state that 10-fold cross-validation is the best of the methods. It is true that nuances expressed by other authors (Refaeilzadeh et al., 2009) could fit into this statement, which in any case does not invalidate that the method is adequate for the characteristics of the study to be carried out.

Thus, cross-validation (CV) was implemented in such a way that the training data is divided into ten random subgroups, each of which represents a fold. Each model is then trained on nine folds and then validated ten times on the remaining fold, with the validation fold changing each time. The standard deviation and mean prediction score are determined using the 10 scores generated by CV.

Regarding the evaluation of the performance of the models, it can be done using various metrics. In this sense, it must be said that this work does not focus on comparing the possible applicable models, since the objective is simply to test some of them to see if precisions are obtained that can be considered sufficient to affirm that they can provide reasonable guidance on the level of the foreseeable damage. Thus, this study focuses fundamentally on the analysis of the confusion matrix, since it is a classification-prediction and therefore other metrics really correspond better to other models that are not the object of study.

## 3. Results

For the development of this section, it has been taken into account at all times that the study focuses on finding out if it is possible to predict the degree of damage to buildings considering only those selected environmental factors whose data can be obtained in virtually any circumstance. On the other hand, it is also intended to find out if, with ML algorithms for multivariate supervised learning, it is possible to achieve a reasonable level of accuracy in predicting the degree of affectation, training these algorithms with a low-density dataset made up of buildings with similar characteristics in a restricted territorial area.

Thus, after indicating in table 1 the most influential factors and among them the easiest to obtain, table 2 specified based on which calculations and initial data the final data used in the models were deduced.

Therefore, Fig. 4 shows the most outstanding factors and the casuistry studied.

Thus, regarding the minimum distance from the building to the bank of the riverbed when it has its maximum ordinary level, in Fig. 5A), this distance can be seen for each dwelling, the furthest being 1549.61 m, and the closest to 29.03 m, and the average of 431.48 m. Regarding the unevenness between the level of the ground floor of the building and the height of the maximum ordinary level of the river, in Fig. 5B), this unevenness can be seen for each dated dwelling, the lowest being −58 ml below the level of the river, and the highest at +10 ml, with the average being −6.44 ml. There is also a clear predominance of buildings located

**Table 2**
Data collected that is more readily available.

| Item | Date |
|---|---|
| **Building location.** | Postal address of the building or buildings.<br>Latitude.<br>Length.<br>Height above sea level (m) of the ground floor of the building. |
| **Location of the nearest channel at its maximum ordinary flood level.** | Latitude.<br>Length.<br>Height above sea level (m) of the level of maximum ordinary flood. |
| **Interest factors.** | A) Minimum distance from the building to the bank of the riverbed when it has its maximum ordinary level (Haversine formula).<br>B) Unevenness between the level of the ground floor of the building and the maximum ordinary level of the river. (Measured considering the point of the minimum distance between the building and the river)<br>C) Existence of possible water communication between the building and the river in case of flooding, regardless of the existence or not of defense elements. |
| **Result** | Damage level of the building. |

at a lower level than the maximum ordinary level of the river. Regarding the level of communication (low, medium, high) of the buildings with the river, in such a way that in case of overflow the water can reach them, Fig. 5C) shows the levels of each case. As can be seen, the number of cases chosen from each level of affection is balanced, that is, practically equal to each other, which is an important issue for the subsequent use of ML algorithms and obtaining correct results that do not fit simply to the frequency of cases.

On the other hand, Fig. 6 shows the relation between the level of affection and A) the minimum distance in meters between the building and the canal; B) the difference in elevation (meters) between the elevation of the ground floor of the building and the ordinary maximum flood elevation; C) the level of communication (1 low, 2 medium, 3 high) of the buildings with the river in such a way that in case of overflow the water can reach them. Several interesting evidences can be seen in this figure. For example, all buildings with high affection are less than 600 m from the river and those with medium damage are less than 700. It can also be seen that in buildings that are above the level -+0.00, their affection is usually medium or low, and from −2, −3 m it is only low. Finally, regarding the existence or not of water communication between the river and the building, it usually coincides with medium or high levels of affectivity in the first case and low or punctually medium in the second. These one-off anomalies maybe because it is a very prominent overflow or because a simple basic communication may cause more damage than expected.

Also, correlation analysis of the data was carried out to understand the relationship between the input and the output parameters table 3 shows the correlation matrix obtained for the parameters.

Thus, the positive Pearson coefficient is 0.706 (communication-affection) that contribute as important factor to the severity of the damage of the building in case of flooding. The distance to the river and the difference in level between its usual channel and that of the ground floor of the building have a negative correlation value of −0.442 and −0.462, respectively. The correlation may be linked to a decrease in the degree of the severity of the damage when there is a long distance or the difference in level is very large.

Correlation analysis gave insight into how different input parameters correlate with the dependent variable. Although some parameters have a minor impact on the target variable, all parameters have been used as input because they do not contribute to the computational complexity of the model.

From the results, it can be inferred that the existence or not of a direct communication of the water from the river to the building and the existence or not of defenses have the maximum impact since they have a positive correlation with respect to the level of affection of the building. With all this information, tests have been conducted by training various well-known machine learning algorithms, using the cross-validation technique, and 10 folds. For them, the R tool (R Foundation. The R Project for Statistical Computing, 2022), Weka (Hall et al., 2009), Knime (KNIME | Open for, 2022 ) and RapidMiner (University of Dortmund, 2001) software have been used fundamentally. Of all the models tested (K-NN Nearest Neighbour Model, Naives Bayes, Random Forest, Gradient Boosted Trees, Rule Induction, among others), the "Decision Tree" was the only one that offered in validation accuracy of <80% (Percentage of correct predictions with respect to the total number of cases-predictions).

The complete results applying the cross-validation in the "Decision

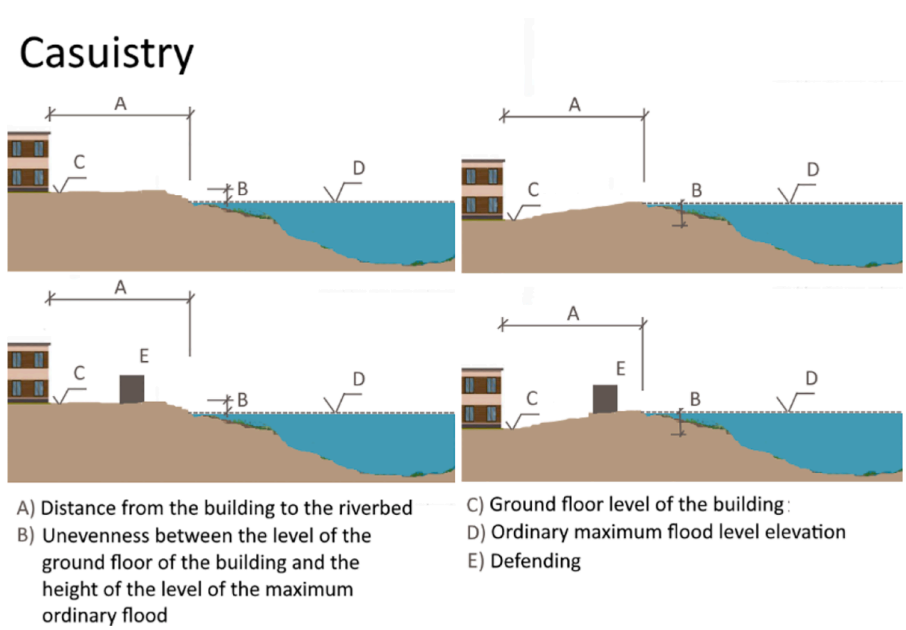



**Fig. 4.** Factors and the casuistry studied.

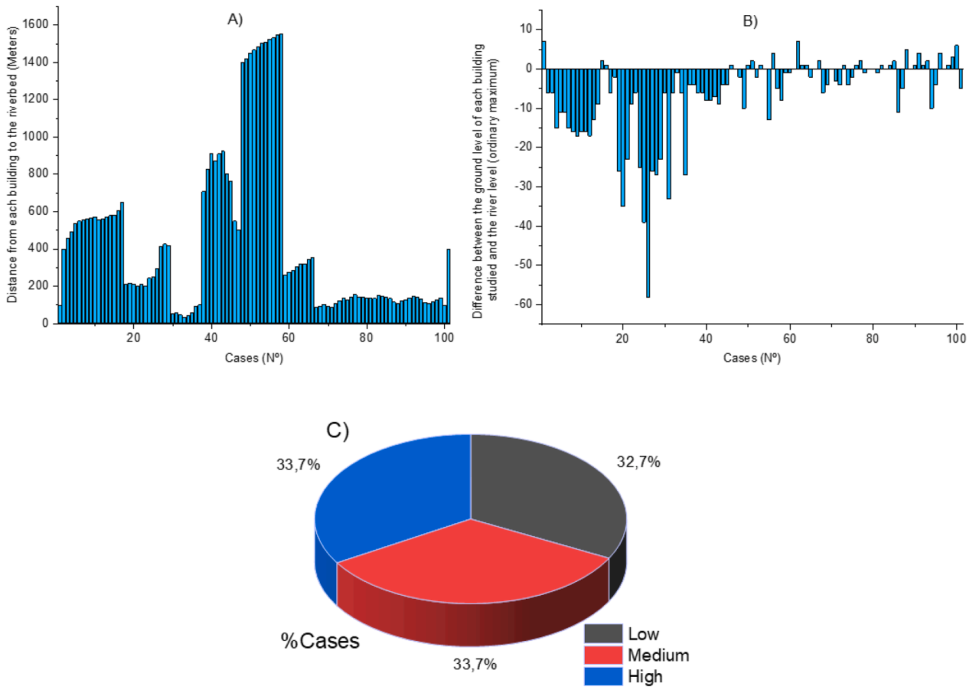



**Fig. 5.** A) Minimum distance (ml) from the building to the bank of the riverbed when it has its maximum ordinary level; B) Unevenness or the difference (ml) between the level of the ground floor of the building and the maximum ordinary level of the river; C) Level of communication (low, medium, high) of the buildings with the river in such a way that in case of overflow the water can reach them.

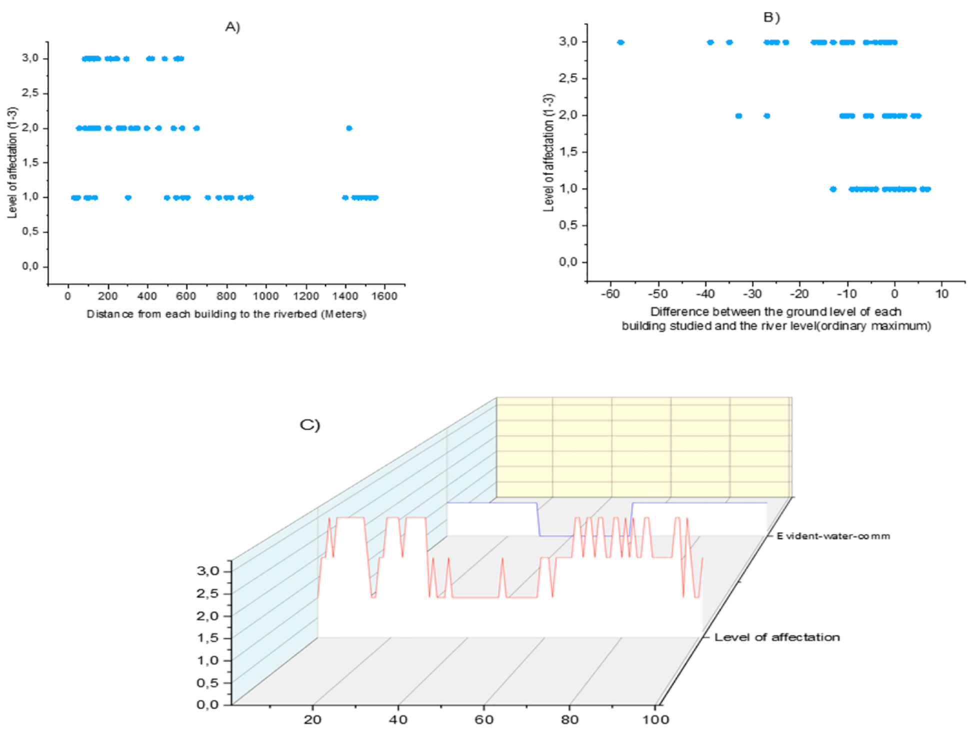



**Fig. 6.** Relation between the level of affection and A) the minimum distance in meters between the building and the channel; B) Unevenness or the difference (ml) between the level of the ground floor of the building and the maximum ordinary level of the river; C) the level of communication (1 low, 2 medium, 3 high) of the buildings with the river in such a way that in case of overflow the water can reach them.

Tree" model can be seen in Table 4.

For the purposes of the validation results, Table 5 shows the "accuracy" and other data of interest obtained directly from the "Decision Tree" model, and applying the cross-validation.

## 4. Analysis and discussion

In the first place, it is necessary to comment that climate change is affecting the water cycle (Da Silva, Alencar, & de Almeida, 2022; Pour

**Table 3**
Correlation matrix of parameters.

| | *Distance* | *Difference* | *Communication* | *Level of affectation* |
|---|---|---|---|---|
| **Distance** | 1 | 0.1 | −0.61 | −0.442 |
| **Difference** | 0.1 | 1 | −0.035 | −0.462 |
| **Communication** | −0.61 | −0.035 | 1 | 0.706 |
| **Level of affectation** | −0.442 | −0.462 | 0.706 | 1 |

et al., 2020), and, in general terms, extreme events are occurring with increasing frequency, for example, in the case with heavy downpours, which increase the risk of flooding.

On the other hand, it should be noted that it is possible to harbor certain doubts regarding the sample size since it is a low-density data set. However, this does not mean that this research is not useful since it is research carried out in a specific area with certain characteristics of the buildings and climatic conditions. Therefore, it applies to this specific area, being of great interest to the methodology to carry out similar studies in other settings with other characteristics.

Focusing the analysis on the results, if these are compared when cross-validation is being applied with those produced without it, it can be deduced that it is a correct and reliable validation, since without it the "accuracy" is 95.5%, that is, too close to 100% to be true, thus casting doubt on whether it really reflects the real performance of the model. On the other hand, with a validation carried out by dividing the data into 70% training and 30% validation, the "accuracy" is 76.67%. However, due to the reasons and documentation exposed in Section 2.5, on the advantages of applying cross-validation, which yields an "accuracy" of 81.09% +/- 13.77%, it can be concluded that these last results better reflect the real performance of the chosen "Decision Tree" model.

Thus, the "Decision Tree" accuracy (Percentage of total correct predictions with respect to the total number of predictions) and the Kappa could be said that they are acceptable values, since if higher values are desired, the number of cases would have to be increased and this is sometimes not possible since the number of cases is usually limited as it is a very specific area and similar buildings. As for the rest of the metrics, it is understood that there is nothing especially remarkable, being the linear relationship and the connection greater than 0.85, and the error metrics in accordance with the "accuracy".

Regarding de result from the observation of the resulting tree, it can be deduced that if there is no clear evidence of water communication between the building, the level of damage is medium or low and it will be one or the other depending on the difference in level with the river. On the other hand, in the event that there is clear communication, if the ground floor of the building is more than 5 m above the highest level that the river usually has, the impact will be low. If this is not the case and it is less, if the unevenness is less than 5.5 but greater than 0.5 m, then the affection will be low if the distance is greater than 448 m and medium if it is less than that distance. On the other hand, if the unevenness is less than 0.5 m, the affection will be medium level if the distance is greater than 572 m and if it is less than this distance, it will be high affection, unless the unevenness is >−7.5 m or the distance <134 m.

Regarding the confusion matrix and the analysis of the result of each class, it is detected that for the cases of "low and high affection", both the "recall" and the "precision" are greater than for the class of "medium affection". In this sense, it should be noted that in the case of "high affection" they are close to 90% (recall and precision=88,24%). This circumstance is of great importance since this class is the most interesting to predict as it is the one that does the most damage.

In short, it is observed that a decision tree such as the one proposed is of great interest to find out the classification in three levels of affectation of future ones, at least for a certain area such as the one studied and for a typology of buildings and similar environment.

**Table 4**
Results of the application of machine learning algorithm "Decision Tree" with cross-validation(University of Dortmund, 2001).

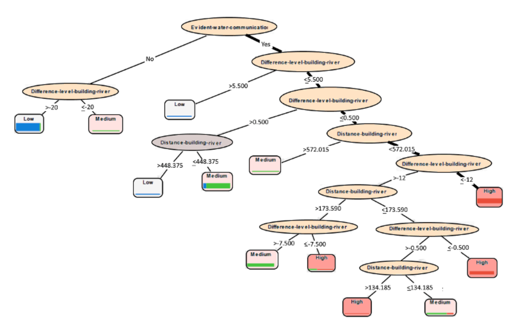
**Training Model** — **Metrics and Results**

**Accuracy: 81.09%** +/- 13.77% (micro average: 81.19%)
**Classification_error: 18.91%** +/- 13.77% (micro average: 18.81%)
**kappa: 0.715** +/- 0.207 (micro average: 0.718)
**Weighted_mean_recall: 80.83%** +/- 13.82% (micro average: 81.22%), weights: 1, 1, 1
**Weighted_mean_precision: 83.00%** +/- 12.83% (micro average: 81.11%), weights: 1, 1, 1
Spearman_rho: 0.889 +/- 0.092 (micro average: 8.892)
Kendall_tau: 0.853 +/- 0.110 (micro average: 8.525)
Absolute_error: 0.209 +/- 0.121 (micro average: 0.208 +/- 0.331)
Relative_error: 20.87% +/- 12.05% (micro average: 20.82% +/- 33.13%)
Relative_error_lenient: 20.87% +/- 12.05% (micro average: 20.82% +/- 33.13%)
Root_mean_squared_error: 0.360 +/- 0.164 (micro average: 0.391 +/- 0.000)
Squared_error: 0.154 +/- 0.116 (micro average: 0.153 +/- 0.314)
Correlation: 0.879 +/- 0.095 (micro average: 0.859)
Squared_correlation: 0.782 +/- 0.159 (micro average: 0.738)
Margin: 0.170 +/- 0.264 (micro average: 0.170)
Soft_margin_loss: 0.209 +/- 0.121 (micro average: 0.208)
Logistic_loss: 0.386 +/- 0.045 (micro average: 0.386)

**TREE**

**Decision tree**

**Evident-water-communication = No**
**| Difference-level-building-river > -20: Low {Low=26, Medium=1, High=0}**
**| Difference-level-building-river ≤ -20: Medium {Low=0, Medium=2, High=0}**
**Evident-water-communication = Yes**
**| Difference-level-building-river > 5.500: Low {Low=3, Medium=0, High=0}**
**| Difference-level-building-river ≤ 5.500**
**| | Difference-level-building-river > 0.500**
**| | | Distance-building-river > 448.375: Low {Low=2, Medium=0, High=0}**
**| | | Distance-building-river ≤ 448.375: Medium {Low=2, Medium=15, High=0}**
**| | Difference-level-building-river ≤ 0.500**
**| | | Distance-building-river > 572.015: Medium {Low=0, Medium=2, High=0}**
**| | | Distance-building-river ≤ 572.015**
**| | | | Difference-level-building-river > -12**
**| | | | | Distance-building-river > 173.590**
**| | | | | | Difference-level-building-river > -7.500: Medium {Low=0, Medium=10, High=0}**
**| | | | | | Difference-level-building-river ≤ -7.500: High {Low=0, Medium=1, High=2}**
**| | | | | Distance-building-river ≤ 173.590**
**| | | | | | Difference-level-building-river > -0.500**
**| | | | | | | Distance-building-river > 134.185: High {Low=0, Medium=0, High=2}**
**| | | | | | | Distance-building-river ≤ 134.185: Medium {Low=0, Medium=3, High=1}**
**| | | | | | Difference-level-building-river ≤ -0.500: High {Low=0, Medium=0, High=12}**
**| | | | Difference-level-building-river ≤ -12: High {Low=0, Medium=0, High=17}**




**Confusion Matrix (Multi-Class Classification)**

| | true Low | true Medium | true High | class precision |
|---|---|---|---|---|
| pred. Low | 28 | 6 | 0 | 82.35% |
| pred. Medium | 5 | 24 | 4 | 72.73% |
| pred. High | 0 | 4 | 30 | 88.24% |
| class recall | 84.85 % | 70.59 % | 88.24 % | |

**DEFINITIONS:**

**TPs/TNs/FPs/FNs:** True Positives / True Negatives /False Positives/False Negative. When is a Multi-Class Classification (not binary), like the case of study, what has to be done here is to find TP, TN, FP and FN for each individual class.
**Accuracy:** Percentage of correct predictions (referred to the total number of predictions). The higher the better. However, the "Precision" in a multiclass ML represents the relationship (in %) between a prediction of a class (for example, the prediction of low damage) and the total number of real cases related to said prediction. On the other hand, the "Recall" represents the relationship (in %) between the true cases of a class (for example, the cases in which minor damage actually occurred) and the number of correct or incorrect predictions for said true case.
**Classification error:** Percentage of incorrect predictions. The alter ego of accuracy. The lower the better. Good results are considered below 10%. However, other issues must be taken into account.
**kappa:** Statistical measure of adjusting for the effect of chance on the observed proportion of agreement. It takes into account that the correct prediction occurs by chance. Therefore, its result is usually of interest.
**Precision:** Only matrix rows because, in the case study, it is calculated for each individual class. It is calculated=TPs/TPs+ another values only in the corresponding row.
**Recall:** Only matrix columns because, in the case study, it is calculated for each individual class. It is calculated=TPs/TPs+ another values only in the corresponding column. It is considered that the model is better the higher the precision and the recall (sensitivity) for all classes. If the model in some class does not achieve high precision or recall, it would be interesting to analyze the causes.
**Weighted mean recall:** The weighted average of all recall measurements by class. Interesting to analyze how recall behaves in general.
**Weighted mean precision:** The weighted average of all precision measurements by class. Interesting to analyze how precision behaves in general.
**Spearman's rho:** measure of the linear relationship between two variables (tag attribute and predict attribute).
**kendall_tau:** rank correlation between the actual and predicted labels.
**Absolut error (Ae):** Average absolute deviation of the prediction from the actual value.
**Relative error:** "Ae" divided by the actual value.
**Relative error lenient:** "Ae" divided by the maximum of the actual value and the prediction
**Root mean squared error:** averaged of measure of differences between values predicted-values observed.
**Squared error:** average squared difference between the estimated values and the actual value.
**Correlation:** correlation coefficient between the label and prediction attributes
**Squared correlation:** he squared correlation coefficient between the label and prediction attributes
**Cross entropy:** the sum over the logarithms of true label's confidences divided by the number of examples.
**Margin:** he minimal confidence for the correct label.
**Soft margin loss:** average of all 1 - confidences for the correct label.
**Logistic loss:** average of ln(1+exp(-[conf(CC)])) where 'conf(CC)' is the confidence of the correct class. It is the negative average of a log of accurately predicted probabilities and indicates the extent to which the prediction probability is similar to its respective actual value.

**Table 5**
Results of main metrics for validation: A) direct; B) with cross-validation.

| Metrics | Direct results | Results applying cross-validation |
|---|---|---|
| **Accuracy:** | 95,5% | **81.09%** +/- 13.77% |
| **Classification_error:** | 4,95 | **18.91%** +/- 13.77% |
| **kappa:** | 0,926 | **0.715** +/- 0.207 |
| **Weighted_mean_recall:** | 95.04% | **80.83%** +/- 13.82% |
| **Weighted_mean_precision:** | 95.12% | **83.00%** +/- 12.83% |

## 5. Conclusions

One of the most interesting and novel issues of this study is related to the methodology presented, which focuses on simplifying the input data, in such a way that they are easy to obtain, based on key factors and limiting the scope to buildings of a typology similar in an area of location also of similar characteristics with respect to the environment of a certain pluvial channel. Since most of the studies in this field are based on the collection of an enormous amount of data, many of which are difficult to obtain, it is understood that this is an innovative contribution.

Thus, the results indicate that with few data, relatively simple to obtain, such as the minimum distance from the building to the river, the difference between the level of the ground floor of the building and the maximum ordinary level of the river, and the existence of possible water communication between the building and the river in case of flooding, it is possible to synthesize the prediction analysis of the condition in a given area with suitable algorithms ML, highlighting the need to avoid communication as one of the most important factors when it comes to improving the vulnerability of buildings. On the other hand, the model still has limitations since it deals with specific buildings in a certain area with similar characteristics. To generalize the model, it would be necessary to increase the cases in other territories with other characteristics and architectural typologies. However, it is understood that the results are of interest to advance in the intended prediction and to deploy strategies and policies that allow them to be used, given their simplicity, to mitigate the damages or to calculate them for future insurance compensation or aid or subsidies to the General State Administration or to make decisions related to contracting an insurance policy at the time of quickly recovering from the economic damages that impact in the dwellings, having adequately calculated the values to be insured.

### Declaration of Competing Interest

This research has not been financed. The authors declare no conflicts of interest.

### Data availability

Research data are not shared.

### References

Adhikari, P., Hong, Y., Douglas, K. R., Kirschbaum, D. B., Gourley, J., Adler, R., et al. (2010). A digitized global flood inventory (1998-2008): Compilation and preliminary results. *Natural Hazards, 55*(2), 405–422. https://doi.org/10.1007/S11069-010-9537-2

Aguirre, M.A., & Caro Gómez, E. (2014). Andalusia: The foundations for a regional Space policy. *12th European Week of Regions and Cities*. http://www.nereus-regions.ovh/wp-content/uploads/2017/11/Open_Days_def_andalusia.pdf (Accessed 10 April 2021).

Ahmadlou, M., Al-Fugara, A., Al-Shabeeb, A. R., Arora, A., Al-Adamat, R., Pham, Q. B., et al. (2021). Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks. *Journal of Flood Risk Management, 14*(1), e12683. https://doi.org/10.1111/JFR3.12683

Amadio, M., Rita Scorzini, A., Carisi, F., Essenfelder, H. A., Domeneghetti, A., Mysiak, J., et al. (2019). Testing empirical and synthetic flood damage models: The case of Italy. *Natural Hazards and Earth System Sciences, 19*(3), 661–678. https://doi.org/10.5194/NHESS-19-661-2019

Arabameri, A., Rezaei, K., Cerdà, A., Conoscenti, C., & Kalantari, Z. (2019). A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. *Science of the Total Environment, 660*, 443–458. https://doi.org/10.1016/j.scitotenv.2019.01.021

Armenakis, C., Du, E. X., Natesan, S., Persad, R. A., & Zhang, Y. (2017). Flood Risk Assessment in Urban Areas Based on Spatial Analytics and Social Factors. *Geosciences 2017, 7*(4), 123. https://doi.org/10.3390/GEOSCIENCES7040123

Bohorquez, P., & del Moral-Erencia, J. D. (2017). 100 Years of Competition between Reduction in Channel Capacity and Streamflow during Floods in the Guadalquivir River (Southern Spain). *Remote Sensing 2017, 9*(7), 727. https://doi.org/10.3390/RS9070727

Chen, Y., Liu, T., Ge, Y., Xia, S., Yuan, Y., Li, W., et al. (2021). Examining social vulnerability to flood of affordable housing communities in Nanjing, China: Building long-term disaster resilience of low-income communities. *Sustainable Cities and Society, 71*, Article 102939. https://doi.org/10.1016/J.SCS.2021.102939

CNIH. (2014). Retrieved., from http://cnih.proteccioncivil.es/cnih/index.php?r=site/inicio (Accessed 10 April 2021).

Da Silva, L. B. L., Alencar, M. H., & de Almeida, A. T. (2022). A novel spatiotemporal multi-attribute method for assessing flood risks in urban spaces under climate change and demographic scenarios. *Sustainable Cities and Society, 76*, Article 103501. https://doi.org/10.1016/J.SCS.2021.103501

Dall'Osso, F., Gonella, M., Gabbianelli, G., Withycombe, G., & Dominey-Howes, D. (2009). A revised (PTVA) model for assessing the vulnerability of buildings to tsunami damage. *Natural Hazards and Earth System Sciences, 9*(5), 1557–1565. https://doi.org/10.5194/NHESS-9-1557-2009

De Moel, H., Van Alphen, J., & Aerts, J. C. J. H. (2009). Flood maps in Europe - methods, availability and use. *Natural Hazards and Earth System Sciences, 9*(2), 289–301. https://doi.org/10.5194/NHESS-9-289-2009

Deckers, P., Kellens, W., Reyns, J., Vanneuville, W., & Maeyer, P. De (2009). A GIS for flood risk management in Flanders. *Geospatial Techniques in Urban Hazard and Disaster Analysis*, 51–69. https://doi.org/10.1007/978-90-481-2238-7_4

Deria, A., Ghannad, P., & Lee, Y. C. (2020). Evaluating implications of flood vulnerability factors with respect to income levels for building long-term disaster resilience of low-income communities. *International Journal of Disaster Risk Reduction, 48*, Article 101608. https://doi.org/10.1016/J.IJDRR.2020.101608

Domínguez Gutiérrez, J., & González Pajaro, A. (2015). Valoración técnica del deterioro de las edificaciones en la zona costera de Santa Fe. *Arquitectura y Urbanismo, 36*, 48–61. scielocu http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1815-58982015000100005&lng=es&nrm=iso. (Accessed 15 April 2021).

Doorga, J. R. S., Magerl, L., Bunwaree, P., Zhao, J., Watkins, S., Staub, C. G., et al. (2021). GIS-based multi-criteria modelling of flood risk susceptibility in Port Louis, Mauritius: Towards resilient flood management. *International Journal of Disaster Risk Reduction, 67*(2212–4209), Article 102683. https://doi.org/10.1016/J.IJDRR.2021.102683

Dutta, D., Herath, S., & Musiake, K. (2003). A mathematical model for flood loss estimation. *Journal of Hydrology, 277*(1–2), 24–49. https://doi.org/10.1016/S0022-1694(03)00084-2

Echeverribar, I., Morales-Hernández, M., Lacasta, A., Brufrau, P., & García-Navarro, P. (2017). Simulación numérica con RiverFlow2D de posibles soluciones de mitigación de avenidas en el tramo medio del río EbroRiverFlow2D numerical simulation of flood mitigation solutions in the Ebro River. *Ingeniería del agua, 21*(1), 53–70. https://doi.org/10.4995/IA.2017.6550

European Commission. (2007). Directive 2007/60 /EC of the European Parliament and of the Council of October 23, 2007 regarding the evaluation and management of the risks of flood. *Official Journal of the European Union (OJ)*. https://ec.europa.eu/environment/water/flood_risk/index.htm. (Accessed 10 April 2021).

Feyen, L., Dankers, R., Bódis, K., Salamon, P., & Barredo, J. I. (2012). Fluvial flood risk in Europe in present and future climates. *Climatic Change, 112*(1), 47–62. https://doi.org/10.1007/S10584-011-0339-7

Galasso, C., & Senarath, S. U. S. (2014). *A Statistical Model for Flood Depth Estimation in Southeast Europe*. Second International Conference on Vulnerability and Risk Analysis and Management (ICVRAM).Sixth International Symposium on Uncertainty, Modeling. *Vulnerability, Uncertainty, and Risk*, 1415–1424. https://ascelibrary.org/doi/10.1061/9780784413609.142.

Galasso, C., Pregnolato, M., & Parisi, F. (2021). A model taxonomy for flood fragility and vulnerability assessment of buildings. *International Journal of Disaster Risk Reduction, 53*, Article 101985. https://doi.org/10.1016/j.ijdrr.2020.101985

García-Prieto Rui, S., de Mazarredo Pampló, F., & cosme, & Serrano Lanzarote, B. (2009). *Guía para la inspección y evaluación de daños en edificios por inundaciones* (G. Valenciana & U. y V. Conselleria de Medio Ambiente, Agua (Eds.)). Instituto Valenciano de la Edificación. https://www.112cv.gva.es/documents/163565706/163566523/47+Guía+para+la+inspección+y+evaluación+de+daños+en+edificios+por+inundaciones.pdf/d3cebddc-f09f-4c78-8bf5-8e239c31c99d.

Government of Spain. Sede Electrónica del Catastro. (2022). Retrieved from https://www.sedecatastro.gob.es/. (Accessed 20 April 2021).

Government of Spain. Sistema Nacional de Cartografía de Zonas Inundables/National Flood Zone Mapping System. (2010). Ministry for the Ecological Transition and the Demographic Challenge. https://www.miteco.gob.es/es/agua/temas/gestion-de-los-riesgos-de-inundacion/snczi/. (Accessed 10 november 2021).

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software. *ACM SIGKDD Explorations Newsletter, 11*(1), 10. https://doi.org/10.1145/1656274.1656278

Jakovljevic, G., Govedarica, M., Alvarez-Taboada, F., & Pajic, V. (2019). Accuracy Assessment of Deep Learning Based Classification of LiDAR and UAV Points Clouds for DTM Creation and Flood Risk Mapping. *Geosciences 2019, 9*(7), 1–21. https://doi.org/10.3390/GEOSCIENCES9070323

Junta de Andalucía. Plan de Prevención de avenidas e inundaciones en cauces urbanos andaluces/Plan for the prevention of avenues and floods in Andalusian urban channels. (2002). DECREE 189/2002. Retrieved from https://www.juntadeandalucia.es/boja/2002/91/5. (Accessed 1 April 2021).

Junta de Andalucía. REDIAM Red de Información Ambiental de Andalucía. Portal Ambiental de Andalucía. (2022). http://www.juntadeandalucia.es/medioambiente/rediam. (Accessed 10 April 2021).

KNIME | Open for Innovation. (2022). Retrieved., from https://www.knime.com/.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Articial Intelligence (IJCAI), 14*(2), 1137–1145. https://doi.org/10.5555/1643031.1643047

Kreibich, H., Piroth, K., Seifert, I., Maiwald, H., Kunert, U., Schwarz, J., et al. (2009). Is flow velocity a significant parameter in flood damage modelling? *Natural Hazards and Earth System Sciences, 9*, 1679–1692. https://doi.org/10.5194/nhess-9-1679-2009

Li, X., Yan, D., Wang, K., Weng, B., Qin, T., & Liu, S. (2019). Flood Risk Assessment of Global Watersheds Based on Multiple Machine Learning Models. *Water 2019, 11*(8), 1–18. https://doi.org/10.3390/W11081654

Manrique Ruíz, A., Nájera Ibáñez, A., Escartín Hernández, C., Moreno Fernández, C., Martínez Bravo, E., Espejo Gil, F. et al. (2017). *Guía para la reducción de la vulnerabilidad de los edificios frente a las inundaciones*.Consorcio de Compensación de Seguros (CCS). 106. https://ebroresilience.com/wp-content/uploads/2018/12/guia_inundaciones_completa_22jun.pdf. (Accessed 10 April 2021).

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article "Assessment of economic flood damage. *Natural Hazards and Earth System Sciences, 10*(8), 1697–1724. https://doi.org/10.5194/NHESS-10-1697-2010

Merz, B., Kreibich, H., Thieken, A., & Schmidtke, R. (2004). Estimation uncertainty of direct monetary flood damage to buildings. *Natural Hazards and Earth System Sciences, 4*, 153–163. https://doi.org/10.5194/nhess-4-153-2004

Meyer, V., Scheuer, S., & Haase, D. (2009). A multicriteria approach for flood risk mapping exemplified at the Mulde river, Germany. *Natural Hazards, 48*(1), 17–39. https://doi.org/10.1007/s11069-008-9244-4

Núñez Collado, A., Salgado Cámara, A., Martín Anta, D., Prieto Leache, I., Montoya Rodríguez, J., Ferreiro Gómez, P., et al. (2019). *Recomendaciones para la construcción y rehabilitación de edificaciones en zonas inundables* (p. 74). Government of Spain. https://www.miteco.gob.es/es/agua/temas/gestion-de-los-riesgos-de-inundacion/guia-recomendaciones-construccion-y-rehabilitacion-edificaciones-zonas-inundables_tcm30-503724.pdf. (Accessed 10 April 2021).

Osés-Eraso, N., & Foudi, S. (2020). Valoración de riesgos por inundaciones. Presupuesto y gasto público, 101, 261-282. https://hdl.handle.net/2454/42326. (Accessed 8 April 2021).

Perles Roselló, M. J., Olcina, J., & Mérida Rodríguez, M. (2018). Balance de las políticas de gestión del riesgo de inundaciones en España: De las acciones estructurales a la ordenación territorial/Balance of flood risk management policies in Spain: From structural actions to territorial planning (197th,, *50. Ciudad y Territorio. Estudios Territoriales* (pp. 417–438). Ministerio de Fomento (España). http://rua.ua.es/dspace/handle/10045/89637. (Accessed 1 April 2021).

Pham, Q. B., Ali, S. A., Bielecka, E., Calka, B., Orych, A., Parvin, F., et al. (2022). Flood vulnerability and buildings' flood exposure assessment in a densely urbanised city: Comparative analysis of three scenarios using a neural network approach. *Natural Hazards, 2022*, 1–39. https://doi.org/10.1007/S11069-022-05336-5

Piyumi, M. M. M., Abenayake, C., Jayasinghe, A., & Wijegunarathna, E. (2021). Urban flood modeling application: assess the effectiveness of building regulation in coping with urban flooding under precipitation uncertainty. *Sustainable Cities and Society, 75*, Article 103294. https://doi.org/10.1016/J.SCS.2021.103294

Pour, S. H., Wahab, A. K. A., Shahid, S., Asaduzzaman, M., & Dewan, A. (2020). Low impact development techniques to mitigate the impacts of climate-change-induced urban floods: Current trends, issues and challenges. *Sustainable Cities and Society, 62*, Article 102373. https://doi.org/10.1016/J.SCS.2020.102373

R Foundation. The R Project for Statistical Computing. (2022). Retrieved., from https://www.r-project.org. (Accessed 10 January 2022).

Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A., et al. (2019). Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis. *Earth Systems and Environment, 3*(3), 585–601. https://doi.org/10.1007/S41748-019-00123-Y

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. *Encyclopedia of Database Systems, 5*, 532–538. https://doi.org/10.1007/978-0-387-39940-9_565

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., & Trigg, M. A. (2016). Perspectives on open access high resolution digital elevation models to produce global flood hazard layers. *Frontiers in Earth Science, 3*. https://doi.org/10.3389/FEART.2015.00085

Topographic-map, Mapa topográfico España, altitud, relieve. (2022). Retrieved., from https://es-es.topographic-map.com/maps/6olb/España/. (Accessed 10 May 2021).

Universidad de Dortmund. RapidMiner Ltd.(2001). https://rapidminer.com/. (Accessed 10 April 2021).

Vales, J., Carpintero, I., Granado, L., Méndez, E., Montoya, G., Pino, I., et al. (2010). Estudio de áreas afectadas de inundación en las cuencas del Guadalquivir y Guadalete a partir de imágenes radar del satélite terrasar-x. REDIAM. *La Información Geográfica Al Servicio de Los Ciudadanos [Recurso Electrónico]: De Lo Global a Lo Local, 1*, 1000–1013. https://idus.us.es/handle/11441/66601.

Vallejo Villalta, I. (2000). Inundaciones en la cuenta del Guadalquivir. *Serie Geográfica, 9*, 133–149. https://idus.us.es/bitstream/handle/11441/73722/Las-Inundaciones-en-la-Cuenca-del-Guadalquivir.pdf?sequence=1. (Accessed 6 April 2021).

Vanneuville, W., Wolters, H., Scholz, M., Uhel, R,. (2016). Flood Risks and Environmental Vulnerability.Exploring synergies between floodplain restoration, water policies and thematic policies. EEA Report,1, 78. https://www.eea.europa.eu/publications/flood-risks-and-environmental-vulnerability. (Accessed 11 April 2021).

Wan Mohtar, W. H. M., Abdullah, J., Abdul Maulud, K. N., & Muhammad, N. S. (2020). Urban flash flood index based on historical rainfall events. *Sustainable Cities and Society, 56*, Article 102088. https://doi.org/10.1016/J.SCS.2020.102088

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology, 527*, 1130–1141. https://doi.org/10.1016/J.JHYDROL.2015.06.008

Yamazaki, D., Sato, T., Kanae, S., Hirabayashi, Y., & Bates, P. D. (2014). Regional flood dynamics in a bifurcating mega delta simulated in a global river model. *Wiley Online Library, 41*(9), 3127–3135. https://doi.org/10.1002/2014GL059744

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., et al. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters, 44*(11), 5844–5853. https://doi.org/10.1002/2017GL072874/FORMAT/PDF