





Article

Systematic Mapping of Global Research on Disaster Damage Estimation for Buildings: A Machine Learning-Aided Study

Dilum Rajapaksha ^{1,2}, Chandana Siriwardana ^{3,*}, Rajeev Ruparathna ⁴, Tariq Maqsood ¹,
Sujeewa Setunge ¹, Lalith Rajapakse ² and Saman De Silva ¹

¹ Department of Civil and Infrastructure Engineering, School of Engineering, RMIT University, Melbourne, VIC 3001, Australia; s3957720@student.rmit.edu.au (D.R.); tariq.maqsood@rmit.edu.au (T.M.); sujeewa.setunge@rmit.edu.au (S.S.); saman.desilva@rmit.edu.au (S.D.S.)

² Department of Civil Engineering, University of Moratuwa, Moratuwa 10400, Sri Lanka; lalith@uom.lk

³ School of Built Environment, Massey University, North Shore, Auckland 0745, New Zealand

⁴ Department of Civil and Environmental Engineering, University of Windsor, 401 Sunset Ave, Windsor, ON N9B 3P4, Canada; rajeev.ruparathna@uwindsor.ca

* Correspondence: c.siriwardana@massey.ac.nz

Abstract: Research on disaster damage estimation for buildings has gained extensive attention due to the increased number of disastrous events, facilitating risk assessment, the effective integration of disaster resilience measures, and policy development. A systematic mapping study has been conducted, focusing on disaster damage estimation studies to identify trends, relationships, and gaps in this large and exponentially growing subject area. A novel approach using machine learning algorithms to screen, categorise, and map the articles was adopted to mitigate the constraints of manual handling. Out of 8608 articles from major scientific databases, the most relevant 2186 were used in the analysis. These articles were classified based on the hazard, geographical location, damage function properties, and building properties. Key observations reveal an emerging trend in publications, with most studies concentrated in developed and severely disaster-affected countries in America, Europe, and Asia. A significant portion (68%) of the relevant articles focus on earthquakes. However, as the key research opportunities, a notable research gap exists in studies focusing on the African and South American continents despite the significant damage caused by disasters there. Additionally, studies on floods, hurricanes, and tsunamis are minimal compared to those on earthquakes. Further trends and relationships in current studies were analysed to convey insights from the literature, identifying research gaps in terms of hazards, geographical locations, and other relevant parameters. These insights aim to effectively guide future research in disaster damage estimation for buildings.

Keywords: damage estimation; fragility curves; vulnerability curves; natural hazards; buildings; machine learning



Citation: Rajapaksha, D.; Siriwardana, C.; Ruparathna, R.; Maqsood, T.; Setunge, S.; Rajapakse, L.; De Silva, S. Systematic Mapping of Global Research on Disaster Damage Estimation for Buildings: A Machine Learning-Aided Study. *Buildings* **2024**, *14*, 1864. <https://doi.org/10.3390/buildings14061864>

Academic Editors: Emlyn Witt, Abdulquadri Ade Bilau and Bin Sun

Received: 11 April 2024

Revised: 19 May 2024

Accepted: 17 June 2024

Published: 20 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The intensity and frequency of natural hazards and resulting disasters have rapidly increased over the past years [1,2]. According to the Emergency Event Database (EM-DAT), in 2023 alone, 399 disastrous events were recorded, claiming 86,473 lives, affecting 93.1 million people, and causing approximately USD 202.7 billion in economic losses [3]. Evidently, the number of recorded disasters has increased by 75% during the past 20 years (2000–2019) compared to the previous period (1980–1999). Climate change can be identified as a key driver behind these increased events, as it affects the temporal patterns of weather conditions, resulting in extreme weather events such as severe heatwaves, inland flooding, and harsh hurricane seasons [4,5]. These extremes result in both direct destruction of the built environment and indirect impacts on it by affecting material deterioration and structural integrity [6–8]. As examples of direct damage, the Tōhoku earthquake (2011)

can be identified as the most damaging disastrous event in recent history, resulting in an estimated USD 253 billion in damage [9]. The event caused significant damage to over 400,000 buildings [10]. Other major earthquake events include the Sichuan earthquake (2008), which resulted in USD 107 billion in damage, and the Chūetsu earthquake (2004), which caused USD 40 billion in damage. In addition to earthquakes, several major hurricanes have caused significant damage in recent years. These include Hurricane Ida (2021), Hurricane Harvey (2018), Hurricane Maria (2018), Hurricane Irma (2018), Hurricane Sandy (2013), and Hurricane Katrina (2005), which have resulted in a combined damage estimate of USD 540 billion. Also, the Thailand Monsoon Floods (2012), which can be identified as the biggest flood event in recent history, also resulted in USD 48 billion in damage [9]. Apart from the direct impacts, accelerated material degradation through carbonation-induced deterioration, corrosion, salt crystallisation cycles, ultraviolet dosage, and freeze–thaw cycles due to climate extremes can aggravate damage to the structural integrity of the built environment, ultimately causing possible collapses [11–14]. In order to reduce and control such impacts, it is essential to consider the effects of climate extremes during the design process and take necessary actions such as updating the design parameters and integrating disaster resilience measures for new structures. Integrating structural health monitoring systems, sustainable retrofitting, and well-designed disaster mitigation strategies would be beneficial for existing structures [15–17].

When dealing with direct impacts, it is beneficial to have an idea of possible damage that could occur due to different magnitudes of disasters. It will assist the designers and policymakers in the process of disaster risk management. Disaster damage estimation methods play a vital role in such occasions. These methods are primarily used to estimate the damage due to a disaster event quantitatively [18]. This knowledge of the estimated damage can be used in different applications, such as assessing the feasibility of disaster resilience, mitigation and adaptation measures, evidence-based planning for disaster response and resource allocation, and decision-making and policy development [18–22]. During the disaster damage estimation process, the probability of occurrence of a hazard with a certain magnitude is obtained from hazard curves. Then, the possible damage to a structure is estimated using damage functions based on the considered hazard parameters and other relevant parameters.

A hazard curve maps the probability of exceeding a certain threshold against the severity of a specific hazard in a given area [23], and a hazard curve can be developed using traditional probabilistic hazard analysis [24]. Damage functions link the hazard parameters with the damage parameters. It allows users to obtain the damage parameters of a structure, such as absolute damage, relative damage, and the damage index, based on the hazard intensities and other relevant parameters. There are different types of damage functions used for this purpose, as introduced below. Fragility functions express the probability of exceeding defined damage levels of a structure or a component at different levels of hazard intensity [25,26]. Vulnerability functions, damage curves, and stage-damage curves instead describe the direct relationship between the probable damage for different severities of a hazard [27,28]. In this study, the “damage function” refers to all the above-introduced functions, as they all express the relationship between damage and hazard characteristics. Also, the scope is limited only to the damage estimations of buildings due to different disasters.

Damage functions for buildings are associated with multiple aspects, such as different hazards focused on, different approaches to development, different building materials, different building types, and different building uses. Damage functions focusing on different hazards such as earthquakes, floods, hurricanes, tsunamis, and landslides can be found in the literature [29–32]. Different approaches, such as empirical approaches, where recorded damage data are used to identify the damage relationship; analytical approaches, where theoretical, experimental, or simulation-based methods are used to identify damage behaviour; expert judgment elicitation, where expert or engineering judgment is used to determine damage behaviour, and the combination of multiple methods

(hybrid approaches), have been identified as the major approaches for damage function development in the literature [33]. In recent years, incorporating machine learning (ML) applications has also been recognised as another approach to developing damage functions. Enabling the derivation of multivariable damage functions and aiming to predict the most accurate results rather than simply capturing the relationships differentiate ML approaches from other identified approaches [34]. Furthermore, the damage functions can be categorised based on building material, including reinforced concrete, masonry, timber, or steel [35]; building use, such as residential, commercial, industrial, educational, or religious [36]; and building type, such as low-rise, medium-rise, or high-rise [37]. Focusing on different parameters, as discussed above, has enabled authors to identify the damage relationships more specifically and has enhanced the reliability of the damage estimations.

As mentioned, it can be identified that the studies on damage functions are distributed over a more extensive scope. Therefore, the idea of current trends, relationships, and gaps in the existing literature will be helpful in future research to direct them through research gaps and efficiently develop damage functions to facilitate building damage predictions. This study aims to map the existing literature on disaster damage estimations and damage functions, capturing the overall scope of the subject area and discussing the insights for future research based on the research trends, relations, and gaps.

2. Background and Related Studies

A literature map covering the overall scope of damage estimation studies on buildings has not been conducted yet. Some of the attempts to identify existing trends in damage estimation studies can be identified from the literature reviews. The review studies below focused on different segments of the scope of damage estimation studies. Maio et al. (2017) assessed the most significant features of fragility curves for seismic risk assessments of buildings in Europe and identified the drawbacks of the current studies by reviewing 39 seismic fragility curves [33]. Mastroberti and Vona (2016) critically reviewed six studies and discussed different analytical methods used to develop fragility models for existing reinforced concrete buildings with moment-resisting frames [38]. Pregolato et al. (2015) reviewed 15 studies on the flood fragility curves in terms of primary data sources, building features, statistical techniques employed in data analysis and reliability, limitations of each study, and recommendations for future studies [39]. Malgwi et al. (2020) conducted a state-of-the-art review of existing flood damage models to develop a new framework for flood damage assessment in data-scarce regions [40]. Jongman et al. (2012) assessed and compared seven flood damage models to develop a harmonised European approach to flood damage assessment [41]. Gerl et al. (2016) reviewed 61 flood damage models and vulnerability relationships to assess the validity and robustness and to present a method for the harmonisation of models for benchmarking and comparison [42]. However, it can be identified that the scope of those studies was restricted by different parameters, such as hazard, location, or other building parameters.

While the above-discussed studies have a more detailed and deeper review of different subsets of damage estimation studies, an overall idea of the subject area cannot be obtained. For instance, insights such as the overall distribution of the damage estimation studies based on hazards, geographical variation of the existing studies for different hazards, and overall trends of integrating different building parameters into damage estimation studies would be beneficial for future research, especially during the initial planning and scope-defining stage. However, considering the whole scope in a manually performed review is unattainable when considering the number of studies available in the subject area. Therefore, manually mapping the existing pool of damage estimation studies would be a practical constraint. However, integrating ML applications to facilitate such situations is an emerging approach. In this case, appropriate ML algorithms can be used to facilitate the mapping study.

As ML applications are used to facilitate most conventional methods, natural language processing (NLP), an ML technology, can be used to facilitate the literature reviews. NLP

works as a bridge between human language and computer-understandable language [43]. NLP applications successfully operate as language translators, email filters, automatic summarisations, text mining, predictive texts, and writing assistants [43–45]. ML and NLP applications can be adopted in scientific article mapping under the concept of a trained computer algorithm that can screen and categorise the records as a human reader does [46]. Some of the previous authors successfully demonstrated the effectiveness of ML in facilitating literature mapping across various subject areas.

Berrang-Ford et al. (2021) used ML techniques to conduct a systematic mapping review in the field of climate and health. The authors identified and mapped 16,049 relevant records from a total of 286,917 query-resulted documents [46]. Similarly, Callaghan et al. (2021) employed ML techniques to analyse a large dataset of 601,677 documents in the field of climate change impacts, resulting in the inclusion and classification of 102,160 relevant studies in their final analysis [47]. Kukushkin, Ryabov, and Borovkov (2022) used ML models to facilitate their literature review on digital twins, including 8693 studies in their analysis [48]. Xiong et al. (2018) integrated ML techniques into their systematic review on diabetes mellitus and atrial fibrillation and used ML models to handle 4177 query-resulted documents [49]. Also, Bannach-Brown et al. (2019) conducted a study evaluating the performance of incorporating ML algorithms into systematic reviews in the field of animal studies. The authors achieved a recall rate of 98.7% and an accuracy rate of approximately 85% and used ML models to handle a document set of 70,365 [50]. Zimmerman et al. (2021) also investigated the performance of incorporating ML techniques into a systematic literature review on diabetes-related studies and were able to predict results with a recall rate of 99.5% by conducting a human review of only 31% of the total articles [51]. These studies highlight that using ML techniques to assist the literature mapping process has successfully identified and analysed relevant records and documents in a large dataset that might be beyond human capabilities.

With the rising trend of artificial intelligence (AI), there are several large language model-based tools available to facilitate literature reviews. Also, publishers are integrating different AI models into search engines. While it has its own advantages, such as providing broader coverage and facilitating exploratory searches, integrating specific ML models, as described in studies, has more specific advantages and is more suitable to targeted and specific tasks. Integrating specific ML models results in improved accuracy and precision, as the models are specifically trained using a document pool of a specific domain. This provides the ability to adapt to the considered domain by enabling them to capture the nuances of that specific domain. Also, as the authors can examine the performance of models, which can be enhanced to a satisfactory level. Therefore, integrating specific ML models is suited for more specific tasks, like the current study [52–54].

In this study, scientific articles related to disaster damage estimations and damage functions of buildings were mapped. Supervised and unsupervised ML applications were used to facilitate screening and categorising, which enabled the whole scope under the subject area to be captured. The main objective is to study different subtopics related to the main subject area and identify their interrelationships. The identified relationships, trends, and research gaps in the selected subject area are presented and discussed in this paper.

3. Methodology

The overall methodology contains four major steps: article searching and filtering, human screening, application of ML models, and finally, the visualisation of results. The overall methodology of the study is graphically presented in Figure 1.

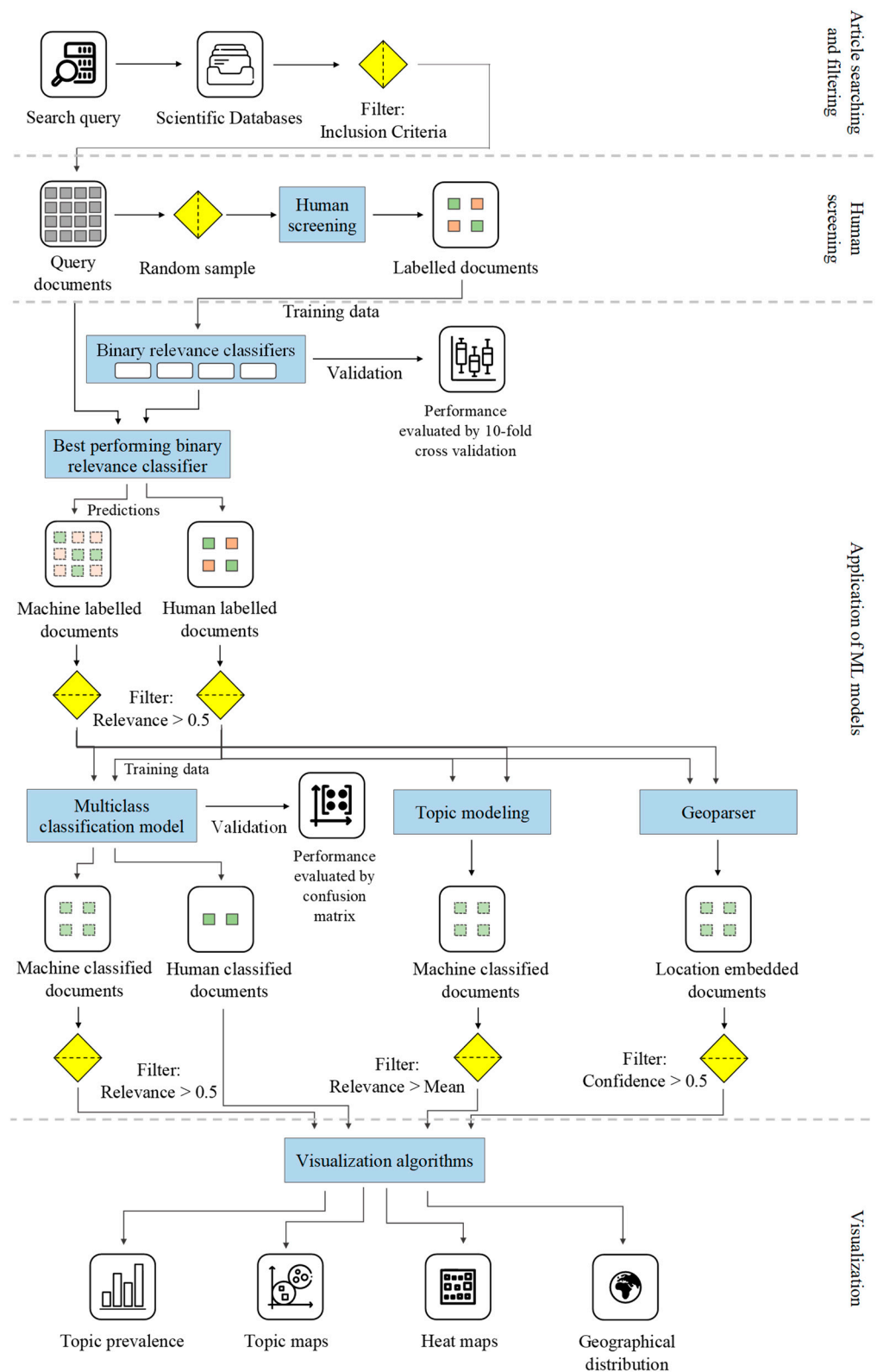


Figure 1. Overall methodology of the study.

The key search items were damage functions, buildings, and natural hazard-induced disasters. The search string was generated by combining the key terms listed in Table 1. The included terms were selected to represent the overall subject area. The selected databases for the article search were Scopus, Web of Science, and Science Direct. A document had

to meet the inclusion criteria of (1) published in or after the year 2000 to 2023 (search was conducted in May 2023) in the English language and (2) a journal article, conference article, review, or book chapter in the discussed databases to be included in the analysis. Only the title, abstract, and keywords were used during the literature search and analysis. The search string with the above inclusion criteria resulted in 8608 records (Scopus—3508 records, Web of Science—3298 records, Science Direct—1802 records), and after the duplication removal, 6061 unique records were used in the analysis.

Table 1. Selected key term to generate a search query.

Key Subject Area	Key Terms
Damage functions	Damage model, damage curves, fragility curves, damage assessment, damage estimation, stage damage curves, vulnerability curves
Buildings	Buildings
Natural hazard-induced disasters	Disasters, flood, earthquake, seismic, tsunami, landslide, hurricane, tornado, typhoon, cyclone, climate change

A random sample of 439 records (approximately 7% of the records) were manually screened and labelled. The exclusion criteria for the screening were (1) studies that did not develop or apply one or more damage functions or disaster damage estimation criteria or (2) studies that were not aimed at buildings. The records that satisfied the above criteria were marked as relevant. After that, the relevant records were subjected to manual classification. The type of natural hazard that the study focused on and the approach used to develop the damage function were identified as critical categories related to disaster damage estimation studies. During the manual classification, it was identified that almost all the identified relevant studies focused on one of the following natural hazards: earthquakes, floods, tsunamis, or hurricanes. Note that all severe wind-related hazards, such as tornados, typhoons, and cyclones, are included under the category of hurricanes. Furthermore, it was identified that the relevant studies mainly adopted analytical, empirical, ML, and hybrid approaches to develop damage functions during the manual classification. Therefore, the identified relevant documents from the random sample were manually classified into the above two categories and forwarded as the training set for the ML algorithms.

Four supervised and unsupervised ML models were incorporated into this study, which are (1) a binary classifier for predicting the relevance of the documents (supervised), (2) a multiclass classifier for predicting the major categories of the documents (supervised), (3) a pre-trained geoparser for extracting the geographical locations of the studies (unsupervised), and (4) a topic-modelling algorithm for identifying the key subtopics discussed (unsupervised). A general introduction to each model is given in the next paragraph, and more detailed technical descriptions for each model are given in the following paragraphs.

As the first model, a binary classifier was used. The objective of this model is to determine the relevance of a given document to the main subject area. During the manual data screening, the relevance to the subject area of each document in the selected random sample was marked as 1 (relevant) or 0 (not relevant). Then, that data set was used to train the binary classifier to predict the relevancy of any given document. Since there are several binary classifiers available, the prediction performance of a few selected classifiers was assessed initially to determine the best-performing classifier for this study, and it was used to predict the relevance of unseen documents. As the second model, multiclass classifiers were used. The objective of these models is to determine the category of each relevant document (e.g., hazard that the particular study focused on). Two separate multiclass classifiers were used in this study to determine the hazard and development method of each document. During the manual data screening, the related categories for each relevant document were labelled. Then, that data set was used to train the multiclass classifiers.

After evaluating the performance of the classifiers by comparing the actual and predicted categories of documents, the trained models were used to determine the categories of unseen documents. A pre-trained geoparser was used as the third model. The objective of this model is to capture the location of a considered document. This model is pre-trained, and if any location-related detail (i.e., city, province, country) is stated in the text, the model captures it. All the relevant documents were forwarded to this model to identify the location. Topic modelling, an unsupervised ML approach, was used as the fourth model. The objective of this model is to identify the subtopics and the relevancy of the documents to the identified subtopics. The algorithm automatically identifies the words that commonly appear together in documents and clusters them into an anonymous topic. The authors review the resulting word clusters and, if it is meaningful, they assign the subtopic names to represent the selected subject area. After that, each document was labelled with the most relevant subtopics.

For the binary classifier, the Scikit-learn package from the Python library was used to develop the required ML algorithms because of the easiness of practice, efficient implementation, and good documentation [55], and those were successfully used for the same purpose by previous authors [46,50,51,56]. Initially, the performance of the four binary classifiers from the Scikit-learn package, namely, the Multi-layer Perceptron Classifier (Neural Net), Multinomial Naive Bayes Classifier (Bayes), Random Forrest Classifier (RandForest), and C-Support Vector Classification (SVM-rbf) [57] for the training data set, was evaluated using 10-fold cross-validation, and the best-performing classifier was used to predict the relevance of the remaining articles. The performance was assessed using the means of three variables (precision, recall, and accuracy). The resulting distribution from 10-fold cross-validation for all four models is presented in Figure 2a. Based on the results, the Multi-layer Perceptron Classifier gave the highest accuracy (85.7%) with the comparatively same distribution for precision and recall, and hence, it was used to predict the relevancy of unseen documents. The model predicts a relevance value between 0 and 1 for each document. By randomly reviewing the resulted relevance values, authors have determined that 0.5 would be suitable as the threshold to determine the relevancy. This means if the relevance value is greater than or similar to 0.5, the document is considered relevant; otherwise, the document is considered non-relevant. The ML model predicted 1941 relevant studies from the unseen documents, and with the manually identified 245 relevant studies, a total of 2186 relevant documents were considered in the analysis.

Two separate multiclass classification models (Support Vector Machine in a One vs Rest setup) were trained to predict the category of each document in terms of the hazard that the document aimed for (earthquake, flood, tsunami, or hurricane) and the approach used to develop the damage function in the study (analytical, empirical, ML, or hybrid). The ML models were able to categorise the documents with mean accuracies of 97.1% and 76.4% for hazard categorisation and approach categorisation, respectively. Also, the performances of the multiclass models were reviewed using the confusion matrixes as given in Figure 2b,c. The horizontal axis denotes the actual category of a document, and the vertical axis denotes the predicted category of a document. The matrix comparatively shows the alignment of actual and predicted categories. Finding the maximum values at diagonals verified that the models were able to predict the categories more accurately. Further, it can be identified that the accuracy for hybrid category predictions was comparatively lower than the other categories. Since the hybrid approach uses multiple approaches to derive a damage function, it is possible to contain keywords from other approach categories in the hybrid approach in subsequent studies, which may cause slightly less accurate predictions. Then, the model was used to predict the categories of unseen documents. Similar to the previous model, after careful consideration, 0.5 was taken as the threshold value for categorisation.

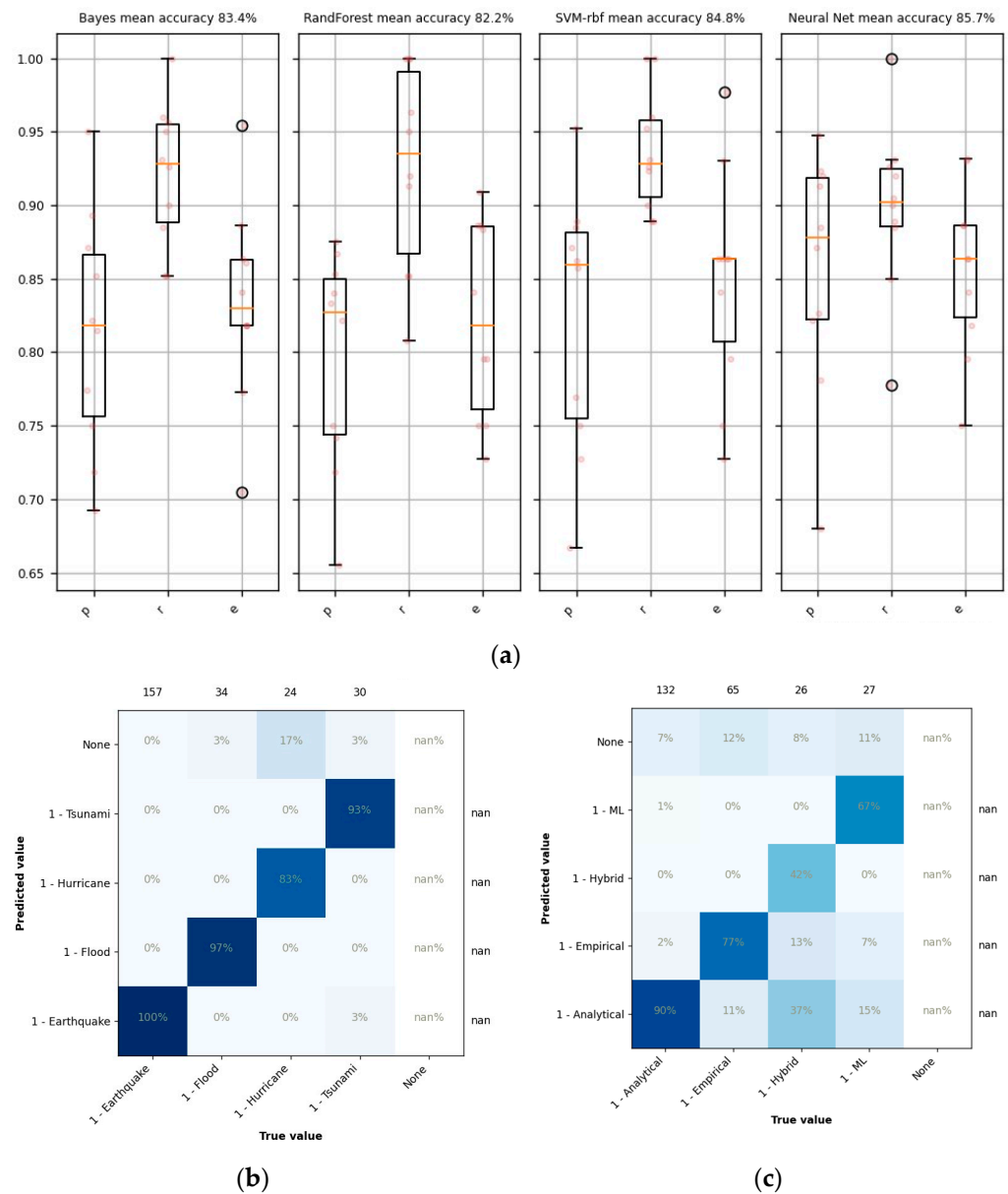


Figure 2. Performance of supervised ML models. (a) Result distribution of 10-fold cross validation for the binary classifier. In the x-axis, ‘p’, ‘r’, and ‘e’ refer to the precision, recall, and accuracy values, respectively; (b) confusion matrix for the multiclass classifier on the hazard; (c) confusion matrix for the multiclass classifier on the development approach.

As the pre-trained geoparser, the Mordecai full-text geoparsing system was used [58]. This model was successfully applied for the same purpose by previous authors [46,47]. All relevant documents were directed to the geoparser to extract locations and obtain the respective country of the study. The model predicts the country with a confidence value, and after careful consideration, it was determined that if the confidence of the country prediction was greater than 0.5, the particular study was related to the predicted country. The results were used to comparatively analyse the study locations with the other disaster parameters.

The latent Dirichlet allocation (LDA) [59] model from the Gensim Python library [60] was used for the topic modelling. The same model was successfully applied by previous authors for the same purpose [48,51,56]. Special concerns, such as enabling bigrams and trigrams and filtering out unwanted repetitive words, were given to achieve more accurate results during the process. The optimum number of topics was selected based

on the coherence value, and 25 topics gave the highest value for coherence. Based on the hyperparameter clusters identified under each topic, the relevant subtopics below were finalised according to the scope of this study. The selected hyperparameters were categorised under key subject areas other than the hazards and approaches, including building materials (reinforced concrete, masonry, precast, timber, and steel), building usage (residential, school, industrial, historical, and religious), building type (low-rise, medium-rise, and high-rise), building location (regional and urban), and life cycle phase (planning, designing, construction, and retrofitting). If the prediction score given by the model per hyperparameter per document was larger than the mean of the prediction values per hyperparameter, it was determined that the considered hyperparameter was related to the particular document.

Finally, the results of the analysis were presented using prevalence charts, choropleth maps, topic maps, and heat maps. Prevalence charts were used to identify the frequency of occurrence of each hyperparameter. Choropleth maps were used to comparatively understand the geographical distribution of the studies against different disaster parameters. A topic map was used to identify the trends of the studies. The relative co-occurrence of the topics was depicted using heat maps. The generated figures, methods of generation, and key observations are discussed in the next section.

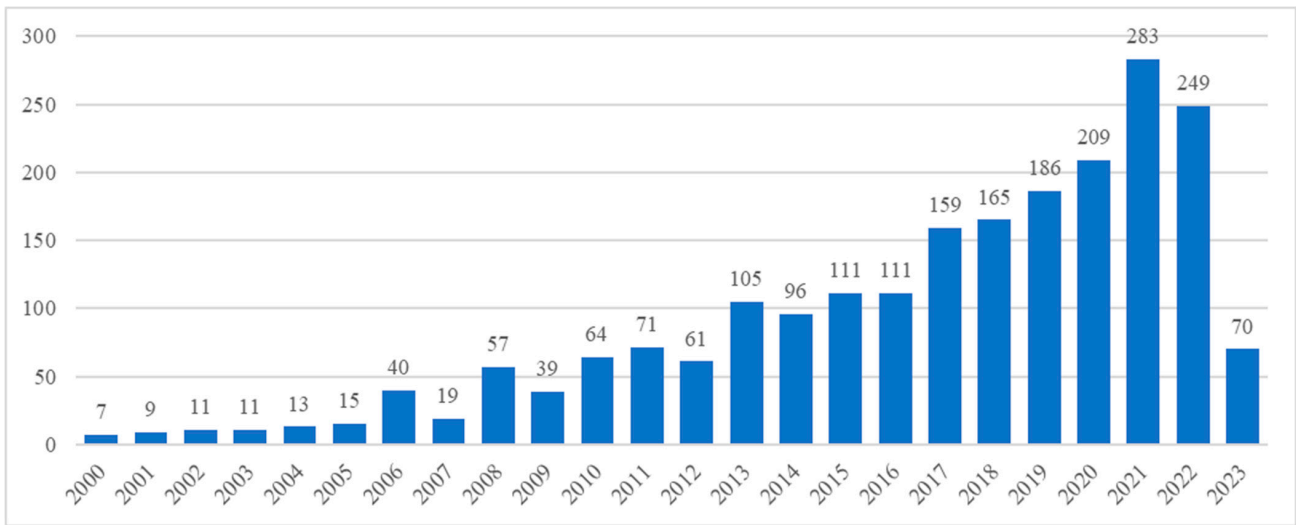
4. Results

4.1. Trends of Publications

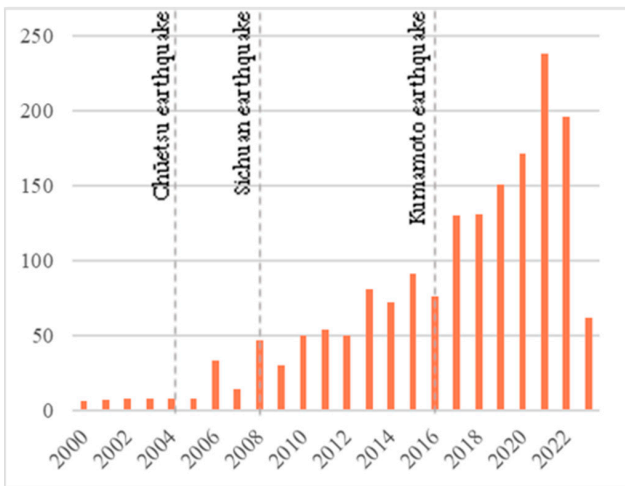
The publication trends in the relevant study area were assessed to understand the focus of researchers over the years. Figure 3 contains the yearly variation of the publications for overall damage estimation studies and separately for hazards considered. The figure was generated by the authors. Identified relevant studies were categorised based on the publication year, and different hazards and the bar plots were generated accordingly. Figure 3a demonstrates a rapid increase in publications related to disaster damage estimation, indicating an increasing interest in this subject area. The publication trends for different types of hazards were analysed along with the influence of the most damaging events on these publications. The data on damaging events were extracted from the EM-DAT database.

Figure 3b presents the publication trend of earthquake-based studies, showing a general trend of rapid increase in recent years. The most damaging earthquake events during the considered time frame were the 2008 Sichuan earthquake (China), the 2016 Kumamoto earthquake (Japan), and the 2004 Chūetsu earthquake (Japan). The number of earthquake-related studies increased after 2004 and continued to show an upward trend after 2008. However, there was a significant increase in publications after 2016, suggesting that major events and simultaneous earthquakes influenced the subject area and subsequent studies.

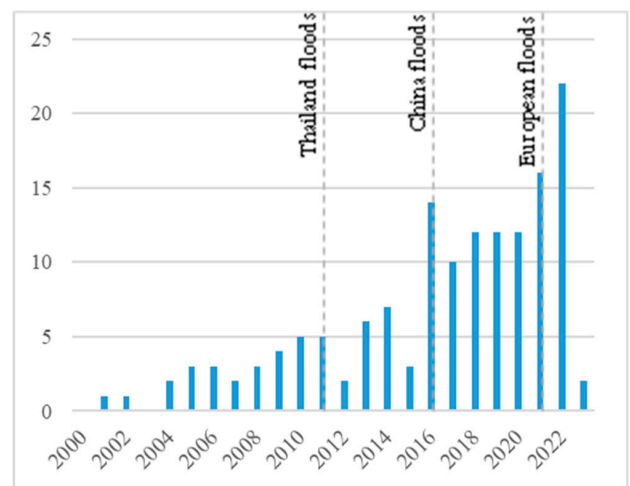
Similarly, Figure 3c illustrates the publication trend of flood-based studies, displaying an increasing pattern over the past few years. The most devastating flood events recorded during the focus period were the 2011 Thailand floods, the 2021 Europe floods, and the 2016 China floods. The trend indicates some engagement in flood-related studies since the early 2000s, but there was increased engagement after the 2016 events. Notably, there was a rapid increase in flood-related studies in 2022 following the European flood events in 2021.



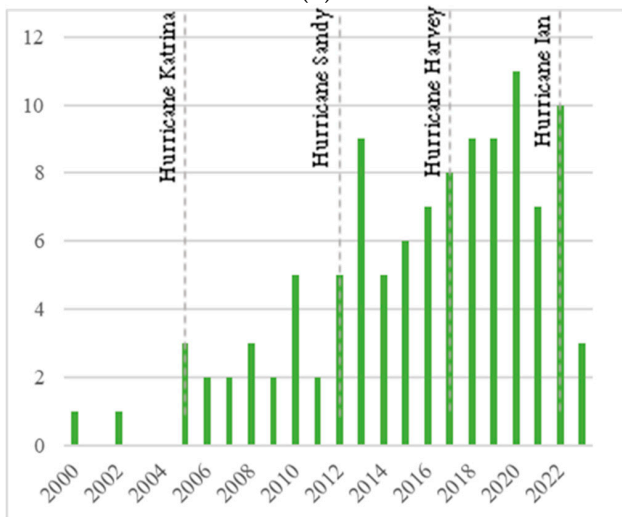
(a)



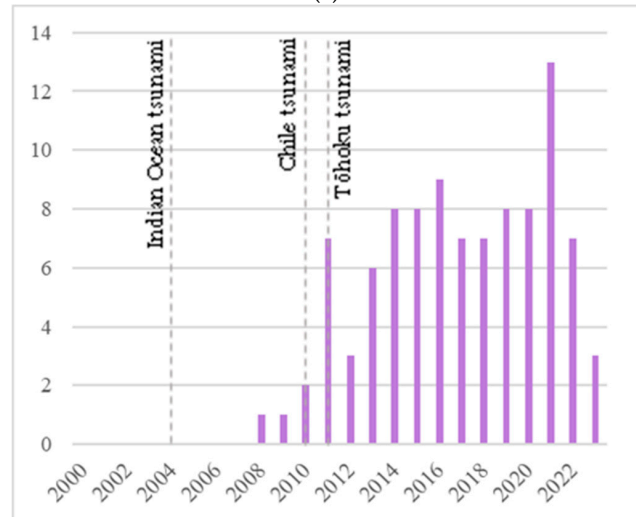
(b)



(c)



(d)



(e)

Figure 3. Publication trend of (a) all disaster damage estimation studies, (b) earthquake-based, (c) flood-based, (d) hurricane-based, and (e) tsunami-based disaster damage estimation studies. The X-axis denotes the years and the Y-axis denotes the number of publications.

The publication trend for hurricane-related studies is displayed in Figure 3d. Although it shows an increasing trend, the total number of studies was relatively low compared to the occurrence and damage caused by hurricanes. The engagement in hurricane-based studies was enabled by the 2005 Hurricane Katrina (USA) event. Moreover, there was a sudden rapid increase in studies after the 2012 Hurricane Sandy (USA) event, which continued due to the recurrent occurrence of hurricanes, such as the 2017 Hurricane Harvey (USA), the 2021 Hurricane Ida (USA), and the 2022 Hurricane Ian (USA). Nevertheless, the built environment remained severely affected by hurricane events. Figure 3e presents the publication trends of tsunami-based studies, revealing a generally increasing trend. The 2004 Indian Ocean tsunami initiated the tsunami-related studies, while the 2010 Chile tsunami and the 2011 Tōhoku tsunami events significantly influenced the study trend.

Overall, as the key observations, the publication trend of the disaster damage estimation studies showed a notable increase over the years. Similarly, the damage estimation studies related to the considered hazards also showed the same increasing trend. The increasing trend in studies related to disaster damage estimation for different disasters underscores the importance of identifying research trends and relationships among existing studies, as the subject area is rapidly expanding and future studies need to direct effectively through research gaps and opportunities. Also, the major disastrous events might have had an influence on the studies, as there were several increases in publications following the major disastrous events. Although the publication trend for each hazard showed an increasing trend, the number of publications focusing on floods, hurricanes, and tsunamis was minimal compared to the publications focusing on earthquakes, highlighting a research opportunity to utilise as per the requirements.

4.2. Geographical Variation of Studies

The geographical variations of the identified relevant studies are given in Figure 4. The figure was developed by the authors by utilising the findings of the geoparser. The relevant documents were then categorised based on the country. Also, the country-based disaster parameters, such as damage due to disasters and the number of recorded disaster events, were obtained from the EM-DAT database. Then, the choropleth maps were generated by indicating the number of publications per country by a circle in the location of the country and the disaster parameters per country by varying the colour intensity, which was filled with the map. Readers can comparatively understand the number of publications of a country with the disaster parameters by looking at the size of the circle and the colour intensity of the filled country. Descriptions and key observations are described in the following paragraphs for each subplot.

Figure 4a compares the number of studies with the damage due to the defined disaster events. The point circle presents the number of studies, and the circle size is proportional to the number. The country is shaded with colour, and the colour intensity represents the cumulative damage due to defined disaster events during the considered period (2000–2023) in the log scale. Note that the damage data were obtained from the EM-DAT database. The adjusted damage quantities due to the disaster events caused by either earthquake, flood, hurricane, or tsunami were used here. The USA and Mexico from North America; China, Japan, and India from Asia; Chile from South America; Germany, the United Kingdom, and Italy from Europe; and Australia and New Zealand from Oceania were the most economically affected countries during the time. Damage functions are the initial point of the damage estimation, which can be utilised to incorporate necessary disaster resilience measures and reduce the damage. Hence, damage estimation studies can be identified as a parameter that indicates the preparedness of a country for probable disasters. Most of them can be found in the USA; the European region, especially Italy and Turkey; and Asian countries, including China, Japan, and India. It was identified that most of these countries are among the list of the most economically damaged countries, which indicates their efforts in damage reduction.

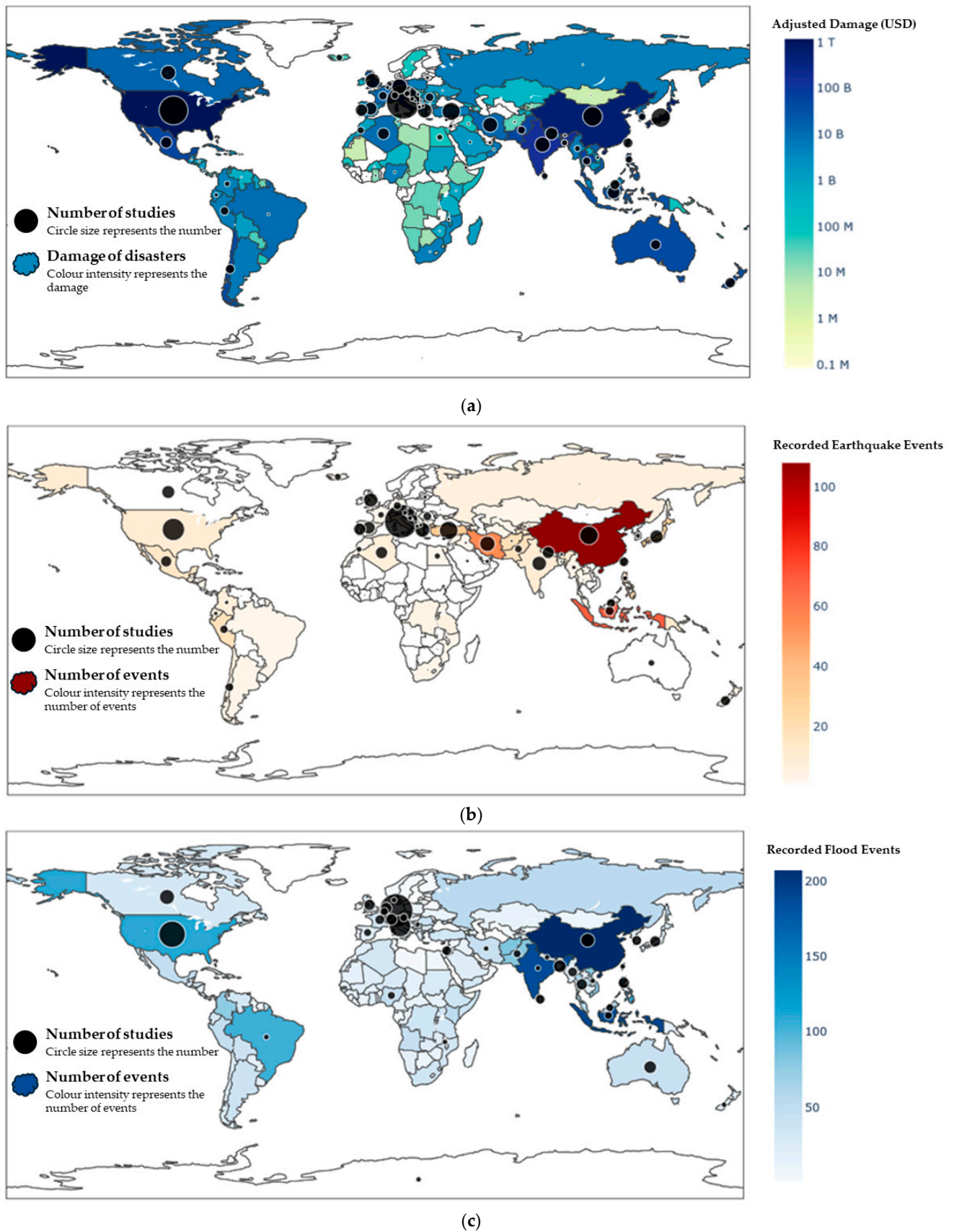


Figure 4. Cont.

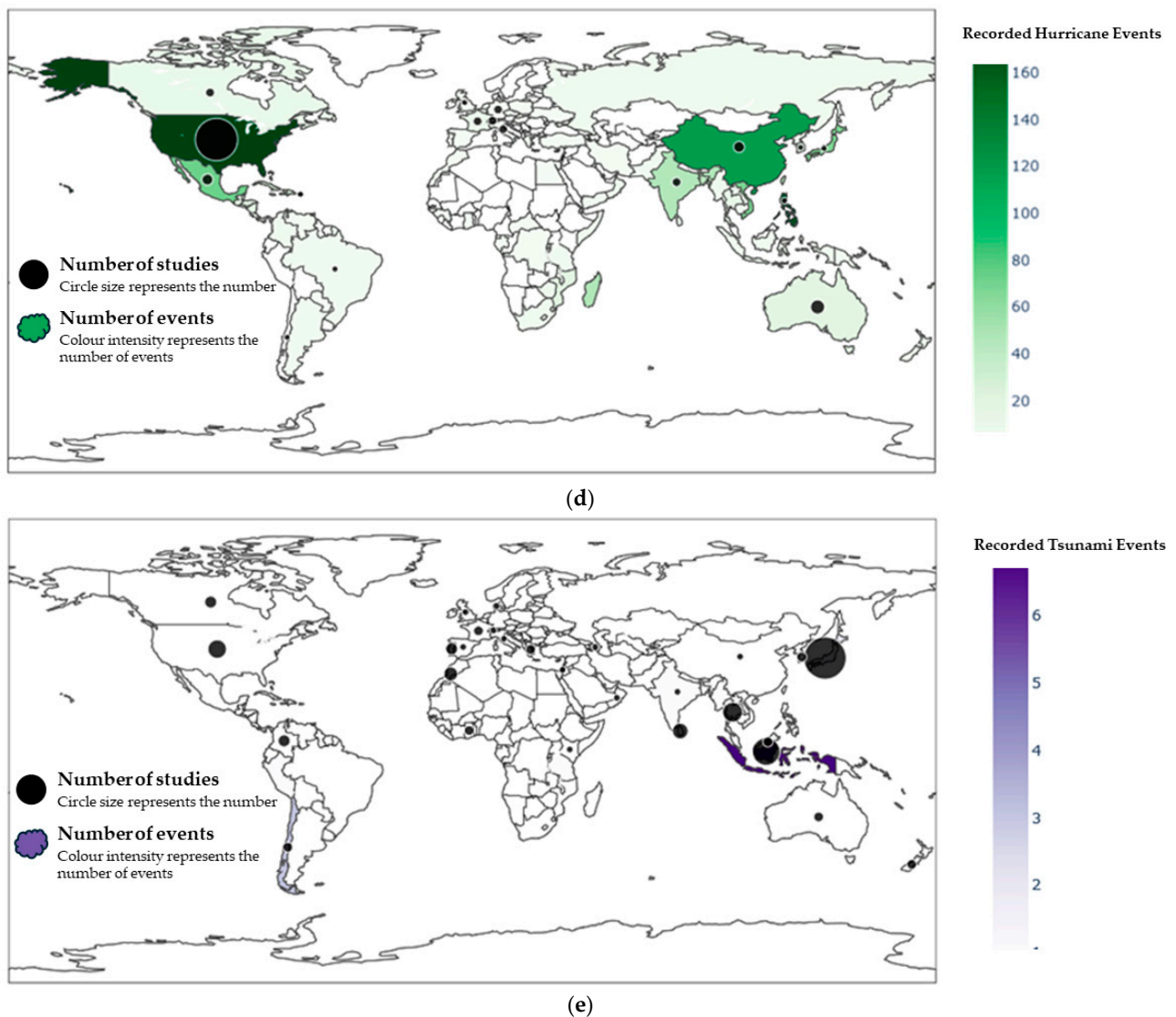


Figure 4. Geographical variation of the (a) disaster damage estimation studies vs adjusted damages, (b) earthquake-based studies vs recorded events, (c) flood-based studies vs recorded events, (d) hurricane-based studies vs recorded events, and (e) tsunami-based studies vs recorded events. Point circles represent the studies, and the circle size is proportional to the number. Colour represents the disaster parameter considered, and colour intensity is proportional to the value.

However, it can be identified that the studies related to damage estimations showed a gap despite the damage that happened in some countries. It can be observed that the studies focusing on the African and South American continents were minimal. Compared with the damage caused by the disaster events, studies can be more focused on Kenya, Zimbabwe, and South Africa from Africa and Venezuela, Brazil, and Argentina from South America. Also, the damage estimation-related studies can be expanded to minimise the damage that occurred due to disasters in severely affected countries, including Russia and Vietnam from Asia; the Czech Republic and Poland from Europe; Cuba, Haiti, and the Bahamas from the Caribbean region; El Salvador and Guatemala from Central America; and Samoa from the Oceania region. The above list of countries was identified based on the high amount of economic damage recorded and the low number of studies identified in the respective countries.

Furthermore, the geographical variation of the studies, specifically on the different disaster types, was assessed. Since different countries are affected by different kinds of disasters. Knowing the specific disaster type is important to identify the gaps and direct

future studies. For this comparison, the number of recorded disaster events was taken since the damage data were not recorded at each disaster event. The number of recorded disasters was chosen to represent the exposure of a country to different hazards.

Figure 4b shows the comparison between recorded earthquakes and the number of earthquake-focused studies. The point circle presents the number of studies, and the circle size is proportional to the number. The country is shaded with a colour, and the colour intensity represents the number of recorded earthquake events in the EM-DAT database between 2000 and 2023. China, Indonesia, Iran, Turkey, and Japan were the countries most exposed to earthquake events during the period. The most earthquake-aimed damage function-related studies were found in Italy, the USA, China, and Turkey. Comparing the difference in recorded earthquake events and the studies identified, the following countries are suggested to have more focus in the studies, as the identified countries were highly exposed to the earthquake events, and hence, there is a significant possibility of catastrophic events. The identified countries include Afghanistan, Tajikistan, Russia, and Kyrgyzstan from Asia; Papua New Guinea from the Oceania region; Guatemala and El Salvador from North America; and Tanzania from Africa.

Figure 4c compares recorded floods and the number of flood-related studies. Similar to the previous images, the colour intensity is proportional to the number of floods recorded, and the circle size is proportional to the number of flood-related studies during the 2000–2023 period. China, Indonesia, India, the USA, the Philippines, and Brazil are among the countries most exposed to floods during that period. However, most flood-related studies were conducted in Germany, the USA, and Italy. Even though flood-related studies were comparatively minimal, considering the exposure, flood-related studies can focus more on severely exposed countries, especially Asian countries, including Indonesia, India, Afghanistan, and Pakistan. Also, it can be identified that engagement in flood-related studies was very much minimal in the medium-exposure countries, including the South American, African, Middle Eastern, and Northern Asian regions.

Figure 4d illustrates the comparison between recorded hurricanes and the number of hurricane-focused damage function-related studies. During the period of 2000–2023, the USA, the Philippines, China, Mexico, Vietnam, and Japan were the countries most exposed to hurricane events. However, only the USA focused on these studies considerably. Considering the exposure severity and the high impacts, it is essential to consider more hurricane-related studies for other countries, especially the highly exposed countries mentioned above. Figure 4e presents a comparison between recorded tsunami events and the number of tsunami-related studies. The occurrence of tsunami events was relatively minimal during the considered period. However, there were a few highly impactful events as well. The Philippines is the country most exposed to tsunami events. Considering the severity of the impacts, the most tsunami-related studies were found in Japan, the Philippines, and Thailand. It is also noticeable that the highly impactful events influenced other non-exposed countries to focus on tsunami-related hazards as well.

Overall, these figures comparatively represent the geographical variation of publications with disaster parameters. Readers can use that comparative understanding to determine the requirements for future studies. For instance, if there is a country with relatively higher damage with a relatively lower number of studies, there is a gap for more studies. In general, it was identified that there is a considerable research gap on the African and South American continents. Further, the hazard-based maps can be used to identify the level of existing studies based on the hazard more specifically. The most affected countries, countries with the highest and lowest studies, and the countries that can focus more on further studies are discussed separately for each hazard in the above descriptions. Although a certain country may not be affected much by a certain disaster type, these maps can be used to get an idea about the variation in the existing literature, as there is a risk of changing the future hazard patterns due to climate change.

4.3. Prevalence of the Identified Subtopics and Hyperparameters

This section aims to identify the distribution of disaster damage estimation studies and the prevalence of studies on different subtopics. The main focus is on the hazards and the approaches used to develop damage functions, with the interpretation of results based on the supervised ML model. Figure 5 presents the prevalence of the considered subtopics in the form of bar plots. The figure was developed by the authors based on the results from multiclass classifiers and topic modelling. The multiclass classifiers were used to identify the hazard and development approach of each relevant document, and the number of documents per category was obtained there. Also, topic modelling was used to identify the relevant subtopics of documents, and then key subtopics were selected and a number per subtopic was obtained. Figure 5 was generated by combining the results.

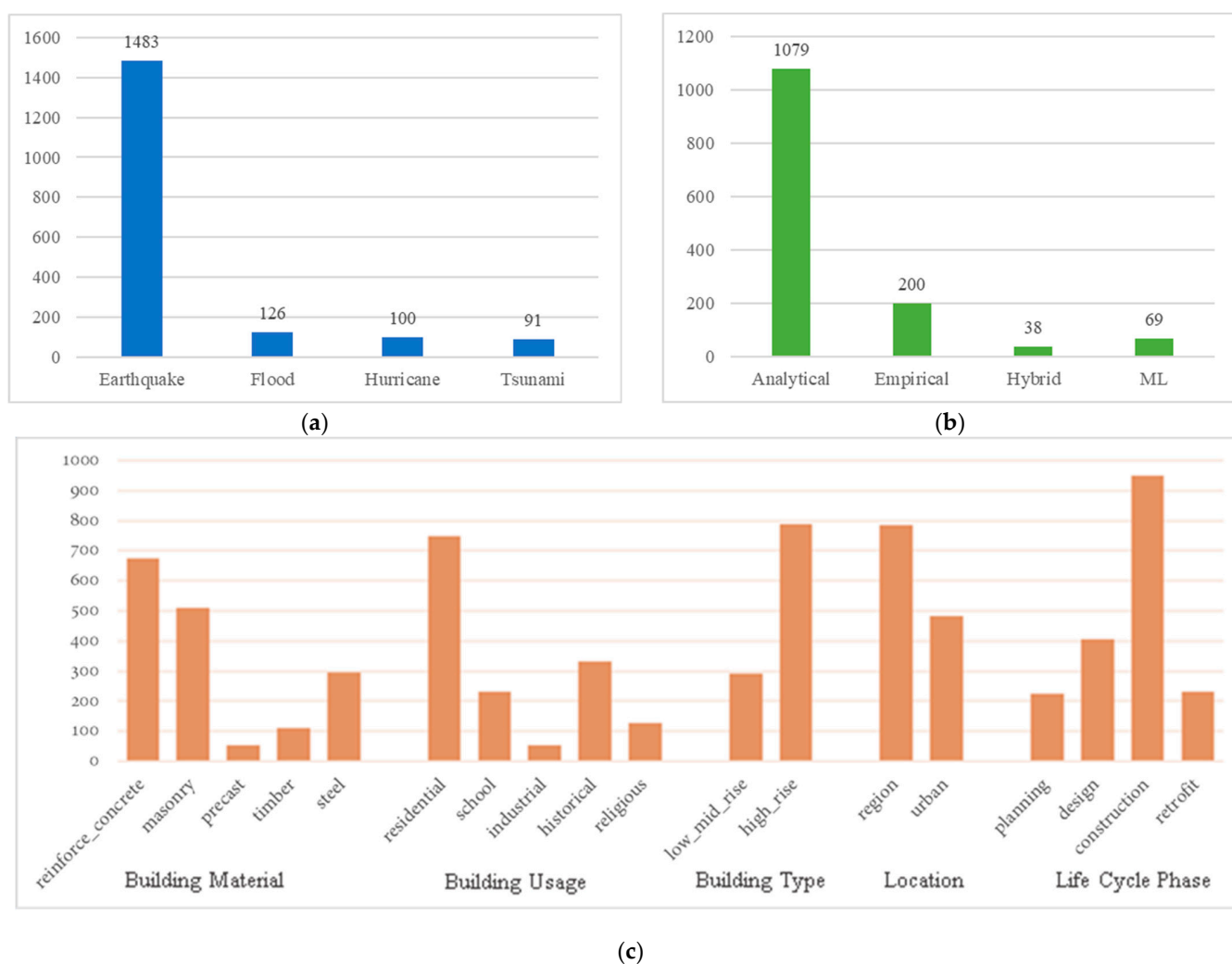


Figure 5. Prevalence of the topics and hyperparameters (a) based on a focused hazard, (b) based on the damage function development approach, and (c) based on common building parameters.

The distribution of focused hazards in damage function-related studies is presented in Figure 5a. The analysis revealed that most studies (68%) focused on earthquake hazards. Specifically, 1483 studies were identified in this subject area. Studies related to floods accounted for 6% (126 studies), while 5% (100 studies) were focused on hurricanes and 4% (91 studies) focused on tsunami hazards. The engagement of disaster damage estimation studies for floods, hurricanes, and tsunamis was relatively low.

Figure 5b illustrates the frequency of different approaches used to develop damage functions. Analytical approaches were employed in 49% (1079 studies) of the 2168 relevant

studies, indicating a higher prevalence than other approaches. Empirical approaches were utilised in 9% (200) of the studies. Hybrid and ML approaches were less commonly used, with 2% (38 studies) and 3% (69 studies), respectively.

Researchers have explored the relationship between damage functions and various building properties to develop more specific and reliable functions. Figure 5c presents the prevalence of these identified hyperparameters. Regarding building materials, reinforced concrete and masonry were the most considered materials, with 675 and 510 related articles, respectively. Steel (296 studies), timber (112 studies), and precast concrete (53 studies) were also included in the damage function studies. The studies also focused on different types of buildings based on their usage. The majority of studies (749) were focused on residential structures, while there was some engagement with school buildings (232) and industrial buildings (53) as well. Historical buildings (333) and religious buildings (128) were also considered in damage assessments due to their unique vulnerabilities. Building types were categorised based on the height of structures, and the results indicate that high-rise buildings received the most attention (789 studies), followed by low-rise and mid-rise buildings combined (291 studies).

Additionally, the geographical scope of the studies was taken into account. The model identified that 785 studies were focused on specific regions, while 482 studies were conducted in urban areas. The damage assessment studies also covered different life cycle phases of the building. The construction phase received the highest number of discussions (949 studies), followed by studies related to planning (226), design (407), and retrofitting (232). The distribution and prevalence of disaster damage estimation studies provide valuable insights into the research landscape, highlighting the prominent hazards, approaches, and hyperparameters considered.

In summary, most of the damage estimation studies focused on earthquakes, while there was a comparatively minimal focus on floods, hurricanes, and tsunamis. Also, the analytical approach was more commonly used in developing damage functions, which are used for damage estimations, while the integration of other approaches, such as empirical, ML, and hybrid, was comparatively minimal. Considering other building parameters, it can be observed that reinforced concrete, masonry, and steel structures received more focus than precast and timber structures. Also, residential and historical structure-focused studies were more prevalent than the other structure types, such as educational, industrial, and religious. These figures comprehensively summarise the status of the current literature of the subject area, which could be helpful to future research. Although research design depends on multiple factors, insights from this study will assist researchers with a comprehensive summary of the existing literature.

4.4. Visualisation of Topic Hyperparameters

The keyword map in Figure 6 provides insights into the distribution and relationships of relevant documents. The figure was generated by the authors based on the results of the topic modelling. As described, the topic modelling algorithms identified subtopics and relevant hyperparameters in the subject area based on the commonly appearing word clusters. Then, a relevance value for each document to each hyperparameter was assigned. This generated a high-dimensional dataset where each document was described by a relevance value for each hyperparameter. The t-distributed stochastic neighbour embedding (t-SNE) technique was used to convert the high-dimensional dataset to a 2D data set. The technique enables similar, related documents to appear closer, whereas dissimilar, non-related documents appear farther apart in a 2D plot. Then, density-based spatial clustering of applications with the noise (DBSCAN) method was used to identify the clusters in the 2D plot. Afterwards, the major hyperparameters that could be found in each cluster were annotated. Since a point on this plot represents a document, readers can identify the groups of more similar studies and gain an overview of the document groups. The documents were visualised based on the hazard identified by the multiclass classifier to give more meaning to the plot.

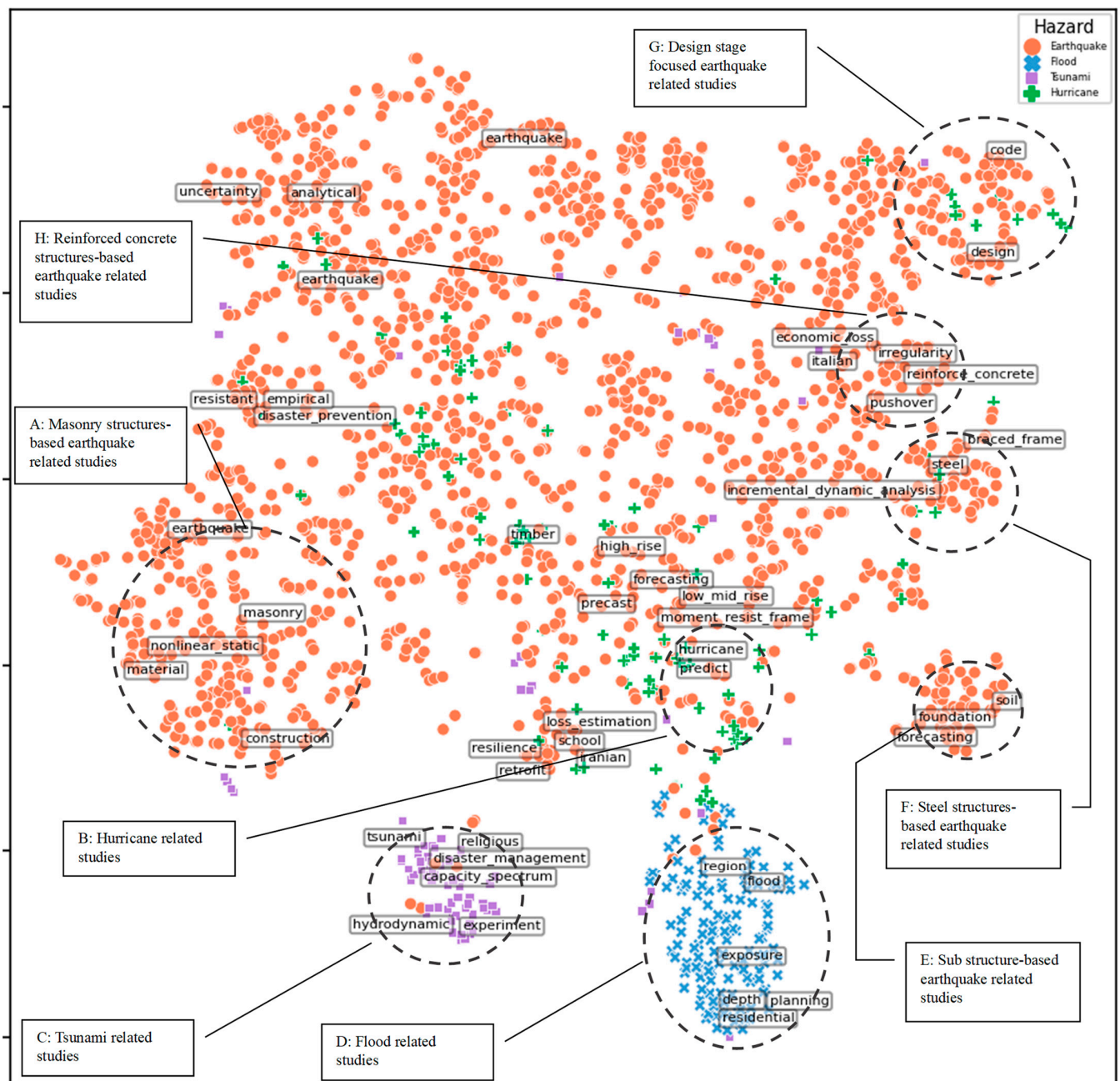


Figure 6. Keyword map of the relevant studies.

Due to most relevant documents being related to earthquakes, data points representing these documents are spread across the plot, resulting in distinct clusters with different relationships. Clusters A, E, F, G, and H exhibit trends related to earthquake-based studies. Cluster A primarily focuses on masonry structures in earthquake-related studies. Nonlinear static methods are commonly used to develop damage functions for masonry structures.

Cluster E represents earthquake-based studies focusing on soil–structure interaction and foundation parameters for damage forecasting. For Cluster F, incremental dynamic analysis methods are frequently employed when developing damage functions for steel structures. Cluster H highlights using the pushover method in developing earthquake-based damage functions for reinforced concrete structures. Cluster G identifies studies that incorporate design principles and design codes.

Cluster B highlights studies focused on hurricane hazards, specifically on damage prediction of residential structures. Cluster C represents damage assessment studies related

to tsunami hazards. Parameters in this cluster indicate the common utilisation of capacity spectrums, hydrodynamic analyses, and experimental tests in developing tsunami-based damage functions. Cluster D presents studies on flood-based damage functions, with parameters indicating the incorporation of flood depth and a primary focus on flood planning for specific regions.

The keyword map analysis provides a comprehensive overview of the distribution and relationships among damage assessment studies, with different clusters representing various hazards and approaches. As the key takeaways, the figure identifies multiple groups that earthquake studies focused on, including different building materials, such as reinforced concrete, masonry, and steel, with a few other groups, such as design-based and substructure-based studies. Also, the figure identifies the relationship of specific analytical techniques that are commonly used in damage function development with different building types, such as non-linear static methods for masonry structures, incremental dynamic analysis for steel structures, and the pushover method for reinforced concrete buildings. Further, the figure highlights other common hyperparameters related to flood and tsunami-based clusters. These findings aid in understanding the research landscape and can guide further exploration and development of damage functions in specific hazard contexts.

4.5. Analysis of Co-Occurrence of Topics and Hyperparameters

The co-occurrence of various topics and hyperparameters was studied using heat maps to identify frequently collaborated elements. The heatmaps were generated by the authors based on the results of multiclass classifiers and topic modelling. After the analysis, each document was marked with the related hazard, development approach, and other relevant hyperparameters. The heatmaps were developed based on the number of documents that collaborated on different parameters in the study. The analysis focused on the relationships between the main subject areas, which included the hazards of the studies and the approaches used to develop damage functions. Additionally, collaborations between these subject areas and other building parameters were assessed.

Figure 7a presents the co-occurrence of topics related to approaches and hazards. It is evident that most damage function-related studies concentrated on earthquakes, with analytical approaches being the most prevalent method used (39% of all relevant studies). Empirical approaches were also utilised in developing damage functions for earthquakes (5%) and floods (3%). Analytical approaches dominated for hurricanes (3%) and tsunamis (2%), while hybrid and ML approaches had limited usage. ML approaches showed some engagement in earthquake-related studies (2%).

Figure 7b explores the collaboration between hazards and building parameters. Earthquake-related studies demonstrated a more comprehensive range of subtopics and engagements with building parameters compared to other hazards. Regarding building materials, earthquake-related studies had a strong focus on reinforced concrete structures (26%), with significant collaborations observed for masonry structures (19%) and steel structures (11%). In terms of building usage, residential structures (15%) received the most attention in earthquake-related studies, with notable engagements in historical structures (10%) and school buildings (7%). High-rise buildings (27%) were a significant focus. These studies primarily addressed the general construction phase (28%) but also considered the design phase (15%) and retrofitting stage (7%). Flood-related studies primarily focused on residential structures (6%) and emphasised location-specific planning within regions and urban areas (6%). Hurricane-related studies collaborated with residential structures (4%), regional studies (3%), and the construction phase (3%). Tsunami-related studies demonstrated significant collaboration with religious structures (4%) and regional studies (3%).

Figure 7c displays the co-occurrence of different approaches to develop damage functions and building parameters. Analytical approach studies showed a higher level of engagement with building parameters, particularly reinforced concrete structures (22%), masonry (12%), and steel (9%) structures. Residential structures (13%) received significant attention in building usage, while high-rise buildings (18%) were the primary focus in

building type. Analytical approach studies considered the design phase (13%) and construction phase (17%) key stages. Empirical approach studies exhibited a regional focus (6%), while hybrid and ML approaches had limited collaborations with the identified building parameters.

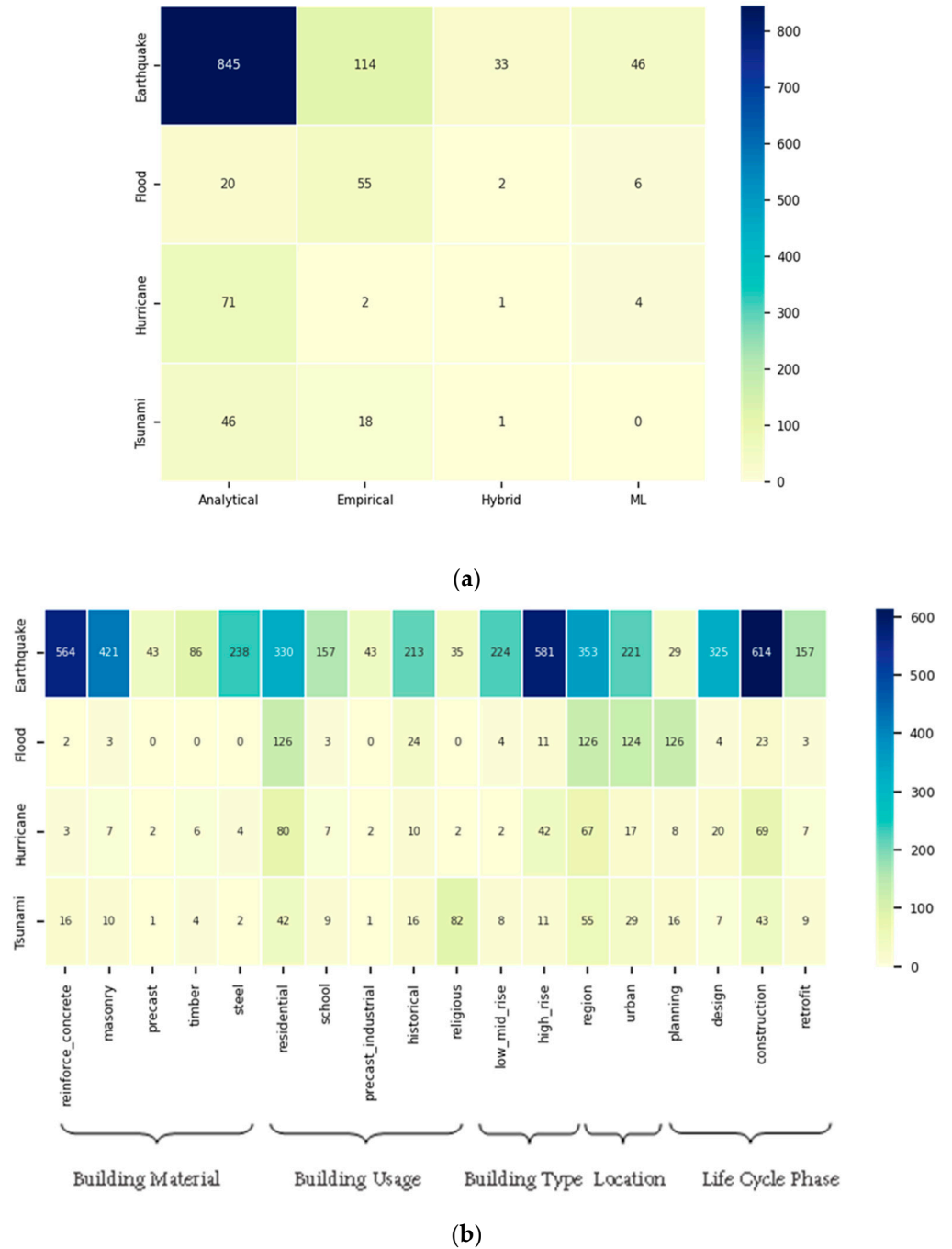


Figure 7. Cont.

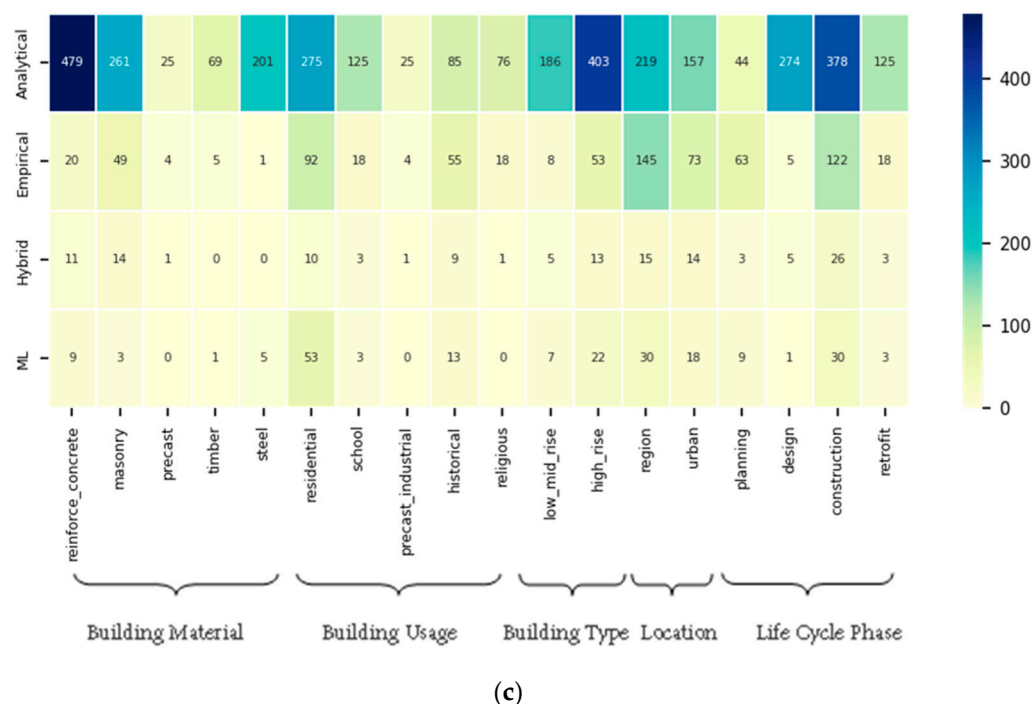


Figure 7. Co-occurrence of topic and hyperparameters. (a) Hazards with a damage function development approach; (b) hazard with common building parameters; (c) damage function development approach with common building parameters.

The co-occurrence analysis provides valuable insights into the relationships and collaborations between different subject areas, approaches, and building parameters within the damage function-related studies. As the key takeaways, most of the relevant documents were earthquake-based documents that followed an analytical approach. Also, only earthquake-related studies made a considerable contribution to other building parameters. Other notable integrations are discussed in the above descriptions. These figures comprehensively summarise the collaborative efforts in the existing literature. The insights of this study can direct future research design because not only do these figures highlight the gaps in collaboration, but they also indicate key trends. Understanding these associations enhances our understanding of research trends and can inform future investigations in specific areas of interest.

5. Discussion

This study aimed at disaster damage estimation studies on buildings and identifying relations among the subtopics discussed in the literature. Supervised and unsupervised ML algorithms were facilitated in the analysis. The current study successfully demonstrated the effectiveness of integrating ML algorithms into the mapping of a large pool of documents. The integration of ML models assisted with the selection of relevant documents in a pool of 6061 documents by only referring to 439 records manually, with an accuracy level of 85.7%, which is acceptable [46,50,56]. Further, the ML models were able to categorise 2186 relevant documents into major categories (hazard, development approach) by only referring to 245 records manually. Additionally, the ML models were able to capture the location of all the applicable studies and identify the related subtopics without any human interference. Therefore, it is evident that the integration of the ML model was able to assist the filtration, categorisation, and data mining process by reducing the human effort considerably. The combined process provided valuable insights into the status of the damage estimation studies, which will assist researchers in multiple ways.

Based on the publication trend, it is evident that the focus of researchers tended towards disaster damage estimation studies of different disasters over the past few years.

The influence of the major disaster events that occurred during the considered period in the publication trend was also assessed. It was observed that there were increases in published studies after significant events, but notably, correlation does not imply causation. Other factors, such as increased awareness, funding opportunities, and advancements in research methodologies, could also have contributed to the observed trends.

Considering the geographical variation, it was identified that the developed countries were the countries most affected economically due to disaster events over the years. This is because the damage that occurred due to the same intense event can be higher in developed countries than in other countries due to the higher property value and repair costs. Also, the properties in developed countries are better insured, which increases the nominal damage compared to less developed countries. Further, it was observed that most of the relevant studies were concentrated in developed countries. It can be highlighted that even though the damage due to disastrous events and exposure was higher, the intended studies were minimal, especially in less developed countries and developing countries in general.

The article analysis was carried out considering different topic areas, such as the different hazards aimed at, the approach followed to develop damage functions, the building materials, and the different building parameters, such as usage, type, location, and applicable life cycle phase of the study. Based on the results and considering the prevalence of the identified topics, the most relevant studies focused on earthquakes, and the analytical methods were the most followed approach. The same trend was also identified in another study by the authors [61].

Earthquake-induced disasters were identified as one of the deadliest, most frequent, and most damaging disaster types [62]. This trend highlights the effort of the researchers to minimise the consequences of such disasters. Damage assessment studies showed some engagement with flood-related disasters, but more studies can focus on floods, as they were the most common disaster type during the past decade. The low engagement with tsunamis and hurricanes can be identified as a clear research gap, because hurricanes were responsible for the most significant economic damage caused by disasters during the past decade [62]. Even though tsunamis are not very frequent, they can be a hazard with severe effects. Hence, future studies can focus on tsunamis and hurricanes.

Analytical approaches were the most used method to develop fragility curves, followed by empirical approaches. Different numerical methods, simulations, and experimental methods were used to develop analytical damage functions, whereas empirical damage functions were developed based on the observed data [63]. The analytical approach enabled the generation of required data that fulfilled the gaps, such as the absence of actual damage data from disasters [64] and the requirement of damage data from higher-magnitude incidents than those observed. The appearance of analytical approaches in most of the related documents reflects the attendance of researchers in such studies. The results showed minimal usage of hybrid and ML methods. Integrating two or more approaches (hybrid) can overcome some data shortage issues and also can be used to verify the results. ML approaches enable the derivation of multivariable damage functions, which is a unique advantage over other approaches [34]. Also, ML approaches can be utilised to reduce uncertainties and irregularities associated with the predictions [65,66]. Therefore, researchers can focus on incorporating hybrid and ML approaches in their future studies.

Considering the different building materials, the results identified a comparatively minor engagement with timber and precast structures. Future studies can focus on the topics mentioned above accordingly. Considering building usage, the topic model only identified the categories of residential, school, industrial, historical, and religious. This demonstrates that the other related types, such as different institutional and commercial structures, can be incorporated into these studies. From the identified categories, industrial and religious structures showed minimal engagement. Considering the applicable life cycle phases, most studies focused on the general construction stage, and some focused on the planning, design, and retrofit stages.

Based on the heatmap results, it was identified that most of the studies related to earthquakes and floods concentrated on analytical and empirical approaches (relative to the identified studies in each category). Hence, as mentioned, there is an opportunity to direct future earthquake- and flood-related studies with hybrid and ML approaches to achieve possible benefits. Considering the hazards, only earthquake-related studies collaborated significantly with other building parameters. Therefore, future studies that aim to study other hazards can focus on different building parameters as applicable to give more specific and accurate assessments. Further, only analytical approach-related studies showed a significant integration with the other building parameters identified, and empirical approach-related studies showed a minor integration. Hence, future studies that use hybrid and ML approaches focused on integrating the study with different building parameters as applicable.

The results of this study aim to provide valuable insights for researchers and professionals working in the field of disaster resilience related to buildings. The study focuses on the field of disaster resilience concerning buildings, an essential aspect of urban planning and disaster management. The study aims to provide a comprehensive understanding of the current state of research in this field by analysing the relationships, trends, and gaps in existing studies. The results of this study comprehensively summarise existing literature in different means, such as geographical variation, categorical prevalence and variation, and status of collaborative studies. It highlights where the current research studies are rich and where the current research is lacking. The insights will greatly help researchers, allowing them to make informed decisions about their future research scopes and methodologies. Furthermore, the findings of this study can also be applied by urban planners, built environment professionals, and civil and structural engineers in their work, helping them to correctly identify the appropriate methods for developing damage functions that meet their specific requirements. This study will provide valuable insights and contribute to the development of future research in the field of disaster resilience in buildings.

Limitations of the Study

The following points can be mentioned as limitations of the current study. In this study, two Python libraries (Scikit-learn package, Gensim) were used to develop the ML algorithms because of the ease of use, effective implementation, and good documentation. However, current research focuses on incorporating more advanced ML applications, such as transformers, into the same purposes, which allows the model to capture contextual information from both preceding and succeeding words, leading to a better understanding and representation of language that can be utilised to generate more accurate predictions efficiently. A trained geoparser was used to extract the geographic locations of the study. The process succeeds only if the location name is presented in the text. Therefore, the model did not capture the studies done in different countries, but the location was not stated. Also, in this study, supervised ML algorithms were only used to categorise the study based on the hazards, and the approach used to develop damage functions and other building parameters was carried out using unsupervised ML algorithms, which can be less accurate. Hence, further research can be conducted focusing on other categories of interest.

6. Conclusions

In this study, literature mapping was conducted with the aim of disaster damage estimation studies on buildings. Supervised and unsupervised ML algorithms were leveraged to overcome the limitation of handling many articles manually. The key observations and takeaways of this study can be summarized as follows:

- This study successfully demonstrates the effectiveness of integrating ML algorithms into mapping a large number of articles.
- The results of the study show that there is an increasing trend in disaster damage estimation-related studies on buildings.

- Most of the studies related to disaster damage estimation are in severely economically affected countries and regions like the USA, the European region (notably, Italy and Turkey), and Asian countries (China, Japan, and India).
- Most of the disaster damage estimation studies focus on earthquakes, while analytical approaches are the most prominent method followed in developing damage functions.
- Considering the integration of disaster damage estimation studies with different building parameters, most studies focus on reinforced concrete, masonry, and steel buildings in terms of material, and residential and historical buildings in terms of usage.
- Earthquake-based studies show wide collaboration with other building parameters.

The key research gaps and opportunities can be listed as follows:

- A notable gap exists in the studies focusing on the African and South American continents despite the economic damage incurred by disasters in these regions.
- The damage estimation studies focusing on floods, hurricanes, and tsunamis are minimal compared to the studies focusing on earthquakes.
- Integration of ML and hybrid methods in the development of damage functions is minimal at the moment, but researchers can consider integrating these approaches, as they offer different advantages that conventional methods do not.
- Current studies focusing on timber and precast building materials, as well as industrial, educational, religious, and commercial buildings, are minimal. Researchers can consider these areas based on the requirements.
- Studies on floods, hurricanes, and tsunamis show minimal collaboration with other building parameters. Researchers can focus on more specific studies in these areas.

This research provides a comprehensive overview of current trends, relationships, and research gaps, offering valuable insights to guide future research studies effectively in the subject area of disaster damage estimation on buildings.

Author Contributions: Conceptualisation, D.R. and C.S.; methodology, D.R.; software, D.R.; validation, C.S., R.R., and T.M.; formal analysis, D.R.; investigation, D.R.; data curation, D.R.; writing—original draft preparation, D.R.; writing—review and editing, C.S., R.R., T.M., S.S., L.R., and S.D.S.; visualisation, D.R.; supervision, C.S., R.R., T.M., S.S., L.R., and S.D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The first author expresses sincere gratitude to the Split PhD program, which is jointly offered by RMIT University, Australia, and the University of Moratuwa, Sri Lanka, for providing an invaluable opportunity to pursue doctoral studies.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Newkirk, R.T. The Increasing Cost of Disasters in Developed Countries: A Challenge to Local Planning and Government. *J. Contingencies Crisis Manag.* **2001**, *9*, 159–170. [[CrossRef](#)]
2. Yabe, T.; Jones, N.K.W.; Rao, P.S.C.; Gonzalez, M.C.; Ukkusuri, S.V. Mobile phone location data for disasters: A review from natural hazards and epidemics. *Comput. Environ. Urban Syst.* **2022**, *94*, 101777. [[CrossRef](#)]
3. CRED. *2023 Disasters in Numbers*; CRED: Brussels, Belgium, 2024.
4. van Aalst, M.K. The impacts of climate change on the risk of natural disasters. *Disasters* **2006**, *30*, 5–18. [[CrossRef](#)] [[PubMed](#)]
5. Stott, P. How climate change affects extreme weather events. *Science* **2016**, *352*, 1517–1518. [[CrossRef](#)]
6. Cavalagli, N.; Kita, A.; Castaldo, V.L.; Pisello, A.L.; Ubertini, F. Hierarchical environmental risk mapping of material degradation in historic masonry buildings: An integrated approach considering climate change and structural damage. *Constr. Build. Mater.* **2019**, *215*, 998–1014. [[CrossRef](#)]
7. de Wilde, P.; Coley, D. The implications of a changing climate for buildings. *Build. Environ.* **2012**, *55*, 1–7. [[CrossRef](#)]
8. Huang, M.; Zhang, J.; Hu, J.; Ye, Z.; Deng, Z.; Wan, N. Nonlinear modeling of temperature-induced bearing displacement of long-span single-pier rigid frame bridge based on DCNN-LSTM. *Case Stud. Therm. Eng.* **2024**, *53*, 103897. [[CrossRef](#)]
9. CRED. *2021 Disasters in Numbers*; CRED: Brussels, Belgium, 2022.

10. Leelawat, N.; Suppasri, A.; Charvet, I.; Imamura, F. Building damage from the 2011 Great East Japan tsunami: Quantitative assessment of influential factors: A new perspective on building damage analysis. *Nat. Hazards* **2014**, *73*, 449–471.
11. Zhang, J.; Huang, M.; Wan, N.; Deng, Z.; He, Z.; Luo, J. Missing measurement data recovery methods in structural health monitoring: The state, challenges and case study. *Measurement* **2024**, *231*, 114528. [[CrossRef](#)]
12. Peng, L.; Stewart, M.G. Spatial time-dependent reliability analysis of corrosion damage to RC structures with climate change. *Mag. Concr. Res.* **2015**, *66*, 1154–1169. [[CrossRef](#)]
13. Johns, J.; Fedeski, M. Adapting Building Construction to the Effects of Climate Change. In *Detecting and Modelling Regional Climate Change*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 605–616. [[CrossRef](#)]
14. Hedayatnia, H.; Van Den Bossche, N. The Impact of Climate Change on Material Degradation: Finding a Feasible Approach for Climate Model Evaluation. In Proceedings of the 15th Edition of the International Conference on Durability of Building Materials and Components, Barcelona, Catalonia, 20–23 October 2020; pp. 325–332.
15. Ladipo, O.; Reichard, G.; McCoy, A.; Pearce, A.; Knox, P.; Flint, M. Attributes and metrics for comparative quantification of disaster resilience across diverse performance mandates and standards of building. *J. Build. Eng.* **2019**, *21*, 446–454. [[CrossRef](#)]
16. Deng, Z.; Huang, M.; Wan, N.; Zhang, J. The Current Development of Structural Health Monitoring for Bridges: A Review. *Buildings* **2023**, *13*, 1360. [[CrossRef](#)]
17. Huang, M.; Ling, Z.; Sun, C.; Lei, Y.; Xiang, C.; Wan, Z.; Gu, J.; Huang, M.; Ling, Z.; Sun, C.; et al. Two-stage damage identification for bridge bearings based on sailfish optimization and element relative modal strain energy. *Struct. Eng. Mech.* **2023**, *86*, 730. [[CrossRef](#)]
18. Zheng, H.; Deng, Z.; Guo, L.; Liu, J.; Liu, L.; Li, T.; Zheng, H.; Zheng, T. Assessment of Building Physical Vulnerability in Earthquake-Debris Flow Disaster Chain. *Int. J. Disaster Risk Sci.* **2023**, *14*, 666–679. [[CrossRef](#)]
19. de Bruijn, J.A.; Daniell, J.E.; Pomonis, A.; Gunasekera, R.; Macabuag, J.; de Ruiter, M.C.; Koopman, S.J.; Bloemendaal, N.; de Moel, H.; Aerts, J.C.J.H. Using rapid damage observations for Bayesian updating of hurricane vulnerability functions: A case study of Hurricane Dorian using social media. *Int. J. Disaster Risk Reduct.* **2022**, *72*, 102839. [[CrossRef](#)]
20. Du, Y.; Ding, Y.; Li, Z.; Cao, G. The role of hazard vulnerability assessments in disaster preparedness and prevention in China. *Mil. Med. Res.* **2015**, *2*, 1–7. [[CrossRef](#)] [[PubMed](#)]
21. Kannangara, K.K.C.L.; Adikariwattage, V.V.; Siriwardana, C.S.A. Development of a Cost-Optimized Model for Evacuation Route Planning for Tsunamis. In Proceedings of the MERCon 2022—Moratuwa Engineering Research Conference, Moratuwa, Sri Lanka, 27–29 July 2022; pp. 1–6.
22. Rathnayaka, B.; Siriwardana, C.; Robert, D.; Amaratunga, D.; Setunge, S. Improving the resilience of critical infrastructures: Evidence-based insights from a systematic literature review. *Int. J. Disaster Risk Reduct.* **2022**, *78*, 103123. [[CrossRef](#)]
23. Noshadravan, A.; Miller, T.R.; Gregory, J.G. A Lifecycle Cost Analysis of Residential Buildings Including Natural Hazard Risk. *J. Constr. Eng. Manag.* **2017**, *143*, 04017017. [[CrossRef](#)]
24. Nakajima, M.; Hirata, K. Seismic Risk Evaluation Method of Structures Based on Deaggregation of Seismic Hazard. In Proceedings of the 13th World Conference on Earthquake Engineering, Vancouver, BC, Canada, 1–6 August 2004; pp. 1–11.
25. Loaiza, M.A.D.; Bricker, J.D.; Meynadier, R.; Duong, T.M.; Ranasinghe, R.; Jonkman, S.N. Development of damage curves for buildings near la Rochelle during storm Xynthia based on insurance claims and hydrodynamic simulations. *Nat. Hazards Earth Syst. Sci.* **2022**, *22*, 345–360. [[CrossRef](#)]
26. Nofal, O.M.; Van De Lindt, J.W. Probabilistic Flood Loss Assessment at the Community Scale: Case Study of 2016 Flooding in Lumberton, North Carolina. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **2020**, *6*, 05020001. [[CrossRef](#)]
27. Enghardt, J.; De Moel, H.; Huyck, C.K.; De Ruiter, M.C.; Aerts, J.C.J.H.; Ward, P.J. Enhancement of large-scale flood risk assessments using building-material-based vulnerability curves for an object-based approach in urban and rural areas. *Nat. Hazards Earth Syst. Sci.* **2019**, *19*, 1703–1722. [[CrossRef](#)]
28. Liu, S.; Zheng, W.; Zhou, Z.; Zhong, G.; Zhen, Y.; Shi, Z. Flood Risk Assessment of Buildings Based on Vulnerability Curve: A Case Study in Anji County. *Water* **2022**, *14*, 3572. [[CrossRef](#)]
29. Magapu, S.; Setia, S. Fragility curves for assessment of seismic vulnerability of buildings on slopes. *Soil Dyn. Earthq. Eng.* **2023**, *173*, 108069. [[CrossRef](#)]
30. Papatoma-Köhle, M.; Schlögl, M.; Dösser, L.; Roesch, F.; Borga, M.; Erlicher, M.; Keiler, M.; Fuchs, S. Physical vulnerability to dynamic flooding: Vulnerability curves and vulnerability indices. *J. Hydrol.* **2022**, *607*, 127501. [[CrossRef](#)]
31. Yum, S.G.; Kim, J.M.; Wei, H.H. Development of vulnerability curves of buildings to windstorms using insurance data: An empirical study in South Korea. *J. Build. Eng.* **2021**, *34*, 101932. [[CrossRef](#)]
32. Charvet, I.; Macabuag, J.; Rossetto, T. Estimating tsunami-induced building damage through fragility functions: Critical review and research needs. *Front. Built Environ.* **2017**, *3*, 267533.
33. Maio, R.; Tsionis, G.; Sousa, M.L.; Dimova, S.L. Review of fragility curves for seismic risk assessment of buildings in Europe. In Proceedings of the 16th World Conference on Earthquake—WCEE, Santiago, Chile, 9–13 January 2017.
34. Amadio, M.; Rita Scorzini, A.; Carisi, F.; Essenfelder, H.A.; Domeneghetti, A.; Mysiak, J.; Castellarin, A. Testing empirical and synthetic flood damage models: The case of Italy. *Nat. Hazards Earth Syst. Sci.* **2019**, *19*, 661–678. [[CrossRef](#)]
35. Lee, J.S.; Song, C.H.; Pradhan, A.M.S.; Ha, Y.S.; Kim, Y.T. Development of structural type-based physical vulnerability curves to debris flow using numerical analysis and regression model. *Int. J. Disaster Risk Reduct.* **2024**, *106*, 104431. [[CrossRef](#)]

36. Galasso, C.; Pregnotato, M.; Parisi, F. A model taxonomy for flood fragility and vulnerability assessment of buildings. *Int. J. Disaster Risk Reduct.* **2021**, *53*, 101985. [[CrossRef](#)]
37. Saruddin, S.N.A.; Nazri, F.M. Fragility Curves for Low- and Mid-rise Buildings in Malaysia. *Procedia Eng.* **2015**, *125*, 873–878. [[CrossRef](#)]
38. Mastroberti, M.; Vona, M. A critical review of fragility curves for existing RC buildings. In Proceedings of the ECCOMAS Congress 2016—Proceedings of the 7th European Congress on Computational Methods in Applied Sciences and Engineering, Crete, Greece, 5–10 June 2016; pp. 5934–5941.
39. Pregnotato, M.; Galasso, C.; Parisi, F. A compendium of existing vulnerability and fragility relationships for flood: Preliminary results. In Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP 2015, Vancouver, BC, Canada, 13–15 July 2015.
40. Malgwi, M.B.; Fuchs, S.; Keiler, M. A generic physical vulnerability model for floods: Review and concept for data-scarce regions. *Nat. Hazards Earth Syst. Sci.* **2020**, *20*, 2067–2090. [[CrossRef](#)]
41. Jongman, B.; Kreibich, H.; Apel, H.; Barredo, J.I.; Bates, P.D.; Feyen, L.; Gericke, A.; Neal, J.; Aerts, J.C.J.H.; Ward, P.J. Comparative flood damage model assessment: Towards a European approach. *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 3733–3752. [[CrossRef](#)]
42. Gerl, T.; Kreibich, H.; Franco, G.; Marechal, D.; Schröter, K. A review of flood loss models as basis for harmonization and benchmarking. *PLoS ONE* **2016**, *11*, e0159791. [[CrossRef](#)] [[PubMed](#)]
43. Priya, B.; Nandhini, J.M.; Gnanasekaran, T. An analysis of the applications of natural language processing in various sectors. *Adv. Parallel Comput.* **2021**, *38*, 598–602. [[CrossRef](#)]
44. Jiang, K.; Lu, X. Natural Language Processing and Its Applications in Machine Translation: A Diachronic Review. In Proceedings of the IEEE 3rd International Conference of Safe Production and Informatization, IICSPI 2020, Chongqing, China, 28–30 November 2020; pp. 210–214.
45. Lai, Y.; Kontokosta, C.E. Topic modeling to discover the thematic structure and spatial-temporal patterns of building renovation and adaptive reuse in cities. *Comput. Environ. Urban Syst.* **2019**, *78*, 101383. [[CrossRef](#)]
46. Berrang-Ford, L.; Sietsma, A.J.; Callaghan, M.; Minx, J.C.; Scheelbeek, P.F.D.; Haddaway, N.R.; Haines, A.; Dangour, A.D. Systematic mapping of global research on climate and health: A machine learning review. *Lancet Planet. Health* **2021**, *5*, e514–e525. [[CrossRef](#)] [[PubMed](#)]
47. Callaghan, M.; Schleussner, C.F.; Nath, S.; Lejeune, Q.; Knutson, T.R.; Reichstein, M.; Hansen, G.; Theokritoff, E.; Andrijevic, M.; Brecha, R.J.; et al. Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. *Nat. Clim. Chang.* **2021**, *11*, 966–972. [[CrossRef](#)]
48. Kukushkin, K.; Ryabov, Y.; Borovkov, A. Digital Twins: A Systematic Literature Review Based on Data Analysis and Topic Modeling. *Data* **2022**, *7*, 173. [[CrossRef](#)]
49. Xiong, Z.; Liu, T.; Tse, G.; Gong, M.; Gladding, P.A.; Smail, B.H.; Stiles, M.K.; Gillis, A.M.; Zhao, J. A machine learning aided systematic review and meta-analysis of the relative risk of atrial fibrillation in patients with diabetes mellitus. *Front. Physiol.* **2018**, *9*, 835.
50. Bannach-Brown, A.; Przybyła, P.; Thomas, J.; Rice, A.S.C.; Ananiadou, S.; Liao, J.; Macleod, M.R. Machine learning algorithms for systematic review: Reducing workload in a preclinical review of animal studies and reducing human screening error. *Syst. Rev.* **2019**, *8*, 23. [[CrossRef](#)]
51. Zimmerman, J.; Soler, R.E.; Lavinder, J.; Murphy, S.; Atkins, C.; Hulbert, L.; Lusk, R.; Ng, B.P. Iterative guided machine learning-assisted systematic literature reviews: A diabetes case study. *Syst. Rev.* **2021**, *10*, 97. [[CrossRef](#)] [[PubMed](#)]
52. Noori, B. Classification of Customer Reviews Using Machine Learning Algorithms. *Appl. Artif. Intell.* **2021**, *35*, 567–588. [[CrossRef](#)]
53. de la Torre-López, J.; Ramírez, A.; Romero, J.R. Artificial intelligence to automate the systematic review of scientific literature. *Computing* **2023**, *105*, 2171–2194. [[PubMed](#)]
54. Wagner, G.; Lukyanenko, R.; Paré, G. Artificial intelligence and the conduct of literature reviews. *J. Inf. Technol.* **2022**, *37*, 209–226.
55. Hao, J.; Ho, T.K. Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. *J. Educ. Behav. Stat.* **2019**, *44*, 348–361. [[CrossRef](#)]
56. Fernando, A.; Siriwardana, C.; Law, D.; Gunasekara, C.; Zhang, K.; Gamage, K. A scoping review and analysis of green construction research: A machine learning aided approach. *Smart Sustain. Built Environ.* **2024**. *ahead of print*.
57. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2012**, *12*, 2825–2830. [[CrossRef](#)]
58. Halterman, A. Mordecai: Full Text Geoparsing and Event Geocoding. *J. Open Source Softw.* **2017**, *2*, 91. [[CrossRef](#)]
59. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
60. Řehůřek, R.; Sojka, P. *Gensim—Statistical Semantics in Python*; NLP Centre, Faculty of Informatics, Masaryk University: Brno, Czech Republic, 2011; Volume 3.
61. Rajapaksha, R.A.D.V.; Siriwardana, C.S.A. A Systematic Review on Different Approaches Used in the Development of Fragility Curves for Buildings. In Proceedings of the 12th International Conference on Structural Engineering and Construction Management; Springer: Singapore, 2023; pp. 407–426. [[CrossRef](#)]
62. CRED; UNDRR. *Human Cost of Disasters, An Overview of the Last 20 Years (2000–2019)*; Centre for Research on the Epidemiology of Disasters; Brussels, Belgium; UN Office for Disaster Risk Reduction: Geneva, Switzerland, 2020.

63. Moussa, A.M.; Christou, P.; Kyriakides, N. The developments of the analytical fragility methods in seismic risk assessment—A review. *J. Sustain. Archit. Civ. Eng.* **2016**, *16*, 70–81. [[CrossRef](#)]
64. Flenga, M.G.; Favvata, M.J. Fragility Curves and Probabilistic Seismic Demand Models on the Seismic Assessment of RC Frames Subjected to Structural Pounding. *Appl. Sci.* **2021**, *11*, 8253. [[CrossRef](#)]
65. Hait, P.; Sil, A.; Choudhury, S. Seismic damage assessment and prediction using artificial neural network of RC building considering irregularities. *J. Struct. Integr. Maint.* **2020**, *5*, 51–69. [[CrossRef](#)]
66. Roeslin, S.; Ma, Q.; Juárez-García, H.; Gómez-Bernal, A.; Wicker, J.; Wotherspoon, L. A machine learning damage prediction model for the 2017 Puebla-Morelos, Mexico, earthquake. *Earthq. Spectra* **2020**, *36*, 314–339. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.