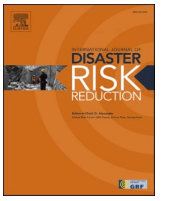
International Journal of Disaster Risk Reduction 98 (2023) 104123

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

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



Explainable artificial intelligence in disaster risk management: Achievements and prospective futures

Saman Ghaffarian [a,*], Firouzeh Rosa Taghikhah [b,**], Holger R. Maier [c]

[a] Institute for Risk and Disaster Reduction, University College London, United Kingdom
[b] Business School, University of Sydney, Australia
[c] School of Architecture and Civil Engineering, University of Adelaide, Australia

ARTICLE INFO

Keywords:
Resilience building
Interpretable artificial intelligence
Transparency
Hazard and disaster type
Data-driven decision making

ABSTRACT

Disasters can have devastating impacts on communities and economies, underscoring the urgent need for effective strategic disaster risk management (DRM). Although Artificial Intelligence (AI) holds the potential to enhance DRM through improved decision-making processes, its inherent complexity and "black box" nature have led to a growing demand for Explainable AI (XAI) techniques. These techniques facilitate the interpretation and understanding of decisions made by AI models, promoting transparency and trust. However, the current state of XAI applications in DRM, their achievements, and the challenges they face remain underexplored. In this systematic literature review, we delve into the burgeoning domain of XAI-DRM, extracting 195 publications from the Scopus and ISI Web of Knowledge databases, and selecting 68 for detailed analysis based on predefined exclusion criteria. Our study addresses pertinent research questions, identifies various hazard and disaster types, risk components, and AI and XAI methods, uncovers the inherent challenges and limitations of these approaches, and provides synthesized insights to enhance their explainability and effectiveness in disaster decision-making. Notably, we observed a significant increase in the use of XAI techniques for DRM in 2022 and 2023, emphasizing the growing need for transparency and interpretability. Through a rigorous methodology, we offer key research directions that can serve as a guide for future studies. Our recommendations highlight the importance of multi-hazard risk analysis, the integration of XAI in early warning systems and digital twins, and the incorporation of causal inference methods to enhance DRM strategy planning and effectiveness. This study serves as a beacon for researchers and practitioners alike, illuminating the intricate interplay between XAI and DRM, and revealing the profound potential of AI solutions in revolutionizing disaster risk management.

1. Introduction

Disasters have increasingly become a significant global challenge, posing threats to lives, infrastructure, and economies. The growing frequency and severity of natural hazard-induced disasters result from numerous factors, including climate change, urbanization, population growth, extensive agglomeration of assets and capital in disaster-prone areas, and environmental degradation. These catastrophic events often lead to the loss of human lives, economic disruption, and long-term damage to infrastructure, and

* Corresponding author.
** Corresponding author.
E-mail addresses: s.ghaffarian@ucl.ac.uk (S. Ghaffarian), Firouzeh.th@gmail.com, Firouzeh.Taghikhah@sydney.edu.au (F.R. Taghikhah), holger.maier@adelaide.edu.au (H.R. Maier).

https://doi.org/10.1016/j.ijdrr.2023.104123
Received 6 April 2023; Received in revised form 31 October 2023; Accepted 2 November 2023
Available online 5 November 2023

social and ecological systems. The outcomes of global initiatives such as the Sendai Framework [1], Paris Agreement [2], and Sustainable Development Goals [3] underscore the urgent need to address the increasing frequency and intensity of such disasters and their devastating impact on people, economies, and the environment. Consequently, effective disaster risk management (DRM) has emerged as a critical component in achieving sustainable development and resilience in the face of evolving risks [4,5].

DRM is the process of identifying, assessing, responding to, recovering from, and mitigating the risks posed by natural hazards, climate change, conflicts, and other types of emergencies or disasters. It requires collaboration among various stakeholders, including governments, non-governmental organizations (NGOs), communities, and individuals. It involves a complex decision-making process that relies on accurate, reliable, and timely information to ensure that appropriate actions are taken to safeguard lives and assets. The growing availability of data from various sources, such as remote sensing, social media, and Internet of Things (IoT) devices, offers unprecedented opportunities to improve DRM decision-making through the application of advanced technologies. In this context, Artificial Intelligence (AI) and Machine Learning (ML) have shown considerable promise in enhancing DRM by providing support in decision-making processes [6–9].

The application of AI in DRM has provided opportunities for more efficient, precise, and effective responses to disasters and emergencies [10–12]. AI-powered tools such as predictive analytics, decision support systems [8,13–15], and early warning systems have enhanced the ability to identify and mitigate potential disasters and their associated risks [8,16,17]. For example, ML algorithms applied to satellite imagery can predict the spread of wildfires, allowing early warning and rapid response to protect affected areas [18]. This directly aligns with SDG 15: Life on Land, which aims to protect, restore, and promote the sustainable use of terrestrial ecosystems. Furthermore, AI can support decision-making by providing real-time, accurate information that can help disaster response teams prioritize actions and allocate resources efficiently [19]. For instance, the use of AI to analyse geospatial data can help authorities identify the extent of damages after a disaster [20] and assess infrastructure risks to design sustainable infrastructure solutions that support SDG 11: Sustainable Cities and Communities [21]. In addition to improving disaster recovery efforts, AI can also contribute to long-term resilience building by identifying areas of vulnerability and recommending sustainable solutions [22,23]. As an example, AI-powered climate modelling can help identify areas at risk of sea-level rise and inform coastal zone management plans to enhance resilience [24,25], aligning with SDG 13: Climate Action.

However, the increasing complexity of AI systems has raised concerns about the interpretability and accountability of the decisions made by these systems in general [26–28] as well as in the DRM field [29]. These concerns stem from the inherent "black box" nature of many AI models, which can make it difficult for users to understand the underlying processes and reasoning behind the model outputs. As the use of AI in DRM continues to grow, the need for transparency and accountability becomes increasingly important [30]. This has led to the development of explainable AI (XAI), which allows for the interpretation and understanding of decisions made by AI models [31,32]. XAI offers transparency, interpretability, and accountability in the application of AI models, which is crucial for their acceptance and adoption. By elucidating the relationships between input data and model outputs, XAI can help decision-makers better understand the reasoning behind model predictions, enhancing their confidence in the AI model and its recommendations [33].

In the context of disaster management, XAI has the potential to improve the effectiveness and efficiency of disaster response and recovery operations [34,35], and support informed decision-making by disaster management authorities ([36–38]. It provides insights into the factors driving disaster risks and the effectiveness of various DRM strategies. This leads to increased trust among stakeholders in AI-based decision support systems and improves their acceptability and adoption [29,39].

Despite the potential benefits of XAI-based DRM (XAI-DRM), the current state of XAI applications, their achievements, and the challenges they face remain underexplored. Recent literature reviews have extensively studied the use of AI, including ML and deep learning methods, in DRM [8,11,12]. However, XAI methods for DRM have yet to be reviewed through this lens. In response to this gap in the literature, this study aims to provide a comprehensive review of the achievements and challenges in XAI-DRM, offer insightful recommendations, and uncover prospective directions for further research.

Our study adopts a well-established systematic literature review approach to review the literature published in online databases to answer predefined research questions and achieve our main objective. This process involved the extraction of 195 publications from the Scopus and ISI Web of Knowledge databases. Out of these, 68 were selected for detailed analysis based on predefined exclusion criteria. Through a rigorous examination of the selected papers, we extracted current trends in publications, applications, lessons learned, weaknesses, and open problems. We examine different risk components, risk measurements, disaster types, the geography of case studies, and AI/XAI methods used or developed in this field. To the best of our knowledge, this is the first review study on the use of XAI methods for DRM.

The study is structured as follows. Section 2 defines fundamental concepts and terms related to DRM, AI, and XAI, providing relevant background information. Section 3 explains the methodology used in conducting and synthesizing the literature review. Section 4 presents and discusses the results, while Section 5 highlights achievements, challenges, recommendations, and future directions in the field. Finally, Sections 6 and 7 outline the study's limitations and conclusions, respectively.

2. Background

2.1. Disaster risk management

Risk as a fundamental concept in disaster management is defined as the likelihood of adverse events occurring and their potential impact on critical outcomes. Adverse outcomes from disasters can have devastating consequences, including significant losses and impacts on human lives, infrastructure, economies, and the environment [40]. A disaster occurs when hazards, vulnerability, and an inability to mitigate potential negative consequences intersect [41]. To avoid, lessen, and transfer the adverse effects of hazards, DRM implements activities and measures for prevention, mitigation, and preparedness [40].

The DRM cycle is a continuous process that is independent of the type of hazard, as depicted in Fig. 1 [3]. The cycle comprises four main phases: prevention-mitigation, preparedness, response, and recovery. Prevention-mitigation aims to prevent or reduce the likelihood of a disaster occurring and to limit its adverse impacts. Preparedness involves developing plans and capacities to anticipate and enhance the response to likely, imminent, or current disasters. Response activities aim to reduce the immediate impact of a disaster, saving lives and properties. Recovery involves actions to restore the community or affected area to normal or improved conditions. The DRM cycle encompasses pre-event and post-event activities, with prevention-mitigation and preparedness being pre-event phases, and response and recovery being post-event phases. Effective DRM requires the integration of all four phases of the cycle, emphasizing the importance of proactive measures, such as prevention and mitigation, rather than reactive measures, such as response and recovery.

In the aftermath of a disaster, the primary goal is to evaluate the impact and damage caused by the event to inform effective decision-making during the response phase and prevent further impact. The recovery process can be lengthy, with its duration dependent on the severity of the damage, as well as the types of hazards and risks present. Therefore, continuous monitoring of the recovery process is essential to ensure successful plan implementation. To effectively mitigate risks and prepare for future disasters, it is critical to assess the vulnerability and resilience of a given area, such as a city. These assessments can provide valuable insights for developing strategies aimed at reducing risks and improving preparedness. Additionally, predictive and exploratory modelling can be utilized to identify potential risks and anticipate associated damages, enabling better preparation for future disasters [42,43,44]. In this context, impact, vulnerability, and resilience assessments, as well as predictive models, are indispensable components of DRM.

XAI methods have significant potential to contribute to various phases of the DRM cycle. For example, they can be employed to automate the processing of data for impact and damage assessment [45,46], providing detailed explanations regarding the features utilized [47]. Additionally, they can be used to extract relevant indicators for assessing resilience and vulnerability [36,37,48,49] and evacuation behavioural modelling [38,50]. Furthermore, predictive models that incorporate various data sources, such as climate data, can be used to map out building and infrastructure damage associated with different types of disaster risks [45,51–53].

2.2. *Explainable artificial intelligence*

In this study, we primarily discuss ML, a subset of AI that enables computers to learn and improve without being explicitly programmed. ML algorithms employ statistical models to analyse vast amounts of data, identifying patterns, trends, and associations within the data. As they receive more data and feedback, these algorithms improve their performance in terms of accuracy. They help automate repetitive tasks and detect anomalies in data, making ML a powerful tool for businesses and organizations seeking insights from large datasets [54,55,56].

There are four main categories of ML (Fig. 2), including.

- **Supervised Learning**: The machine is trained on labelled data, which means that the data is already classified. The goal of supervised learning is to learn a function that can map input data to the correct output data. The machine is provided with input data and corresponding labels during the training phase. It learns from the labelled data and can then predict the output for new, unlabelled data. There are primarily two categories of supervised learning, which are:
  - Classification: The output variable is a categorical or discrete variable. The goal of classification is to build a model that can accurately predict the class of a new observation based on a set of input variables. There are many different classification algorithms, including logistic regression, support vector machines (SVM) [57], decision trees (DT) [58], random forest (RF) [59], and Extreme/Light Gradient Boosting (XGBoost/LGBoost) [60,61].
  - Regression: The output variable is a continuous variable. The goal of regression is to build a model that can accurately predict the value of a new observation based on a set of input variables. There are many different regression algorithms, including linear regression [62], polynomial regression [63], K-nearest neighbours (KNN) [64], and neural networks (NN) [65].
- **Unsupervised Learning:** The machine is trained on unlabelled data, meaning that the data is not pre-classified. The goal of unsupervised learning is to discover patterns and relationships in the data without prior knowledge of what the data represents [66]. The machine is presented with data and must identify patterns independently. The discovered patterns can then be employed to classify new data. Unsupervised learning can be further divided into three subcategories:

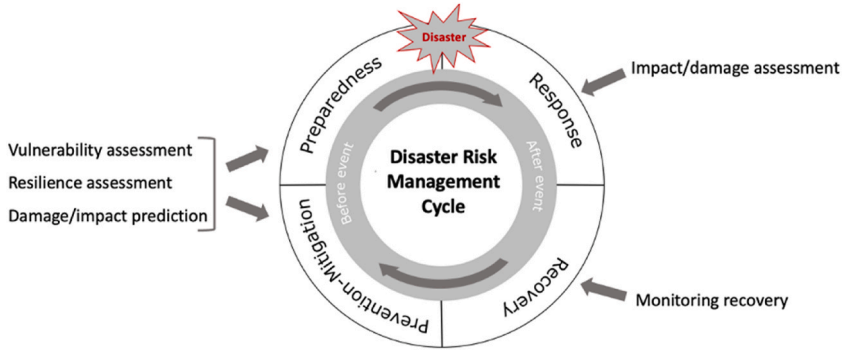


Fig. 1. DRM cycle and associated assessment types in which XAI methods can be used.

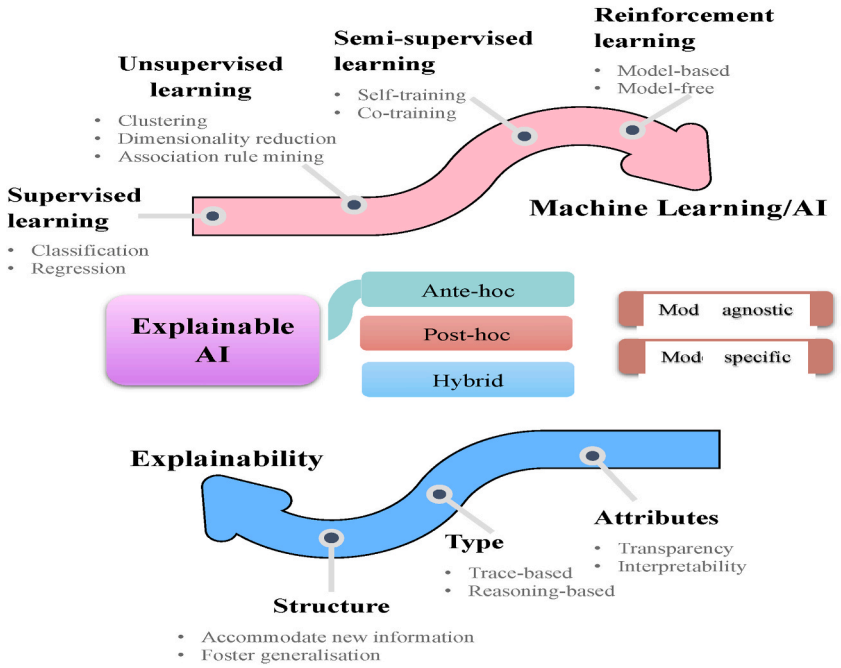


Fig. 2. An overview of the main categories of ML algorithms, explainability, and XAI.

- Clustering: The objective is to group similar data points together into clusters. Clustering algorithms do not rely on labelled data but instead identify patterns and structures within the data. Some prevalent clustering algorithms include k-means [67], hierarchical clustering [68], and density-based clustering, such as DBSCAN [69].
- Dimensionality Reduction: The aim is to decrease the number of variables in a dataset while preserving the most crucial information. This process involves identifying the underlying structure of the data and pinpointing the most significant features. Common dimensionality reduction algorithms include Principal Component Analysis (PCA) [70], t-distributed Stochastic Neighbours Embedding (t-SNE) [71], and Autoencoders [66].
- Association Rule Mining: The goal is to identify interesting relationships or patterns within a dataset. This is accomplished by detecting items that frequently co-occur in the dataset. Some common association rule mining algorithms include Apriori [72], Eclat [73], and FP-Growth [74].

- **Semi-supervised learning:** This type of ML utilizes a combination of labelled and unlabelled data to train a model. In semi-supervised learning, the model is provided with a small amount of labelled data and a large quantity of unlabelled data, leveraging the unlabelled data to enhance its performance on the task [75]. There are primarily two categories of semi-supervised learning, which are:
  - Self-training: The algorithm employs the labelled data to make predictions on the unlabelled data, and subsequently adds high-confidence predictions to the labelled data. This process is carried out iteratively, with the model being retrained on the updated labelled data [76].
  - Co-training: The algorithm trains two distinct models on separate sets of features, and then utilizes the unlabelled data to facilitate learning for both models [77]. Each model employs the labelled data to make predictions on the unlabelled data, and these predictions are used to improve the performance of both models.
- **Reinforcement Learning (RL)**: is a domain within machine learning where an agent learns by interacting with an environment. In this process, the agent receives rewards for making correct decisions and faces penalties for making incorrect ones. The goal of RL is to learn a strategy or policy that optimizes the cumulative rewards from the environment [78]. The machine learns by taking actions and observing the rewards or penalties it receives for each action. It then adjusts its policy based on the rewards and penalties received. Reinforcement learning can be divided into two main categories:
  - Model-based RL: The agent learns a model of the environment, which it can then use to plan future actions. This involves learning the transition probabilities between states and the reward function, and using these to make decisions [79]. It is often more sample-efficient than model-free RL but can be more computationally expensive.
  - Model-free RL: The agent learns a policy directly, without first learning a model of the environment. These algorithms employ trial-and-error learning to update the policy, based on the rewards received by the agent [80]. There are two main types of model-free RL: value-based and policy-based.

Explainability (also known as interpretability) in the context of ML refers to a model's ability to elucidate the reasons behind its decisions and recommendations in a manner that humans can comprehend. It is important to note that in this study interpretability is

one facet of explainability. While 'interpretability' pertains to the degree to which humans can understand the algorithm's processes or outcomes, 'explainability' encompasses a broader concept. It includes not only the understanding of the inner workings of the model, but also its relationship with the real world, its reliability, and its implications. Therefore, while all interpretable models are explainable, not all explainable models are inherently interpretable. This viewpoint aligns with those of scholars such as Doshi-Velez [81], and Gilpin [31], who proposed that interpretability should be seen as a subset of explainability. Explainability is essential for fostering trust with users and enhancing their confidence in a system [31]. Besides trust, explainability yields other benefits, such as humanizing the system, increasing the justifiability of decisions, enhancing transparency, improving accuracy and efficiency, and facilitating the extraction of novel knowledge [26].

Various attributes, types, and structures related to explainability in AI have been proposed by scholars.

- **Attributes of explainability** pertain to the criteria and characteristics used to define the concept. Despite its importance, there is no universal, objective criterion for constructing and validating explanations.
  - Transparency is a primary criterion for defining explainability, aiming to achieve simulatability, decomposability, and algorithmic transparency [82]. Simulatability refers to a model's ability to enable users to thoroughly understand its structure and functioning. Decomposability refers to the extent to which a model can be dissected into its individual components, such as input, parameters, and output, and the comprehensibility of these components. Algorithmic transparency denotes the level of confidence in a learning algorithm's sensible behaviour.
  - Interpretability is another criterion, referring to the degree to which a human observer can understand the rationale behind a decision or prediction made by the model [83]. An explanation consists of the features that contributed to a prediction. Requirements for interpretability include fidelity, diversity, and grounding. These requirements have been expanded to encompass the contrastive nature of explanations, selectivity of explanations, social nature of explanations, and irrelevance of probabilities to explanations [84].
- **Types of explanation** refer to the various ways scholars have presented explanations for their ad-hoc applications, as well as the information included or excluded. The most common questions that an explainability method should address are "why" and "how" the model generates its predictions. However, additional questions may necessitate different types of explanations, and factors such as the user type or the problem being solved can influence the required explanation type [84].
  - Trace-based explanations are designed for system designers and illustrate the reasoning within a model. Reconstructive explanations, on the other hand, are intended for end-users and strive to construct a narrative that elucidates the input features contributing to a prediction [85]. These can also be classified as mechanistic explanations (trace-based), ontological explanations (reconstructive), and operational explanations.
  - Reasoning domain knowledge, communication domain knowledge, and domain communication knowledge are classifications based on the intrinsic knowledge embedded in an explanation [86]. They focus on communication within a specific domain while considering the recipient's prior knowledge and cognitive state.
- **The structure of explanation** encompasses components such as causes, context, and consequences of a model's prediction and their arrangement [85]. Two properties of an explanation's structure can significantly impact learning: the ability to accommodate new information and promote generalization. For instance, in a study examining textual explanations in a conversation between end-users and an explanatory tool, the most suitable and effective structures involved dividing the dialogue into three stages (opening, explanation, and closing) and adhering to a set of rules to ensure successful knowledge transfer [87]. However, in task planning systems, information should be provided about why a specific action was chosen, why other actions were not, and why the planner's decisions are the best among the alternatives [88].

There are several methods to achieve explainability in AI, which can be broadly classified, according to their stage, into ante-hoc (model-based or white-box models), post-hoc (model-agnostic and model-specific), and mixed approaches.

- **Ante-hoc approaches** to XAI aim to build AI systems that are inherently interpretable and transparent. Designed to create an explainable ML model from the start of the training process, they strive to achieve high accuracy or low error [89]. These approaches often rely on simple, interpretable models such as DTs or rule-based systems, which can provide clear explanations for their decisions. Referred to as "white-box models," these models' output format depends on their architecture and input format. Designed to be easily interpretable, their primary goal is to make the decision-making process transparent. Examples include Bayesian Case Models (BCMs), Gaussian Process Regression (GPR), Generalized Additive Models (GAMs), and Mind the Gap Models (MGMs) [90].
- **Post-hoc approaches** to XAI, on the other hand, aim to provide explanations for the decisions made by existing AI systems that are not inherently interpretable or transparent. They leave the trained model unchanged and instead use an external tool to mimic or explain the model behaviour during testing. These methods do not attempt to build explainability into the model itself but rather provide a way to understand the model's decisions after training [91]. These approaches often rely on techniques such as sensitivity analysis or visualization techniques like saliency maps and layer-wise relevance propagation. Post-hoc methods can be further divided into model-agnostic and model-specific methods.
  - Model-agnostic methods do not consider the internal details of a model and can be applied to any black-box model [92]. There is a range of model-agnostic methods for explainability that generate numerical, visual, rule-based, and mixed outputs in global and local scopes for different input types in both classification and regression problems, including numerical, pictorial, textual, and time series data.

- Model-specific methods, on the other hand, are limited to certain types of models and may only work with specific types of interpretation, such as the weights of a linear regression model or the interpretation of NNs [93]. In contrast, ante-hoc methods are designed to make the functioning of a model transparent, and as a result, they are often model-specific. Many studies about model-specific methods for explainability focus on interpreting NNs at global and local scope for various input types. This is not unexpected given the widespread use of deep neural networks (DNN) in different fields [94]. These methods often produce visual explanations, such as salient masks, scatter plots, and other visual aids. Some methods also produce rules, textual and numerical explanations, or a combination of these different types of explanations.

- **Hybrid approaches** to XAI combine elements of both ante-hoc and post-hoc approaches, aiming to provide transparent and interpretable models as well as post-hoc explanations for the AI system's decisions. These approaches seek to merge the benefits of both by incorporating some level of explainability into the model itself while also offering a means to understand the model's decisions after training [26]. For example, a hybrid approach might involve training a NN with a particular type of architecture that is inherently more interpretable, such as a DT or a rule-based system [94]. The model could then be further enhanced with the use of post-hoc techniques, such as feature importance or sensitivity analysis, to better understand its decision-making process [91].

  Another hybrid approach might involve using an ante-hoc technique, such as a relevance propagation method, to train a NN with a more interpretable structure [95]. Subsequently, a post-hoc technique, such as a local surrogate model, can be employed to further explain the model's decisions [92]. Some emerging hybrid methods include the use of modular NNs, where a model consists of multiple smaller, interpretable NNs combined in a hierarchical structure [96]. Another promising hybrid approach is the combination of feature selection and feature engineering techniques with post-hoc explanation methods [97].

3. Methodology

We employed a systematic review process to identify and evaluate the literature on XAI-DRM, which was adapted from Ref. [98] to ensure a rigorous review. The process begins by defining a set of research questions aligned with the objective of our study: reviewing and investigating XAI-DRM. We then formulated a search scope and strategy to create a search string for finding relevant papers. This string was refined through iterative trial and error to ensure its accuracy and effectiveness in producing appropriate search results. We selected the final set of papers based on specific exclusion criteria used to determine the quality and relevance of the studies concerning the study's objective. A well-designed search strategy is essential for obtaining suitable search results that meet both sensitivity and precision criteria.

We chose the final set of papers for the systematic review based on predefined exclusion criteria (EC). We defined these criteria to identify the most appropriate studies according to their contribution to the study's objective and their quality. Consequently, we manually screened the papers to apply these criteria.

We developed a data extraction strategy to elicit the necessary information from each selected paper. The extracted data was then synthesized into a data extraction form, and the results were presented and discussed. The following sections provide detailed explanations of the steps taken during this review process. By adhering to this systematic review process, we ensured the reliability and validity of our findings.

3.1. Research questions

We addressed the objective of this study through the identification of five main research questions (RQs). These RQs were carefully selected to understand the current state of the art and identify notable achievements, open questions, or challenges in this research domain. Our structured analysis of the SLR is based on the following RQs.

RQ1. What case studies demonstrate the use of explainable AI-based DRM (XAI-DRM)?

RQ2. What specific DRM objectives have been addressed with XAI?

RQ3. What XAI-based approaches have been applied to DRM?

RQ4. What datasets or data types have been used to address XAI-DRM?

RQ5. What are the existing research directions, achievements, and challenges in XAI-DRM?

3.2. Search strategy

This study does not impose a time constraint on the publication date, and the search is conducted on the ISI Web of Knowledge and Scopus databases, which include a wide range of high-quality publications. The search query targets the title, abstract, and keywords of the papers. The search is executed automatically by inputting the following search string into the search engine of the platforms:

((disaster OR hazard OR typhoon OR hurricane OR earthquake OR storm OR erosion OR flood OR tsunami OR landslide OR subsidence OR drought OR tornado OR asteroid OR volcan* OR cyclone OR *fire OR seism* OR "ground deformation" OR "slope stability" OR rockfall OR "debris flow" OR hydraulic OR hydrological OR drainage OR meteorology* OR "land sink*" OR "subsurface compaction" OR "groundwater depletion" OR "soil consolidation" OR "land surface deformation") AND ("explainable artificial intelligence" OR "explainable deep learning" OR "explainable machine learning" OR "explainable ai" OR "interpretable artificial intelligence" OR "interpretable deep learning" OR "interpretable machine learning" OR "interpretable ai" OR xai OR exai OR "Local Interpretable Model-agnostic Explanations" OR "SHapley Additive exPlanations" OR "contrafactual explanation" OR "explainable boosting machine" OR "decision tree interpreter" OR "integrated gradients" OR "class activation mapping" OR deeplift OR "occlusion testing" OR "partial dependence plots" OR "individual conditional expectation" OR "global surrogate models" OR rulefit OR interpretml OR fairml) AND

(risk OR damage OR recovery OR reconstruction OR relief OR vulnerability OR impact OR "adaptive capacity" OR "coping capacity" OR resilien* OR susceptibility) NOT (health OR pandem* OR epidem*))

The search string is designed to retrieve as many relevant publications as possible for the defined objective and search scope. It consists of three primary parts, separated by the term "AND." The first part targets publications relevant to the research subject, using keywords associated with different types of disasters (e.g., earthquake). The second part targets publications that used relevant methods, such as "xai." The third part targets publications relevant to the research task, using keywords associated with different components of risk (e.g., impact).

3.3. Study selection criteria

Initially, the papers resulting from the execution of the search query were manually filtered to select the final list of the most appropriate papers. Both general and specifically defined exclusion criteria for the study's objective were employed during this process (as outlined in Table 1).

3.4. Data extraction

The data extraction step is a vital part of the process, as it gathers the information needed to answer the research questions from the selected papers. We created a preliminary data extraction form based on the initial screening of the papers to collect all the information required to address the research questions. This form includes attributes such as the study identification, publisher, and publication year. After thoroughly reviewing the selected papers, we created and filled out a more detailed final data extraction form, which includes the specific objectives addressed in the papers and the XAI algorithms used.

3.5. Data synthesis

The synthesis of extracted data from selected papers is a critical step in the systematic literature review process. This step involves answering the research questions and presenting the extracted data and results. The results are then summarized and visualized, with the papers grouped into distinct categories for each research question. This approach enables a clear and concise understanding of the key findings and insights, allowing for a comprehensive analysis of the research area under investigation. Thus, this step is crucial for deriving meaningful conclusions and recommendations aligned with the study's objective. Additionally, we discuss the results in detail, identifying the key points from the selected studies. Finally, we provide the current research directions, achievements, challenges, and recommendations for future studies.

4. Results and discussion

In total, 196 papers were identified using the search strategy, out of which 68 were selected for detailed review based on the exclusion criteria. This section presents an overview of the main statistics pertaining to the selected papers, followed by a comprehensive discussion of the results associated with each research question in the subsequent sections.

The publisher and journal names, along with the publication years of the selected studies, were extracted. Researchers began using XAI methods for DRM in 2018, and accordingly, the selected papers were published from 2018 to 2023. Fig. 3 illustrates the year-wise distribution of the selected papers. There was a significant increase in the number of XAI-DRM papers in 2022, which continued into 2023, especially considering that the search from the online database was conducted in August 2023.

The 68 papers on XAI-DRM considered for detailed review were published in 49 journals. However, as shown in Fig. 4, only 9 journals and 4 publishers published two or more of these papers. The most popular journal is "Remote Sensing" with 6 publications, followed by "Journal of Hydrology" and "Science of the Total Environment" with 4 publications each. Seven more journals published two papers (Fig. 4), while the remaining 40 journals only published one of the papers selected for detailed review. These statistics highlight the diversity of perspectives and research domains tackling and addressing DRM using XAI methods. This also demonstrates that researchers from various disciplines attempt to address DRM using XAI from multidisciplinary and interdisciplinary perspectives.

RQ1. What case studies demonstrate the use of explainable AI-based DRM (XAI-DRM)?

Fig. 5 displays the global distribution of study areas and their respective countries. The findings indicate that XAI-DRM publications analysed case studies from 21 countries across 6 continents, while 12 studies did not focus on specific cases and either utilized databases generated globally [37,48,99] or developed generic methods without focusing on a specific cases study [100,101]. China, with 14 publications, had the highest number of studied disaster cases [102,103], followed by the United States with 12 publications [35,

Table 1
Exclusion criteria (EC) for paper selection in the literature review process.

ID	Criterion
EC1.	Papers with unavailable full text
EC2.	Papers not written in English
EC3.	Papers not directly contributing to DRM
EC4.	Papers not explicitly discussing or connecting the study to DRM
EC5.	Papers not directly using XAI or interpretable AI methods
EC6.	Papers without validation of the proposed study
EC7.	Papers providing a general summary without a clear contribution
EC8.	Review and editorial papers

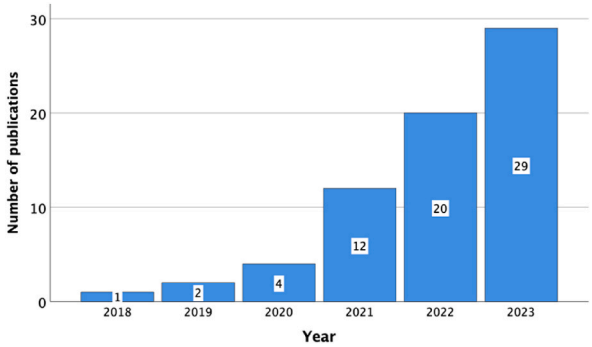


Fig. 3. Temporal distribution of the 68 selected papers on XAI-DRM.

Fig. 4. Distribution of articles on XAI-DRM categorized by journal and publisher. This chart solely incorporates the 26 articles that appeared in journals which have published at least two out of the chosen 68 papers.

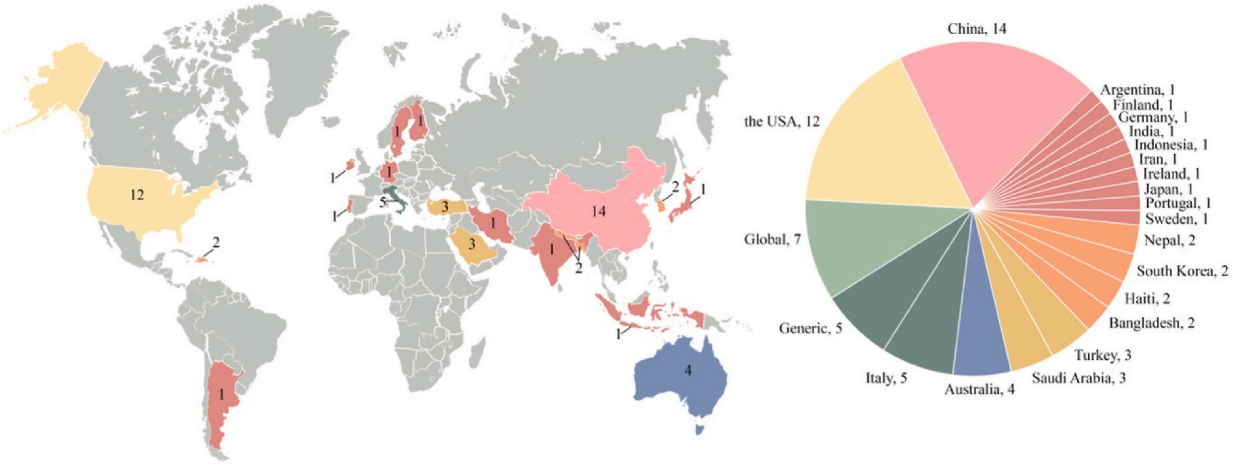


Fig. 5. Worldwide distribution of case studies in XAI-DRM included in the 68 selected papers.

45]. Notably, XAI-DRM case studies from Italy, Australia, Saudi Arabia, and Turkey were studied in 5, 4, 3, and 3 articles, respectively [46,104–106]. However, no case studies from the African continent have been included.

RQ2. What specific DRM objectives have been addressed with XAI?

The majority of the selected papers (27 publications) addressed DRM from a general perspective without focusing on a specific DRM component ([107–109] (Fig. 6). The response component of the DRM cycle was studied in 20 publications [110–112], while the recovery component was not considered in any of the selected papers. Furthermore, the risk mitigation ([113–115], and preparedness [36,116,117] components of DRM were addressed in 14 and 8 papers, respectively. In terms of hazard types, floods and landslides were studied most frequently, with 13 publications each. Earthquake and wildfire followed, with 9 publications each, while drought, hurricane, soil erosion, and land subsidence were the subject of 7, 4, 3, and 2 publications, respectively. Moreover, cyclone, volcano, typhoon, tornado, and crown snow hazards were only investigated in one publication each. Overall, the selected papers demonstrate a strong focus on general disaster risk assessment, mapping and response, particularly for floods, landslides, earthquakes, and wildfires. However, there is a need for more research on the recovery, risk mitigation, and preparedness components of DRM, as well as on less frequently studied hazard types such as volcano and wind-borne hazards.

RQ3. What XAI-based approaches have been applied to DRM?

RQ3.1. What type of ML algorithms are used?

As can be seen from the horizonal axis of Fig. 7, a wide range of ML methods have been applied in the selected papers. Supervised learning is the predominant type of learning utilized in DRM. While multiple algorithms have been employed in some studies, we only consider those that have demonstrated superior performance over other tested algorithms, which are subsequently used for interpretation.

Regarding the architecture of ML models, ***decision tree-based and boosting models*** are the most commonly adopted. These models represent decisions and their possible consequences in tree-like structures and frequently undergo boosting (an optimization technique) to enhance their efficiency and accuracy. RF emerges as the most applied method, being featured in 18 studies for predicting wildfire susceptibility and threat, evacuation rate during hurricane and cyclone, flood and drought susceptibility, crown snow load outage risk, spatial landslide susceptibility, soil susceptibility, volunteer rescue requests during hurricanes, mining induced subsidence, and wind property damage caused by tornadoes [34,36,37,105,114,118–128]. Following closely are 11 studies utilizing XGBoost for predicting landslide events and dam stability, vegetation transpiration variations, turbidity (amount of fine sediment in water), flood and run-off susceptibility, vulnerability of vegetation during volcano, relief operations during crisis, and species richness and abundance of post-larval fish affected by storms [102,106,107,109,111,117,129–135]. LGBoost is a similar algorithm that is the focus of studies by Ref. [136] to assess earthquake spatial probability and [137] to predict landslide susceptibility. CatBoost is used to develop forest fire risk index [132], and GBoost to assist with predicting flood susceptibility [106].

NNs and deep learning models are the second most applied method. These models simulate the human brain and learn from data by adjusting the weights of connections between neurons. They recognize patterns and interpret sensory data through machine perception, labelling, and clustering. In this regards, Convolutional Neural Networks (CNN) are the most commonly used method, appearing in 13 studies, including [51] for spatial drought prediction [138], for hurricane damage estimation [139], for post-earthquake structural damage assessment [140], for global flood susceptibility mapping [141], for quantifying infrastructure damage [142], for locating and quantifying the degree of the damage post-events [143], for predicating wildfire response [144], for flood susceptibility mapping, and [108] for global wildfire susceptibility modelling, while Artificial Neural Networks (ANNs) are used less extensively for purposes , such as soil erosion probability prediction and drought prediction [50, 142]. Deep Feedforward NN and Bayesian Regularized NN are used by Ref. [53] for building risk assessment model for proactive hurricane response and [145] to identify flood classes. Long Short-Term Memory (LSTM) networks feature in 3 studies, including [146] for drought forecasting [147], to forecast flood risk and [103] for streamflow prediction. Beyond these mainstream algorithms, specialized ML models, such as Transformers, are explored by Ref. [100] for identifying wildfire ignition points in before the fire spread and ensemble deep learning models are used by Ref. [148] for mapping land subsidence and [149] for developing a wildfire susceptibility prediction model.

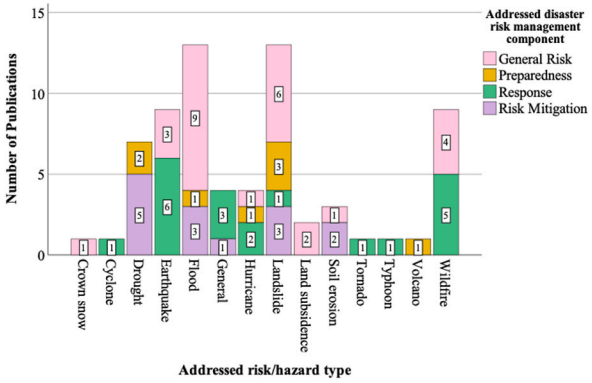


Fig. 6. DRM cycle components (Fig. 1) and hazard types being addressed in the selected XAI-DRM publications.

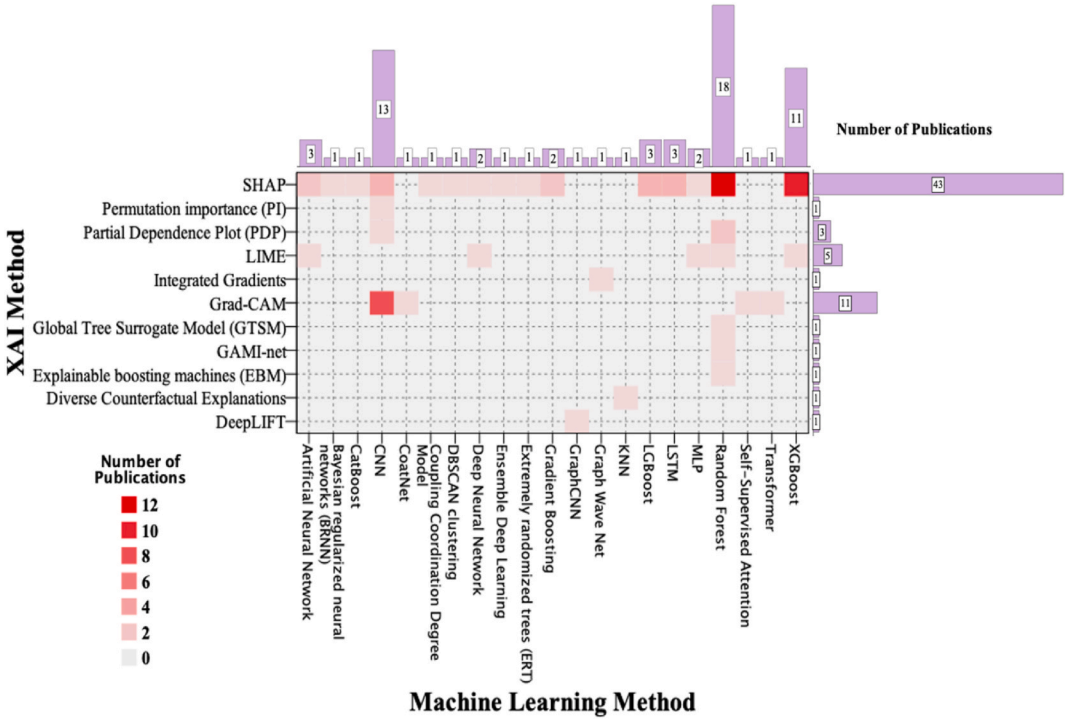


Fig. 7. Trends in application of XAI and ML methods in DRM research based on the selected publications.

Examples of other NNs used in risk management include GraphCNN for modelling community resilience to extreme events [113], Graph wave net for flood risk predication [150], as well as Multi-Layer Perceptron (MLP) for earthquake-induced building-damage mapping, and disaster relief analysis during natural hazard crises [46,110]. Other supervised algorithms that appear less frequently but are worth mentioning include K-Nearest Neighbours (KNN) for predicting the impacts of drought, especially crop growth in an uncertain climate future (refer to Ref. [151], and CoatNet for building damage detection (refer to Ref. [152].

Unsupervised and Semi-supervised learning have been used least in the selected papers, with one study using DBSCAN clustering, which is an unsupervised algorithm for landslide displacement prediction [116] and one study using a self-supervised attention foreground-aware pooling algorithm for forest fire segmentation [153].

RQ3.2. Which notions of explainability are addressed?

Regarding the attribute of AI explainability, most studies exhibit a medium level of algorithmic transparency. While ensemble methods such as RF, XGBoost, and LGBoost fall into the medium category since due to their reliance on decision trees (DTs) that are generally comprehensible, as noted in almost half of the relevant literature, it is essential to mention that the aggregation techniques in these ensemble methods can enhance their inherent complexity. Convolutional Neural Networks, LSTM, ANNs, and DFFNN/MLPs are categorized as having a low level of explainability because their internal mechanisms are complex and not easily understood. In contrast, the explainability of KNN is classified as high since it makes predictions based on the most similar historical data, rendering its predictions understandable and transparent. However, this approach was only used in one study.

Studies focusing on transparency, whether predicting wildfire susceptibility in the Mediterranean countries of Southern Europe or assessing soil erosion probabilities in Saudi Arabia, strive for algorithmic transparency. This ensures stakeholders can trust the model predictions by understanding the factors it considers, from vegetation dryness to past weather patterns. Post-disaster efforts like earthquake-induced building damage mapping and damage estimation from aerial imagery require decomposable models. By breaking down these models, analysts can pinpoint specific model components responsible for damage predictions, aiding timely and effective interventions. Whether detecting typhoons or assessing hurricane risk, these models emphasize simulatability. Such transparency ensures that meteorologists and emergency response units understand the factors contributing to the model's decisions and can thus prepare and respond effectively. In studies focusing on interpretability such as spatial drought prediction, streamflow prediction, and soil susceptibility to wind erosion, interpretability ensures that geographical and ecological experts can tie predictions back to familiar environmental factors, such as rainfall patterns and soil composition. Studies leveraging social media for disaster relief, like sentiment analysis during the COVID-19 crises and resource management via time series sentiment analysis, thrive on interpretability. Clear interpretations help response units prioritize areas of intervention based on public sentiment and specific cries for help. Whether predicting landslide events or assessing slope failure susceptibility, these studies underscore the importance of grounding explanations. Geologists and urban planners can understand predictions in the context of known geological factors, ensuring safe infrastructural developments.

Based on types of explanation, most hazard risk management studies have reported using trace-based explanations for tree-based algorithms, which fall under mechanistic explanations for wildfire, cyclone, flood, drought, crown snow, landslide, and erosion predication, as well as damage detection caused by hurricane, and tornado [34,36,111,118,129,130,154]; [35,37,38,47,48,53,99,102,105–107,109,114,117,119–121,123–128,131–136,155–159]. In the case of predicting events such as landslides, cyclones, or droughts, decision-makers can easily trace back and understand which factors or conditions led to a particular prediction. This aids in building trust and facilitating communication with stakeholders. For damage detection after hurricanes or tornadoes, the ease of visualizing decision trees can help in rapid dissemination and understanding of findings amongst emergency responders. However, in dynamic environments with many interacting variables, like in the case of predicting cyclone paths or flood intensities, tree-based models might miss capturing intricate patterns or feature interactions. On the other hand, neural network-based algorithms, such as CNN and LSTM models which are mainly used for post-crisis damage detection post-crisis and susceptibility predication of wildfire drought and flood, fall under reconstructive explanations or ontological explanations [51,101,113,138,141]. By understanding which features or inputs the model deemed significant in predicting hazards, researchers can prioritize collecting relevant data and fine-tuning their models for better accuracy. Even with trace-based explanations, the complexity of these models can make them challenging to interpret. In scenarios like flood prediction, while you might know which features influenced a prediction, understanding the intricate interactions between those features can be tough.

RQ3.3. Which methods are used for explainability?

As shown in Fig. 7, the post-hoc approach is the only utilized approach for XAI in DRM. This choice aligns with the inherent complexity of DRM problems, where models often need to be highly accurate and capable of handling a broad range of variables, thus making the need for transparency and understanding of decisions equally critical. Regarding sub-categories, the use of model-agnostic methods was predominant with 89 % of studies using them. Given the diversity of modelling techniques employed in DRM—from ML methods for predicting natural disasters to deep learning algorithms for assessing infrastructure risks—the flexibility offered by model-agnostic methods becomes an invaluable asset.

Regarding model-agnostic methods, SHAP (SHapley Additive exPlanations) was the most preferred method (used in 43 studies), followed by LIME (Local Interpretable Model-Agnostic Explanations) (used in 5 studies).

The SHAP method calculates feature importance scores to explain how different features contribute to a prediction. It does so by approximating the Shapley values from game theory, taking into account all possible combinations of features for assigning importance, rather than just evaluating each feature in isolation [91,160]. In DRM articles, SHAP identified the most important features in (i) RF models for predicting wildfire susceptibility and threat, evacuation rate during hurricanes and cyclones, flood and drought susceptibility, spatial landslide susceptibility, soil susceptibility, volunteer rescue requests during hurricanes, and mining induced subsidence [34,36,37,105,112,118,121,122,125–128,155,161], (ii) XGBoost for predicting landslide events and dam stability, vegetation transpiration variations, turbidity, flood and run-off susceptibility, and vulnerability of vegetation to volcanoes [102,107,109,117,129–131,133–135], (iii) LSTM for drought forecasting [146], (iv) ANN for drought prediction [52], (v) CNN for spatial drought prediction ([51], (vi) Gboost for predicting flood susceptibility ([106], (vii) LightGboost for landslide susceptibility mapping [137,158], (viii) LSTM for flood and drought forecasting [147], (ix) BRNN for flood classification [145], (x) CDDM for predicting soil erosion [115], (xi) DL for wildfire susceptibility prediction [149], (xii) MLP for earthquake-induced building-damage mapping [46], (xiii) DBSCAN for landslide displacement prediction [116], and (xiv) GBoost for predicting flood susceptibility [162]. This method can be computationally demanding and may not be ideal for analysing high-dimensional data typical in DRM. While it can identify the importance of individual features, it may not always accurately represent the intricate dependencies between features, crucial in DRM scenarios like interdependent factors in landslide or flood susceptibility. In dynamic DRM situations, such as evolving wildfire threats or cyclone paths, the granular insights SHAP provides for each prediction might be too detailed, potentially obscuring broader trends or patterns.

LIME is another model-agnostic explanation method that approximates a complex, black-box model with a simple, interpretable model in the vicinity of the prediction being explained. It provides individualised explanations for predictions and has been used in two studies on disaster management to improve the transparency of XGboost models for time series analysis for relief operations [111], MLP models for classification of tweets during earthquakes [110], DFFNN models for building risk assessment in hurricanes [53], RF models for rescue request prediction [124], and NN models for risk assessment of road networks in earthquakes [163]. However, its applicability is limited due to certain challenges; for example, approximation relies on linear models. However, many DRM scenarios, like risk assessment of road networks during earthquakes or building risk during hurricanes, might involve non-linear relationships. This can result in oversimplified or even misleading explanations. In scenarios such as classification of tweets during earthquakes, the data might be high-dimensional (e.g., word embeddings). In such high-dimensional spaces, LIME local approximations in such high-dimensional spaces can sometimes be less precise, possibly overlooking intricate patterns. In addition, the reliance of LIME on independent perturbations might not capture inter-feature dependencies effectively. In DRM contexts, like rescue request predictions, where factors are interrelated (e.g., weather conditions and infrastructure damage), this could lead to incomplete or suboptimal insights.

Other methods such as CE (Counterfactual Explanations), Partial Dependence Plots (PDP), and Permutation Importance (PI), appeared only in a few studies. CE can be both model-specific and model-agnostic, depending on the approach used for generating the explanations. This method answers questions about what would have happened if an input feature had a different value, providing insight into the causal impact of individual features on the model's prediction. In the reviewed articles, it was used to explain KNN prediction about factors impacting crop growth in an uncertain climate future [151]. However, the intricacies of climate's effect on crop growth often involve cascading impacts, where one factor indirectly affects another. CE, by looking at direct feature

modifications, might miss these intricate, multi-step causal chains. By focusing on the change in one feature at a time, the method might give the impression that factors operate in isolation, whereas, in reality, climate factors impacting crop growth are deeply interconnected. PDP depicts the relationship between a single feature and the model's prediction. It offers insight into the impact of changes in a particular feature on the model's prediction. It was applied to the interpretation of the RF model for landslide space-time forecasting and evacuation preparation time during a cyclone [105,126]. This method assumes that the distribution of one feature is independent of that of others. This might not hold true in real-world scenarios like landslides and cyclones, where features often co-vary. PI works by randomly permuting the value of a single feature and measuring the impact of this permutation on the model's performance. This gives an indication of how "important" that particular feature is for accurate prediction. In the studies reviewed, PI was specifically utilized to explain CNN model predictions of global wildfire susceptibility [108]. It is to be noted that for a CNN, specific regions (e.g., areas with frequent wildfires) might overly influence the importance score, giving a skewed perspective if not accounted for properly. However, the method is computationally more efficient compared to techniques like CE, making it relatively easier to deploy even on large datasets or models with many features. While PI can provide a global understanding of feature importance, it might not be suitable for local interpretations. Given the spatial and temporal dynamics of wildfires, local variations might be crucial In a complex scenario like wildfire susceptibility, interactions between features (e.g., temperature, humidity, wind speed) can be pivotal, and PI might overlook these interactions.

In the sub-category of ***NN model-specific explanation*** the GRAD-CAM (Gradient-weighted Class Activation Mapping) was the most popular method (11 studies). It operates by tracing the gradient of the predicted class score back to the activations in the last convolutional layer of a CNN. The activations are then weighted by their corresponding gradients, resulting in a heatmap that highlights the regions of an input image that had the greatest impact on the model prediction. This approach has been used to explain CNN and transformer models for wildfire susceptibility prediction, hurricane infrastructure damage estimation, and typhoon detection [100,138,139,141–143,164,165] in the DRM literature. In images with infrastructure, the surrounding environment (like trees, vehicles, shadows, etc.) might influence the gradients. This could lead to GRAD-CAM highlighting areas that are not necessarily related to the damage but had strong activations. It might also not provide the granular detail necessary to pinpoint specific areas of damage, especially on large structures. Moreover, this method of visualizations highlights only the positive influences on a decision and discards negative contributions. This can be limiting when trying to understand factors that reduce risks, such as areas that are less susceptible to wildfires. For events like typhoons which have temporal sequences, capturing the sequence's importance or changes over time might not be straightforward with GRAD-CAM, especially if the model is primarily image-based and not sequence-based. Other methods such as Integrated Gradients (IG), Deep Learning Important FeaTures (DeepLIFT), and Class Activation Mapping (CAM) were among the least utilized techniques. DeepLIFT, specifically, has been applied to Graph CNNs for understating factors contributing to community resilience to extreme events [113]. The method quantifies the contribution of each neuron to every other neuron's output, providing a more detailed understanding of feature importance within the network. As community resilience often involves multi-faceted factors, the method might not provide a holistic view of interdependent features, especially when there are non-linear relationships or when nodes have heterogeneous characteristics. IG is particularly interesting for its application to Graph WaveNet models to identify spatial-temporal factors influencing flood risk [150]. It assigns an importance score to each feature by integrating the gradients of the model output with respect to each feature along a straight path from the feature's baseline value to its actual value. This results in a more comprehensive understanding of how each feature contributes to the model's prediction. Nevertheless, flood risk assessments often require very precise spatial information, and IG might not always be adept at capturing the details of micro-level spatial changes, especially when gradients are sparse or inconsistent across areas. CAM has seen specific application in self-supervised attention foreground-aware pooling mechanisms in Ref. [153] to explain the segmentation of forest fire. However, this method might not capture the subtle variances of forest terrains and fire intensities. Forest fires have complex propagation patterns influenced by factors like wind, humidity, and vegetation type and CAM may not adapt well to these rapidly changing conditions. Being designed specifically for architectures with a global average pooling layer before the fully connected layer, its adaptability can be questioned. Nevertheless, in forest fire contexts, the information heatmaps it produces are instrumental in quickly identifying wildfire hotspots.

In the sub-category focusing on ***tree-based model-specific explanation***, Explainable Boosting Machines (EBM), Global Tree Surrogate Models, and GAMI-Net (Generalized Additive Models with Interactions Network) stand out for their application to explain RF models. EBM has been used to explain a RF model of slope failure predication in landslide events [114]. It is a glass-box model that combines the interpretability of generalized additive models (GAMs) with the performance of gradient-boosted machines. As EBM tries to maintain a balance between interpretability and performance, it might sometimes sacrifice some predictive power in favour of clarity. Additionally, capturing all the intricate details of slope failures, which can be influenced by various unpredictable factors, might be challenging. Similarly, a Global Tree Surrogate Model has been used to explain a predictive model of wind property damages caused by tornado [123]. It serves as an approximation of the more complex model it seeks to explain—in this case RF. By generating a simplified, single decision tree that closely mimics the behaviour of the RF, it offers a more understandable representation of the underlying model. However, it is worth noting that this simplification can come at the cost of granularity. By reducing the complex RF model into a single decision tree, some details are inevitably lost. This can result in the model overlooking subtle patterns or relationships in the data that might be crucial in predicting the specific damages caused by tornadoes, especially in regions with varied infrastructure and building practices. Finally, GAMI-Net has been used to interpret an RF model for spatial landslide susceptibility prediction in Karst mining areas [120]. This approach adds a layer of complexity by incorporating interaction terms to generalized additive models. This allows for capturing more complex relationships among features and produces highly interpretable models. However, in Karst mining areas, where data might be scarce or inconsistent, GAMI-Net might become too sensitive to noise or outliers, potentially reducing its predictive accuracy.

This snapshot of XAI techniques in DRM provided above reveals a strong lean towards post-hoc and model-agnostic methods. This is

likely driven by the need for both flexibility and depth in explanations, as DRM involves a multitude of complex, interdependent variables that require nuanced interpretation for effective decision-making.

RQ4. What datasets or data types have been used to address XAI-DRM?

Remote sensing data including multi-spectral satellite images [34,129], SAR data [36,154], and drone images [45], are the most frequently used data sets in XAI-DRM studies, with 45 papers utilizing them (Fig. 8). For example [100], developed a Transformer-based model to detect fire flames from images and combined with Grad-CAM to find the causes of the object detection results [46]. developed a ML-based method to detect earthquake-induced building damages from post-event satellite images. They used the SHAP method to reveal the impact of each feature descriptor included in the model for building-damage assessment and to examine the reliability of the model. Earth data, including geological [130], geomorphological [99], and rainfall data [51], are the second most common dataset, with 36 papers employing them in their studies. As an example [166], processed soil data using ML and SHAP methods to identify the best soil erodibility indices. Climate and weather data [52], geographical information systems (GIS) [129], IoT [104], socio-economic data (e.g., demographic data) [38], and social media (e.g., tweets) [111], traffic data [163], and simulated data [123] are other data sources used in the selected papers. For instance Ref. [121], produced a wildfire susceptibility map for Turkey using ML and GIS data, and further, demonstrated that elevation, temperature, and slope factors were the most contributing factors as the result of IG and SHAP methods.

RQ5. What are the existing research directions, achievements, and challenges in XAI-DRM?

XAI-DRM is an emerging field experiencing rapid growth, encompassing a diverse array of research directions, accomplishments, and challenges. Although still in its infancy, significant progress has been made in developing XAI techniques. However, their application in DRM remains limited. Consequently, it is crucial for researchers and practitioners to collaborate, tackle challenges, and seize the opportunities that XAI offers. This continuous process of learning and improvement will enable the DRM community to stay ahead of the curve and effectively manage evolving risks associated with disasters.

Table 2 summarizes the AI methods used, insights extracted, and gaps identified in the reviewed publications. XAI methods are primarily employed to identify the most contributing factors, such as those affecting susceptibility to various disasters, and to enhance result accuracy, as exemplified in object detection from images. The reliability of the developed AI methods was analysed in only one study, which was also based on ranking contributing factors. Nevertheless, one of the primary contributions of XAI methods is to enhance transparency in developed AI methods, by assessing their reliability. Thus far, in the DRM field, XAI methods have predominantly been employed to extract additional insights in DRM field.

In the context of wildfire, volcano, and drought disasters, the current employment of XAI methods exhibits several notable gaps. For instance, there is a lack of time-series-based models for predicting the progression of these disasters. Furthermore, geospatial-tailored models, particularly those focused on GIS data, are underrepresented. In studies pertaining to flood, soil erosion, and landslides, certain key elements are lacking. Specifically, there is an absence of probabilistic models for in-depth uncertainty analysis. The need for hybrid models that combine multiple techniques is evident and warrants more attention and research. For earthquake risk management, there is a clear need for the incorporation of ensemble techniques and combined prediction models. In the realm of windborne disasters, the research focus appears to be mainly on predicting the extent of impacts, with no studies addressing the prediction of the path and intensity of these hazards. In a more general context, socio-economic data has yet to be effectively harnessed. There is a pressing need to leverage this data, with a particular emphasis on its relevance to the business and operational aspects of DRM.

Most components of DRM have been examined, with the focus predominantly on specific disasters like earthquakes. While the body of research on the application of XAI in DRM is still growing, preliminary studies suggest potential benefits. However, more comprehensive evaluations are needed to ascertain the true impact and significance of XAI methods in DRM. Additionally, numerous disaster types, including tsunamis, typhoons, coastal flooding, and heatwaves, warrant further investigation. It is imperative to assess the potential and applicability of XAI methods across various disasters and stages of the DRM cycle. Expanding the scope of research to include a diverse range of disasters is crucial for developing a comprehensive understanding of the effectiveness of XAI methods in DRM. Moreover, exploring the application of XAI methods across different earthquake case studies can provide valuable insights into each region's unique geophysical characteristics and improve our understanding of disaster and geophysical science.

Although various datasets have been effectively used to address XAI-DRM, there remains significant untapped potential when compared to the broader applications of AI and XAI capabilities. For example, remote sensing can be employed to study virtually all

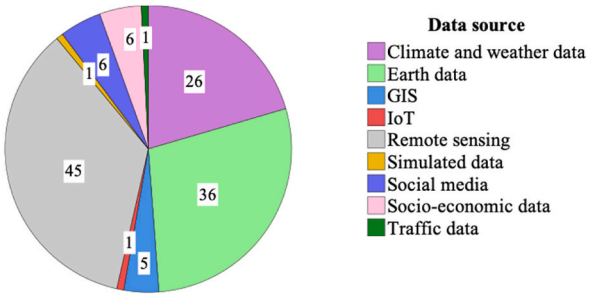


Fig. 8. The number of publications on the employed data for XAI-DRM.

Table 2
AI Methods, Extracted XAI Insights, and Identified Gaps in reviewed publications.

Disaster	AI Methods Used	Insights from XAI	Identified Gaps
Wildfire, Volcano, Drought	Deep learning, Transformer, RF, CatBoost, CNN, self-supervised, XGBoost.	- Improve result accuracy. - Identify the most contributing factors.	- Lack of time-series based models for progression prediction. - No unsupervised models for anomaly detection. - Absence of geospatial-tailored models.
Flood, Soil erosion, and Landslide	GB, ERT, RF, CNN, LSTM, XGBoost, Gboost, Graph wave net, LightGBoost, ensemble deep learning models, ANN.	- Improve result accuracy. - Identify the most contributing factors.	- Absence of probabilistic models for uncertainty. - No models for changing environmental variables due to climate change. - No models considering land use changes and agricultural practices. - Lack of hybrid models combining multiple techniques.
Earthquake	MLP, XGBoost, LGBoost, CNN, CoatNet, NN.	- Reliability analysis. - Improve result accuracy. - Identify the most contributing factors.	- Absence of models utilizing seismic wave data in real-time. - Lack of ensemble techniques for combined predictions. - No models predicting secondary effects like tsunamis or aftershocks.
Cyclone, Typhoon, Hurricane, and Tornado	RF, CNN, deep feedforward neural network, Random Forest.	- Improve result accuracy. - Identify the most contributing factors.	- Lack of models predicting path and intensity over longer horizons. - Absence of models considering sea temperature and oceanographic data.
Others	GraphCNN, CNN, RF, Bayesian regularized neural networks (BRNN), KNN.	- Improve result accuracy. - Identify the most contributing factors.	- Absence of models considering socio-economic factors.

types of disasters. With advancements in big data collection and analysis, there is a growing need to better integrate diverse data types and leverage AI and XAI methods for processing and elucidating results. Incorporating XAI techniques in the analysis of diverse datasets can help address challenges posed by the volume, variety, and velocity of data generated during disaster events.

From our analysis (as detailed in Q3.1), decision tree-based models are the predominant model type represented in DRM applications. While the results of our analysis underscores their frequent use, the underlying reasons for this choice are not explicitly detailed in the majority of the studies reviewed. It can be speculated that the attributes of interpretability and transparency of DTs might contribute to their common usage. These characteristics make them more accessible and comprehensible to both domain experts and non-experts alike, facilitating communication and collaboration among stakeholders. However, they may not always provide the best predictive performance compared to more complex models like NNs. There is a pressing need for more research on improving the explainability of complex NN-based models in DRM, as these models become increasingly relevant for handling large-scale, high-dimensional, and dynamic data generated during disaster events.

Our in-depth analysis of XAI methods reveals a lack of ante-hoc approaches in DRM. While these methods have been frequently applied to tasks like assessing earthquake damages, predicting typhoons, and evaluating the risks of coastal flooding, our findings indicate a disparity in their widespread acceptance. Possible reasons for this include a lack of technical capacity, insufficient research, or inadequate awareness of their potential benefits. Interestingly, despite their widespread use in certain disaster scenarios, comprehensive evaluations of the success rates and impact of XAI methods in DRM remain sparse. For instance, while decision tree-based models have been popular due to their inherent transparency, especially in earthquake damage predictions, there is limited evidence to suggest they consistently outperform other techniques in varied disaster contexts. Additionally, our analysis indicates that the DRM field may be more focused on using well-established, interpretable models rather than exploring novel, less transparent models that might require ante-hoc approaches for explainability. However, it is crucial to recognize that model-agnostic methods might have inherent limitations and may not be universally applicable to all ML systems employed in disaster management. This calls for further multidisciplinary research and development in XAI, specifically targeting DRM applications, to determine not just the most frequently used, but the most effective methods in real-world disaster scenarios.

One significant challenge faced by the DRM community is the narrow scope of XAI methods in elucidating neural network-based models. As our analysis has shown, NNs have primarily been employed for tasks such as prediction of flood extents and typhoon trajectories. However, the inherent opacity of these models poses severe risks in disaster situations, given the potential ramifications of inaccuracies. For instance, a lack of transparency in predicting flood extents might lead to inadequate evacuation measures, putting lives at risk. Similarly, misinterpreting a typhoon's path due to an opaque model could lead to massive infrastructure and economic losses. The current state of XAI in DRM, with Grad-CAM being the most widely applied method, highlights the urgent need for more research and development in this area. Failure to address this gap may lead to suboptimal decision-making due to stakeholders' limited understanding of the underlying factors influencing the likelihood and potential impact of disasters. This lack of comprehension can, in turn, result in inappropriate resource allocation, inadequate planning, and ineffective risk management strategies, ultimately exacerbating the consequences of disasters. As such, the development of innovative methods to enhance the transparency and

interpretability of NNs must be considered a top priority for researchers and practitioners in DRM.

The potential advantages of employing hybrid approaches to XAI in DRM are immense. Combining model-specific and model-agnostic methods allows researchers and practitioners to achieve a more comprehensive and nuanced understanding of the intricate factors that influence disaster risk. This in-depth knowledge can be translated into highly effective, targeted risk management strategies capable of safeguarding lives and preserving critical infrastructure. Additionally, it is vital to emphasize the need for continuous evaluation and refinement of hybrid XAI methods as the DRM landscape evolves. By staying abreast of emerging risks, technological advancements, and lessons learned from past disasters, researchers and practitioners can adapt and optimize these approaches to remain effective in a rapidly changing world.

5. Synthesized insights for future research

In previous sections, the results are discussed within the context of existing research directions, and the groupings were based on current literature. However, these groupings do not cover the entire DRM domain's problems and challenges that can be addressed using XAI methods. Additionally, they do not encompass all state-of-the-art methods that can be employed for DRM. Hence, in this subsection, the results are further discussed, and insights from the broader DRM and AI domains are provided to develop guidelines for future studies.

***Recommendation 1**: multi-hazard risk analysis for XAI-DRM*

Approaching DRM from a multi-hazard risk perspective has gained significant importance due to the complex and interconnected nature of disasters. Multi-hazard risk management considers the potential impacts of various hazards, such as earthquakes, floods, and landslides, and their cascading effects within a single framework. This approach allows for a more comprehensive assessment of risk and can improve the efficiency and effectiveness of disaster risk reduction efforts.

The application of AI and XAI methods in multi-hazard risk management can significantly improve risk assessments by offering precise predictions and assessments of potential impacts and their underlying factors. For instance, modelling approaches for multi hazards can be conceptually distinct, making integration challenging. The utilization of AI methods can facilitate an integrated approach to the development of multi-hazard models, allowing for the exploration and understanding of interactions using XAI methods. Furthermore, XAI methods can offer transparency and interpretability of model results, which can improve the decision-making process and the acceptance of the model by stakeholders. For example, XAI methods can be used to develop models that predict the impact of multiple hazards on infrastructure, such as buildings and transportation systems. These models can be used to identify the most vulnerable areas and prioritize the allocation of resources for risk reduction measures. XAI methods can also be applied to optimize the allocation of resources for response and recovery actions in the event of a multi-hazard situation.

However, it is crucial to ensure the explainability, interpretability, accountability, and trustworthiness of the AI models with XAI techniques used for multi-hazard risk assessment with the aid of XAI techniques. This can be achieved by designing AI/XAI models that can explain their decisions and by validating the model performance using real-world data. Furthermore, involving stakeholders in the model development and validation process can increase the acceptance and adoption of XAI methods in multi-hazard risk management.

***Recommendation 2**: explainability for disaster early warning systems and digital twins*

The DRM field is becoming increasingly reliant on complex models such as Digital Twins and Early Warning Systems (EWS) to identify and mitigate risks. These models use large amounts of data from various sources to model imminent hazards, elements at risk, and simulate disaster scenarios and predict potential outcomes. However, these models are often too complex for humans to fully understand, making it challenging to interpret and act upon the predictions made by these models. This is where XAI comes into play.

XAI techniques can provide insight into how complex models such as Digital Twins and EWS work and can help decision-makers understand how and why certain predictions are made. This is especially important in DRM, where accurate and timely decisions can mean the difference between life and death. XAI techniques can also help improve the accuracy and reliability of complex models by identifying biases and errors in the data used to train them. In addition, they can provide more transparency and accountability in decision-making, as stakeholders can better understand the reasoning behind certain decisions. The use of XAI in DRM can also lead to increased public trust and confidence in these systems, which is particularly vital in EWS. When the decision-making process is transparent and easily understood, people are more likely to trust and accept the recommendations made by these systems.

***Recommendation 3**: incorporating causal inference methods in DRM decision making*

Currently, the use of XAI is gaining traction as a tool to analyse data and support the decision-making process in DRM. However, XAI may be limited in providing accurate inferences to identify the causal factors of a disaster, which is crucial for effective DRM. Instead, we suggest incorporating causal inference methods, the process of determining cause-and-effect relationships between different variables or events. Causal inference can help identify factors contributing to disaster occurrence and guide the development of effective response, recovery, prevention/mitigation, and preparedness strategies [167].

In the response stage, a causal inference approach can be used to determine the impact of factors such as communication channels, transportation networks, and medical supplies availability on response times and emergency services' success. In the recovery stage, potential causal effects of different recovery interventions [168] can be identified using data on the affected population's characteristics, the type and severity of the disaster, and the implemented recovery interventions. For instance, causal inference methods could be employed to determine the impact of interventions like temporary housing, medical care, access to clean water, and mental

health support services on recovery speed after a flood.

In the prevention/mitigation stage, XAI can identify factors that increase disaster occurrence probability. For example, it can be used to determine factors contributing to the occurrence and severity of floods in a particular area by analyzing historical data related to the degree of proximity to rivers or the level of deforestation in the area. In the preparedness/adaptation stage, this information can be used to inform future planning and preparedness efforts, ensuring that the most effective measures are put in place to minimize the impact of future disasters. For instance, if an analysis of past disasters reveals that a lack of early warning systems contributed to the disaster's severity, then a focus on developing and implementing effective early warning systems would be a key adaptation measure. For the case of flooding [169], measures such as reforestation, building flood barriers, or relocating populations away from high-risk areas can be designed.

Causal inference methods can also be used to evaluate the effectiveness of DRM strategies, allowing practitioners to determine whether a particular strategy has been successful in reducing disaster risk. By incorporating causal inference methods in DRM decision-making processes, policymakers and practitioners can better understand the complex relationships between various factors and disasters, ultimately leading to more effective and targeted interventions to reduce disaster risk and enhance resilience.

***Recommendation 4**: exploring the potentials of generative AI for transforming the DRM landscape*

Generative AI has the potential to revolutionize DRM by leveraging deep learning capabilities. This technology can generate synthetic data, simulate disaster scenarios, and propose innovative solutions for disaster response and recovery, significantly improving our ability to plan for and respond to disasters [170].

For example, generative AI can simulate various disaster scenarios, enabling disaster response teams to test and refine their strategies in a safe, controlled environment. This can help improve response times and minimize the risk of errors or missteps during actual disaster events. Moreover, generative AI can propose innovative solutions for disaster response and recovery, such as new evacuation routes or shelter designs that better meet the needs of disaster victims.

This technology can also generate synthetic data, which can be used to augment existing datasets and improve the accuracy of predictive models [171]. This can help identify areas at higher risk for disasters, enabling better planning and preparation. Despite these potential benefits, it is crucial to consider the ethical implications of generative AI. There are concerns about using synthetic data in predictive models, which may not accurately reflect real-world conditions [172]. Furthermore, the applications of generative AI in disaster response may raise questions around data privacy and security [173].

***Recommendation 5**: collaborating with domain experts and incorporating their knowledge and expertise*

Reliance on ML algorithms alone is insufficient, as they can often lack contextual knowledge and may not capture the nuances of DRM [31]. Domain experts possess a wealth of knowledge and expertise that can inform and improve the accuracy and explainability of AI models. One approach involves incorporating domain experts in the design and development of AI models from the outset, collaborating to identify relevant data sources, select appropriate modelling techniques, and develop necessary algorithms [174]. For instance, engaging experts from various domains, such as computer science, disaster management, and psychology, will facilitate the development of XAI methods that are not only technically sound but also user-friendly and contextually relevant. This ensures that the AI model is tailored to the specific context of DRM and is transparent and understandable to its users.

Another approach employs XAI to highlight areas where domain experts can provide additional insight and expertise [175]. For example, XAI can identify areas of uncertainty or conflicting data that require further investigation or validation from domain experts. Domain experts can then provide additional data or input to refine the AI model and improve its accuracy and explainability.

Moreover, involving domain experts in the validation and testing of AI models can help ensure that the models are reliable and effective within the context of DRM. Domain experts can provide feedback on the relevance and accuracy of the data, as well as the model's appropriateness and effectiveness in predicting and mitigating disaster risks. Involving domain experts in AI and XAI system development and testing can improve acceptance and adoption. When experts see their knowledge and expertise being utilized and recognize their role in developing the technology, they are more likely to trust and embrace it [176]. This, in turn, can lead to increased efficiency and effectiveness in DRM decision-making. Integrating domain experts into the development, testing, and validation of AI and XAI models can substantially improve their performance and trustworthiness in DRM. By leveraging the unique insights and experience of domain experts, AI models can be tailored to the specific challenges and provide more accurate, transparent, and understandable predictions and recommendations.

6. Limitations

Limitations that may impact the systematic literature review study include publication bias, data extraction, and classification. In this study, the main limitations and threats to validity are discussed under four categories: construct, internal, external, and conclusion validities.

Construct validity concerns the extent to which the study accurately represents the target concept. In this study, only high-quality studies were selected by employing the ISI Web of Knowledge and Scopus databases as sources to find relevant publications. However, this approach may lead to missing other pertinent publications not indexed in these databases. Nonetheless, indexing in an ISI journal and Scopus is an accepted way to find and extract high-quality papers. Although there may be missing terms that could impact the final results, the search was kept broad, and the search query was refined several times to reduce the possibility of missing any relevant studies. Therefore, the impact of missing relevant papers in the final results is low.

Internal validity is concerned with the extent to which the study accurately addresses the research questions. The research

questions in this study were designed to investigate and extract all the necessary information and components for XAI-DRM. The questions were based on precisely defined hazard and disaster types, risk components, AI and XAI methods, and other relevant information. Therefore, the findings of this study are well-explained and linked to the extracted results.

External validity is concerned with the extent to which the findings of a study can be generalized to other contexts. In this study, we reviewed publications that utilized XAI methods for DRM, focusing on various disaster types and risk cycle components. However, it is important to note that not all existing XAI methods and potential disaster risk types and components are covered in the reviewed papers. Therefore, the generalizability of the findings to other contexts may be limited. However, we have provided insights from broader perspectives in the AI and DRM domains in a separate section (Section 5), which may contribute to the external validity of the study.

Regarding **conclusion validity**, this review followed the widely accepted structure and protocol for systematic literature reviews outlined by Ref. [98]. All the steps, including defining research questions, search strategies, exclusion criteria, and synthesizing results, were performed according to this structure. Furthermore, the search string and data extraction form used in this study are provided in Appendix A, and a full list of the extracted publications is included in the supplementary materials. Therefore, the results of this study can be easily reproduced.

7. Conclusions

This review has critically examined the current achievements and challenges in utilizing explainable artificial intelligence (XAI) techniques for disaster risk management (DRM), with the primary objective of suggesting future research directions. By extracting 195 publications from the Scopus and ISI Web of Knowledge databases and selecting 68 for detailed review based on predefined exclusion criteria, our study has provided a comprehensive understanding of the applications, challenges, and opportunities of XAI in this crucial domain. The selected papers were analysed and classified to address our research questions, and additional analyses were conducted to identify research achievements and challenges in the field. We observed a significant increase in the use of XAI techniques for DRM in 2022, emphasizing the growing importance of explainability and transparency in disaster decision-making.

Our study has synthesized and analysed the existing literature, highlighting the various hazard and disaster types, risk components, AI and XAI methods, and the inherent challenges and limitations of these approaches in DRM. We have identified key research directions and provided synthesized insights for enhancing explainability and effectiveness in disaster decision-making. Some of the key recommendations identified in this review include the adoption of multi-hazard risk analysis for XAI-DRM, the incorporation of explainability in disaster early warning systems and Digital Twins, and the use of causal inference methods in DRM decision-making. Additionally, this review highlights the potential of generative AI for transforming the DRM landscape and emphasizes the importance of collaborating with domain experts to incorporate their knowledge and expertise in the development and validation of AI and XAI models for DRM.

Despite the limitations of this review, our rigorous methodology ensures that the findings are well-explained and linked to the extracted results. This review has unearthed the intricate interplay between XAI and DRM, revealing the profound potential of AI solutions and emphasizing the necessity to enhance the transparency, interpretability, and effectiveness of disaster-related predictive models. By considering the recommendations and insights provided herein, researchers and practitioners can work towards the development and implementation of innovative XAI solutions that are not only more accurate and reliable but also more understandable and trustworthy for stakeholders. Ultimately, the integration of XAI in DRM can lead to more efficient and effective decision-making, helping to save lives, reduce losses, and build more resilient communities in the face of ever-increasing disaster risks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix-A Data Extraction Form

Table A.1
Data extraction form.

#	Extraction element	Contents
General information		
1	ID	Unique ID for the study
2	Title	Full title of the article
3	Authors	The authors of the article
4	Year	The publication year
5	Publisher name	The publisher name (e.g., Nature)

(*continued on next page*)

Table A.1 (*continued*)

#	Extraction element	Contents
6	Journal name	The journal name (e.g., Nature Communications)
Study description		
7	Case Study (study area)	The location (country name) of the case study
8	Hazard risk type	The natural hazard type (e.g., Earthquake)
9	Main objective of the study (addressed DRM component)	☐Response ☐Recovery ☐Risk mitigation ☐Preparedness ☐General risk
10	Details about the study	E.g., any interesting findings or problems
11	Directly address DRM	☐Yes ☐ No
12	XAI method/type	The name of the used XAI method (e.g., SHAP)
13	AI method/type	The name of the used AI method (e.g., Random Forest)
14	Data source/type	Data source/type used (e.g., Remote sensing)
15	Additional notes	E.g., the opinions of the reviewer about the study

References

[1] C. Wannous, G. Velasquez, United nations office for disaster risk reduction (unisdr)—unisdr's contribution to science and technology for disaster risk reduction and the role of the international consortium on landslides (icl), in: Advancing Culture of Living with Landslides: Volume 1 ISDR-ICL Sendai Partnerships 2015-2025, 2017.

[2] P. Agreement, United Nations Framework Convention on Climate Change, 2015, United Nations Treaty Series, 2015. December, 12.

[3] D.P. Coppola, Chapter 1 - cche management of disasters, in: D.P. Coppola (Ed.), Introduction to International Disaster Management, third ed., Butterworth-Heinemann, 2015, pp. 1–39, https://doi.org/10.1016/B978-0-12-801477-6.00001-0.

[4] J.P. Newman, H.R. Maier, G.A. Riddell, A.C. Zecchin, J.E. Daniell, A.M. Schaefer, H. van Delden, B. Khazai, M.J. O'Flaherty, C.P. Newland, Review of literature on decision support systems for natural hazard risk reduction: current status and future research directions, Environ. Model. Software 96 (2017) 378–409.

[5] G.A. Riddell, H. van Delden, H.R. Maier, A.C. Zecchin, Tomorrow's disasters–Embedding foresight principles into disaster risk assessment and treatment, Int. J. Disaster Risk Reduc. 45 (2020), 101437.

[6] S. Guha, R.K. Jana, M.K. Sanyal, Artificial neural network approaches for disaster management: a literature review, Int. J. Disaster Risk Reduc. 81 (2022), 103276, https://doi.org/10.1016/j.ijdrr.2022.103276.

[7] J. Latvakoski, R. Öörni, T. Lusikka, J. Keränen, Evaluation of emerging technological opportunities for improving risk awareness and resilience of vulnerable people in disasters, Int. J. Disaster Risk Reduc. 80 (2022), 103173, https://doi.org/10.1016/j.ijdrr.2022.103173.

[8] W. Sun, P. Bocchini, B.D. Davison, Applications of artificial intelligence for disaster management, Nat. Hazards 103 (3) (2020) 2631–2689.

[9] F. Taghikhah, E. Erfani, I. Bakhshayeshi, S. Tayari, A. Karatopouzis, B. Hanna, Artificial intelligence and sustainability: solutions to social and environmental challenges, in: Artificial Intelligence and Data Science in Environmental Sensing, Elsevier, 2022, pp. 93–108.

[10] H.S. Munawar, M. Mojtahedi, A.W.A. Hammad, A. Kouzani, M.A.P. Mahmud, Disruptive technologies as a solution for disaster risk management: a review, Sci. Total Environ. 806 (2022), 151351, https://doi.org/10.1016/j.scitotenv.2021.151351.

[11] R.I. Ogie, J.C. Rho, R.J. Clarke, Artificial intelligence in disaster risk communication: a systematic literature review, in: 2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), 2018, 4-7 Dec. 2018.

[12] S. Thekdi, U. Tatar, J. Santos, S. Chatterjee, Disaster risk and artificial intelligence: A framework to characterize conceptual synergies and future opportunities (2022), https://doi.org/10.1111/risa.14038.

[13] S. Ghaffarian, D. Roy, T. Filatova, N. Kerle, Agent-based modelling of post-disaster recovery with remote sensing data, Int. J. Disaster Risk Reduc. 60 (2021), 102285, https://doi.org/10.1016/j.ijdrr.2021.102285.

[14] A. Khan, S. Gupta, S.K. Gupta, Multi-hazard disaster studies: monitoring, detection, recovery, and management, based on emerging technologies and optimal techniques, Int. J. Disaster Risk Reduc. 47 (2020), 101642, https://doi.org/10.1016/j.ijdrr.2020.101642.

[15] L. Tan, J. Guo, S. Mohanarajah, K. Zhou, Can we detect trends in natural disaster management with artificial intelligence? A review of modeling practices, Nat. Hazards 107 (3) (2021) 2389–2417, https://doi.org/10.1007/s11069-020-04429-3.

[16] V. Deparday, C.M. Gevaert, G. Molinario, R. Soden, S. Balog-Way, Machine Learning for Disaster Risk Management, 2019.

[17] M. Motta, M. de Castro Neto, P. Sarmento, A mixed approach for urban flood prediction using Machine Learning and GIS, Int. J. Disaster Risk Reduc. 56 (2021), 102154, https://doi.org/10.1016/j.ijdrr.2021.102154.

[18] M. Mohajane, R. Costache, F. Karimi, Q. Bao Pham, A. Essahlaoui, H. Nguyen, G. Laneve, F. Oudija, Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area, Ecol. Indicat. 129 (2021), 107869, https://doi.org/10.1016/j.ecolind.2021.107869.

[19] G. Baryannis, S. Validi, S. Dani, G. Antoniou, Supply chain risk management and artificial intelligence: state of the art and future research directions, Int. J. Prod. Res. 57 (7) (2019) 2179–2202, https://doi.org/10.1080/00207543.2018.1530476.

[20] S. Ghaffarian, N. Kerle, E. Pasolli, J. Jokar Arsanjani, Post-disaster building database updating using automated deep learning: an integration of pre-disaster OpenStreetMap and multi-temporal satellite data, Rem. Sens. 11 (20) (2019) 2427. https://www.mdpi.com/2072-4292/11/20/2427.

[21] S. Yousefi, H.R. Pourghasemi, S.N. Emami, S. Pouyan, S. Eskandari, J.P. Tiefenbacher, A machine learning framework for multi-hazards modeling and mapping in a mountainous area, Sci. Rep. 10 (1) (2020), 12144, https://doi.org/10.1038/s41598-020-69233-2.

[22] N. Algiriyage, R. Prasanna, K. Stock, E.E.H. Doyle, D. Johnston, Multi-source multimodal data and deep learning for disaster response: a systematic review, SN Computer Science 3 (1) (2021) 92, https://doi.org/10.1007/s42979-021-00971-4.

[23] N. Kerle, S. Ghaffarian, R. Nawrotzki, G. Leppert, M. Lech, Evaluating resilience-centered development interventions with remote sensing, Rem. Sens. 11 (21) (2019) 2511. https://www.mdpi.com/2072-4292/11/21/2511.

[24] N. Adebisi, A.-L. Balogun, A deep-learning model for national scale modelling and mapping of sea level rise in Malaysia: the past, present, and future, Geocarto Int. 37 (23) (2022) 6892–6914, https://doi.org/10.1080/10106049.2021.1958015.

[25] V. Nieves, C. Radin, G. Camps-Valls, Predicting regional coastal sea level changes with machine learning, Sci. Rep. 11 (1) (2021) 7650, https://doi.org/10.1038/s41598-021-87460-z.

[26] A. Adadi, M. Berrada, Peeking inside the black-box: a survey on explainable artificial intelligence (XAI), IEEE Access 6 (2018) 52138–52160.

[27] H.R. Maier, S. Galelli, S. Razavi, A. Castelletti, A. Rizzoli, I.N. Athanasiadis, M. Sànchez-Marrè, M. Acutis, W. Wu, G.B. Humphrey, Exploding the myths: an introduction to artificial neural networks for prediction and forecasting, Environ. Model. Software (2023), 105776.

[28] F. Taghikhah, A. Voinov, T. Filatova, J.G. Polhill, Machine-assisted agent-based modeling: opening the black box, J. Comp. Sci. 64 (2022), 101854.

[29] C.M. Gevaert, M. Carman, B. Rosman, Y. Georgiadou, R. Soden, Fairness and accountability of AI in disaster risk management: opportunities and challenges, Patterns 2 (11) (2021), 100363, https://doi.org/10.1016/j.patter.2021.100363.

[30] F. Taghikhah, A. Voinov, N. Shukla, T. Filatova, Shifts in consumer behavior towards organic products: theory-driven data analytics, J. Retailing Consum. Serv. 61 (2021), 102516.

[31] L.H. Gilpin, D. Bau, B.Z. Yuan, A. Bajwa, M. Specter, L. Kagal, Explaining explanations: an overview of interpretability of machine learning, in: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), 2018.

[32] A. Rai, Explainable AI: from black box to glass box, J. Acad. Market. Sci. 48 (2020) 137–141.
[33] C. Meske, E. Bunde, J. Schneider, M. Gersch, Explainable artificial intelligence: objectives, stakeholders, and future research opportunities, Inf. Syst. Manag. 39 (1) (2022) 53–63.
[34] R. Cilli, M. Elia, M. D'Este, V. Giannico, N. Amoroso, A. Lombardi, E. Pantaleo, A. Monaco, G. Sanesi, S. Tangaro, Explainable artificial intelligence (XAI) detects wildfire occurrence in the Mediterranean countries of Southern Europe, Sci. Rep. 12 (1) (2022), 16349.
[35] K. Demertzis, K. Kostinakis, K. Morfidis, L. Iliadis, An interpretable machine learning method for the prediction of R/C buildings' seismic response, J. Build. Eng. 63 (2023), 105493.
[36] H.A.H. Al-Najjar, B. Pradhan, G. Beydoun, R. Sarkar, H.J. Park, A. Alamri, A novel method using explainable artificial intelligence (XAI)-based Shapley Additive Explanations for spatial landslide prediction using Time-Series SAR dataset, Gondwana Res. (2022), https://doi.org/10.1016/j.gr.2022.08.004.
[37] W. Li, M. Migliavacca, M. Forkel, J.M. Denissen, M. Reichstein, H. Yang, G. Duveiller, U. Weber, R. Orth, Widespread increasing vegetation sensitivity to soil moisture, Nat. Commun. 13 (1) (2022) 3959.
[38] X. Zhao, R. Lovreglio, D. Nilsson, Modelling and interpreting pre-evacuation decision-making using machine learning, Autom. ConStruct. 113 (2020), 103140.
[39] L. Hancox-Li, Robustness in machine learning explanations: does it matter? FAT*, in: 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020.
[40] UNISDR, Terminology on Disaster Risk Reduction, 2009.
[41] I. Kelman, Disaster by Choice: How Our Actions Turn Natural Hazards into Catastrophes, Oxford University Press, 2020.
[42] D.A. Radford, T.C. Lawler, B.R. Edwards, B.R. Disher, H.R. Maier, B. Ostendorf, J. Nairn, H. van Delden, M. Goodsite, A framework for the mitigation and adaptation from heat-related risks to infrastructure, Sustain. Cities Soc. 81 (2022), 103820.
[43] G.A. Riddell, H. van Delden, H.R. Maier, A.C. Zecchin, Exploratory scenario analysis for disaster risk reduction: considering alternative pathways in disaster risk assessment, Int. J. Disaster Risk Reduc. 39 (2019), 101230.
[44] S. ayari, F. Taghikhah, G. Bharathy, A. Voinov, Designing a conceptual framework for strategic selection of bushfire mitigation approaches, J. Environ. Manag. 344 (2023) 118486.
[45] C.S. Cheng, A.H. Behzadan, A. Noshadravan, Uncertainty-aware convolutional neural network for explainable artificial intelligence-assisted disaster damage assessment [Article], Struct. Control Health Monit. 29 (10) (2022), https://doi.org/10.1002/stc.3019. Article e3019.
[46] S.S. Matin, B. Pradhan, Earthquake-induced building-damage mapping using Explainable AI (XAI), Sensors 21 (13) (2021) 4489.
[47] S. Mangalathu, K. Karthikeyan, D.-C. Feng, J.-S. Jeon, Machine-learning interpretability techniques for seismic performance assessment of infrastructure systems, Eng. Struct. 250 (2022), 112883.
[48] M. Midwinter, C. Yeum, E. Kim, Explainable machine learning for seismic vulnerability assessment of low-rise reinforced concrete buildings, in: Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021: CSCE21 Structures Track Volume 2, 2022.
[49] F.R. Taghikhah, D. Baker, M.T. Wynn, M.B. Sung, S. Mounter, M. Rosemann, A. Voinov, Resilience of Agri-Food Supply Chains: Australian Developments After a Decade of Supply and Demand Shocks, in: Supply Chain Risk and Disruption Management: Latest Tools, Techniques and Management Approaches, Springer Nature Singapore, Singapore, 2023, pp. 173–192.
[50] F. Taghikhah, A. Voinov, N. Shukla, T. Filatova, Exploring consumer behavior and policy options in organic food adoption: insights from the Australian wine sector, Environ. Sci. Pol. 109 (2020) 116–124.
[51] A. Dikshit, B. Pradhan, Interpretable and explainable AI (XAI) model for spatial drought prediction, Sci. Total Environ. 801 (2021), 149797.
[52] E. Felsche, R. Ludwig, Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations [Article], Nat. Hazards Earth Syst. Sci. 21 (12) (2021) 3679–3691, https://doi.org/10.5194/nhess-21-3679-2021.
[53] S. Gao, Y. Wang, Explainable deep learning powered building risk assessment model for proactive hurricane response, Risk Anal. 43 (6) (2023) 1222–1234.
[54] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT press, 2016.
[55] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553) (2015) 436–444.
[56] I. Bakhshayeshi, E. Erfani, F.R. Taghikhah, S. Elbourn, A. Beheshti, M. Asadnia, An Intelligence Cattle Re-Identification System over Transport by Siamese Neural Networks and YOLO, IEEE Internet Things J. (2023).
[57] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[58] J.R. Quinlan, Induction of decision trees, Mach. Learn. 1 (1986) 81–106.
[59] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[60] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016.
[61] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, Lightgbm: a highly efficient gradient boosting decision tree, Adv. Neural Inf. Process. Syst. 30 (2017).
[62] F. Galton, Regression towards mediocrity in hereditary stature, J. Anthropol. Inst. G. B. Ireland 15 (1886) 246–263.
[63] M.G. Kendall, The Advanced Theory of Statistics. *The Advanced Theory of Statistics*, second ed., 1946.
[64] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theor. 13 (1) (1967) 21–27.
[65] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, Psychol. Rev. 65 (6) (1958) 386.
[66] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, science 313 (5786) (2006) 504–507.
[67] J. MacQueen, Classification and analysis of multivariate observations, in: 5th Berkeley Symp. Math. Statist. Probability, 1967.
[68] P.H. Sneath, R.R. Sokal, Numerical Taxonomy. The Principles and Practice of Numerical Classification, 1973.
[69] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, kdd (1996).
[70] A. Maćkiewicz, W. Ratajczak, Principal components analysis (PCA), Comput. Geosci. 19 (3) (1993) 303–342.
[71] S. Hinton, N. Roweis, t-SNE van der Maaten & Hinton, JMLR 9 (2008) 2579–2605.
[72] R. Agrawal, T. Imieliński, A. Swami, Mining association rules between sets of items in large databases, in: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, 1993.
[73] M.J. Zaki, Scalable algorithms for association mining, IEEE Trans. Knowl. Data Eng. 12 (3) (2000) 372–390.
[74] J. Han, J. Pei, Y. Yin, Mining frequent patterns without candidate generation, ACM sigmod record 29 (2) (2000) 1–12.
[75] O. Chapelle, B. Scholkopf, A. Zien, Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews], IEEE Trans. Neural Network. 20 (3) (2009), 542-542.
[76] H. Scudder, Probability of error of some adaptive pattern-recognition machines, IEEE Trans. Inf. Theor. 11 (3) (1965) 363–371.
[77] A. Blum, T. Mitchell, Combining labeled and unlabeled data with co-training, in: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, 1998.
[78] R.S. Sutton, A.G. Barto, Reinforcement Learning: an Introduction, MIT press, 2018.
[79] T.M. Moerland, J. Broekens, A. Plaat, C.M. Jonker, Model-based reinforcement learning: a survey, Foundations and Trends® in Machine Learning 16 (1) (2023) 1–118.
[80] A.L. Strehl, L. Li, E. Wiewiora, J. Langford, M.L. Littman, PAC model-free reinforcement learning, in: Proceedings of the 23rd International Conference on Machine Learning, 2006.
[81] F. Doshi-Velez, B. Kim, Towards a Rigorous Science of Interpretable Machine Learning, 2017 *arXiv preprint arXiv:1702.08608*.
[82] Z.C. Lipton, The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery, Queue 16 (3) (2018) 31–57.
[83] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115.
[84] T. Miller, Explanation in artificial intelligence: insights from the social sciences, Artif. Intell. 267 (2019) 1–38.

[85] T. Lombrozo, The structure and function of explanations, Trends Cognit. Sci. 10 (10) (2006) 464–470.
[86] O. Biran, C. Cotton, Explanation and justification in machine learning: a survey, IJCAI-17 workshop on explainable AI (XAI) (2017).
[87] A. Cawsey, Generating Interactive Explanations. AAAI, Cf, O. (2015). Transforming Our World: the 2030 Agenda for Sustainable Development, United Nations, New York, NY, USA, 1991.
[88] M. Fox, D. Long, D. Magazzeni, Explainable Planning, 2017 *arXiv preprint arXiv:1709.10256*.
[89] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell. 1 (5) (2019) 206–215.
[90] B. Kim, R. Khanna, O.O. Koyejo, Examples are not enough, learn to criticize! criticism for interpretability, Adv. Neural Inf. Process. Syst. 29 (2016).
[91] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).
[92] M.T. Ribeiro, S. Singh, C. Guestrin, Why should i trust you?, in: Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.
[93] G. Montavon, W. Samek, K.-R. Müller, Methods for interpreting and understanding deep neural networks, Digit. Signal Process. 73 (2018) 1–15.
[94] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, D. Shen, Deep convolutional neural networks for multi-modality isointense infant brain image segmentation, Neuroimage 108 (2015) 214–224.
[95] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, W. Samek, On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, PLoS One 10 (7) (2015), e0130140.
[96] S. Masoudnia, R. Ebrahimpour, Mixture of experts: a literature survey, Artif. Intell. Rev. 42 (2) (2014) 275.
[97] C. Molnar, Interpretable Machine Learning, Lulu. com, 2020.
[98] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman, Systematic literature reviews in software engineering – a systematic literature review, Inf. Software Technol. 51 (1) (2009) 7–15, https://doi.org/10.1016/j.infsof.2008.09.009.
[99] S.K. Kuntla, M. Saharia, P. Kirstetter, Global-scale characterization of streamflow extremes [Article], J. Hydrol. 615 (2022), https://doi.org/10.1016/j.jhydrol.2022.128668. Article 128668.
[100] J.-W. Baek, K. Chung, Swin transformer-based object detection model using explainable meta-learning mining, Appl. Sci. 13 (5) (2023) 3213.
[101] Y.S. Patel, S. Banerjee, R. Misra, S.K. Das, Low-latency energy-efficient cyber-physical disaster system using edge deep learning, in: Proceedings of the 21st International Conference on Distributed Computing and Networking, 2020.
[102] W. Guo, S. Huang, Q. Huang, D. She, H. Shi, G. Leng, J. Li, L. Cheng, Y. Gao, J. Peng, Precipitation and vegetation transpiration variations dominate the dynamics of agricultural drought characteristics in China, Sci. Total Environ. 898 (2023), 165480.
[103] Y. Lin, D. Wang, G. Wang, J. Qiu, K. Long, Y. Du, H. Xie, Z. Wei, W. Shangguan, Y. Dai, A hybrid deep learning algorithm and its application to streamflow prediction, J. Hydrol. 601 (2021), 126636.
[104] F.M. Aswad, A.N. Kareem, A.M. Khudhur, B.A. Khalaf, S.A. Mostafa, Tree-based machine learning algorithms in the Internet of Things environment for multivariate flood status prediction [Article], J. Intell. Syst. 31 (1) (2021) 1–14, https://doi.org/10.1515/jisys-2021-0179.
[105] N. Nocentini, A. Rosi, S. Segoni, R. Fanti, Towards landslide space-time forecasting through machine learning: the influence of rainfall parameters and model setting, Front. Earth Sci. 11 (2023), 1152130.
[106] Q.B. Pham, Ö. Ekmekcioğlu, S.A. Ali, K. Koc, F. Parvin, Examining the role of class imbalance handling strategies in predicting earthquake-induced landslide-prone regions, Appl. Soft Comput. 143 (2023), 110429.
[107] H.M. Lyu, Z.Y. Yin, Flood susceptibility prediction using tree-based machine learning models in the GBA, Sustain. Cities Soc. 97 (2023), https://doi.org/10.1016/j.scs.2023.104744. Article 104744.
[108] G. Zhang, M. Wang, K. Liu, Deep neural networks for global wildfire susceptibility modelling, Ecol. Indicat. 127 (2021), 107735.
[109] J. Zhang, X. Ma, J. Zhang, D. Sun, X. Zhou, C. Mi, H. Wen, Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model, J. Environ. Manag. 332 (2023), 117357.
[110] S. Behl, A. Rao, S. Aggarwal, S. Chadha, H. Pannu, Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises, Int. J. Disaster Risk Reduc. 55 (2021), 102101.
[111] G. Bhullar, A. Khullar, A. Kumar, A. Sharma, H. Pannu, A. Malhi, Time series sentiment analysis (SA) of relief operations using social media (SM) platform for efficient resource management, Int. J. Disaster Risk Reduc. 75 (2022), 102979.
[112] Ö. Ekmekcioğlu, K. Koc, M. Özger, Z. Işık, Exploring the additional value of class imbalance distributions on interpretable flash flood susceptibility prediction in the Black Warrior River basin, Alabama, United States, J. Hydrol. 610 (2022), 127877.
[113] H. Hao, Y. Wang, Modeling dynamics of community resilience to extreme events with explainable deep learning, Nat. Hazards Rev. 24 (2) (2023), 04023013.
[114] A.E. Maxwell, M. Sharma, K.A. Donaldson, Explainable boosting machines for slope failure spatial predictive modeling, Rem. Sens. 13 (24) (2021) 4991.
[115] J. Wang, Y. Wang, L. Liu, H. Yin, N. Ye, C. Xu, Weakly supervised forest fire segmentation in uav imagery based on foreground-aware pooling and context-aware loss, Rem. Sens. 15 (14) (2023) 3606.
[116] Q. Ge, H.Y. Sun, Z.Q. Liu, X. Wang, A data-driven intelligent model for landslide displacement prediction, Geol. J. 58 (6) (2023) 2211–2230, https://doi.org/10.1002/gj.4675.
[117] N. Shi, Y. Li, L. Wen, Y. Zhang, Rapid prediction of landslide dam stability considering the missing data using XGBoost algorithm, Landslides 19 (12) (2022) 2951–2963.
[118] A.E. Brower, B. Corpuz, B. Ramesh, B. Zaitchik, J.M. Gohlke, S. Swarup, Predictors of evacuation rates during hurricane laura: weather forecasts, twitter, and COVID-19, Weather, Climate, and Society 15 (1) (2023) 177–193.
[119] Ö. Ekmekcioğlu, K. Koc, Explainable step-wise binary classification for the susceptibility assessment of geo-hydrological hazards, Catena 216 (2022), 106379.
[120] H. Fang, Y. Shao, C. Xie, B. Tian, C. Shen, Y. Zhu, Y. Guo, Y. Yang, G. Chen, M. Zhang, A new approach to spatial landslide susceptibility prediction in Karst mining areas based on explainable artificial intelligence, Sustainability 15 (4) (2023) 3094.
[121] M.C. Iban, A. Sekertekin, Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: a case study of Adana and Mersin provinces, Turkey, Ecol. Inf. 69 (2022), 101647.
[122] S.K. Kuntla, M. Saharia, P. Kirstetter, Global-scale characterization of streamflow extremes, J. Hydrol. 615 (2022), 128668.
[123] A. Maillart, C.Y. Robert, Tail index partition-based rules extraction with application to tornado damage insurance, ASTIN Bulletin: .J.the IAA 53 (2) (2023) 258–284.
[124] V.V. Mihunov, K. Wang, Z. Wang, N.S. Lam, M. Sun, Social media and volunteer rescue requests prediction with random forest and algorithm bias detection: a case of Hurricane Harvey, Environ. Res. Commun. (2023).
[125] R. Otto, S. Susanne, H. Jouni, L. Jukka, Crown snow load outage risk model for overhead lines, Appl. Energy 343 (2023), 121183.
[126] M.A. Rahman, A. Hokugo, N. Ohtsu, Household evacuation preparation time during a cyclone: random Forest algorithm and variable degree analysis, Progress in disaster science 12 (2021), 100209.
[127] C. Xu, K. Zhou, X. Xiong, F. Gao, Y. Lu, Prediction of mining induced subsidence by sparrow search algorithm with extreme gradient boosting and TOPSIS method, Acta Geotechnica (2023) 1–17.
[128] W. Yue, C. Ren, Y. Liang, J. Liang, X. Lin, A. Yin, Z. Wei, Assessment of wildfire susceptibility and wildfire threats to ecological environment and urban development based on GIS and multi-source data: a case study of guilin, China, Rem. Sens. 15 (10) (2023) 2659.
[129] S. Biass, S.F. Jenkins, W.H. Aeberhard, P. Delmelle, T. Wilson, Insights into the vulnerability of vegetation to tephra fallouts from interpretable machine learning and big Earth observation data, Nat. Hazards Earth Syst. Sci. 22 (9) (2022) 2829–2855.
[130] E. Collini, L.I. Palesi, P. Nesi, G. Pantaleo, N. Nocentini, A. Rosi, Predicting and understanding landslide events with explainable AI, IEEE Access 10 (2022) 31175–31189.
[131] H. Jaonalison, J.-D. Durand, J. Mahafina, H. Demarcq, N. Teichert, D. Ponton, Predicting species richness and abundance of tropical post-larval fish using machine learning, Mar. Ecol. Prog. Ser. 645 (2020) 125–139.

[132] Y. Kang, E. Jang, J. Im, C. Kwon, S. Kim, Developing a new hourly forest fire risk index based on catboost in South Korea, Appl. Sci. 10 (22) (2020) 8213.
[133] J. Park, J.C. Joo, I. Kang, W.H. Lee, The use of explainable artificial intelligence for interpreting the effect of flow phase and hysteresis on turbidity prediction, Environ. Earth Sci. 82 (15) (2023) 375.
[134] S. Wang, H. Peng, Q. Hu, M. Jiang, Analysis of runoff generation driving factors based on hydrological model and interpretable machine learning method, J. Hydrol.: Reg. Stud. 42 (2022), 101139.
[135] X. Zhou, H. Wen, Z. Li, H. Zhang, W. Zhang, An interpretable model for the susceptibility of rainfall-induced shallow landslides based on SHAP and XGBoost, Geocarto Int. 37 (26) (2022) 13419–13450.
[136] R. Jena, A. Shanableh, R. Al-Ruzouq, B. Pradhan, M.B.A. Gibril, M.A. Khalil, O. Ghorbanzadeh, G.P. Ganapathy, P. Ghamisi, Explainable artificial intelligence (XAI) model for earthquake spatial probability assessment in Arabian peninsula, Rem. Sens. 15 (9) (2023) 2248.
[137] D. Sun, D. Chen, J. Zhang, C. Mi, Q. Gu, H. Wen, Landslide susceptibility mapping based on interpretable machine learning from the perspective of geomorphological differentiation, Land 12 (5) (2023) 1018.
[138] R.G. Franceschini, J. Liu, S. Amin, Damage estimation and localization from sparse aerial imagery, in: 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), 2021.
[139] P.D. Ogunjinmi, S.-S. Park, B. Kim, D.-E. Lee, Rapid post-earthquake structural damage assessment using convolutional neural networks and transfer learning, Sensors 22 (9) (2022) 3471.
[140] J. Liu, K. Liu, M. Wang, A residual neural network integrated with a hydrological model for global flood susceptibility mapping based on remote sensing datasets, Rem. Sens. 15 (9) (2023) 2447.
[141] X. Li, D. Caragea, H. Zhang, M. Imran, Localizing and quantifying infrastructure damage using class activation mapping approaches, Social Network Analysis and Mining 9 (2019) 1–15.
[142] Y.S. Patel, S. Banerjee, R. Misra, S.K. Das, M. Assoc Comp, Low-Latency energy-efficient cyber-physical disaster system using edge deep learning, in: [Proceedings of the 21st International Conference on Distributed Computing and Networking (Icdcn 2020)]. 21st International Conference on Distributed Computing and Networking (ICDCN), Jadavpur Univ, Kolkata, INDIA, 2020. Jan 04-07.
[143] M. Park, D.Q. Tran, S. Lee, S. Park, Multilabel image classification with deep transfer learning for decision support on wildfire response, Rem. Sens. 13 (19) (2021) 3985.
[144] B. Pradhan, S. Lee, A. Dikshit, H. Kim, Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model, Geosci. Front. 14 (6) (2023), 101625.
[145] W. Yang, H. Yang, D. Yang, Classifying floods by quantifying driver contributions in the Eastern Monsoon Region of China, J. Hydrol. 585 (2020), 124767.
[146] A. Dikshit, B. Pradhan, M.E. Assiri, M. Almazroui, H.-J. Park, Solving transparency in drought forecasting using attention models, Sci. Total Environ. 837 (2022), 155856.
[147] M. Jalal Uddin, Y. Li, M. Abdus Sattar, S. Mistry, Climatic water balance forecasting with machine learning and deep learning models over Bangladesh, Int. J. Climatol. 42 (16) (2022) 10083–10106.
[148] A. Mohammadifar, H. Gholami, S. Golzari, Stacking-and voting-based ensemble deep learning models (SEDL and VEDL) and active learning (AL) for mapping land subsidence, Environ. Sci. Pollut. Control Ser. 30 (10) (2023) 26580–26595.
[149] A. Abdollahi, B. Pradhan, Explainable artificial intelligence (XAI) for interpreting the contributing factors feed into the wildfire susceptibility prediction model, Sci. Total Environ. 879 (2023), 163004.
[150] A.Y. Sun, P. Jiang, M.K. Mudunuru, X. Chen, Explore spatio-temporal learning of large sample hydrology using graph neural networks, Water Resour. Res. 57 (12) (2021), e2021WR030394.
[151] M. Temraz, E.M. Kenny, E. Ruelle, L. Shalloo, B. Smyth, M.T. Keane, Handling climate change using counterfactuals: using counterfactuals in data augmentation to predict crop growth in an uncertain climate future, in: Case-Based Reasoning Research and Development: 29th International Conference, ICCBR 2021, Salamanca, Spain, September 13–16, 2021, Proceedings 29, 2021.
[152] S.T. Seydi, M. Hasanlou, J. Chanussot, P. Ghamisi, BDD-Net+: a building damage detection framework based on modified coat-net, IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. (2023).
[153] J.Y. Wang, Z. Wang, K.K. Li, C. Li, F. Wen, Z.H. Shi, Factors affecting phase change in coupling coordination between population, crop yield, and soil erosion in China's 281 cities, Land Use Pol. 132 (2023), https://doi.org/10.1016/j.landusepol.2023.106761. Article 106761.
[154] G. Antzoulatos, I.-O. Kouloglou, M. Bakratsas, A. Moumtzidou, I. Gialampoukidis, A. Karakostas, F. Lombardo, R. Fiorin, D. Norbiato, M. Ferri, Flood hazard and risk mapping by applying an explainable machine learning framework using satellite imagery and GIS data, Sustainability 14 (6) (2022) 3251.
[155] M. Müller, P.-O. Olsson, L. Eklundh, S. Jamali, J. Ardö, Features predisposing forest to bark beetle outbreaks and their dynamics during drought, For. Ecol. Manag. 523 (2022), 120480.
[156] M. Naser, CLEMSON: an automated machine-learning virtual assistant for accelerated, simulation-free, transparent, reduced-order, and inference-based reconstruction of fire response of structural members, J. Struct. Eng. 148 (9) (2022), 04022120.
[157] S.N. Somala, S. Chanda, K. Karthikeyan, S. Mangalathu, in: Explainable Machine Learning on New Zealand Strong Motion for PGV and PGA. Structures, 2021.
[158] D. Sun, X. Wu, H. Wen, Q. Gu, A LightGBM-based landslide susceptibility model considering the uncertainty of non-landslide samples, Geomatics, Nat. Hazards Risk 14 (1) (2023), 2213807.
[159] A. Tahmassebi, M. Motamedi, A.H. Alavi, A.H. Gandomi, An explainable prediction framework for engineering problems: case studies in reinforced concrete members modeling, Eng. Comput. 39 (2) (2022) 609–626.
[160] E. Štrumbelj, I. Kononenko, Explaining prediction models and individual predictions with feature contributions, Knowl. Inf. Syst. 41 (2014) 647–665.
[161] Y. Zhao, G. Gao, G. Ding, L. Wang, Y. Chen, Y. Zhao, M. Yu, Y. Zhang, Assessing the influencing factors of soil susceptibility to wind erosion: a wind tunnel experiment with a machine learning and model-agnostic interpretation approach, Catena 215 (2022), 106324.
[162] H.E. Aydin, M.C. Iban, Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations, Nat. Hazards 116 (3) (2023) 2957–2991.
[163] R. Silva-Lopez, J.W. Baker, A. Poulos, Deep learning–based retrofitting and seismic risk assessment of road networks, J. Comput. Civ. Eng. 36 (2) (2022), 04021038.
[164] W. Qiao, L. Shen, J. Wang, X. Yang, Z. Li, A weakly supervised semantic segmentation approach for damaged building extraction from postearthquake high-resolution remote-sensing images, Geosci. Rem. Sens. Lett. IEEE 20 (2023) 1–5.
[165] X. Yang, Z. Zhan, J. Shen, A deep learning based method for typhoon recognition and typhoon center location, in: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, 2019.
[166] S. Alqadhi, J. Mallick, S. Talukdar, M. Alkahtani, An artificial intelligence-based assessment of soil erosion probability indices and contributing factors in the Abha-Khamis watershed, Saudi Arabia, Frontiers in Ecology and Evolution 11 (2023), 1189184.
[167] J. Pearl, D. Mackenzie, AI Can't reason why, Wall St. J. (2018).
[168] W.N. Adger, T.P. Hughes, C. Folke, S.R. Carpenter, J. Rockstrom, Social-ecological resilience to coastal disasters, Science 309 (5737) (2005) 1036–1039.
[169] S.L. Cutter, L. Barnes, M. Berry, C. Burton, E. Evans, E. Tate, J. Webb, A place-based model for understanding community resilience to natural disasters, Global Environ. Change 18 (4) (2008) 598–606.
[170] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Nets Advances in Neural Information Processing Systems, 2014 *arXiv preprint arXiv:1406.2661*.
[171] H.A. Al-Najjar, B. Pradhan, Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks, Geosci. Front. 12 (2) (2021) 625–637.
[172] M. Coeckelbergh, AI Ethics, Mit Press, 2020.
[173] L. Taylor, L. Floridi, B. Van der Sloot, Group Privacy: New Challenges of Data Technologies, vol. 126, Springer, 2016.

[174] A. Holzinger, C. Biemann, C.S. Pattichis, D.B. Kell, What Do We Need to Build Explainable AI Systems for the Medical Domain?, 2017 *arXiv preprint arXiv: 1712.09923*.
[175] D. Gunning, D. Aha, DARPA's explainable artificial intelligence (XAI) program, AI Mag. 40 (2) (2019) 44–58.
[176] A. Selbst, J. Powles, "Meaningful Information" and the Right to Explanation. Conference on Fairness, Accountability and Transparency, 2018.