



Review Article: Leveraging Social Media for Managing Natural Hazard Disasters: A Critical Review of Data Collection Strategies and Actionable Insights

Lakshmi S. Gopal¹, Rekha Prabha¹, Hemalatha Thirugnanam¹, Maneesha Vinodini Ramesh¹, and Bruce D. Malamud²

¹Center for Wireless Networks & Applications (WNA), Amrita Vishwa Vidyapeetham, Amritapuri, India

²Institute of Hazard, Risk and Resilience (IHRR), Durham University, Durham, DH1 3LE, UK

Correspondence: Lakshmi S. Gopal (lakshmisgopal@am.amrita.edu)

Abstract. This paper critically reviews 250 articles from 2010 to September 2023, analyzing how social media data is utilized in disaster management, addressing challenges in relevance filtering and noise reduction to extract actionable disaster information and enhance decision-making efficiency. The results of our critical analysis are given in a Social Media Literature Database where we categorize each article's information into 7 main categories and 27 subcategories, covering article details, case study regions, disaster events, social media data specifics, data collection and analysis methods, and evaluation methodologies. To assess the effectiveness of social media in providing actionable disaster information, we further classify the articles into 9 categories, covering public discourse analysis, temporal and spatial insights, relevance filtering methods, community/stakeholder collaborations, disaster trends, and resource identification. We also illuminate historical disaster events within the review period and discuss the results through graphical visualizations. Our findings show that natural language processing methods, particularly content analysis, were commonly utilized in the literature, and contribute significantly to basic data filtering by removing noise. Commonly used advanced robust analysis machine learning methods included Support Vector Machines, Naive Bayes, and Neural Networks. We found that proficiency in temporal and spatial analysis of social media data is widespread among the studies, with varying success in implementing effective relevance filtering. Our actionable information categorization revealed a need for further exploration into community interactions and resource identification using social media data during and after disasters. Based on the literature study and our own experience on the subject, we propose six best practices for social media usage in disaster situations for the community and five best practices for researchers to enhance disaster management strategies.

1 Introduction

In the age of information, social media has become a powerful platform for communication and rapid information dissemination (McCormick et al., 2017; Wang et al., 2018; Li et al., 2018b; Fauzi, 2023). Social media platforms introduced a new direction in assisting in disaster management, enhanced situation awareness, analysing emotions, and community interaction analysis discovering solutions unified with current technologies (Bruns and Liang, 2012; Gerlitz and Rieder, 2013; Kryvasheyev et al.,



2016; Martínez-Rojas et al., 2018). This critical review explores the multifaceted relationship between social media and disaster management, aiming to identify gaps, provide insights, and offer potential future directions.

25 While traditional media sources like newspapers, television, and radio offer reliable information, social media provides distinct advantages, including convenient access to information, interactive community engagement, and diverse situational insights from various perspectives and locations (Chatfield and Brajawidagda, 2013; Dashti et al., 2014; Li et al., 2015; Stieglitz et al., 2018; Wang and Ye, 2018). However, the challenge lies in sifting through the abundance of information to identify trustworthy and pertinent data (Smith et al., 2017; Gulnerman and Karaman, 2020; Srivastava et al., 2020).

30 This is particularly critical in disaster scenarios where the spread of rumours and misinformation is unacceptable (Cenni et al., 2017; Yan et al., 2017). It is also important that the data extracted from social media platforms must be actionable for disaster response, recovery, relief, and rapid decision-making by authorities (Sriram et al., 2010; Li et al., 2015; Cenni et al., 2017). This critical review focuses on the process of discerning relevant and actionable data from social media to enhance disaster response and recovery efforts.

35 Existing literature reviews on social media data (SMD) platform evaluations, data collection tools, and analysis methods over time (Cheng et al., 2016; Shibuya and Tanaka, 2019; Kitazawa and Hale, 2021). These reviews address the utility of social media data across various phases of disaster management. However, limited attention has been devoted to the collection and analysis of topic-relevant data with an emphasis on noise reduction for method enhancement. Even when literature explores topic discovery methods (Volkova, 2014; Čišija et al., 2018; Qarabash and Qarabash, 2018), less focus is placed on assessing the actionability of discovered data in disaster scenarios. This critical review meticulously examines the literature, aiming to
40 establish a classification system for actionable information, thereby assessing the practical value of social media data in disaster management.

The purpose of this critical review is twofold. First, we seek to evaluate the existing literature on the topic of social media usage for managing disasters where we discuss the key findings, and methodologies used for relevance filtering of social media data.

45 Second, we aim to perform an in-depth analysis of how the existing solutions aided in bringing out ‘actionable information’ from social media data. By performing this critical review we aim to shed light on the various methods of social media data analysis to identify pertinent data and to suggest future directions.

Throughout the following sections, we discuss the methodologies used in the existing body of literature, major disaster events in the past decade, and emerging trends, and offer recommendations for future studies. By doing so, we hope to gain a deeper
50 understanding of how social media data analysis can play a relevant role in improving rapid decision-making during a disaster scenario by assisting policymakers, emergency responders, researchers, and the general community.

The manuscript is organised as follows. In section 2 we describe the background. In Section 3 we present the critical review methodology which includes sub-sections detailing research question identification and the steps in constructing our Social Media Literature Database (Gopal et al. (2023)). In Section 4, we bring in the results of the critical review methodology. In
55 Section 5, we critically discuss all the categories in our Social Media Literature Database (Gopal et al. (2023)) to present insightful information and proposes best practices to utilise social media data for the community and researchers to improve disaster management strategies. Finally, in section 6, we summarise our analysis based on the lessons learned.



2 Background

In the last decade, computer scientists have made significant contributions to organizing and evaluating data in disaster management. In this background we focus on critical and systematic review articles within the context of social media on data collection, analysis, relevant data identification, as well as applications, opportunities, and challenges in various phases of disaster management.

In the late 2010s, a seminal work by Hristidis et al. (2010) investigated data integration, information extraction, retrieval, filtering, data mining, and decision support methods in disaster management. This article, among the earlier ones (Imran et al., 2015; Granell and Ostermann, 2016), emphasised the necessity of establishing a disaster management dataspace and outlined associated challenges. Concurrently, Veil et al. (2011) systematically reviewed articles that addressed the creation of risk and crisis management rules and processes, enabling community engagement in decision-making systems.

Several authors emphasise the potential of social media to enhance community interaction across all disaster management phases (Tim et al., 2017; Anson et al., 2017; Nazer et al., 2017). A pivotal work by Landwehr and Carley (2014) extensively explores the roles of the community and organisations in disaster management. They illuminate how the public not only seeks life-saving information but can also contribute to effective information dissemination, fostering community awareness. Additionally, the authors critically review how first responder organisations, including government agencies, increasingly rely on social media data to identify areas in need of assistance during crises.

In recent years, some authors conducted comprehensive reviews and discussions on advanced data acquisition, characterization, and preparation methods (Houston et al., 2015; Spence et al., 2016; Eriksson, 2018; Zhou et al., 2018; Luna and Pennock, 2018; Saroj and Pal, 2020). They examined technical approaches, including API calls and social media querying. These articles also emphasised the importance of data pre-processing, geolocation identification, and geocoding for effectively identifying affected areas. One of the early articles in the literature database, Imran et al. (2015), provided detailed insights into disaster event detection using social media data, covering the aforementioned methods.

One of the prominent challenges addressed in these review articles is the security and quality of data, particularly its relevance in disaster management (Simon et al., 2015; Lin et al., 2016; Said et al., 2019; Acikara et al., 2023). Data quality is often associated with data credibility, where concerns about rumours or false information arise (Reuter and Kaufhold, 2018; Jurgens and Helsloot, 2018). A significant article authored by Haworth and Bruce (2015) discusses the risks associated with data from untrained individuals with diverse agendas and expertise, emphasising the lack of quality assurance. The article underscores the critical importance of accurate information in disaster scenarios, as a delay due to waiting for official notification could have life-threatening consequences and pose a significant danger.

In their study, Nazer et al. (2017) focused on various misleading content, including rumours, spam, and bot-generated data, emphasising the necessity of filtering such content before its utilisation in any phase of disaster management. The article also observed that social media users' language changes due to distress. The victim's cognitive processing and social orientation during or after a disaster lead to changes in language at the sentence or topic level. The authors argued that probabilistic methods, such as Latent Dirichlet Allocation (LDA), can be employed to extract the probability of each topic in a document.



Recent reviews have offered comprehensive insights into managing disaster scenarios using advanced AI technologies and social media data (Pender et al., 2014; Yu et al., 2018; Goswami et al., 2018; Eckert et al., 2018; Akter and Wamba, 2019; Vongkusolkiet and Huang, 2021; Aboualola et al., 2023). In a pivotal work by Imran et al. (2020), the article highlights the mul-
95 timodal nature of SMD, encompassing text, images, videos, relevant URLs, and additional metadata. These modalities often provide complementary information that, when collectively analysed, can significantly enhance the understanding of a crisis. The article delves into AI and ML-based solutions for effectively handling multimodal data.

Several authors conducted bibliometric analyses in the aforementioned field. In a significant article by Tang et al. (2021), it was noted that the use of social media data for disaster management achieved stability during the period from 2015 to 2019. The
100 authors observed that advanced technologies in NLP, ML, and computer vision emerged as key focal points in disaster management. Additionally, Fauzi (2023) conducted a bibliometric analysis to analyse and predict future trends and developments in social media applications.

An important finding derived from the review articles is that social media significantly facilitates community interactions in crises (Alexander, 2014; Reuter and Kaufhold, 2018). However, the primary concern lies in the unreliability of the data, which
105 hinders its practicality as actionable information during disasters (Beigi et al., 2016; Palen and Hughes, 2018; Zhang et al., 2019a; Imran et al., 2020).

3 Critical Literature Review Methodology

To construct the Social Media Literature Database (Gopal et al. (2023)), we conducted a critical review of pertinent English-language articles using 'social media' and 'disaster management' related keywords, primarily sourcing content from Google
110 Scholar. The time period covered was from 2010 to September 2023. Section 3.2 details the specific search criteria employed in building the literature database. A two-stage screening process was implemented: an initial assessment based on titles and abstracts to shortlist relevant articles, followed by a critical review of the selected articles to confirm their relevance to the research topic.

We have taken elements from Boaz et al. (2002) to follow a specific protocol for the critical literature review:

- 115 i Focusing on answering a specific question(s)
- ii Seeking to identify relevant research
- iii Synthesising the research findings in the studies included
- iv Aiming to be as objective as possible about research to remove bias

In this paper, we followed a critical literature review with four major stages as shown in figure 1 and each stage is described in
120 the following sub-sections.

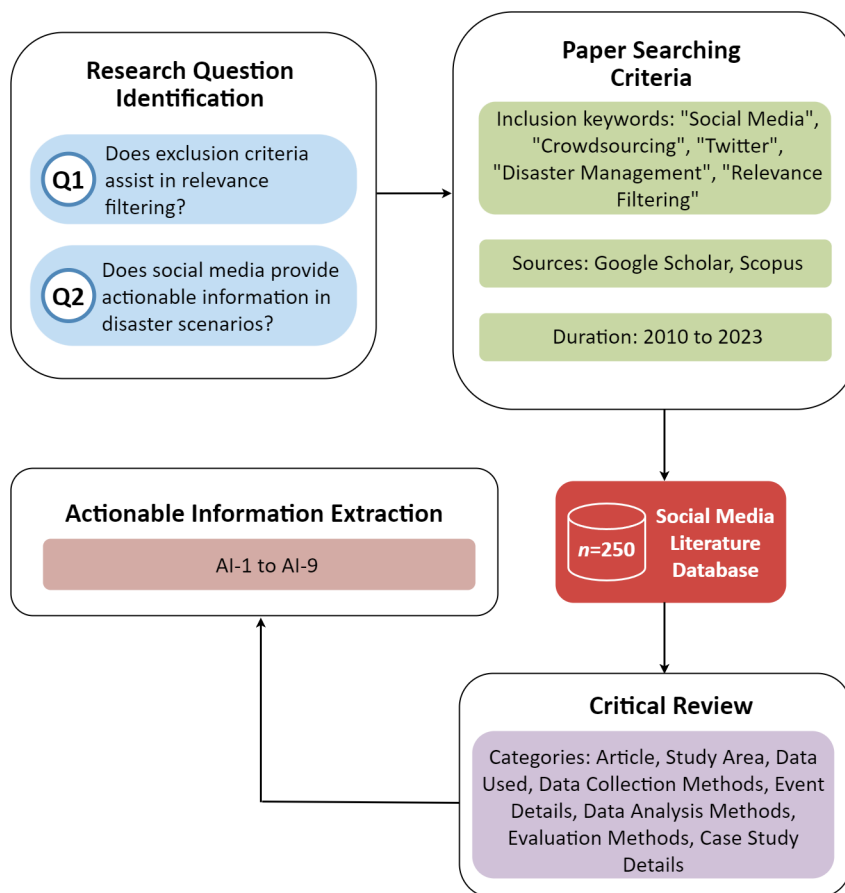


Figure 1. The block diagram summarises the critical review methodology with four major stages and their respective details.

3.1 Research Question Identification

In a hazard scenario, Volunteered Generated Information (VGI) generated through social media is advantageous, but due to lack of reliability and increased generation of data, rapid decision-making is affected (Black et al., 2012; Ashktorab et al., 2014; Radianti et al., 2016; Yan et al., 2017; Kankanamge et al., 2020b). Considering these issues, we have derived the following
125 research questions.

- Q1: Do exclusion criteria assist in relevance filtering of social media data?

Exclusion criteria are conditions used to eliminate irrelevant data, facilitating relevance filtering. These criteria often consist of keywords or phrases related to topics to be excluded from large datasets and are considered a straightforward method for reducing irrelevant data.

130 We analyse articles in the literature that utilise social media data to identify the methods applied for relevance filtering. Our specific focus is on articles that have experimented with exclusion criteria in data collection. Our objective is to



determine whether Natural Language Processing (NLP) methods can effectively aid in relevance filtering for basic data collection or if Machine Learning (ML) methods are required.

– Q2: Does social media provide actionable information in disaster scenarios?

135 A significant drawback of social media data is its credibility (Win and Aung, 2017; Ravi Shankar et al., 2019; Nair et al., 2022; Loynes et al., 2022). Social media users encompass various categories, including public users, government organisations, NGOs, public figures, and news media. During a disaster scenario, government, non-government organisations, and news media typically provide trustworthy information about the crisis. However, public posts may also include valuable emergency information from actual victims, often in the form of photos or videos (Khaleq and Ra, 2018; Banujan et al., 2018).

140 Inaccurate information may be disseminated, whether intentionally or unintentionally, including the spread of rumours or discussions about similar disaster events occurring elsewhere (Remy et al., 2013; Musaev et al., 2018; Arapostathis, 2021). This challenge underscores the difficulty in identifying relevant data that can be considered actionable. In this context, actionable information is defined as data that facilitates prompt decision-making in disaster scenarios.

145 We have defined various forms of actionable information from social media data, as detailed in Section 4.3. We reviewed the articles in the database to ascertain if they proposed solutions for extracting actionable information. Our objective is to gain a comprehensive understanding and determine whether social media indeed contributes to effective disaster management by providing pertinent information for rapid decision-making. By addressing these research questions, we also aim to offer optimal guidance for investigators regarding the extent to which social media contributes to disaster management research.

150 To address the above research questions, we bring in 7 main categories in our critical review literature database where data related to the following questions will be placed:

- (a) What are the methods opted to collect disaster-related social media data?
- (b) What are the existing methods of relevance or domain filtering of social media data, within and outside disaster scenarios?
- (c) What are the methods of exclusion criteria usage for relevance filtering?
- (d) Does the literature further analyse the exclusion criteria to avoid missing data and not to include irrelevant data?
- (e) What are the existing data analysis methods used, specifically using ML and NLP?
- (f) Does the literature address the issue of false information dissemination?
- (g) What approaches have the articles introduced to identify, analyse, and extract actionable information?

160

3.2 Paper Searching Criteria

To construct the Social Media Literature Database (Gopal et al. (2023)), we searched articles in Google Scholar and Scopus online platforms using the keywords related to disaster management and data science (see figure 1) and created five Boolean

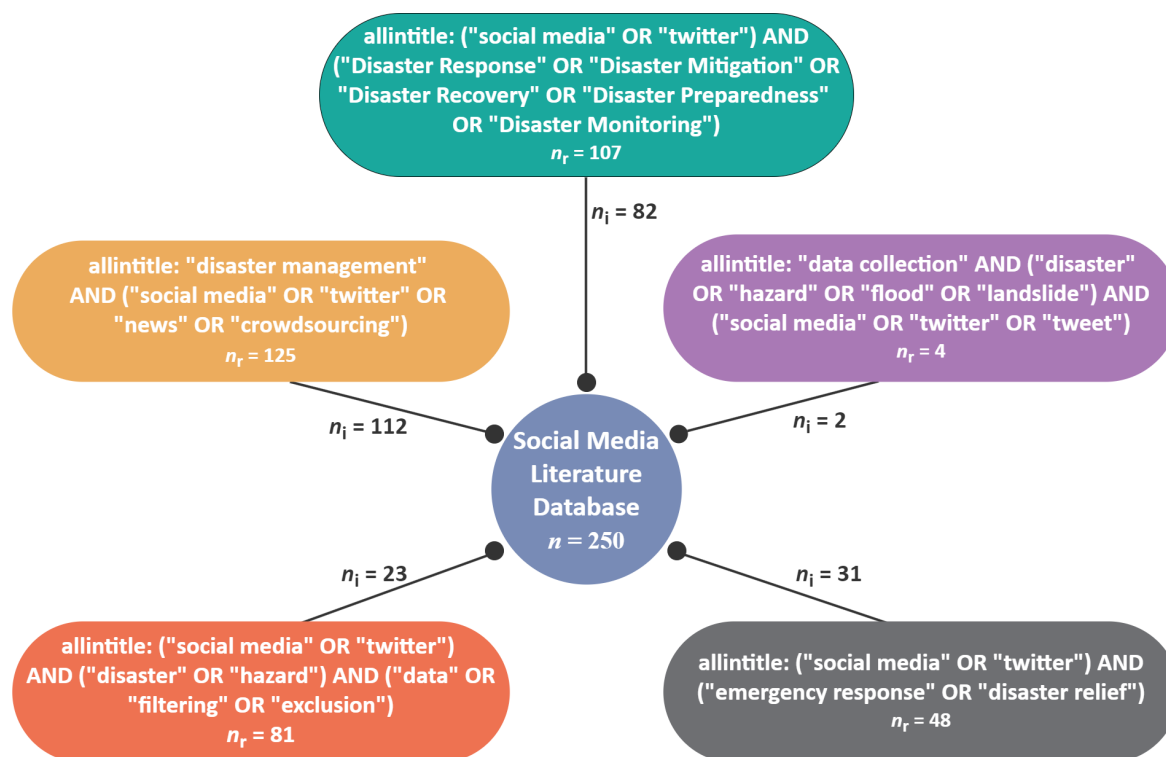


Figure 2. Boolean search strings used for article collection to list in the literature database - The search strings are applied on the article title for searching in Google Scholar (Last queried on September 2023). n_r represents the number of resultant articles of each Boolean search string, n_i represents the number of articles included in the literature database from each search string and n represents the total number of articles in the literature database.

search strings applied to the title of the articles. Figure 2 shows the Boolean search strings used to collect the articles to create a literature database. We have used five boolean search strings and each string generated a certain number of articles, depicted as n_r . After an analysis of these articles, we included a certain number of articles in the literature database, depicted as n_i . Following the application of search terms, each resulting article was manually reviewed to determine its relevance, specifically focusing on social media, disaster management, or relevance filtering. It is important to note that we did not implement any exclusions during the collection process. In total, we have included 250 articles in our database, selecting various data types for extraction to compile the database.

We compiled the database by considering peer-reviewed journals, conference papers, and reports. The articles included in the database span from 2010 to September 2023. The majority of these articles are associated with disaster management, with a subset related to social media data analytics and social science, where methodologies for relevance filtering using social media data are employed.

During phase I of paper selection for the critical review, we used the disaster-related keywords in the titles as described in the



boolean search strings (see figure 2), and then reading the abstract of the article for applicability. There may however be other keywords that we did not use which would have identified further relevant literature. For instance, a seminal work by Niles et al. (2019) is indexed in Google Scholar and can be found when searched with the keywords 'social media' and 'natural hazards'. However, using the developed boolean search strings, this article did not get listed in the database, as we missed it
180 due to the absence of relevant keywords with plural combinations.

For example, the keyword search in Google Scholar for (("Social Media") AND ("Natural Hazard" OR "Natural Hazards")) identifies 30 articles. Of these 30, after reading the abstracts, 7 would have been included in our Database, but only one actually was included. Examples of those not included are "Rapid flood inundation mapping using social media, remote sensing and topographic data" (Rosser et al., 2017), "Sub-event discovery and retrieval during natural hazards on social media data" (Wu
185 et al., 2016), "Detecting Natural Hazard-Related Disaster Impacts with Social Media Analytics: The Case of Australian States and Territories" (Yigitcanlar et al., 2022), and "Public Attention to Natural Hazard Warnings on Social Media in China" (Hu et al., 2019).

3.3 Synthesis of Research Findings

We extracted seven major categories, each containing multiple sub-categories, which were added to the database, as depicted
190 in Figure 3. For each of the 250 papers, we conducted reviews to identify information that could be assigned to these seven categories and their respective subcategories. In addition to this data, we also conducted an actionable information analysis of the articles, as detailed in Section 4. In the following subsections, we provide descriptions of the collected information under each category and its subcategories, along with an explanation of the data grouping procedure for each category.

3.3.1 Category A: Article Description

195 In the Social Media Literature Database, article description (A) comprises nine unique sub-categories, as depicted in Figure 3. Subcategory A1, 'ID,' provides a unique identification number for each article, while subcategory A2, 'Title,' lists the article's title in the database. Subcategory A3, 'Author(s),' denotes the authors of each article, with a range of 1 to 4 authors per article. Subcategory A4, 'Theme,' categorises each article into one of the themes: 'Disaster Management,' 'Social Media Analytics,' or 'Social Science,' with each article assigned to a specific category.

200 Subcategory A5, 'Year,' indicates the publication year of the article. Subcategories A6 and A7 specify the kind and type of article, respectively. 'Kind of article' classifies an article as 'Journal,' 'Conference,' 'Report,' or 'Book,' while 'Article Type' designates an article as either 'Survey' or 'Other.' In this context, 'Survey' includes survey or review papers related to each theme, and 'Other' encompasses articles with analytical, experimental, or interpretative characteristics. Subcategory A8, 'Publication Name,' details the title of the journal, conference, or book series in which the article was published. Finally,
205 subcategory A9, 'Citations,' reflects the number of citations an article has accumulated as of September 2023.

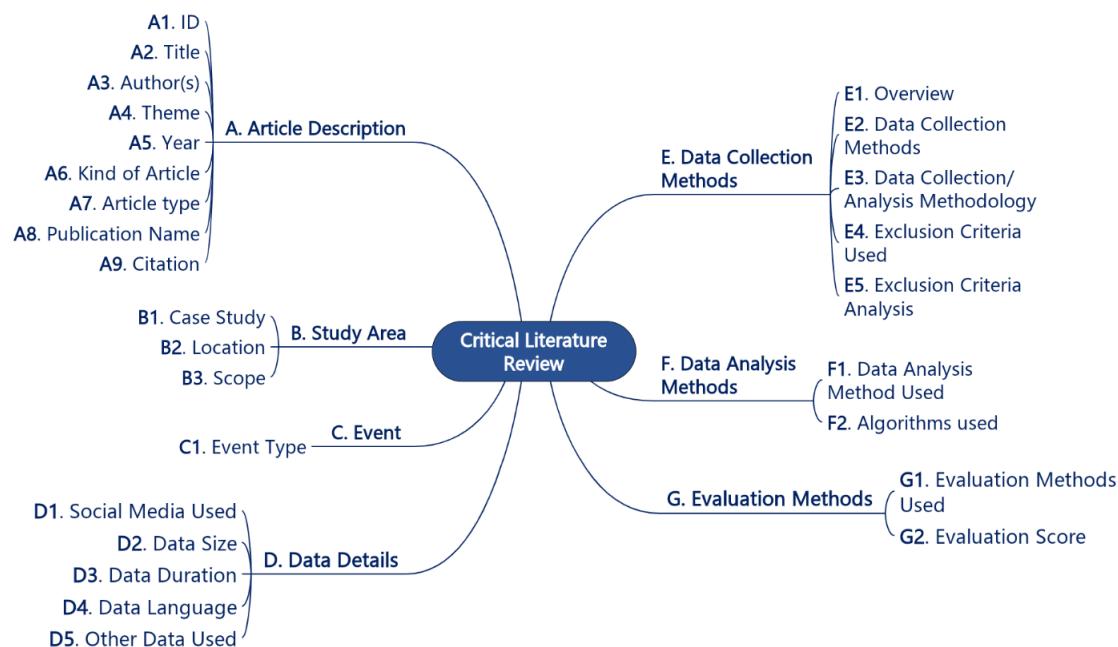


Figure 3. The mind map summarises the seven main categories and their respective subcategories populated in the critical review database.

3.3.2 Category B: Study Area

In the literature database, Category B provides study area details with three subcategories. Subcategory B1, 'Case Study,' is a binary field (yes or no) indicating whether an article employs a case study to support its methodology. If an article uses a case study, we fill subcategory B2, 'Location,' with the relevant place names in focus.

210 An article may incorporate case studies related to disaster events affecting areas at various spatial scales, as illustrated in Figure 4. This figure delineates the spatial scales considered in the study, along with associated terminology and their characteristics. Case studies are categorized into subcategory B3, labeled 'Scope,' which characterizes geographic scales using the terminologies 'national,' 'regional,' and 'local.' In the 'national' category, articles focus on cases where the disaster event impacts one or more countries. The 'regional' and 'local' categories group articles with case studies specific to disaster events within states, 215 districts, large counties, cities, and towns. Figure 4 offers an overview of data characteristics and spatial scales corresponding to each scope terminology. For categorization, 'regional' encompasses countries, states, districts, counties, and provinces within a range of a million km². A 'local' scope includes towns, cities, or villages within a range of ten thousand km².

3.3.3 Category C – Event

We categorise the type of disaster event in subcategory C1, 'Event Type,' which includes categories like 'Landslide,' 'Flood,' 220 'Earthquake,' 'Hurricane,' or other natural or man-made disaster event names. Some articles focus on disaster management

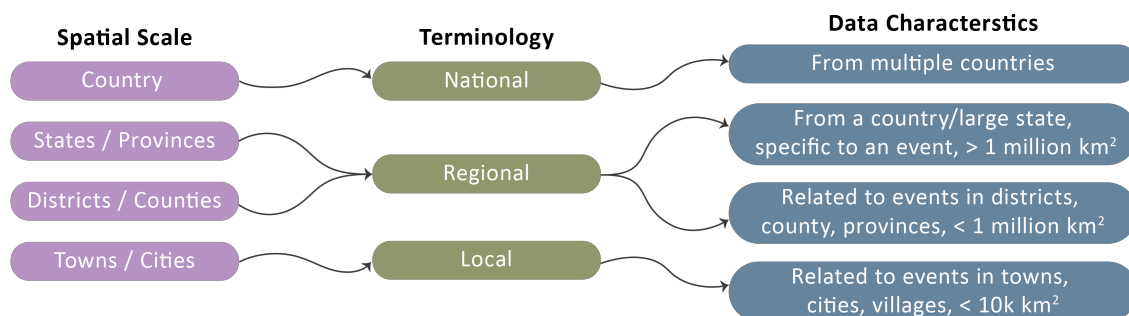


Figure 4. Defining study area scope categorically by geographic scale namely, national, regional and local with their respective spatial scale and data characteristics.

methodologies applicable to any disaster and are labeled as 'General Emergency.' Additionally, articles related to social media analytics and social science, not necessarily in the disaster management domain, are categorised as 'Other' in the event type field.

3.3.4 Category D - Data Details

225 In this category, we record the details of data used in the articles within the database under five subcategories, as depicted in
 Figure 3. Subcategory D1, 'Social Media Used,' indicates whether an article utilises social media data, with options 'Yes' or
 'No.' If an article uses social media data, we populate the amount of data, duration under consideration, and the data language
 in subcategories D2, 'Data Size,' D3, 'Data Duration,' and D4, 'Language,' respectively. Some articles combine social media
 data with official government records of a disaster event as supporting data, and details of such data are entered in subcategory
 230 D5, 'Other Data Used'.

3.3.5 Category E - Data Collection Methods

This category extensively examines the data collection and filtering methods employed in each article. We begin by providing
 an overview of the methodology in subcategory E1, 'Overview.' Subcategory E2, 'Data collection method used,' receives a
 'Yes' or 'No' value, indicating whether an article employs a specific data collection methodology. Some articles utilise existing
 235 social media corpora rather than developing their own data collection methodology, and this is noted. When an article defines
 a data collection methodology, a summary is provided in subcategory E3.

Given our focus on relevance filtering of social media data, subcategories E4 and E5, 'Exclusion Criteria Used' and 'Exclusion
 Criteria Analysis,' indicate whether an article utilises criteria to exclude irrelevant data and whether they analyse their filtering
 methodology to address the challenge of potentially missing relevant data.



240 3.3.6 Category F - Data Analysis Methods

We document the data analysis methods employed in each article in subcategory F1, 'Data Analysis Method Used,' which may encompass 'NLP,' 'ML,' 'Statistical,' and other methods. The specific algorithms used for each analysis method are detailed in subcategory F2, 'Algorithms Used'.

3.3.7 Category G - Evaluation Methods

245 For each of the algorithms mentioned in Category F2, we document the evaluation methods employed in subcategory G1, 'Evaluation Method Used,' which may include commonly used scoring metrics. The respective average scores for these evaluation methods are recorded in subcategory G2, 'Evaluation Score'.

In the following section, we further analyse the results of the Critical Review Literature Database.

4 Results

250 In this section, we present the results and findings of the social media literature database construction. In the following subsections, we present a detailed analysis across several key dimensions, including early works, publication trends, article classification, data collection methodologies, relevance filtering strategies, and actionable information extraction.

4.1 Overview of Literature Database Construction

Our critical review methodology identified a total of 250 articles published in journals, conferences, reports, and books, all of which are listed in the Social Media Literature Database. Figure 5 provides an overview of the total number of papers and the total citations per year from 2010 to September 2023. Approximately 90% of the articles were sourced from Google Scholar, with the remaining 10% obtained from Scopus.

Over the past decade, many authors (Sakaki et al., 2012; Carter et al., 2014; Gunawong and Butakhieo, 2016; Stephenson et al., 2018; Bunney et al., 2018; Brangbour et al., 2019; Brangbour et al., 2020; Podhoranyi, 2021) have conducted experiments in social media data collection and analysis as depicted in figure 5. Initially, while the number of papers was relatively low, there were a significant number of citations. However, in the literature database, we observe a substantial increase in both papers and citations from 2014 to 2018. During the last ten years, a wide range of articles, including journals, conference proceedings, reports, and book chapters, have been published due to the growing use of web data in various phases of the disaster management cycle.

265 Our critical review encompasses not only journal articles but also conference proceedings, reports, and book series chapters. This choice is driven by the fact that these sources often provide insights into the development of social media data collection, which includes filtering, a core aspect of our review. Figure 6 illustrates the distribution of articles among the categories: 'Journal,' 'Conference,' 'Report,' and 'Book,' with the majority of articles falling under the 'Journal' category. We can also observe that 2018 is the year when journals and conference papers were maximum. Reports and book chapters are comparatively less

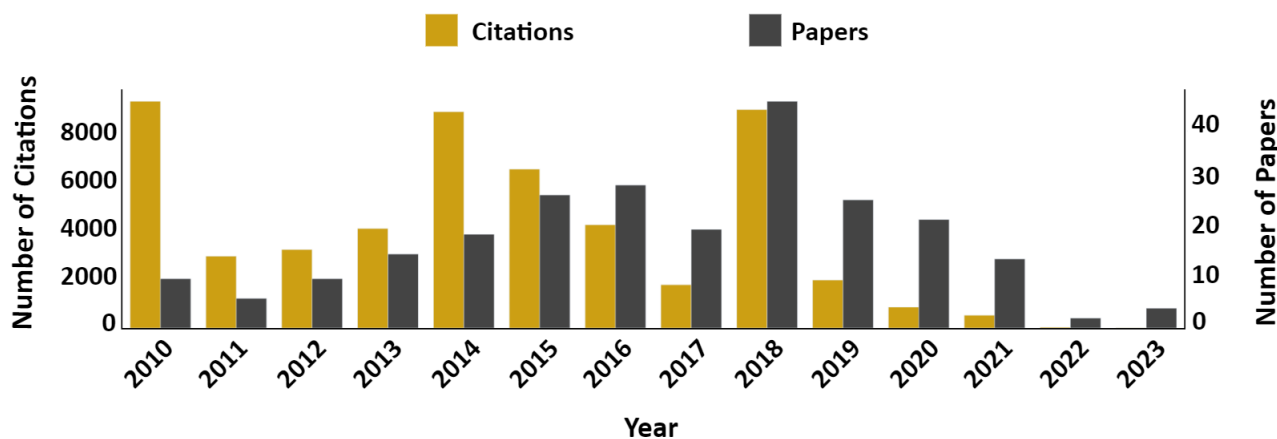


Figure 5. The figure shows the number of papers (dark grey, secondary y-axis) and total citations (dark yellow, primary y-axis) per year, in the review duration (2010 to September 2023).

270 but provide insights into data collection and analysis strategies.

4.2 Early Works

This section provides an overview of the early works and discusses the current state of research regarding the use of relevance filtering methods in social media data.

275 4.2.1 Social Media for Disaster Management

From the literature, we have identified four major categories of articles that obtain and utilise User Generated Information (UGI) for disaster management.

I Articles that obtained UGI based on surveys and questionnaires

II Articles that discuss why social media data can be considered an effective UGI for managing disasters

280 III Articles that obtained and utilised social media data for disaster management, particularly Twitter data and,

IV Articles that use social media data and perform advanced analysis using ML and NLP

In category I, we examine articles that collect UGI through on-site surveys and questionnaires conducted in disaster-affected areas (LÓPEZ-MARRERO, 2010; Aisha et al., 2015; Anson et al., 2017). These responses are subsequently used for enhancing future preparedness, mitigation, and awareness initiatives. A substantial body of research began leveraging public data for disaster management analysis across all phases, primarily through post-disaster surveys and questionnaires (Islam and Walkerden, 285

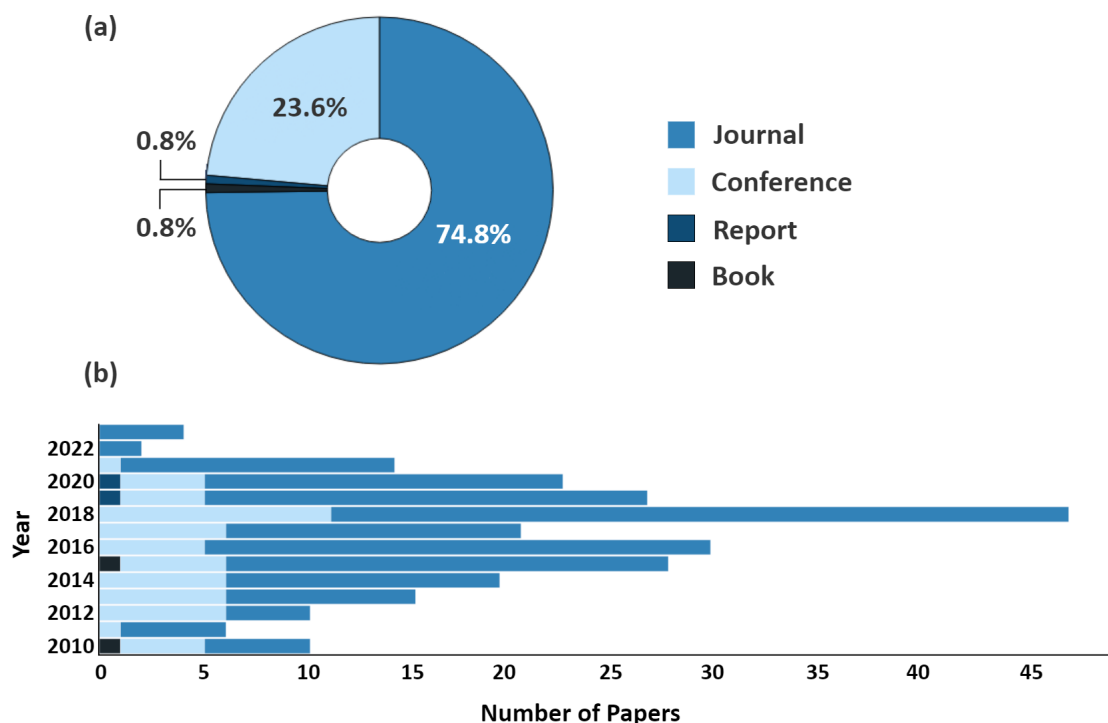


Figure 6. The figure shows the number of articles under the categories ‘Journal’, ‘Conference’, ‘Report’, and ‘Book’. (a) - The pie chart shows the total percentage distribution of the categories listed in the database; (b) - The horizontal bar chart shows the number of articles under each category year-wise.

2015; Ferris et al., 2016). The overall objective was to generate data that enables government and non-government organisations to collaborate, plan, and execute post-disaster missions and enhance future mitigation efforts. Such data is considered credible when compared to the data extracted from social media platforms as may include false information.

Generally, the questionnaire surveys were administered among local residents and government officials, and school authorities, to quantitatively assess public awareness of natural disasters and propose recommendations to improve awareness and their impacts. They focused on schools and the local communities to help in future mitigation and preparedness. They also performed damage estimations using the data acquired through surveys and interviews of the affected population and discussed the preventive measures they undertook and the damages caused at their homes.

They also highlighted the potential future directions where such data can be used for effective disaster mitigation and preparedness. The articles under this category indicated that data from actual victims, and government or non-government authorities collected through a questionnaire or a survey is reliable and can be used for future preparedness (Tandoc Jr and Takahashi, 2017; Albris, 2018; Delilah Roque et al., 2020).

In Category II, we explored articles that examined the potential of social media data as a source of UGI for disaster management. During the late 2010s to early 2013s, authors discussed methodologies for collecting social media data and its role in



300 enhancing post-disaster scenarios (Gao et al., 2011; Abel et al., 2012; Imran et al., 2013b). Notably, the use of platform-specific APIs for data collection was a common practice, alongside the development of open-source platforms that gathered social media data and generated crisis maps, facilitating response and recovery efforts.

These articles also emphasised the active engagement of people on platforms like Twitter and Weibo during disasters and their role in information dissemination. They showcased the utilisation of social media and various free web-based platforms to
305 enhance awareness of humanitarian assistance and disaster relief. In this category, the focus primarily revolved around considering social media data for disaster management due to its widespread availability, user accessibility, its role in raising hazard awareness and enhancing situational awareness.

In Category III, we reviewed articles that gathered and analysed social media data to enhance disaster management. These articles investigated the utility of social media data through historical disaster case studies, demonstrating its potential for
310 improving disaster management. Notably, in 2011 and 2012, research began delving into the use of social media data for the analysis of disaster events, with a particular focus on early warning, response, and recovery.

Several authors (Choi and Bae, 2015; Huang and Xiao, 2015; Ogie et al., 2019; Son et al., 2019; Singh et al., 2019; Fan et al., 2020) developed frameworks for collecting and analysing VGI, particularly Twitter data, to improve the reliability of data in crisis management. They explored current trends, identified hotspots, and gathered community-level insights. Their analysis
315 revealed increased Twitter activity during and after hazard events, with higher post frequencies in proximity to these events. These studies also highlighted the role of emergency communication within the context of social networks, serving as a responsive channel during disasters and fostering a platform for an emergent community of responders and victims.

Since 2013, a substantial body of research (Olteanu et al., 2015; Wang et al., 2016; Han et al., 2020) has concentrated on collecting and analysing social media data to enable rapid decision-making in emergencies. This led to the creation of category IV,
320 where advanced ML and NLP methods were applied to extract insightful inferences from social media data. During this period, researchers recognized the significance of deep text analysis in identifying pertinent data, considering the diverse languages, informal writing styles, dialects, and the need to preserve contextual nuances in social media posts.

Figure 7 illustrates the evolution of research on the use of social media and web data for disaster management over the years. The articles in consistently affirm the utility of social media data in disaster management while acknowledging its primary
325 limitation – reliability. In 2010 and early 2011, there was active use of surveys and questionnaires for data collection, which decreased after 2011. Similarly, during this period, articles discussed the potential of social media data for disaster management, with a decline observed after 2011. Categories 3 and 4, focusing on the effective use of social media data and the implementation of ML and NLP methods for deeper analysis, remained active from late 2011 to 2022. It is important to note that some articles may belong to multiple categories, discussing both the relevance of social media and its practical use. Figure
330 7 provides a detailed representation of this analysis.

4.2.2 Previous Works on Social Media Analytics

As part of our research focus on examining the existing relevance filtering methodologies for social media data, we expanded our scope beyond disaster management-related articles to encompass those related to social media analytics. This approach

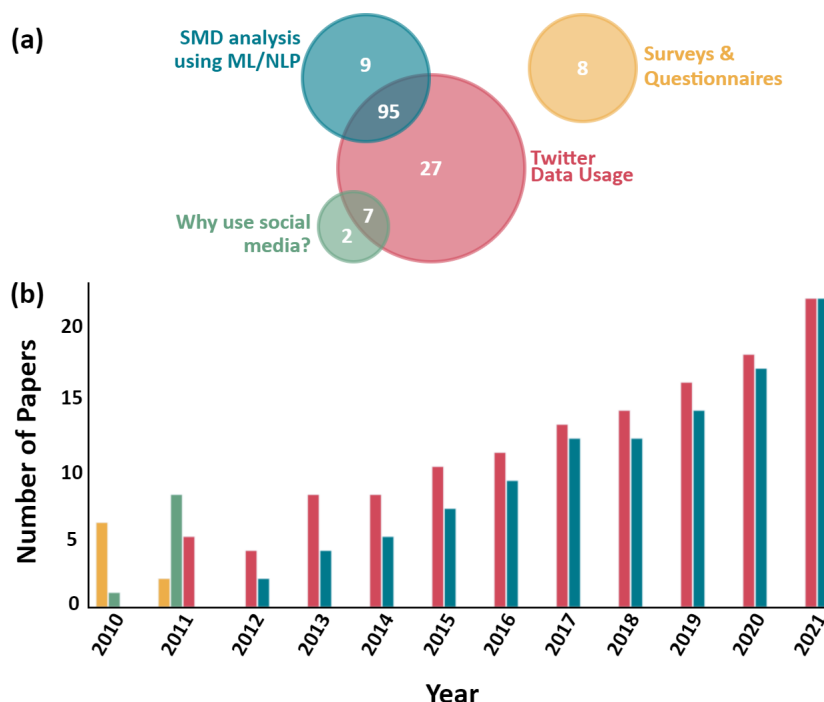


Figure 7. The figure summarises the existing usage of social media data for disaster management. (a) - The Venn diagram shows the number of papers in one or more categories; (b) - The grouped bar chart shows the number of papers in each category year-wise.

allows us to cover both the technical aspects of data collection and analysis. Among the 250 articles in the literature database, 15.6% (39 of 250) fall within this category. In this section, we provide a brief overview of a few key articles.

Since 2010, several authors conducted experiments utilising advanced methods for identifying, acquiring, filtering, and analysing relevant data (Gerlitz and Rieder, 2013; Tang et al., 2015; Batrinca and Treleaven, 2015; Steiger et al., 2015; Eilander et al., 2016; Ilieva and McPhearson, 2018; Zhang et al., 2018; Bukar et al., 2022). Approximately 95% of the articles within the literature database relied on Twitter data for their experiments, while the remaining 5% utilised data from other web-based sources (Facebook, YouTube, Flickr, Weibo) and incorporated public interviews or surveys obtained through manual efforts.

Several authors (Cameron et al., 2012; Black et al., 2012; Oussalah et al., 2013; Schempp et al., 2019; Kejriwal and Gu, 2019; St Denis et al., 2020; de Oliveira and Guelpeli, 2020) developed software frameworks capable of querying the Twitter Streaming API through a series of independent search jobs on an ongoing basis. The extracted data is stored in a well-structured database. These frameworks also describe the query creation process, enabling software to search and collect Twitter data based on specific criteria, including topic-specific keywords, user-specific posts, location-specific data (defined by bounding boxes), and date-specific parameters. Such architecture proves invaluable in disaster scenarios where precise location-specific data acquisition is crucial (Abel et al., 2012; Muhammad et al., 2018; Yang et al., 2019a; Ghawana et al., 2021).

In the later years, a few authors (Gautam and Yadav, 2014; Wachowicz et al., 2016; Huang et al., 2016; Huang et al., 2018;



Mazoyer et al., 2018; Suzuki, 2019; Brena et al., 2019; Hao and Wang, 2020; Domala et al., 2020; Yuan et al., 2021; Akhter
350 et al., 2021; Jiang et al., 2022) intensively examined the usage of NLP and AI technologies to analyse the social media data to
extract insightful information that can be applied in various disciplines.

These articles explored the use of public posts that contain opinions and emotions to facilitate solution development. They
conducted feature extraction by examining adjectives in tweets, which reflected the sentiment of individuals as positive, nega-
355 tive, or neutral. The articles employed probabilistic classifiers like Naive Bayes and maximum entropy to determine sentiment.
Sentiment analysis holds great relevance in disaster management scenarios as it contributes to community analysis and feed-
back during or after a disaster (Mandel et al., 2012; Neppalli et al., 2017; Ragini et al., 2018; Wu and Cui, 2018; Reynard and
Shirgaokar, 2019; Pourebrahim et al., 2019; Yabe and Ukkusuri, 2019; Karimiziarani and Moradkhani, 2023).

Similarly, some authors investigated the behaviours of social media users engaged in the consumption and dissemination of
news items (Lachlan et al., 2010; Houston et al., 2012; Liu and Stevenson, 2013; Kaewkitipong et al., 2016; Valenzuela et al.,
360 2017; Kim et al., 2018; Verma et al., 2019; Yeo et al., 2022). They provided detailed insights into the distribution and statistics
of Twitter users who shared or retweeted topic-relevant news articles. Behavioural analysis of the public on social media holds
significance in crisis management, as it enables an examination of information dissemination patterns (Mendoza et al., 2010;
Kim, 2014; Chae et al., 2014; Spence et al., 2015; Hara, 2015; Kibanov et al., 2017; Jitkajornwanich et al., 2018).

While investigating public behaviour, several researchers deemed the language usage of social media users as relevant (Lee
365 et al., 2011; Abel et al., 2012; Reuter and Schröter, 2015; Carley et al., 2016a; Xu et al., 2019b). They argued that enhancing
methods for the analysis of non-English social media data is essential to improve global community interaction. Various en-
sembles of ML models and NLP algorithms have been utilised to analyse non-English content. Each of these articles discussed
how every language varies in writing style, grammar, and local usage. In disaster scenarios, the inclusion of posts written in
local languages aids in analysing community-level needs.

370 We observed that, over the past decade, researchers have increasingly advocated the use of NLP and AI technologies for social
media data acquisition and analysis across various disciplines. However, a recurring challenge in every article is the issue of
data reliability, which has prompted the inclusion of supplementary data sources such as government records or news reports
to substantiate their findings.

4.3 Actionable Information (AI) Analysis

375 To address our research question “Does social media provide actionable information (AI) in disaster scenarios?”, we analyse
the articles listed in the Social Media Literature Database under the theme of “Disaster Management”. By studying various
articles (Palen et al., 2010; Sakaki et al., 2010; Zhou et al., 2013; Jongman et al., 2015; Musaev et al., 2018; Phengsuwan et al.,
2019; Guntha et al., 2020b; Gopal et al., 2020; Guntha et al., 2020a; Gopal et al., 2022; Aswathy et al., 2022) and based on
our experience, we have defined nine generic actionable information categories which will be assigned to each article under
380 the “Disaster management” theme listed in the literature database.

An article can fall into multiple AI categories, and Table 1 outlines these nine categories. These classifications center around
data collection methods, geolocation identification, relevance filtering strategies, community and stakeholder collaborations,



and software development. Table 1 displays the various categories with their respective descriptions, detailing the methods and applications considered within each AI category in this study. Additionally, we include references for articles under each AI category that have garnered higher citations compared to others in the same category.

Table 1: Actionable Information (AI) analysis - table describes the 9 AI categories. Each article listed in the Social Media Literature Database belonging to the “Disaster Management” theme gets grouped under one or more AI categories. The ‘References’ column shows the articles with high citations under an AI category for reference.

AI	Category	Description	References
1	Disaster Data Collection	APIs or other programming used for data collection, temporal or distribution analysis performed, explores communication of the public over the course of the disaster.	Middleton et al. (2013), Radianti et al. (2016), Granell and Ostermann (2016)
2	Geolocation Detection and Analysis	Methods of geolocation identification or geocoding from tweet content or user details or from geotagged data, performs spatial analysis of their respective research.	Kryvasheyev et al. (2016), Cenni et al. (2017), Kankanamge et al. (2020b)
3	Relevance Filtering	Uses appropriate keywords for collecting data, irrelevant data exclusion (advertisements, social media posts discussing older disaster events that has no relevance in the current events, exclude topics with word similarity, but context dissimilarity), identifies fake data (rumoured disaster events, retweets of rumours posts)	Campan et al. (2018), Abedin and Babar (2018)



Table 1: Actionable Information (AI) analysis - table describes the 9 AI categories. Each article listed in the Social Media Literature Database belonging to the “Disaster Management” theme gets grouped under one or more AI categories. The ‘References’ column shows the articles with high citations under an AI category for reference.

AI	Category	Description	References
4	Community Collaborations	Methods of SMD usage in improving community awareness, disseminating or identifying information such as methods of preparation - rescue camps, sources of food and medication, financial needs, financial claims, emergency information, examining community interactions over social media, tracking people’s emotions pre/during/post a disaster occurrence.	Plachouras et al. (2013), Olteanu et al. (2014), Verma et al. (2019)
5	Disaster trends and hotspot identification	Analyses previous disaster events of various locations from social media data, Studies current landscape of a location based on geography, demography or development/urbanisation, predicts the probability of the disaster to occur again in future, Identifying disaster events or performs disaster relevant topic modelling or event classifications.	De Albuquerque et al. (2015), Fohringer et al. (2015)



Table 1: Actionable Information (AI) analysis - table describes the 9 AI categories. Each article listed in the Social Media Literature Database belonging to the “Disaster Management” theme gets grouped under one or more AI categories. The ‘References’ column shows the articles with high citations under an AI category for reference.

AI	Category	Description	References
6	Stakeholder Identification and Collaboration	Identifies community stakeholders, government stakeholders, non-government organisation stakeholders, inter-governmental organisations and volunteers; establishes a strategy for stakeholder network creation; defining methods of collaborative crisis management within stakeholders; generates network graphs and analyses communications between stakeholders/public.	Imran et al. (2013a), Ragini et al. (2018)
7	Open-source software development	Software tool/dashboards/websites/apps, Tool that provides real time or near real time alerts, warnings	Lachlan et al. (2010), Ashktorab et al. (2014), Jitkajornwanich et al. (2018)
8	Resource Identification	Extracts the needs of the public from social media posts, establishes a strategy to be employed by the organisations to allocate resources based on the needs - includes rescue missions, food/water/medication/clothing requirements, performs damage/risk assessment.	Lee et al. (2011)
9	Community Response Analysis	Response/feedback messages from the public post a requirement fulfilment from concerned authorities, behavioural analysis of public, interviews or surveys taken.	Chen et al. (2014), Reuter and Schröter (2015), Huang et al. (2016)

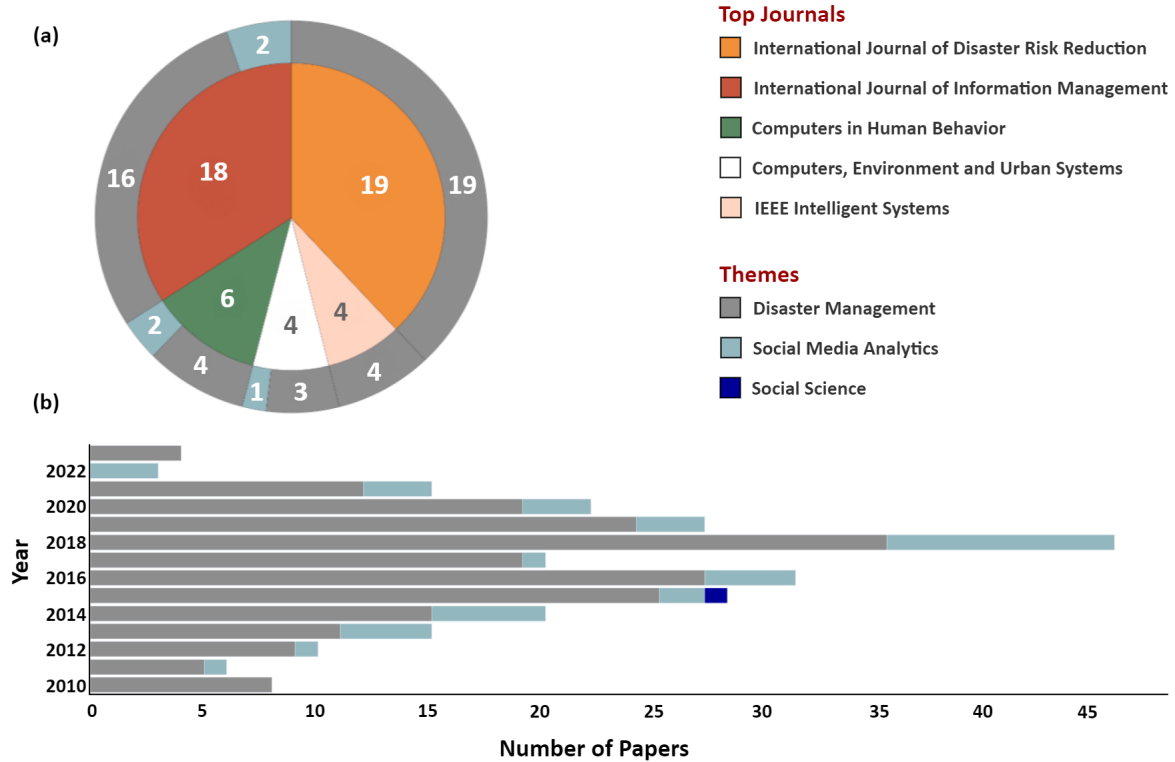


Figure 8. The figure shows the classification of articles based on themes ‘Disaster Management’, ‘Social Media Analytics’, and ‘Social Science’ and based on top journals. (a) - The sunburst chart shows the number of articles under the top 5 journals and is further classified under each theme category. The inner circle shows the former and the outer circle shows the latter. (b) - The horizontal bar chart shows the number of articles under each theme category year-wise.

4.4 Journal Distribution and Theme Analysis

Among the 250 articles included in the literature database, 184 were published in journals. Figure 8 provides insight into the top 5 journals, which account for 2.98% of the journals included in the database, and the number of articles published in each of these journals. We have further classified these articles under the themes ‘Disaster Management,’ ‘Social Media Analytics,’ and ‘Social Science.’

390

Figure 8 presents the statistics for this classification as well, where the ‘Disaster Management’ theme is maximum. The figure also shows the number of articles on a particular theme falling under a top journal. We can observe that the ‘Social Science’ theme is the least in comparison to the other categories. However, an article belonging to the ‘Social Science’ theme aids in understanding how demographic study can be done using social media data. Similarly, the articles belonging to the ‘Social Media Analytics’ show various methods of social media data collection from a development perspective.

395

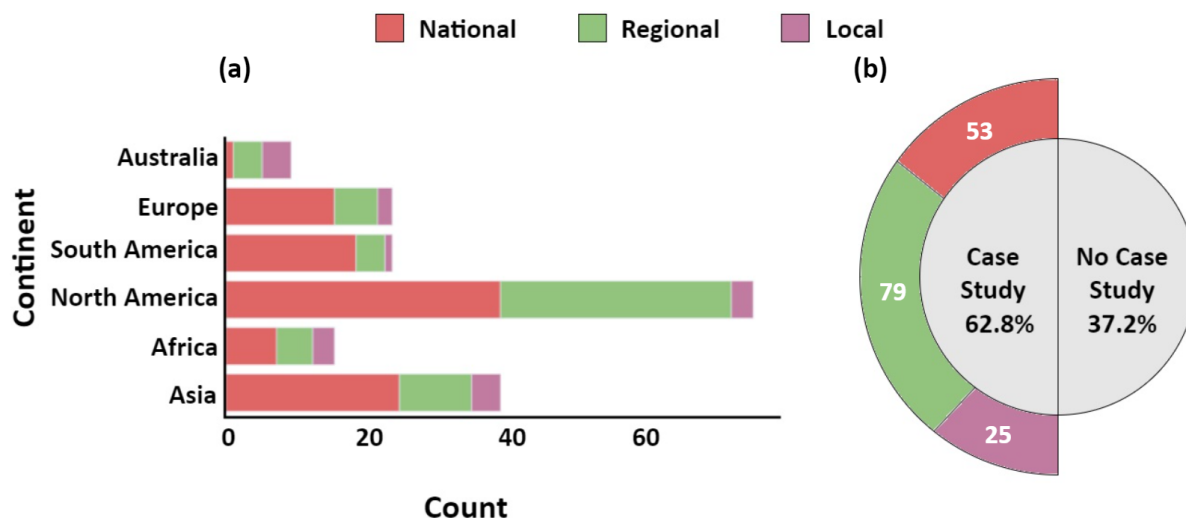


Figure 9. The figure shows the articles that use a case study or not. (a) - The bar graph shows the number of articles that use a case study area categorised continent-wise. (b) - In the sunburst chart the inner circle represents the percentage of papers that use a case study or not; the outer circle represents the number of articles that use a case study categorised under each scope.

4.5 Case Studies and Geographic Scope

Over 60% of the articles (Jung et al., 2015; Chen et al., 2016; Bala et al., 2017; Kurkcu et al., 2017; Bala et al., 2017; Liu et al., 2018; Brangbour et al., 2020; Rahmadan et al., 2020; Liu et al., 2020) categorised under the 'Disaster Management' theme employed case studies to evaluate their methodologies.

400 Figure 9 provides statistics on the use of case studies and their geographical scope. Notably, approximately 50% of the articles (Yuan and Liu, 2018; Abirami et al., 2019; Yang et al., 2019b; Kankanamge et al., 2020a) utilised regional case studies, which was the most prevalent among the different geographical scopes. It is worth mentioning that North America, particularly events such as Hurricane Sandy (2012), Hurricane Matthew (2016), and the Red River Valley Flood (2009), was the most frequently used region in these case studies (Ferris et al., 2016; Martín et al., 2017). We can also observe from the figure that around 37%
405 of the articles do not use a case study to validate their respective methodologies.

4.6 Disaster Events

Articles categorised under the 'Disaster Management' theme were further classified under the 'Event' section, indicating the specific disaster events studied by the respective authors. Figure 10 shows the statistics of the 'Event' category of the literature database. Our examination of these case studies revealed that 'Flood' was the most frequently studied disaster event, followed
410 by 'Hurricane' event (Middleton et al., 2013; Freberg et al., 2013; Gupta et al., 2013; Guan and Chen, 2014; Xiao et al., 2015; Yoo et al., 2016; Jamali et al., 2019; Wang et al., 2019). We can also observe that the 'Earthquake' event was studied every year of the review period. The least studied events were 'Storm', 'Volcanoes' and 'Cyclone'.

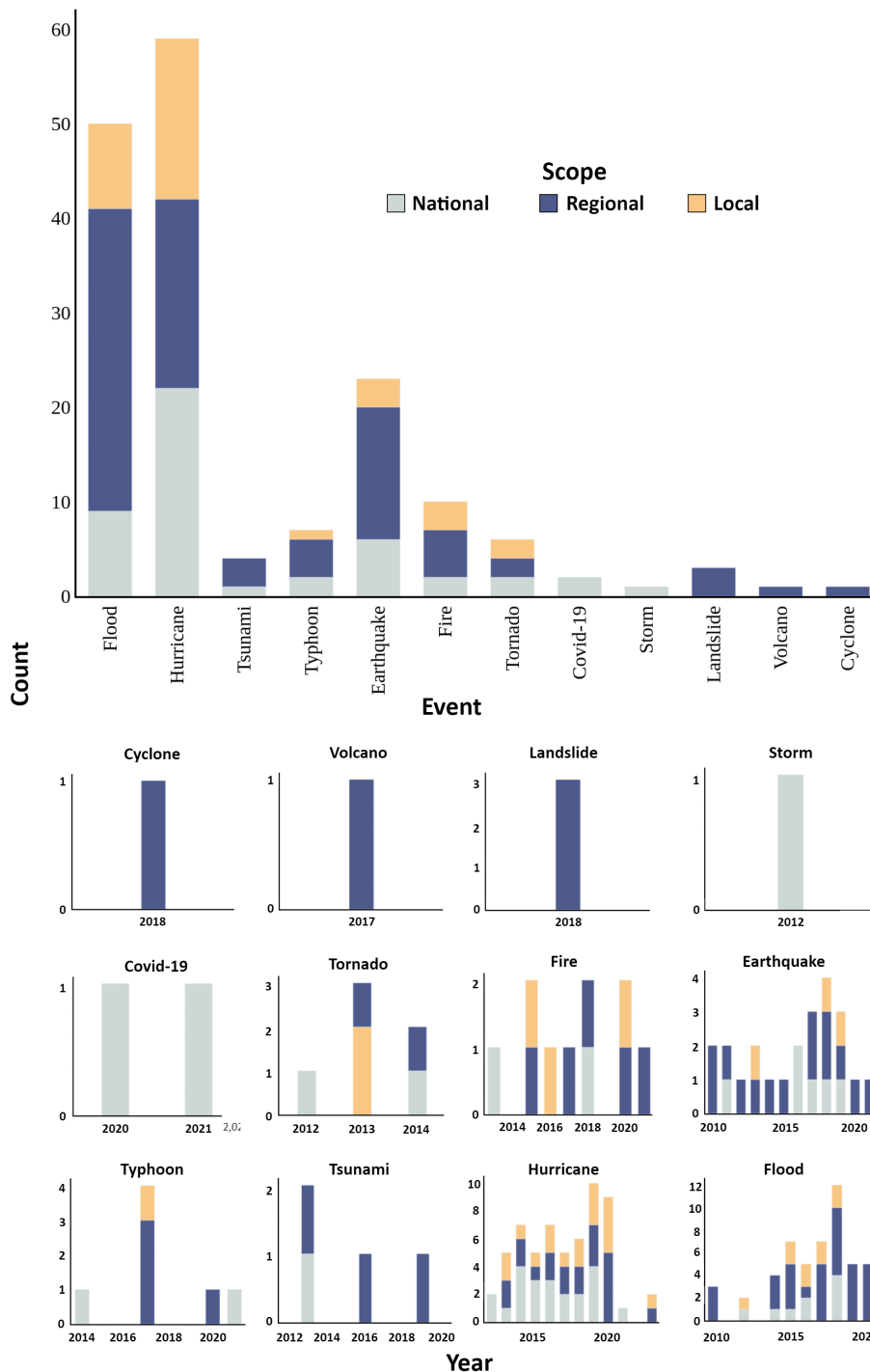


Figure 10. The figure shows the analysis of the articles categorised under the ‘Disaster Management’ theme – The large bar chart represents the number of articles under various disasters used as a case study. The smaller bar charts show the year-wise statistics of each disaster event.



4.7 Data Sources and Collection Methods

Among the articles listed in the literature database, excluding the review articles, approximately 72% (182 out of 250) utilised
415 social media data from various platforms as their input data (Gautam and Yadav, 2014; Uchida et al., 2016; Branz and Brock-
mann, 2018; Alampay et al., 2018). Within this category, 70% of the articles developed their own methodologies for collecting
social media data tailored to their specific needs (Driscoll and Walker, 2014; Gaspar et al., 2016; Mac Kim et al., 2016; Healy
et al., 2017; Campan et al., 2018). They frequently employed APIs, such as the Twitter Streaming API and REST API. The
remaining 2% of the articles utilised social media data available as online resources from various portals (Ai et al., 2016;
420 Madichetty and Sridevi, 2021).

Out of the 250 articles, excluding the review papers, nearly 13% (34 articles) sourced their data from government authority
portals (Ofii et al., 2016; Williams et al., 2018). Frequently accessed portals included FEMA (Federal Emergency Manage-
ment Agency, USA) and USGS (United States Geological Survey), which offered valuable disaster-related social information,
satellite image data, and historical event damage data. Additionally, approximately 4% (11 articles) of the total used manually
425 collected interview or survey data (Adam et al., 2012; Aisha et al., 2015; Le Coz et al., 2016; Lin et al., 2018; Lu and Yuan,
2021).

4.8 Data Relevance Filtering

The identification of relevant data presents a significant challenge in social media data collection. The majority of articles,
around 70%, employed NLP-based methods, particularly text analysis, to address this challenge (Starbird et al., 2010; Terpstra
430 et al., 2012; Panagiotopoulos et al., 2016; Laylavi et al., 2017; Lin et al., 2018). These methods involved the use of inclusion
keywords specific to their topics of interest during data collection. While this approach aids in identifying topic-relevant data,
it may also introduce a considerable amount of noise.

The use of exclusionary criteria proved valuable in noise reduction, with approximately 12% of the articles adopting this
approach (Joseph et al., 2014; Radiani et al., 2016; McCormick et al., 2017). These articles utilised NLP and ML-based
435 solutions to exclude irrelevant data. Exclusionary criteria are often constructed based on assumptions, emphasising the need
for rigorous evaluation before concluding. However, only a small percentage, approximately 2% of the articles, conducted such
evaluations before proceeding with the data analysis (Spinsanti and Ostermann, 2013; Herfort et al., 2014; Li et al., 2018a;
Ahmad et al., 2019).

In Figure 11, we present a summary of the relevance filtering analysis from the articles in the literature database. We can
440 observe that only 14% of the articles used exclusionary criteria to perform relevance filtering. Notably, the majority of the
articles employed NLP methods to perform filtering in comparison to ML methods. This analysis allowed us to answer our
research question (Q1), demonstrating that performing relevance filtering is vital for improving data quality and application
effectiveness. We recommend a thorough study of input data and the implementation of NLP or ML methods for effective
relevance filtering strategies.

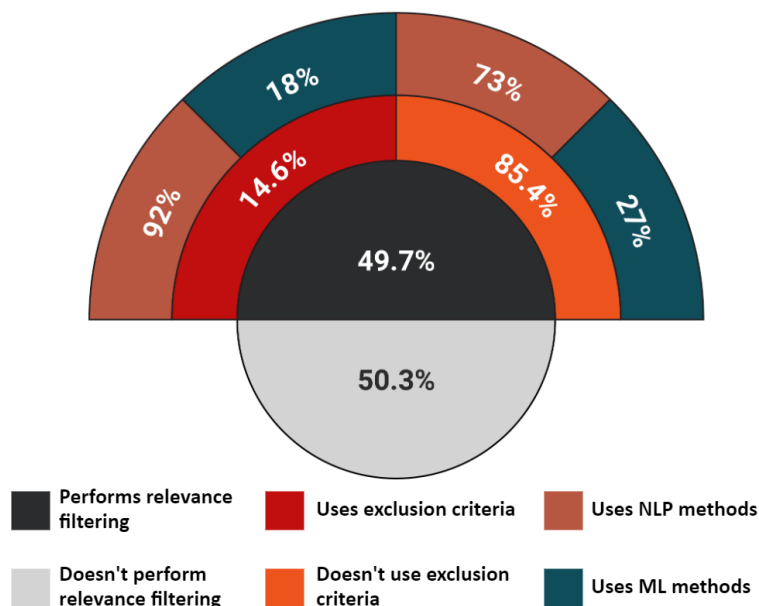


Figure 11. Result of research question Q1 - The chart shows the percentage of articles under each legend category summarising the data filtering methods employed in the articles listed in the literature database.

445 4.9 Data Analysis Methodologies

The methodologies employed in the articles within the literature database encompass a range of techniques in the fields of NLP and ML. These methodologies include text analysis, Named Entity Recognition (NER), Bag-of-Words (BoW), Part-of-speech Tagging (PoS), and various feature extraction methods. Data analysis is carried out using both supervised and unsupervised ML models, employing algorithms such as Logistic Regression (LR), Support Vector Machines (SVM), Naive Bayes (NB), K
450 Nearest Neighbors (KNN), Convolutional Neural Networks (CNN), Decision Trees (DT), Random Forest (RF), Latent Dirichlet Allocation (LDA), and more (Plachouras et al., 2013; De Albuquerque et al., 2015; Zhang et al., 2019b).

Some articles employ statistical techniques, including correlation analysis (e.g., Pearson's and Kendall's), distribution analysis (e.g., Poisson and Binomial), and Generalised Additive Models (GAM) (LÓPEZ-MARRERO, 2010; Liu and Lee, 2010; Lu and Yang, 2011; Yin et al., 2012; Westerman et al., 2014). Others explore methodologies that establish relationships among
455 stakeholders in disaster scenarios and conduct network analyses to enhance decision-making in the wake of disasters (Kogan et al., 2015; Wang et al., 2016; Htein et al., 2018; Kim and Hastak, 2018; Rajput et al., 2020; Wang et al., 2021). Figure 12 provides statistics on the technologies featured in the reviewed articles.

From figure 12 we can observe that NLP methods were employed the most where text analysis was in the majority. Analysing the text of the social media post helps in identifying topic-relevant keywords, event location, duration of the event, and senti-
460 ment of the user. ML methods were also used for analysis and the SVM algorithm was found frequently used by the investiga-

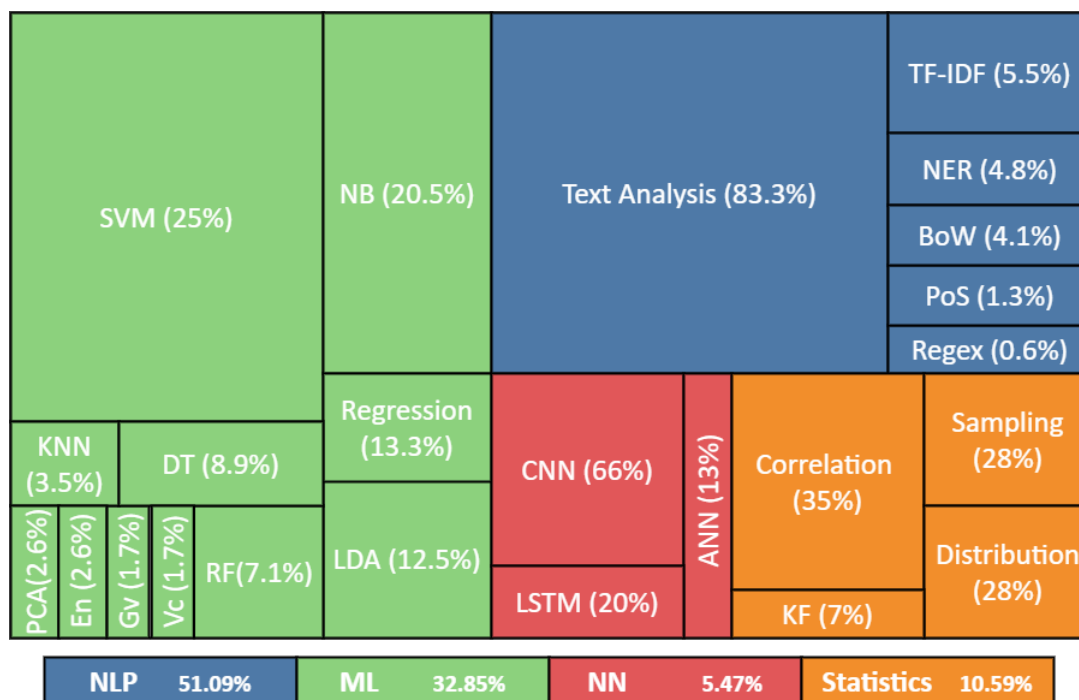


Figure 12. The treemap shows the algorithms used in the methodologies defined in the articles listed in the database. The percentage represents the percent of articles that use a particular algorithm (for abbreviations, refer to appendix A).

tors. However, neural network algorithms were not used much in the literature duration.

Roughly 65% of the articles in the literature database conduct performance evaluations using a range of methods. Articles employing ML algorithms often rely on scoring metrics like accuracy, precision, recall, and F-score (Imran et al., 2013a; Olteanu et al., 2014; Wang et al., 2016; Nguyen et al., 2017). Those exploring sentiment analysis in SMD typically utilise polarity scores for evaluation (Bala et al., 2017; Yuan et al., 2021). Some articles also employ statistical tests, such as ANOVA, chi-square, correlation values, and invariance tests to validate their methodologies (Steelman et al., 2015; Reuter and Spielhofer, 2017). Additionally, a few authors opt for manual evaluations (Stephenson et al., 2018; Liu et al., 2020).

4.10 Actionable Information

The articles categorised under the ‘Disaster Management’ theme were categorised further based on the Actionable Information (see table 1) classes to address our research question Q2. Figure 13 shows the number of articles assigned to each AI category year-wise and on an overall basis.

From figure 13 we can observe that AI-1 (95 of 211) (Howe et al., 2011; Chae et al., 2012; Fohringer et al., 2015), AI-2 (91 of 211) (McClendon and Robinson, 2013; Abedin and Babar, 2018; Boas et al., 2020) and AI-3 (121 of 211) (Castillo et al., 2013; Neppalli et al., 2017; Madichetty, 2020) was the most prevalent where articles focused on the development and testing of social

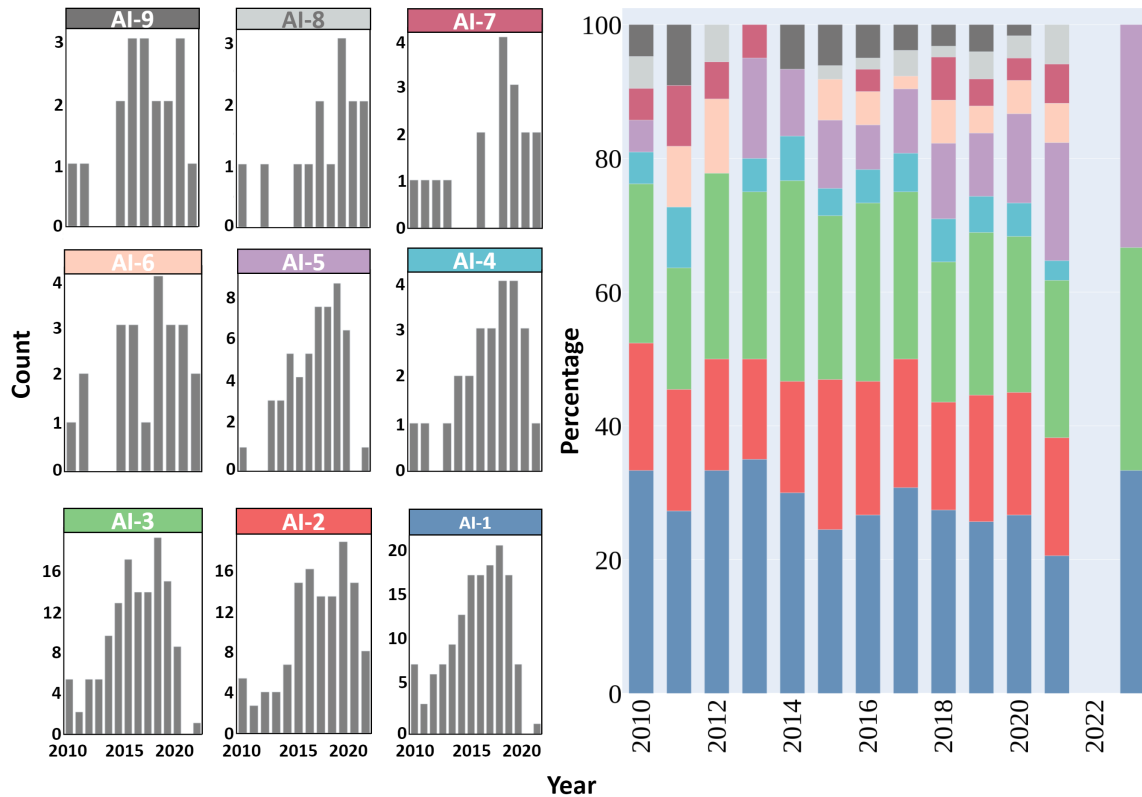


Figure 13. The figure shows the analysis of the disaster management-related articles categorised under the ‘Disaster Management’ theme - The stacked bar chart summarises the percentage of AI classes yearwise and the smaller bar charts show individual summaries of each AI class yearwise.

475 media data collection methodologies and geolocation identification methodologies and conducting spatial analyses. Notably, 57% of the articles were classified under AI-3, emphasising the significance of relevance filtering in disaster scenarios, and investigating methods to enhance data quality and reduce noise.

Around 11% (25 of 211) fall under AI-4 which studied how SMD can be utilised for community collaborations (Yang et al., 2019b; Yuan et al., 2021). Approximately 23% (50 of 211) and 10% (22 of 211) fell under AI-5 (Podhoranyi, 2021; Karim-
 480 iziarani and Moradkhani, 2023) and AI-6 (Htein et al., 2018; Delilah Roque et al., 2020) respectively, which focused on the identification of disaster trends, disaster hotspots, and stakeholder collaborations. Around 8% articles (17 of 211) developed open-source software and were categorised under AI-7 (Yuan and Liu, 2018; Podhoranyi, 2021).

AI-8 (14 of 211), involving resource identification methodologies, received the least attention (LÓPEZ-MARRERO, 2010; Houston et al., 2012; Delilah Roque et al., 2020). Articles categorised under AI-9 (18 of 211) focused on studying the commu-
 485 nity response (Jitkajornwanich et al., 2018; Ahmad et al., 2019).

The analysis indicates that current methods, such as NLP and ML, effectively aid in filtering social media data for relevance,



reducing noise, and excluding irrelevant content. However, challenges related to data reliability, including rumours and false information, persist. Many data collection methods employ inclusion keywords for relevance, which can introduce noise. The use of exclusion criteria proves valuable in enhancing efficiency by eliminating specific data.

490 Each article categorised under the 'Disaster Management' theme fulfilled at least one actionable information category. Several articles (Cervone et al., 2016; Schempp et al., 2019; Podhoranyi, 2021) met more than five actionable information categories, demonstrating their valuable contributions to efficient disaster management.

5 Discussion

In this section, we analyse and discuss the different categories and the corresponding information within the Social Media Literature Database (Gopal et al. (2023)). We organise this section into subsections to address the various categories within the Social Media Literature Database. We discuss the data collection methods used in the articles (section 5.1), major disaster events used as case studies in the articles (section 5.2), social media data reliability and external data usage in the article methodologies (section 5.3), algorithms used in the article methodologies (Section 5.4), actionable information in the articles (section 5.5), methodological biases (section 5.7), best practices of social media usage (section 5.8) and the practical applications of the Social Media Literature Database (section 5.9). Additionally, in section 5.6, we delve into a methodology based on our previous work for effectively collecting social media data through the use of exclusion criteria and other NLP techniques.

5.1 Data Collection Methods Used in Database Articles

The majority of the 250 articles in our Social Media Literature Database (Gopal et al. (2023)), roughly 70%, devised their methods for collecting social media data, primarily relying on topic-relevant inclusion keywords to sift through the extensive web resources (Wendt et al., 2016; Henry, 2021). The investigators used the approach of in-depth studying of the topic of interest to identify related keywords which helps to identify relevant content. In such an approach, the major challenge is noise filtering.

For instance, one of the most frequently studied disaster events in the database was Hurricane Sandy, which struck the USA in 2012 (Neppalli et al., 2017; Wang et al., 2019). Many of the articles that focused on Hurricane Sandy as their case study employed a set of inclusion keywords, including terms like 'Sandy,' 'Hurricane,' 'rainstorm,' 'New York,' 'Caribbean,' 'USA,' and '2012,' along with relevant hashtags. While these keyword sets were effective in gathering data related to Hurricane Sandy from social media platforms, they also resulted in significant noise, such as posts mentioning 'hurricane of emotions' (Spence et al., 2015; Kogan et al., 2015).

Some authors experimented with data collection using exclusion keyword sets, effectively reducing noise. For instance, McCormick et al. (2017) collected Twitter data to analyse demographics for social science research, and they used an exclusion keyword set related to 'TV shows' to filter out unrelated data. Similarly, Aswathy et al. (2022) collected tweets and news reports related to disasters and employed an exclusion keyword set that included terms like 'Songs,' 'Election,' and 'Victory' to eliminate irrelevant data, such as mentions of 'Landslide Victory' (part of previous work, discussed in detail in section 5.6).



This approach demonstrates the effectiveness of exclusion keywords in improving data collection efficiency.

520 Various authors have employed machine learning methods for noise filtering. They utilized supervised classification algorithms to determine the relevance of social media posts to a specific topic. However, a significant amount of data is needed for training, along with experts to prepare the data, making the process time-consuming and requiring ongoing maintenance (Chen et al., 2014; Ghani et al., 2019; Fan et al., 2021).

Our analysis has highlighted the importance of avoiding noise in social media data collection through the use of exclusionary
525 criteria. However, it is crucial to evaluate these methods to ensure that relevant data is not missed. We recommend that researchers thoroughly study the topic and related data before constructing their methodologies and consider including exclusion criteria to filter out noise at the initial stage. Excluding irrelevant data can significantly impact both time and space complexities and enhance the overall efficiency of the model.

5.2 Major Disaster Events in Database Articles

530 Around 74% of the 250 articles categorised under the 'Disaster Management' theme in the Social Media Literature Database used various disaster events around the globe as their case studies to evaluate their respective methodologies. Figure 14 shows some of the significant historical disaster events in the database. These articles devised methods to extract location information from social media data for conducting case studies. Social media platforms often offer API support that allows for the retrieval of location-specific data using bounding boxes (Purohit et al., 2014; Neubaum et al., 2014). However, challenges arise when
535 social media users input non-existent place names in their content or profiles, leading to data collection inaccuracies. To mitigate this, some authors exclusively gathered geotagged social media posts, which are considered reliable as they represent the user's present location based on their device settings (Steelman et al., 2015; Resch et al., 2018; Leon et al., 2018; Wang et al., 2021).

Figure 14 illustrates that hurricane events were frequently chosen as case studies by multiple authors. In the 'Disaster Manage-
540 ment' theme, several articles (Gupta et al., 2013; Neubaum et al., 2014; Steelman et al., 2015; Olteanu et al., 2015; Mukkamala and Beck, 2016; Jamali et al., 2019) utilised hurricane Sandy, a 2012 disaster in the US, as a case study due to the extensive social media activity related to this event. However, assessing the reliability of the content in such scenarios presented challenges, leading to research endeavours aimed at identifying and analysing fake data circulated on Twitter and examining their impact on the public (Pourebrahim et al., 2019; Wang et al., 2019).

545 Smith et al. (2017) presented a significant article that introduced a real-time monitoring framework for identifying flooded areas using social media data. They evaluated their framework by applying it to the 2012 Tyne and Wear flood as a case study. Another notable event was the devastating Nepal earthquake of 2015. In a pivotal study, Radianti et al. (2016) devised a data collection methodology, that enables the collection, filtering, and categorization of tweets in multiple languages. This methodology aimed to identify disaster response issues, including resource requirements and damages, associated with the
550 Nepal earthquake.

The 2018 Woolsey fire, which ravaged areas in Los Angeles and Ventura counties, was a significant fire event documented in the database. St Denis et al. (2020) conducted a study on this event, focusing on the behaviour and content of local residents in

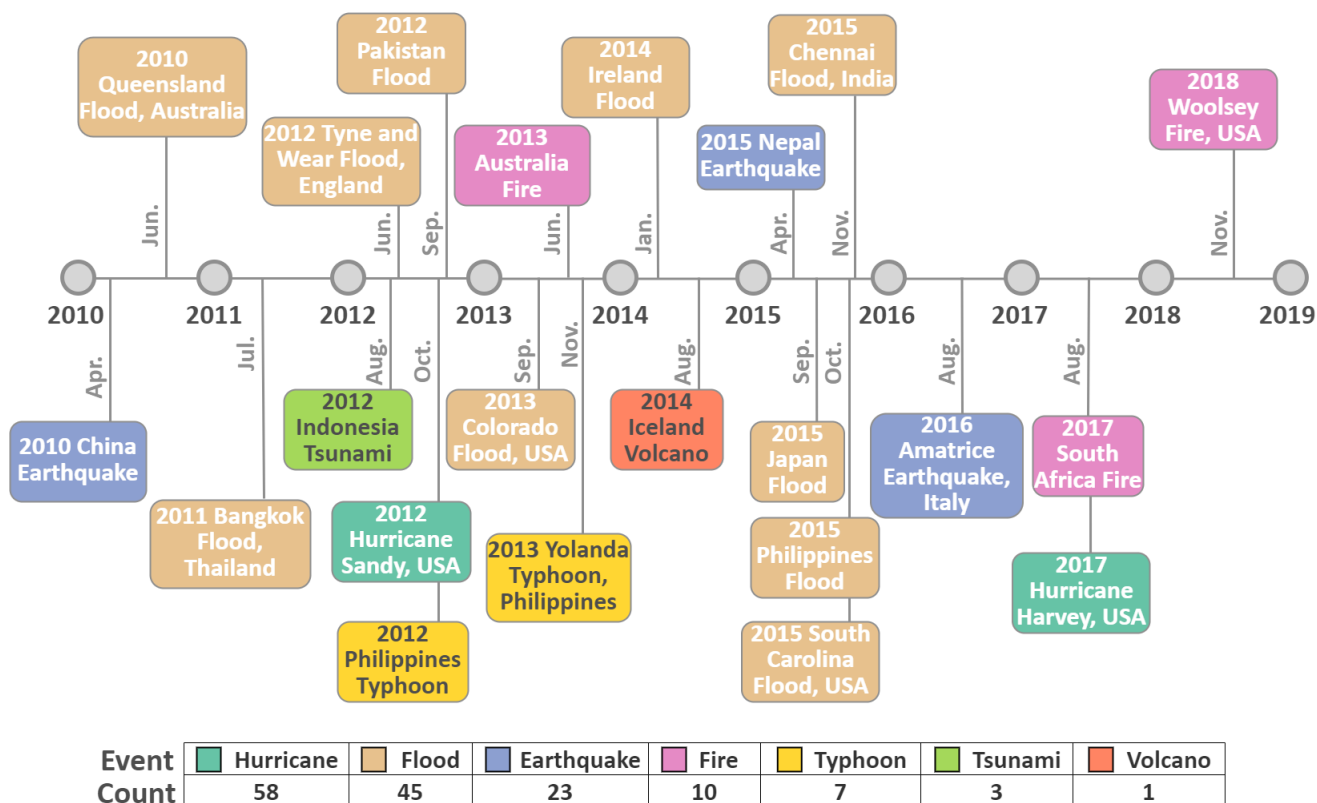


Figure 14. The timeline shows some of the significant disaster events that occurred from 2010 to 2019. The legend shows the number of articles that used a particular disaster as their case study.

the affected area. Their research aimed to capture valuable information that could otherwise go unnoticed.

The events depicted in Figure 14 have had significant impacts on the affected populations. To gain a better understanding of the scale of these disasters, we collected and analysed data from the EM-DAT database to assess the number of affected individuals. Figure 15 presents the statistical findings, revealing that the 2012 Hurricane Sandy in the USA and the 2010 China earthquake each affected more than 20 lakh people.

As demonstrated in Figure 9 (refer to Section 4.5), our analysis of continent-based case studies revealed that North America was the most frequently utilised region by the authors in the Social Media Literature Database. Furthermore, it was evident that major disaster events generated more data and garnered increased attention on social media platforms. In light of these observations, we recommend that researchers explore methodologies using local disaster events, enabling a more comprehensive examination of data to identify relevant content, manage unfamiliar place names, and enhance disaster management efforts effectively.

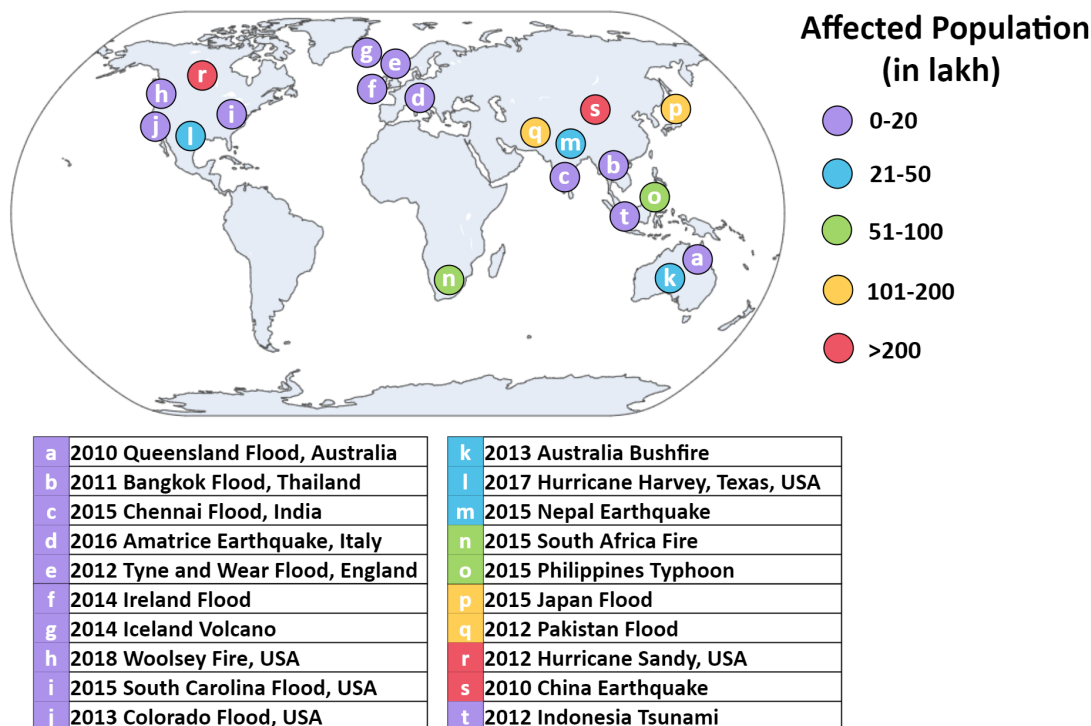


Figure 15. The figure shows the statistics of the affected population of various historical disaster events. The colour legend represents the affected population and the alphabet legend shows the details of the disaster event plotted on the map.

5.3 Social media reliability and usage of external data in Database Articles

565 The issue of reliability is a significant concern when using social media data for disaster management (Mazoyer et al., 2018; Liu et al., 2020). To address this concern, many authors incorporated external data sources alongside social media data, aiming to enhance the efficiency and reliability of their methodologies (Chatfield and Brajawidagda, 2013; Joseph et al., 2014; Musaev et al., 2018).

570 External data sources encompass information available on government portals, such as damage-related data, satellite images of affected areas, disaster event statistics, precipitation data, GIS data, news reports about disaster events, and interview or survey data, among others. Out of the 250 articles in the Social Media Literature Database, 26% (65 articles) integrated external data to enhance their methodologies. Notably, the majority of external data utilisation was observed in articles focusing on events in the USA, where FEMA and USGS data were frequently employed (Hodas et al., 2015; Liu et al., 2018; Musaev et al., 2018). External data played a vital role in not only collecting additional trustworthy information but also in validating the methodologies by enabling comparisons of results (Earle et al., 2011; Li et al., 2018b). This approach helped investigators demonstrate the effectiveness of their methodologies in disaster management. Furthermore, it allowed them to identify and refine keyword sets for data collection and uncover place names and other location-specific information that may not be readily available in social

575



media posts (Dashti et al., 2014; Kryvasheyev et al., 2016). As a recommendation, we suggest that researchers continue to leverage trustworthy external data sources in conjunction with social media data to obtain more authentic and reliable insights for disaster management.

5.4 Algorithms used by Database Articles

In section 4.9, we explored the algorithms utilised in the methodologies of the articles within the database, as illustrated in Figure 12. These algorithms were categorised into 'NLP,' 'ML,' 'Statistical,' and 'Neural Networks.' Notably, NLP algorithms were the most frequently employed and were consistently utilised across the review period. NLP methods played a crucial role in content analysis during data collection and filtering (Gupta et al., 2013; Madichetty, 2020).

Statistical methods were primarily used for distribution and correlation analysis, providing valuable insights into the data (Htein et al., 2018; Wang et al., 2019). Starting in 2013, ML methods gained popularity among researchers and were extensively applied to classification, clustering, and filtering tasks. Of the various classification algorithms, SVM was commonly employed in the literature and was found effective. Additionally, neural network algorithms began to gain traction around the same time and exhibited impressive performance in various applications (Neppalli et al., 2017; Reynard and Shirgaokar, 2019).

Figure 16 presents statistics on the methodology categories, offering an overview of overall percentages and a year-wise analysis. Notably, there was an increase in the usage of these categories after 2015, indicating more extensive research experiments during this period. By analysing the literature, we observe that ML algorithms like SVM, NB, and RF were often used for classification problems (Nair et al., 2017; Srivastava et al., 2020).

Researchers also employed these algorithms to address relevance filtering problems, utilising simple binary classifications to identify content related to a specific topic. Subsequently, this relevant data was used as input for further analysis (Khaleq and Ra, 2018; Loynes et al., 2022). ML and NN methods were primarily adopted for data classification, particularly for the detection of disaster events within the content of social media posts. These approaches have yielded promising results (Zhang et al., 2019b; Madichetty, 2020). However, it is worth noting that there was relatively less focus on analysing posts in the pre-event, during-event, and post-event phases. Pre-event posts typically contain information related to warnings, alerts, or precursor events (Chatfield and Brajawidagda, 2013; Carley et al., 2016b).

Analysing such posts could provide valuable insights for early warning and preparedness efforts. During-event posts often contain critical information about the immediate needs of the public, such as requests for rescue, evacuation, food, or information about rescue camps. Analysing these posts can significantly aid disaster relief efforts (Jongman et al., 2015; Jitkajornwanich et al., 2018). Post-event posts are valuable for recovery analysis, particularly in terms of damage assessment and understanding the aftermath of a disaster. Many articles in the database employed NLP, ML, or NN techniques to classify and analyse post-event data, contributing to damage assessment and recovery efforts (Shi et al., 2019; Rahmadan et al., 2020).

In our recommendations, we emphasise the importance of investigating pre-event posts, as they can provide critical information for early warning systems, helping to save lives and reduce the impact of disasters before they strike a location. Analysing pre-event posts can contribute to better disaster preparedness and timely responses.

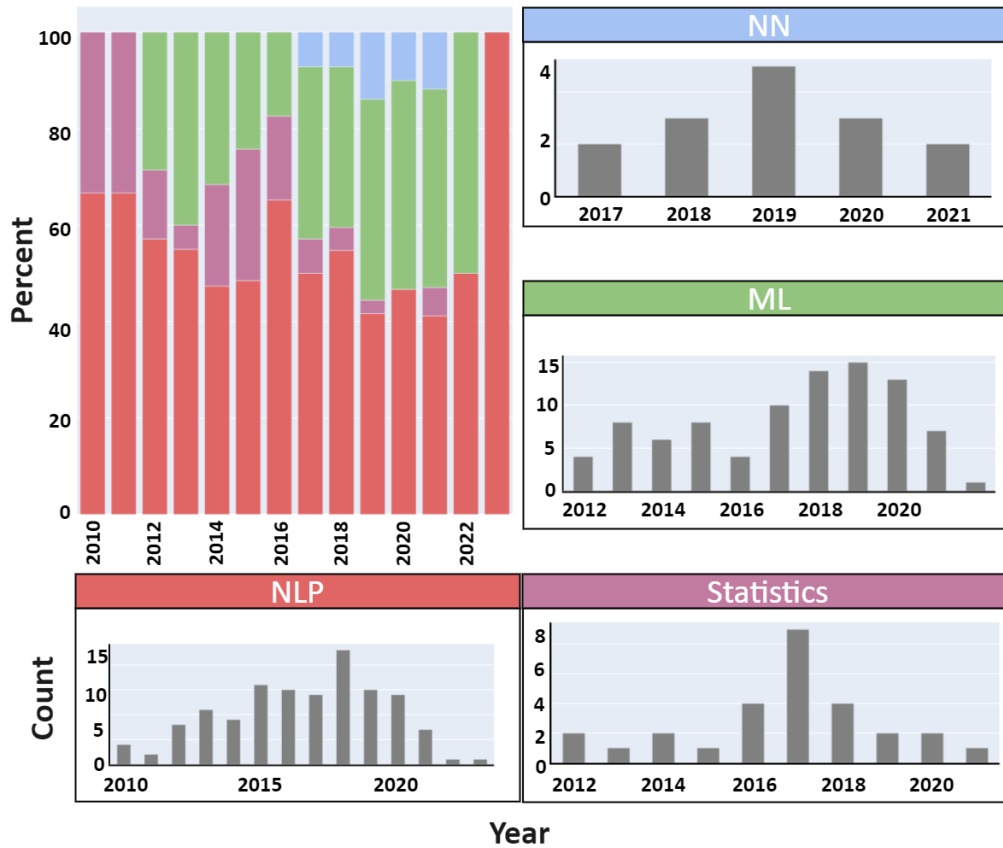


Figure 16. The figure shows the analysis of the methodology categories employed in the articles listed in the database - The stacked bar chart shows the percentage of ‘NLP’, ‘ML’, ‘Statistics’, and ‘NN’ categories year-wise. The sub-bar plot shows the number of articles that used a particular category year-wise.

5.5 Actionable Information in Database Articles

The articles categorised under the ‘Disaster Management’ theme was further classified into nine actionable information (AI) classes to address our research question Q2. Our goal was to analyse whether the literature provides actionable data from social media content to improve disaster management. Our analysis revealed that each article fell into at least one AI class, with AI-1, AI-2 and AI-3 being the most commonly addressed categories in the literature (Li et al., 2018b; Wang et al., 2018). Conversely, AI-8 received the least attention, indicating that researchers have explored the identification of resources from social media posts less during event classification. This highlights the relatively limited exploration of during-event classifications in the reviewed literature (Houston et al., 2012; Kryvasheyeu et al., 2016).

AI-7 is the next category in line, and it received relatively less attention from researchers, possibly due to a lack of development and availability of open-source platforms during the review period (LÓPEZ-MARRERO, 2010; Carley et al., 2016a). We



recommend that researchers consider creating more platforms or applications for making disaster-relevant data and real-time analysis available to the public. Such platforms can significantly contribute to rapid disaster relief, response, and recovery efforts.

625 AI-8, which focuses on community interaction analysis, received limited attention from researchers. Studying how communities respond and interact during and after a disaster is essential for gaining a broader perspective on disaster management. It can aid in creating effective strategies for mitigating disasters in the future (Valenzuela et al., 2017; Delilah Roque et al., 2020). We encourage researchers to delve deeper into the study of community interactions and analyse behavioural changes in pre-event, during-event, and post-event phases through social media data.

630 While most articles in the literature database fell into one or two AI classes, a select few managed to be categorised under five or more AI classes, achieving significant results. The articles that achieved 5 or more AIs are described in table 2. Articles achieving 5+ AI classifications focused primarily on floods and hurricanes. They consistently satisfied AI-1, AI-2, and AI-3 by collecting and filtering social media data and considering temporal and spatial analysis. Some also addressed AI-4, emphasising community engagement to enhance preparedness and response.

Table 2: Analysis of the AI categories in the literature database - Table shows the key articles that gained 5 or more AI classifications (AI-1: Data Collection; AI-2: Geolocation Detection; AI-3: Relevance Filtering; AI-4: Community Collaborations; AI-5: Disaster Trends; AI-6: Stakeholder Collaborations; AI-7: Software Development; AI-8: Resource Identification; AI-9: Community Response).

Article	AIs	Purpose	Event Type
Fohringer et al. (2015)	1, 2, 3, 5, 8	Discusses a new methodology for using social media data to task the collection of remote-sensing imagery during disasters or emergencies, specifically for damage assessment of transportation infrastructure.	Flood
Haworth and Bruce (2015)	1, 2, 3, 5, 7	Proposes a method based on the "wisdom of the crowd" principle for decision support during disasters. The goal is to create a flood probability map by combining multiple observations over the affected area, each assessed for its reliability.	Flood



Table 2: Analysis of the AI categories in the literature database - Table shows the key articles that gained 5 or more AI classifications (AI-1: Data Collection; AI-2: Geolocation Detection; AI-3: Relevance Filtering; AI-4: Community Collaborations; AI-5: Disaster Trends; AI-6: Stakeholder Collaborations; AI-7: Software Development; AI-8: Resource Identification; AI-9: Community Response).

Article	AIs	Purpose	Event Type
Yu et al. (2018)	1, 2, 3, 4, 6, 7, 8	Introduces an interdisciplinary approach to disaster relief management. The proposed framework combines dynamic and static databases, including social media and authoritative data, to model rescue demand during a disaster.	Hurricane
Guy et al. (2010)	1,2,3,4,7,8	Presents an architecture for emergency situational awareness by employing natural language processing and data mining techniques to extract situational awareness information from Twitter messages generated during disasters and crises.	General Emergency
Chae et al. (2012)	1,2,3,4,5	Explores connections and patterns created by the aggregated interactions in Facebook pages during disaster responses. Analyses social roles and key players using social network analysis.	Flood



Table 2: Analysis of the AI categories in the literature database - Table shows the key articles that gained 5 or more AI classifications (AI-1: Data Collection; AI-2: Geolocation Detection; AI-3: Relevance Filtering; AI-4: Community Collaborations; AI-5: Disaster Trends; AI-6: Stakeholder Collaborations; AI-7: Software Development; AI-8: Resource Identification; AI-9: Community Response).

Article	AIs	Purpose	Event Type
Bala et al. (2017)	1,2,3,5,7	Introduces an approach to improve the identification of relevant social media messages during disasters.	Flood
Carley et al. (2016a)	1,2,3,4,5,6	Aims to categorise social media messages into different themes within different disaster phases during a disaster.	Hurricane

Analysing the literature, we found that investigator-defined, topic-specific social media data collection strategies are valuable for gathering relevant data, but challenges related to noise and data reliability persist. We recommend researchers place greater emphasis on AI-8 and AI-9 to extract resource requirements from social media data, enabling informed resource allocation decisions during and after disasters.

5.6 Exclusionary Criteria - reducing noise in data

In a previous study, Aswathy et al. (2022), we conducted Twitter data collection to gather tweets related to disaster events, utilising inclusion keyword sets to identify relevant data. Upon analysing the initial data collected, it became evident that data containing the inclusion keyword set but unrelated to the disaster theme was present, contributing to the noise. A comprehensive examination of the data led us to create a set of exclusion keyword sets. A few examples may include "landslide election", "landslide victory", "flood of emotions", "landslide lyrics", "market flooded" and more. This effort resulted in the identification of approximately 56 keywords related to elections, music, life quotes, market, and emotions.

As previously mentioned, it is important to evaluate implementing exclusion criteria to ensure that no relevant data is inadvertently excluded. In our methodology, we sampled 1000 tweets and assessed the performance. Figure 17 presents the results of this evaluation, indicating the number of tweets correctly excluded and the number of relevant tweets that were potentially missed when utilising exclusionary criteria.

From figure 17 we can observe that around 80% of data was filtered correctly and around 14% of noise was collected, combining results from the 4 exclusion criteria sets. From the music exclusion criteria, we missed 20% of relevant data. This method can be employed for first-level filtering to reduce noise. The limitation of this work is that the exclusion keyword set is static

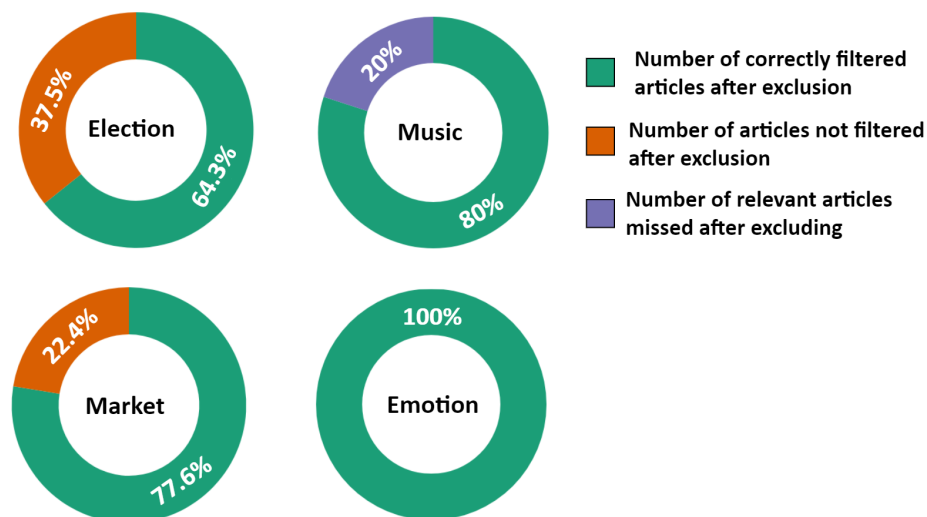


Figure 17. Evaluation results of the exclusionary criteria methodology: The donut charts show the percentage of tweets that were filtered from a sample data of 1000 tweets. Each sub-donut chart represents the different exclusion categories.

and needs constant maintenance when new exclusionary criteria are identified manually.

Figure 18 provides a word cloud visualisation that illustrates the impact of using exclusionary criteria on the data. As the exclusion keywords are applied, it is evident that noise is significantly reduced, and disaster-relevant data becomes more prominent in the dataset. Our observations indicate that the utilisation of exclusionary criteria is effective in noise reduction. However, one drawback is that a thorough manual analysis is needed to initially study the data and create a keyword set that can be used to formulate an exclusionary strategy.

In disaster situations, where accuracy is paramount, ML can play a pivotal role in identifying and eliminating outliers and noise. By leveraging both basic NLP and advanced ML, researchers can aspire to achieve a comprehensive strategy for data collection, ensuring that the information extracted from social media during crises is both accurate and actionable.

5.7 Methodological Biases in Disaster-Related Social Media Studies in Database Articles

This section describes seven biases within the Social Media Literature Database (Gopal et al. (2023)) articles regarding geographic, methodological, and data-related tendencies. We recognised these biases as the critical review methodology proceeded, solidified through extensive discussions among the authors. By identifying these biases, we aim to enhance the transparency of our analysis and provide a foundation for future research.

1. Geographic location of the case studies used in disaster-related articles

A notable geographic bias was found in the case studies employed by researchers in the literature, with a predominant focus on North America (see figure 9). Around 40% of the articles (60 of 154) used Hurricanes as the case study event, among which around 60% of the articles used events from North America (Kryvasheyev et al., 2016; Mukkamala and

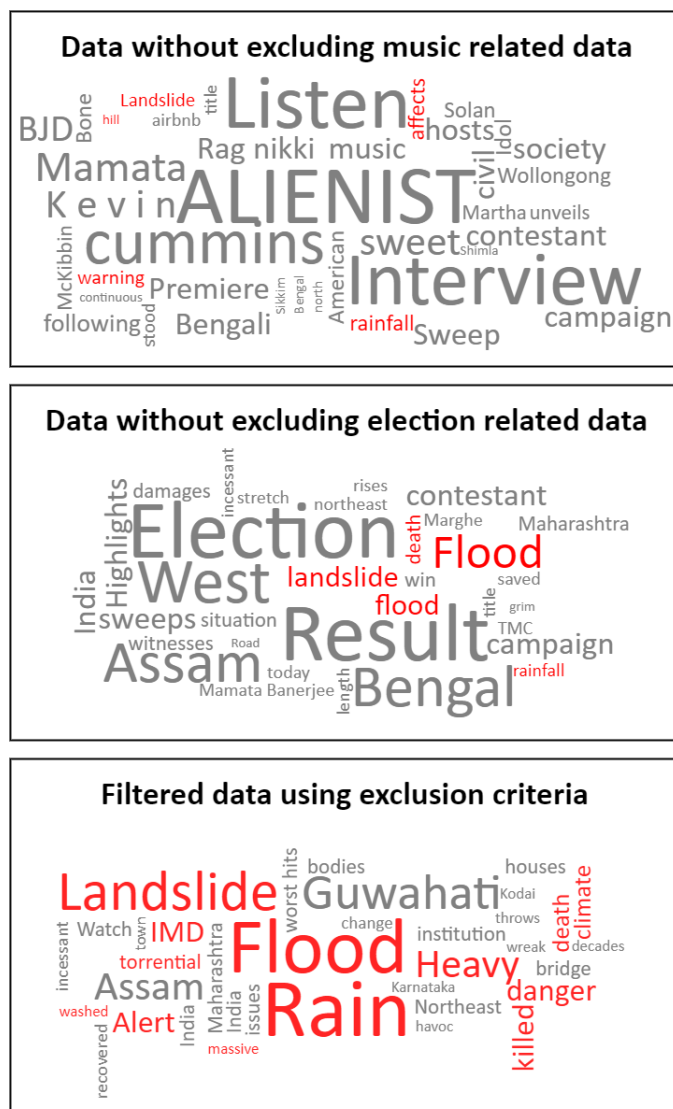


Figure 18. The word clouds show frequently occurring words before and after filtering the data using exclusionary criteria. The words highlighted in red are related to disasters. The size of each word represents the frequency of occurrence of a word in the sample data.



670 Beck, 2016; Jamali et al., 2019). This raises concerns about the generalisation of the findings in a global context.
The majority of studies exhibit a bias towards regional and national investigations, overshadowing the importance of
local studies (see figure 9). One of the reasons is the availability of data is limited from a local scope when compared
to a national or regional disaster event. The amount of population that uses social media platforms also varies based
on the area scope. Such biases may limit the applicability of findings to specific contexts (Liu and Stevenson, 2013;
675 Kankanamge et al., 2020a; Li et al., 2021).

2. *External data used in disaster-related articles for methodology validation*

Various articles in the literature use external data such as EM-DAT, FEMA, USDS, AIDR, and more as supporting data
to validate the methodologies employed (Spinsanti and Ostermann, 2013; Hodas et al., 2015; Liu et al., 2018). This may
introduce a bias as these datasets may not comprehensively represent the effects of a disaster that occurred in a specific
680 region.

3. *Social media data language preference in the article*

The articles that used social media data predominantly focused on the English language, which raises a linguistic bias
potentially excluding valuable insights from non-English sources. Around 6% (15 of 250) articles used a language that
is regional and relevant to their respective case studies (Lee et al., 2011; Jongman et al., 2015; Radianti et al., 2016).

685 4. *Social Media Platform preference for data collection methodology in the articles*

A platform bias is evident, with the majority of studies drawing from Twitter data (Earle et al., 2011; Sakaki et al.,
2012). This bias may limit the understanding of disaster dynamics on other social media platforms. The Twitter platform
provides user-friendly API usages and sufficient meta-data making it easy for researchers to extract data (Jitkajornwanich
et al., 2018). It is also important to emphasize that platforms such as Facebook and Weibo were also utilised in a few
690 articles in the database (Xu et al., 2016; Li et al., 2018a; Fang et al., 2019).

5. *Disaster events used in the articles for case studies*

The articles listed in the literature database predominantly explore hurricanes and floods (Le Coz et al., 2016; Madichetty,
2020), neglecting other impactful events such as pandemics, landslides, storms, and cyclones, which are few (Islam and
Walkerden, 2015; Musaev et al., 2018). This may overlook crucial aspects of disaster dynamics (see section 5.2). It is
695 also relevant to analyse precursor events, such as heavy rain as a precursor of a flood or a landslide, which aids in early
warning and mitigation.

6. *Preference of disaster management phase in the articles for case studies*

A bias emerges towards post-disaster phases such as response and recovery, with limited exploration of early warning
and mitigation phases. Around 4% (7 of 157) articles experimented with early warning methodologies (Leon et al., 2018;
700 Wu and Cui, 2018; Kitazawa and Hale, 2021). This raises the concern about social media data availability in real-time
from the social media platforms to develop solutions for early warning and mitigation.

Figure 19 shows the number of articles categorised under each disaster management phase, yearwise, and we can observe

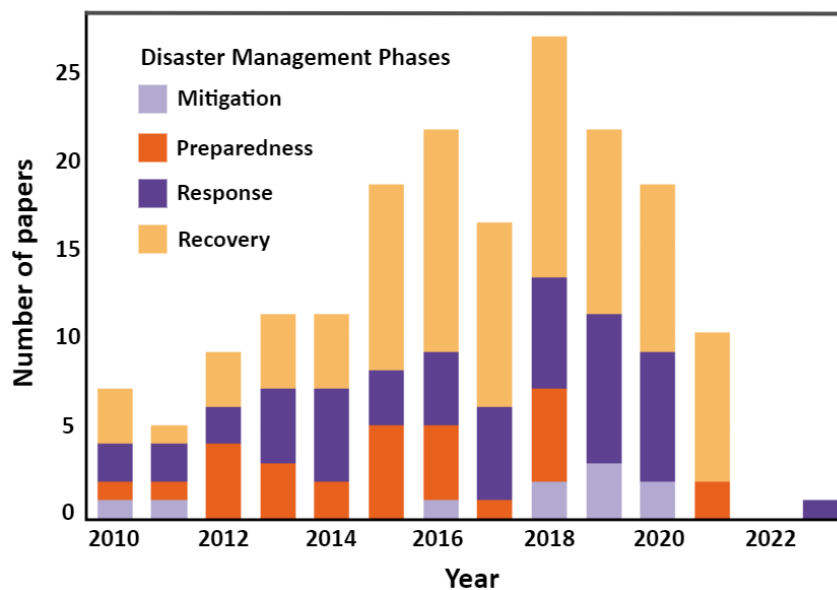


Figure 19. The stacked bar chart shows the number of articles listed in the literature database categorised under each disaster management phase, yearwise.

that post-disaster phases, which include response and recovery are discussed more when compared to mitigation and preparedness. We recommend the investigators develop early warning solutions using social media by analysing the precursor events of a disaster.

705

7. Actionable Information the methodologies of disaster-related articles

While researchers excel in temporal and spatial analysis (Kryvasheyev et al., 2016; Wang et al., 2018), there is a noticeable bias with limited attention given to community interaction analysis, stakeholder engagement, and resource allocation strategies hindering a holistic approach to actionable information (see figure 13 and section 4.10). Around 24% of the articles (49 of 212) discuss methods of community and stakeholder engagement to understand the needs of the public during and post a disaster event (Chae et al., 2014; Wang et al., 2016).

710

This critical review has systematically examined biases within the analysed articles and the methodology employed. The identification and acknowledgment of these biases are necessary for ensuring transparency in the analysis and for future research in the field. These biases bring attention to potential limitations in the existing methodologies, emphasising the need for a more inclusive approach in future studies.

715

5.8 Best Practices of Social Media Usage for Community and Researchers

Social media has emerged over the years as a tool for information dissemination during disasters offering real-time communication aiding the public, government and non-government organisations, volunteer groups, and other actors in disaster



720 management (Bruns and Liang, 2012; Smith et al., 2017; Kankanamge et al., 2020b). As the public becomes increasingly
 725 reliant on these platforms to share or acquire information, it is necessary to explore the best practices to be followed by the
 community and the researchers to harness the potential of these platforms responsibly (Lin et al., 2016). Based on the literature
 study, we propose guidelines for the public to share information on social media and discuss the ways to acquire disaster-
 relevant information for the researchers. By implementing these best practices, communities and researchers can contribute to
 a more effective response, ultimately aiding in the mitigation and recovery efforts.

725 The community plays a crucial role in disaster response by providing valuable information to first responders (Stephenson
 et al., 2018; Kankanamge et al., 2020a). Social media platforms are widely utilised for data acquisition during disasters, but
 the major challenge is to identify reliable information (Khaleq and Ra, 2018; Loynes et al., 2022). Table 3 shows a few best
 practices identified from the literature that can be followed by the community to provide credible information on social media.

Table 3: The proposed best practices of social media usage for the community for effective disaster response.

#	Best Practices	Description
i)	Social Media Platform Selection for Effective Disaster Communication	Leveraging location-specific popular platforms enhances reach and improves information dissemination during disasters (e.g., Twitter in the USA, Facebook in India and the UK, and Weibo in China).
ii)	Mitigating Rumours and Misinformation	Ensuring accuracy is crucial to prevent the spread of rumors during disaster management. Avoiding assumptions and speculations in social media sharing is vital for effective information dissemination.
iii)	Tagging official social media handles for effective response	Every major disaster-managing government or non-governmental organisations use social media handles to share information. During or post a disaster, the public may have water, food, shelter, and rescue requirements that need immediate attention. The public can tag these handles while sharing information which aids in informing the first responders easily.
iv)	Provide location information in the social media post	For first responders, acquiring accurate location information is a challenge. The public can contribute effectively by sharing the location details of the affected area by geotagging or by providing landmarks or street names for efficient response.



Table 3: The proposed best practices of social media usage for the community for effective disaster response.

#	Best Practices	Description
v)	Disaster event description in the social media post	Public-provided detailed descriptions during disasters, including emergency type, severity, and visible hazards, aid first responders in assessing the situation. Using relevant hashtags of official authorities helps consolidate information for easier tracking of updates.
vi)	Contribute multimedia data in the social media post	Image and video information provides a better understanding of the disaster situation. By not compromising on safety and privacy, if the public can share such data, it will assist the authorities in rapid decision-making. It also provides additional credibility to the social media posts which encourages other users to forward it further.

As researchers increasingly turn to social media data to gain insights about disasters, it is necessary to consider a few best practices to be followed so that data can be acquired and analysed efficiently (Branz and Brockmann, 2018; Campan et al., 2018). Drawing from our comprehensive examination of the literature and our own experience in the subject, we offer recommendations to researchers as described in table 4.

Table 4: The proposed best practices of social media usage for the investigators for effective research in the field of social media and disaster management.

#	Best Practices	Description
i)	Optimise data collection with clear data requirements	Conduct thorough data requirement analysis in the initial stages to devise an efficient data collection strategy. Developing a well-defined keyword set, especially for temporal and spatial-specific data, necessitates a deep understanding of the topic of interest.
ii)	Look beyond metadata for precise location extraction	Pay particular attention to the content within social media posts when extracting location information. This approach may yield more accurate and contextually relevant location data compared to relying solely on metadata.
iii)	Validate social media data with external data sources	Detecting rumours can be challenging. The usage of valid data (such as news reports, and verified social media handles, government reports) along with social media data can be experimented with for validation.



Table 4: The proposed best practices of social media usage for the investigators for effective research in the field of social media and disaster management.

#	Best Practices	Description
iv)	Improve stakeholder engagement	Stakeholder identification and network creation are highly necessary for the effective management of disasters. Through social media, a spatial analysis may assist in identifying necessary stakeholders which can in turn help in rapid communication pre-, during, and post a disaster event.
v)	Language inclusivity in disaster data extraction	Be language independent – focusing only on a single language could be ineffective. The community-level public may post information in local languages which may contain relevant information.

These recommendations can enhance the effectiveness of data collection and analysis methodologies when working with social media data for disaster management.

735 5.9 Utilizing the Social Media Literature Database: Practical Applications and Recommendations

Our Social Media Literature Database is available in the form of an Excel file that is open-access on GitHub (Gopal et al. (2023)). Upon accessing the database, users can employ various functionalities to facilitate their research. Following are a few examples:

- 740 i *Search and Filter*: Researchers can search for articles based on specific criteria such as year, keyword, or journal using the search option. Additionally, the filtering option enables users to view articles based on particular conditions (e.g., articles published in a specific year).
- ii *Sort Data*: The sorting option allows users to organize the data in ascending or descending order based on parameters such as year, citations, and the number of data used.
- 745 iii *Advanced Data Extraction*: Advanced users with proficiency in Excel can utilize formulas to perform complex data extractions. For example, researchers can identify articles that utilize Natural Language Processing (NLP) as a methodology within a specified timeframe.
- 750 iv *Reuse for Review Articles*: In the last decade, various authors contributed critical and systematic reviews in the domain of social media and disaster management (Tang et al., 2021; Tsao et al., 2021; Bukar et al., 2022). Researchers interested in conducting review articles in their domain can follow the paper searching criteria and Boolean search string formation methodologies outlined in the database. This enables them to search for relevant papers and extract pertinent information for their review.



v *Usage for social media researchers:* While the columns in the database are tailored for social media relevance filtering in disaster management, researchers from the social media domain can adapt the database to their needs. By excluding irrelevant columns and focusing on relevant ones such as article source details, researchers can redefine the database for their specific domain.

755

vi *Usage for disaster management researchers:* Researchers in the field of disaster management can leverage the "Event" and "Case Study" columns to perform basic searching and sorting techniques. This allows for a detailed analysis of various disaster events in different years and locations.

6 Conclusions

760 In conclusion, the surge in social media data usage as a real-time information source has had a transformative impact on the field of disaster management (Valenzuela et al., 2017; Mazoyer et al., 2018). To leverage social media data usage for improving disaster management, the identification of relevant and credible information is the main priority (Schempp et al., 2019; Domala et al., 2020). Our critical review of 250 articles, spanning from 2010 to 2023, is available as a Social Media Literature Database (Gopal et al. (2023)) and has unveiled the methodologies, challenges, and actionable insights on how to harness the potential of social media data.

765

Our findings highlight the usage of diverse technological approaches employed by researchers over the years, mainly focusing on NLP (Houston et al., 2012; de Oliveira and Guelpeli, 2020), ML (Hodas et al., 2015; Domala et al., 2020), and statistical approaches (Middleton et al., 2013; Lu and Yuan, 2021) to address the challenges in identifying relevant and actionable information from social media to apply in the various phases of disaster management. We discussed various algorithms used over the years to collect and analyse social media data. These methodologies offer the means to identify noise which improves the data and relevance filtering (De Albuquerque et al., 2015; Jiang et al., 2022).

770

Our review also focused on the influence of historical disaster events on the researchers (Gupta et al., 2013; Ferris et al., 2016) and observed that major disaster events were considered case studies (see section 5.2). Such events will contain vast amounts of data which helps in gaining a wider perspective from multiple dimensions. By categorising the articles into nine actionable information classes (see section 4.3 and section 5.5), we observed the multifaceted usage of social media data in various applications. Notably, some researchers have achieved classification into multiple AI classes, as shown in table 2. This success points to the potential usage of social media data in disaster response, preparedness, and relief efforts.

775

The studies included in the critical review that employed a spatiotemporal analysis mostly studied hurricanes and floods (Rossi et al., 2018; Madichetty, 2020). We observed that Hurricane Sandy was one of the key events that was used as a case study by the researchers (Pourebahim et al., 2019; Wang et al., 2019). Across the majority of the articles used in this review, Twitter was the most prevalent platform. Other platforms such as Facebook and Weibo were also used, but in limited numbers (Xu et al., 2016; Han et al., 2020).

780

Through this critical review, we were able to conclude that exclusionary criterias implemented using the current technologies such as NLP and ML, certainly aid in relevance filtering of social media data. Moreover, it highlights that while actionable in-



785 formation for disaster management can indeed be extracted from social media, there is a need for more emphasis on addressing
data reliability (Bruns and Liang, 2012; Muhammad et al., 2018). A predominant challenge identified in social media usage for
disaster management is the spreading of rumours and misinformation, a concern that holds critical implications in emergencies
(Mendoza et al., 2010; Zhang et al., 2019b). Notably, our observation underscores a relatively limited focus on the analysis
of community and stakeholder interactions, which holds significant potential as a major contribution to first responders during
790 disasters.

Our review also proposed best practices for the usage of social media to the community and researchers. We suggested meth-
ods of posting disaster-related content on social media to gain maximum reach and attention. We suggested including account
tagging and hashtags of concerned authority accounts to receive attention. We also observed through the critical review that
clarity in the post content and inclusion of multimedia improves credibility (Muhammad et al., 2018; Alam et al., 2018). We
795 suggested methods of extracting social media data to the researchers and the good practices to utilise them.

In essence, this review not only provides a comprehensive overview of the existing literature but also aims to contribute to
future studies to explore various disciplines in leveraging social media data to fortify disaster management efforts.

Appendix A: Appendix A

800 Following is the list of acronyms and abbreviations used in this article.

Acronym	Description
ANN	Artificial Neural Networks
BoW	Bag-of-Words
CNN	Convolutional Neural Network
DT	Decision Trees
En	Entropy
Gv	Glove
KNN	K Nearest Neighbours
LDA	Latent Dirichlet Allocation
LSTM	Long Short-Term Memory
ML	Machine Learning
NB	Naive Bayes
NER	Named Entity Recognition
NLP	Natural Language Processing
NN	Neural Networks
PCA	Principal Component Analysis



Acronym	Description
PoS	Part-of-Speech
Regex	Regular Expression
RF	Random Forest
SVM	Support Vector Machines
TF-IDF	Term Frequency-Inverse Document Frequency
UGI	User Generated Information
Vc	Vectorisation
VGI	Volunteered Geographic Information

Author contributions. L.S Gopal and R. Prabha conceptualized and devised the methodology; L.S Gopal worked on the visualization and D. Malamud provided guidance. D. Malamud and H. Thirugnanam provided crucial supervision, review, and editing. M.V. Ramesh contributed significantly to the conceptualization and supervision phases. The manuscript was prepared by L.S. Gopal, integrating contributions from all co-authors.

805 *Competing interests.* At least one of the (co-)authors is a member of the editorial board of Natural Hazards and Earth System Sciences.

Acknowledgements. We would like to express our immense gratitude to our beloved Chancellor Sri Mata Amritanandamayi Devi (AMMA) for providing the motivation and inspiration for doing this research work. I would like to express my deepest gratitude to the late Dr. Rekha Prabha for her invaluable guidance and significant contributions to this work. We thank Ms. Emma Bee (Senior Geospatial Analyst, British Geological Survey) for her valuable input. This review was done as part of the Landslide multi-hazard Risk Assessment, preparedness and Early Warning in South Asia: Integrating Meteorology, landscape, and Society (LANDSLIP) project. I would like to thank the team members, Mr. Ramesh Guntha, Mr. Sudarshan Navada, Mr. Y. V. Rayudu, Ms. Divya Pullarkatt, Ms. Aswathy A and Ms. Krishnendu K for their contributions. We also thank Mr. Subhilash Sadanandan, Mr. Sibun N., and Mr. Sravan Thampan for their valuable suggestions in graphical visualizations. Additionally, we acknowledge the use of AI tools for checking grammar, spelling, and rephrasing of sentences wherever required.



815 References

- Abedin, B. and Babar, A.: Institutional vs. non-institutional use of social media during emergency response: A case of twitter in 2014 Australian bush fire, *Information Systems Frontiers*, 20, 729–740, <https://doi.org/10.1007/s10796-017-9789-4>, 2018.
- Abel, F., Hauff, C., Houben, G.-J., Stronkman, R., and Tao, K.: Semantics+ filtering+ search= twitcident. exploring information in social web streams, in: *Proceedings of the 23rd ACM conference on Hypertext and social media*, pp. 285–294, 2012.
- 820 Abirami, S., Chitra, P., et al.: Real time twitter based disaster response system for indian scenarios, in: *2019 26th International Conference on High Performance Computing, Data and Analytics Workshop (HiPCW)*, Hyderabad, India, December 17-20, 2019, pp. 82–86, IEEE, 2019.
- Aboualola, M., Abualsaud, K., Khattab, T., Zorba, N., and Hassanein, H. S.: Edge Technologies for Disaster Management: A Survey of Social Media and Artificial Intelligence Integration, *IEEE Access*, 11, <https://doi.org/10.1109/ACCESS.2023.3293035>, 2023.
- 825 Acikara, T., Xia, B., Yigitcanlar, T., and Hon, C.: Contribution of Social Media Analytics to Disaster Response Effectiveness: A Systematic Review of the Literature, *Sustainability*, 15, 8860, <https://doi.org/10.3390/su15118860>, 2023.
- Adam, N. R., Shafiq, B., and Staffin, R.: Spatial computing and social media in the context of disaster management, *IEEE Intell Syst*, 27, 90–96, <https://doi.org/10.1109/MIS.2012.113>, 2012.
- Ahmad, K., Pogorelov, K., Riegler, M., Conci, N., and Halvorsen, P.: Social media and satellites: Disaster event detection, linking and 830 summarization, *Multimed Tools Appl*, 78, 2837–2875, <https://doi.org/10.1007/s11042-018-5982-9>, 2019.
- Ai, F., Comfort, L. K., Dong, Y., and Znati, T.: A dynamic decision support system based on geographical information and mobile social networks: A model for tsunami risk mitigation in Padang, Indonesia, *Safety Sci.*, 90, 62–74, <https://doi.org/10.1016/j.ssci.2015.09.022>, 2016.
- Aisha, T. S., Wok, S., Manaf, A. M. A., and Ismail, R.: Exploring the use of social media during the 2014 flood in Malaysia, *Procd Soc Behv*, 835 211, 931–937, <https://doi.org/10.1016/j.sbspro.2015.11.123>, 2015.
- Akhter, M. P., Zheng, J., Afzal, F., Lin, H., Riaz, S., and Mehmood, A.: Supervised ensemble learning methods towards automatically filtering Urdu fake news within social media, *PeerJ Computer Science*, 7, e425, <https://doi.org/10.7717/peerj-cs.425>, 2021.
- Akter, S. and Wamba, S. F.: Big data and disaster management: a systematic review and agenda for future research, *Ann Oper Res.*, 283, 939–959, <https://doi.org/10.1007/s10479-017-2584-2>, 2019.
- 840 Alam, F., Ofli, F., and Imran, M.: Crisismmd: Multimodal twitter datasets from natural disasters, in: *Proceedings of the international AAAI conference on web and social media*, California, USA, June 25-28, 2018, vol. 12, 2018.
- Alampay, E. A., Asuncion, X. V., and Santos, M. d.: Management of social media for disaster risk reduction and mitigation in Philippine local government units, in: *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance*, Galway, Ireland, April 04-06, 2018, pp. 183–190, 2018.
- 845 Albris, K.: The switchboard mechanism: How social media connected citizens during the 2013 floods in Dresden, *J Conting Crisis Man*, 26, 350–357, <https://doi.org/10.1111/1468-5973.12201>, 2018.
- Alexander, D. E.: Social media in disaster risk reduction and crisis management, *Sci Eng Ethics*, 20, 717–733, <https://doi.org/10.1007/s11948-013-9502-z>, 2014.
- Anson, S., Watson, H., Wadhwa, K., and Metz, K.: Analysing social media data for disaster preparedness: Understanding the opportunities 850 and barriers faced by humanitarian actors, *Int J Disast Risk Re.*, 21, 131–139, <https://doi.org/10.1016/j.ijdr.2016.11.014>, 2017.



- Arapostathis, S. G.: Utilising Twitter for disaster management of fire events: steps towards efficient automation, *Arab J Geosci*, 14, 667, <https://doi.org/10.1007/s12517-021-06768-2>, 2021.
- Ashktorab, Z., Brown, C., Nandi, M., and Culotta, A.: Tweedr: Mining twitter to inform disaster response., in: *ISCRAM 2014 Conference Proceedings - 11th International Conference on Information Systems for Crisis Response and Management*, Pennsylvania, USA, May 18-21, 2014, 142, pp. 269–272, 2014.
- 855 Aswathy, A., Prabha, R., Gopal, L. S., Pullarkatt, D., and Ramesh, M. V.: An efficient twitter data collection and analytics framework for effective disaster management, in: *2022 IEEE Delhi Section Conference (DELCON)*, pp. 1–6, IEEE, 2022.
- Bala, M. M., Navya, K., and Shruthilaya, P.: Text mining on real time Twitter data for disaster response, *Int. J. Civ. Eng. Technol*, 8, 20–29, 2017.
- 860 Banujan, K., Kumara, B. T., and Paik, I.: Twitter and online news analytics for enhancing post-natural disaster management activities, in: *2018 9th International Conference on Awareness Science and Technology (iCAST)*, Fukuoka, Japan, September 19-21, 2018, pp. 302–307, IEEE, 2018.
- Batrinca, B. and Treleaven, P. C.: Social media analytics: a survey of techniques, tools and platforms, *Ai & Society*, 30, 89–116, <https://doi.org/10.1007/s00146-014-0549-4>, 2015.
- 865 Beigi, G., Hu, X., Maciejewski, R., and Liu, H.: An overview of sentiment analysis in social media and its applications in disaster relief, *Sentiment analysis and ontology engineering: An environment of computational intelligence*, pp. 313–340, https://doi.org/10.1007/978-3-319-30319-2_13, 2016.
- Black, A., Mascaro, C., Gallagher, M., and Goggins, S. P.: Twitter zombie: Architecture for capturing, socially transforming and analyzing the Twittersphere, in: *Proceedings of the 2012 ACM International Conference on Supporting Group Work*, Florida, USA, October 27-31, 2012, pp. 229–238, 2012.
- 870 Boas, I., Chen, C., Wiegel, H., and He, G.: The role of social media-led and governmental information in China’s urban disaster risk response: The case of Xiamen, *Int J Disast Risk Re.*, 51, 101 905, <https://doi.org/10.1016/j.ijdr.2020.101905>, 2020.
- Boaz, A., Ashby, D., Young, K., et al.: *Systematic reviews: what have they got to offer evidence based policy and practice?*, ESRC UK Centre for Evidence Based Policy and Practice London, 2002.
- 875 Brangbour, E., Bruneau, P., Marchand-Maillet, S., Hostache, R., Matgen, P., Chini, M., and Tamisier, T.: Extracting localized information from a Twitter corpus for flood prevention, *arXiv preprint arXiv:1903.04748*, 2019.
- Brangbour, E., Bruneau, P., Marchand-Maillet, S., Hostache, R., Chini, M., Matgen, P., and Tamisier, T.: Computing flood probabilities using Twitter: application to the Houston urban area during Harvey, *arXiv preprint arXiv:2012.03731*, 2020.
- Branz, L. and Brockmann, P.: Sentiment analysis of Twitter data: towards filtering, analyzing and interpreting social network data, in: *Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems*, Hamilton, New Zealand, June 25-29, 2018, pp. 238–241, 2018.
- 880 Brena, G., Brambilla, M., Ceri, S., Di Giovanni, M., Pierri, F., and Ramponi, G.: News sharing user behaviour on twitter: A comprehensive data collection of news articles and social interactions, in: *Proceedings of the International AAAI Conference on Web and Social Media*, München, Germany, June 11-14, 2019, vol. 13, pp. 592–597, 2019.
- 885 Bruns, A. and Liang, Y. E.: Tools and methods for capturing Twitter data during natural disasters, *First Monday*, 17, 1–8, <https://doi.org/10.5210/fm.v17i4.3937>, 2012.



- Bukar, U. A., Sidi, F., Jabar, M. A., Nor, R. N. H., Abdullah, S., Ishak, I., Alabadla, M., and Alkhalifah, A.: How Advanced Technological Approaches Are Reshaping Sustainable Social Media Crisis Management and Communication: A Systematic Review, *Sustainability*, 14, 5854, <https://doi.org/10.3390/su14105854>, 2022.
- 890 Bunney, S., Ward, S., and Butler, D.: Inter-organisational resilience for flood focussed emergency planning: examining multi-agency connectedness through Twitter, *Water Practice & Technology*, 13, 321–327, <https://doi.org/10.2166/wpt.2018.043>, 2018.
- Cameron, M. A., Power, R., Robinson, B., and Yin, J.: Emergency situation awareness from twitter for crisis management, in: Proceedings of the 21st international conference on world wide web, Lyon, France, April 16 - 20, 2012, pp. 695–698, 2012.
- Campan, A., Atnafu, T., Truta, T. M., and Nolan, J.: Is data collection through twitter streaming api useful for academic
895 research?, in: 2018 IEEE international conference on big data (big data), December 10-13, 2018, pp. 3638–3643, IEEE, <https://doi.org/10.1109/BigData44402.2018>, 2018.
- Carley, K. M., Malik, M., Landwehr, P. M., Pfeffer, J., and Kowalchuck, M.: Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia, *Safety Sci.*, 90, 48–61, <https://doi.org/10.1016/j.ssci.2016.04.002>, 2016a.
- Carley, K. M., Malik, M., Landwehr, P. M., Pfeffer, J., and Kowalchuck, M.: Crowd sourcing disaster management: The complex nature of
900 Twitter usage in Padang Indonesia, *Safety Sci.*, 90, 48–61, <https://doi.org/10.1016/j.ssci.2016.04.002>, 2016b.
- Carter, L., Thatcher, J. B., and Wright, R.: Social media and emergency management: Exploring state and local tweets, in: 2014 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, January 6-9, 2014, pp. 1968–1977, IEEE, 2014.
- Castillo, C., Mendoza, M., and Poblete, B.: Predicting information credibility in time-sensitive social media, *Internet Res*, 23, 560–588, <https://doi.org/10.1108/IntR-05-2012-0095>, 2013.
- 905 Cenni, D., Nesi, P., Pantaleo, G., and Zaza, I.: Twitter vigilance: A multi-user platform for cross-domain Twitter data analytics, NLP and sentiment analysis, in: 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CB-DCOM/IOP/SCI), California, USA, August 4-8, 2017, pp. 1–8, IEEE, 2017.
- Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J., and Waters, N.: Using Twitter for tasking remote-sensing data
910 collection and damage assessment: 2013 Boulder flood case study, *International Journal of Remote Sensing*, 37, 100–124, <https://doi.org/10.1080/01431161.2015.1117684>, 2016.
- Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D. S., and Ertl, T.: Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition, in: 2012 IEEE conference on visual analytics science and technology (VAST), Seattle, WA, October 14-19, 2012, pp. 143–152, IEEE, 2012.
- 915 Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., and Ebert, D. S.: Public behavior response analysis in disaster events utilizing visual analytics of microblog data, *Computers & Graphics*, 38, 51–60, <https://doi.org/10.1016/j.cag.2013.10.008>, 2014.
- Chatfield, A. T. and Brajawidagda, U.: Twitter early tsunami warning system: A case study in Indonesia’s natural disaster management, in: 2013 46th Hawaii international conference on system sciences, Maui, USA, January 7-10, 2013, pp. 2050–2060, IEEE, 2013.
- Chen, M., Mao, S., and Liu, Y.: Big data: A survey, *Mobile networks and applications*, 19, 171–209, <https://doi.org/10.1007/s11036-013-0489-0>, 2014.
- 920 Chen, X., Elmes, G., Ye, X., and Chang, J.: Implementing a real-time Twitter-based system for resource dispatch in disaster management, *GeoJournal*, 81, 863–873, <https://doi.org/10.1007/s10708-016-9745-8>, 2016.



- Cheng, J. W., Mitomo, H., Otsuka, T., and Jeon, S. Y.: Cultivation effects of mass and social media on perceptions and behavioural intentions in post-disaster recovery—The case of the 2011 Great East Japan Earthquake, *Telemat Inform*, 33, 753–772, <https://doi.org/10.1016/j.tele.2015.12.001>, 2016.
- Choi, S. and Bae, B.: The real-time monitoring system of social big data for disaster management, in: *Computer Science and its Applications: Ubiquitous Information Technologies*, Saida, Algeria, May 20–21, 2015, pp. 809–815, Springer, 2015.
- Čišija, M., Žunić, E., and Đonko, D.: Collection and sentiment analysis of twitter data on the political atmosphere, in: *2018 14th symposium on neural networks and applications (NEUREL)*, Belgrade, Serbia, November 20–21, 2018, pp. 1–5, IEEE, 2018.
- 925 Dashti, S., Palen, L., Heris, M. P., Anderson, K. M., Anderson, T. J., and Anderson, S.: Supporting disaster reconnaissance with social media data: A design-oriented case study of the 2013 Colorado floods, in: *ISCRAM 2014 Conference Proceedings - 11th International Conference on Information Systems for Crisis Response and Management*, pp. 632–641, USA, 2014.
- De Albuquerque, J. P., Herfort, B., Brenning, A., and Zipf, A.: A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management, *Int J Geogr Inf Sci.*, 29, 667–689, <https://doi.org/10.1080/13658816.2014.996567>, 2015.
- 935 de Oliveira, H. B. and Guelpeli, M. V. C.: Performance analysis of the Oráculo framework for data collection from Twitter, *Brazilian Journal of Development*, 6, 100 969–100 986, <https://doi.org/10.34117/bjdv6n12-549>, 2020.
- Delilah Roque, A., Pijawka, D., and Wutich, A.: The role of social capital in resiliency: Disaster recovery in Puerto Rico, *Risk, Hazards & Crisis in Public Policy*, 11, 204–235, <https://doi.org/10.1002/rhc3.12187>, 2020.
- 940 Domala, J., Dogra, M., Masrani, V., Fernandes, D., D’souza, K., Fernandes, D., and Carvalho, T.: Automated identification of disaster news for crisis management using machine learning and natural language processing, in: *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, July 2–4, 2020, pp. 503–508, IEEE, <https://doi.org/10.1109/ICESC48915.2020>, 2020.
- Driscoll, K. and Walker, S.: Big data, big questions! working within a black box: Transparency in the collection and production of big twitter data, *Int J Commun-US*, 8, 20, <https://doi.org/1932-8036/20140005>, 2014.
- 945 Earle, P. S., Bowden, D., and Guy, M.: Twitter earthquake detection: earthquake monitoring in a social world, *Ann Geophys*, 54, 708–715, <https://doi.org/10.4401/ag-5364>, 2011.
- Eckert, S., Sopory, P., Day, A., Wilkins, L., Padgett, D., Novak, J., Noyes, J., Allen, T., Alexander, N., Vanderford, M., et al.: Health-related disaster communication and social media: mixed-method systematic review, *Health Commun*, 33, 1389–1400, <https://doi.org/10.1080/10410236.2017.1351278>, 2018.
- 950 Eilander, D., Trambauer, P., Wagemaker, J., and Van Loenen, A.: Harvesting social media for generation of near real-time flood maps, *Procedia Engineer*, 154, 176–183, <https://doi.org/10.1016/j.proeng.2016.07.441>, 2016.
- Eriksson, M.: Lessons for crisis communication on social media: A systematic review of what research tells the practice, *International Journal of Strategic Communication*, 12, 526–551, <https://doi.org/10.1080/1553118X.2018.1510405>, 2018.
- 955 Fan, C., Esparza, M., Dargin, J., Wu, F., Oztekin, B., and Mostafavi, A.: Spatial biases in crowdsourced data: Social media content attention concentrates on populous areas in disasters, *Comput Environ Urban.*, 83, 101 514, <https://doi.org/10.1016/j.compenvurbsys.2020.101514>, 2020.
- Fan, C., Zhang, C., Yahja, A., and Mostafavi, A.: Disaster City Digital Twin: A vision for integrating artificial and human intelligence for disaster management, *Int J Inform Manage.*, 56, 102 049, <https://doi.org/10.1016/j.ijinfomgt.2019.102049>, 2021.



- 960 Fang, J., Hu, J., Shi, X., and Zhao, L.: Assessing disaster impacts and response using social media data in China: A case study of 2016
Wuhan rainstorm, *Int J Disast Risk Re.*, 34, 275–282, <https://doi.org/10.1016/j.ijdr.2018.11.027>, 2019.
- Fauzi, M. A.: Social media in disaster management: review of the literature and future trends through bibliometric analysis, *Nat Hazards*,
118, 953–975, <https://doi.org/10.1007/s11069-023-06079-7>, 2023.
- Ferris, T., Moreno-Centeno, E., Yates, J., Sung, K., El-Sherif, M., and Matarrita-Cascante, D.: Studying the usage of social media and mobile
965 technology during extreme events and their implications for evacuation decisions: A case study of Hurricane Sandy, *International Journal
of Mass Emergencies & Disasters*, 34, 204–230, <https://doi.org/10.1177/028072701603400202>, 2016.
- Fohringer, J., Dransch, D., Kreibich, H., and Schröter, K.: Social media as an information source for rapid flood inundation mapping, *Nat
Hazard Earth Sys*, 15, 2725–2738, <https://doi.org/10.5194/nhess-15-2725-2015>, 2015.
- Freberg, K., Saling, K., Vidoloff, K. G., and Eosco, G.: Using value modeling to evaluate social media messages: The case of Hurricane
970 Irene, *Public Relat Rev*, 39, 185–192, <https://doi.org/10.1016/j.pubrev.2013.02.010>, 2013.
- Gao, H., Barbier, G., and Goolsby, R.: Harnessing the crowdsourcing power of social media for disaster relief, *IEEE Intell Syst*, 26, 10–14,
<https://doi.org/10.1109/MIS.2011.52>, 2011.
- Gaspar, R., Pedro, C., Panagiotopoulos, P., and Seibt, B.: Beyond positive or negative: Qualitative sentiment analysis of social media reactions
to unexpected stressful events, *Comput Hum Behav*, 56, 179–191, <https://doi.org/10.1016/j.chb.2015.11.040>, 2016.
- 975 Gautam, G. and Yadav, D.: Sentiment analysis of twitter data using machine learning approaches and semantic analysis, in: 2014 Seventh
international conference on contemporary computing (IC3), Noida, India, August 7-9, 2014, pp. 437–442, IEEE, 2014.
- Gerlitz, C. and Rieder, B.: Mining one percent of Twitter: Collections, baselines, sampling, *M/C Journal*, 16, <https://doi.org/10.5204/mcj.620>,
2013.
- Ghani, N. A., Hamid, S., Hashem, I. A. T., and Ahmed, E.: Social media big data analytics: A survey, *Comput Hum Behav*, 101, 417–428,
980 <https://doi.org/10.1016/j.chb.2018.08.039>, 2019.
- Ghawana, T., Pashova, L., and Zlatanova, S.: Geospatial data utilisation in national disaster management frameworks and the
priorities of multilateral disaster management frameworks: case studies of India and Bulgaria, *ISPRS Int Geo-Inf.*, 10, 610,
<https://doi.org/10.3390/ijgi10090610>, 2021.
- Gopal, L. S., Prabha, R., Pullarkatt, D., and Ramesh, M. V.: Machine learning based classification of online news data for disaster manage-
985 ment, in: 2020 IEEE Global Humanitarian Technology Conference (GHTC), pp. 1–8, IEEE, 2020.
- Gopal, L. S., Prabha, R., and Ramesh, M. V.: Developing information extraction system for disaster impact factor retrieval from Web News
Data, in: *Information and Communication Technology for Competitive Strategies (ICTCS 2021) Intelligent Strategies for ICT*, pp. 357–
365, Springer, 2022.
- Gopal, L. S., Prabha, R., Thiruganam, H., Ramesh, M. V., and Malamud, B. D.: Social Media Literature Database,
990 <https://doi.org/10.5281/zenodo.10803017>, 2023.
- Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A., and Chakraborty, B.: A review on application of data mining techniques to combat
natural disasters, *Ain Shams Engineering Journal*, 9, 365–378, <https://doi.org/10.1016/j.asej.2016.01.012>, 2018.
- Granell, C. and Ostermann, F. O.: Beyond data collection: Objectives and methods of research using VGI and geo-social media for disaster
management, *Comput Environ Urban.*, 59, 231–243, <https://doi.org/10.1016/j.compenvurbsys.2016.01.006>, 2016.
- 995 Guan, X. and Chen, C.: Using social media data to understand and assess disasters, *Nat Hazards*, 74, 837–850,
<https://doi.org/10.1007/s11069-014-1217-1>, 2014.



- Gulnerman, A. G. and Karaman, H.: Spatial reliability assessment of social media mining techniques with regard to disaster domain-based filtering, *ISPRS Int Geo-Inf.*, 9, 245, <https://doi.org/10.3390/ijgi9040245>, 2020.
- Gunawong, P. and Butakhieo, N.: Social media in local administration: An empirical study of twitter use in flood management, in: 2016
1000 Conference for E-Democracy and Open Government (CeDEM), Krems, Austria, May 18-20, 2016, pp. 77–83, IEEE, 2016.
- Guntha, R., Rao, S. N., and Ramesh, M. V.: Architectural considerations for building a robust crowdsourced disaster relief application, in: 2020 international conference on COMMunication Systems & NETworkS (COMSNETS), pp. 638–641, IEEE, 2020a.
- Guntha, R., Rao, S. N., and Shivdas, A.: Lessons learned from deploying crowdsourced technology for disaster relief during Kerala floods, *Procedia Comput Sci*, 171, 2410–2419, <https://doi.org/10.1016/j.procs.2020.04.313>, 2020b.
- 1005 Gupta, A., Lamba, H., Kumaraguru, P., and Joshi, A.: Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy, in: Proceedings of the 22nd international conference on World Wide Web, pp. 729–736, 2013.
- Guy, M., Earle, P., Ostrum, C., Gruchalla, K., and Horvath, S.: Integration and dissemination of citizen reported and seismically derived earthquake information via social network technologies, in: Advances in Intelligent Data Analysis IX: 9th International Symposium, IDA 2010, Tucson, AZ, USA, May 19-21, 2010. Proceedings 9, pp. 42–53, Springer, 2010.
- 1010 Han, X., Wang, J., Zhang, M., and Wang, X.: Using social media to mine and analyze public opinion related to COVID-19 in China, *Int J ENV RES PUB HE*, 17, 2788, <https://doi.org/10.3390/ijerph17082788>, 2020.
- Hao, H. and Wang, Y.: Leveraging multimodal social media data for rapid disaster damage assessment, *Int J Disast Risk Re.*, 51, 101 760, <https://doi.org/10.1016/j.ijdr.2020.101760>, 2020.
- Hara, Y.: Behaviour analysis using tweet data and geo-tag data in a natural disaster, *Transp Res Proc*, 11, 399–412, <https://doi.org/10.1016/j.trpro.2015.12.033>, 2015.
- 1015 Haworth, B. and Bruce, E.: A review of volunteered geographic information for disaster management, *Geography Compass*, 9, 237–250, <https://doi.org/10.1111/gec3.12213>, 2015.
- Healy, E., Siddiqui, H., and Olmsted, A.: Classifying influenza outbreaks by analyzing and filtering Twitter data, in: 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), Cambridge, December 11-14, 2017, pp. 205–207, IEEE, <https://doi.org/10.23919/ICITST41748.2017>, 2017.
- 1020 Henry, D.: Twiscraper: a collaborative project to enhance twitter data collection, in: Proceedings of the 14th ACM international conference on web search and data mining, pp. 886–889, 2021.
- Herfort, B., de Albuquerque, J. P., Schelhorn, S.-J., and Zipf, A.: Exploring the geographical relations between social media and flood phenomena to improve situational awareness: A study about the river Elbe flood in June 2013, *Connecting a digital Europe through location and place*, pp. 55–71, https://doi.org/10.1007/978-3-319-03611-3_4, 2014.
- 1025 Hodas, N. O., Ver Steeg, G., Harrison, J., Chikkagoudar, S., Bell, E., and Corley, C. D.: Disentangling the lexicons of disaster response in twitter, in: Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, May 18-22, 2015, pp. 1201–1204, 2015.
- Houston, J. B., Pfefferbaum, B., and Rosenholtz, C. E.: Disaster news: Framing and frame changing in coverage of major US natural disasters, 2000–2010, *Journalism & Mass Communication Quarterly*, 89, 606–623, <https://doi.org/10.1177/1077699012456022>, 2012.
- 1030 Houston, J. B., Hawthorne, J., Perreault, M. F., Park, E. H., Goldstein Hode, M., Halliwell, M. R., Turner McGowen, S. E., Davis, R., Vaid, S., McElderry, J. A., et al.: Social media and disasters: a functional framework for social media use in disaster planning, response, and research, *Disasters*, 39, 1–22, <https://doi.org/10.1111/disa.12092>, 2015.
- Howe, A. W., Jennex, M. E., Bressler, G. H., and Frost, E.: Exercise 24: Using social media for crisis response, *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, 3, 36–54, <https://doi.org/10.4018/jisrcm.2011100103>, 2011.



- 1035 Hristidis, V., Chen, S.-C., Li, T., Luis, S., and Deng, Y.: Survey of data management and analysis in disaster situations, *J Syst Software*, 83, 1701–1714, <https://doi.org/10.1016/j.jss.2010.04.065>, 2010.
- Htein, M. K., Lim, S., and Zaw, T. N.: The evolution of collaborative networks towards more polycentric disaster responses between the 2015 and 2016 Myanmar floods, *Int J Disast Risk Re.*, 31, 964–982, <https://doi.org/10.1016/j.ijdr.2018.08.003>, 2018.
- Hu, X., Zhang, X., and Wei, J.: Public attention to natural hazard warnings on social media in China, *Weather, climate, and society*, 11, 183–197, <https://doi.org/10.1175/WCAS-D-17-0039.1>, 2019.
- 1040 Huang, P.-Y., Liang, J., Lamare, J.-B., and Hauptmann, A. G.: Multimodal filtering of social media for temporal monitoring and event analysis, in: *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, Yokohama, Japan, June 11–14, 2018, pp. 450–457, 2018.
- Huang, Q. and Xiao, Y.: Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery, *ISPRS Int Geo-Inf.*, 4, 1549–1568, <https://doi.org/10.3390/ijgi4031549>, 2015.
- 1045 Huang, Y., Guo, D., Kasakoff, A., and Grieve, J.: Understanding US regional linguistic variation with Twitter data analysis, *Comput Environ Urban.*, 59, 244–255, <https://doi.org/10.1016/j.compenvurbsys.2015.12.003>, 2016.
- Ilieva, R. T. and McPhearson, T.: Social-media data for urban sustainability, *Nature Sustainability*, 1, 553–565, <https://doi.org/10.1038/s41893-018-0153-6>, 2018.
- 1050 Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., and Meier, P.: Extracting information nuggets from disaster-Related messages in social media., *Iscram*, 201, 791–801, 2013a.
- Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., and Meier, P.: Practical extraction of disaster-relevant information from social media, in: *Proceedings of the 22nd international conference on world wide web*, Rio de Janeiro, Brazil May 13 - 17, 2013, pp. 1021–1024, 2013b.
- Imran, M., Castillo, C., Diaz, F., and Vieweg, S.: Processing social media messages in mass emergency: A survey, *ACM Comput Surv.*, 47, 1–38, <https://doi.org/10.1145/2771588>, 2015.
- 1055 Imran, M., Offi, F., Caragea, D., and Torralba, A.: Using AI and social media multimodal content for disaster response and management: Opportunities, challenges, and future directions, <https://doi.org/10.1016/j.ipm.2020.102261>, 2020.
- Islam, R. and Walkerden, G.: How do links between households and NGOs promote disaster resilience and recovery?: A case study of linking social networks on the Bangladeshi coast, *Nat Hazards*, 78, 1707–1727, <https://doi.org/10.1007/s11069-015-1797-4>, 2015.
- 1060 Jamali, M., Nejat, A., Ghosh, S., Jin, F., and Cao, G.: Social media data and post-disaster recovery, *Int J Inform Manage.*, 44, 25–37, <https://doi.org/10.1016/j.ijinfomgt.2018.09.005>, 2019.
- Jiang, L., Shi, L., Liu, L., Yao, J., and Ali, M. E.: User interest community detection on social media using collaborative filtering, *Wireless networks*, pp. 1–7, <https://doi.org/10.1007/s11276-018-01913-4>, 2022.
- Jitkajornwanich, K., Kongthong, C., Khongsoontornjaroen, N., Kaiyasuan, J., Lawawirojwong, S., Srestasathiern, P., Srisonphan, S., and 1065 Vateekul, P.: Utilizing Twitter Data for Early Flood Warning in Thailand, in: *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, USA, December 10–13, 2018, pp. 5165–5169, IEEE, 2018.
- Jongman, B., Wagemaker, J., Revilla Romero, B., and Coughlan de Perez, E.: Early flood detection for rapid humanitarian response: harnessing near real-time satellite and Twitter signals, *ISPRS Int Geo-Inf.*, 4, 2246–2266, <https://doi.org/10.3390/ijgi4042246>, 2015.
- Joseph, K., Landwehr, P. M., and Carley, K. M.: An approach to selecting keywords to track on twitter during a disaster., in: *ISCRAM 2014 Conference Proceedings - 11th International Conference on Information Systems for Crisis Response and Management*, Pennsylvania, USA, May 18–21, 2014, 38, 2014.
- 1070



- Jung, C.-T., Tsou, M.-H., and Issa, E.: Developing a real-time situation awareness viewer for monitoring disaster impacts using location-based social media messages in Twitter, in: International Conference on Location-Based Social Media Data, Athens, GA, USA March, pp. 12–14, 2015.
- 1075 Jurgens, M. and Helsloot, I.: The effect of social media on the dynamics of (self) resilience during disasters: A literature review, *J Conting Crisis Man*, 26, 79–88, <https://doi.org/10.1111/1468-5973.12212>, 2018.
- Kaewkitipong, L., Chen, C. C., and Ractham, P.: A community-based approach to sharing knowledge before, during, and after crisis events: A case study from Thailand, *Comput Hum Behav*, 54, 653–666, <https://doi.org/10.1016/j.chb.2015.07.063>, 2016.
- Kankanamge, N., Yigitcanlar, T., and Goonetilleke, A.: How engaging are disaster management related social media channels? The case of
1080 Australian state emergency organisations, *Int J Disast Risk Re.*, 48, 101 571, <https://doi.org/10.1016/j.ijdr.2020.101571>, 2020a.
- Kankanamge, N., Yigitcanlar, T., Goonetilleke, A., and Kamruzzaman, M.: Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets, *Int J Disast Risk Re.*, 42, 101 360, <https://doi.org/10.1016/j.ijdr.2019.101360>, 2020b.
- Karimizarani, M. and Moradkhani, H.: Social response and Disaster management: Insights from twitter data Assimilation on Hurricane Ian,
1085 *Int J Disast Risk Re.*, 95, 103 865, <https://doi.org/10.1016/j.ijdr.2023.103865>, 2023.
- Kejriwal, M. and Gu, Y.: A pipeline for rapid post-crisis twitter data acquisition, filtering and visualization, *Technologies*, 7, 33, <https://doi.org/10.3390/technologies7020033>, 2019.
- Khaleq, A. A. and Ra, I.: Cloud-based disaster management as a service: A microservice approach for hurricane twitter data analysis, in: 2018 IEEE Global Humanitarian Technology Conference (GHTC), San Jose, CA, October 18–21, 2018, pp. 1–8, IEEE, 2018.
- 1090 Kibanov, M., Stumme, G., Amin, I., and Lee, J. G.: Mining social media to inform peatland fire and haze disaster management, *Social Network Analysis and Mining*, 7, 1–19, <https://doi.org/10.1007/s13278-017-0446-1>, 2017.
- Kim, J. and Hastak, M.: Social network analysis: Characteristics of online social networks after a disaster, *Int J Inform Manage.*, 38, 86–96, <https://doi.org/10.1016/j.ijinfomgt.2017.08.003>, 2018.
- Kim, J., Bae, J., and Hastak, M.: Emergency information diffusion on online social media during storm Cindy in US, *Int J Inform Manage.*,
1095 40, 153–165, <https://doi.org/10.1016/j.ijinfomgt.2018.02.003>, 2018.
- Kim, T.: Observation on copying and pasting behavior during the Tohoku earthquake: Retweet pattern changes, *Int J Inform Manage.*, 34, 546–555, <https://doi.org/10.1016/j.ijinfomgt.2014.03.001>, 2014.
- Kitazawa, K. and Hale, S. A.: Social media and early warning systems for natural disasters: A case study of Typhoon Etau in Japan, *Int J Disast Risk Re.*, 52, 101 926, <https://doi.org/10.1016/j.ijdr.2020.101926>, 2021.
- 1100 Kogan, M., Palen, L., and Anderson, K. M.: Think local, retweet global: Retweeting by the geographically-vulnerable during Hurricane Sandy, in: Proceedings of the 18th ACM conference on computer supported cooperative work & social computing, Vancouver BC Canada, March 14 - 18, 2015, pp. 981–993, 2015.
- Kryvasheyev, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., and Cebrian, M.: Rapid assessment of disaster damage using social media activity, *Science advances*, 2, e1500 779, <https://doi.org/10.1126/sciadv.1500779>, 2016.
- 1105 Kurkcu, A., Zuo, F., Gao, J., Morgul, E. F., and Ozbay, K.: Crowdsourcing incident information for disaster response using twitter, in: Proceedings of the 65th Annual Meeting of Transportation Research Board, 2017.
- Lachlan, K. A., Westerman, D. K., and Spence, P. R.: Disaster news and subsequent information seeking: Exploring the role of spatial presence and perceptual realism, *Electronic News*, 4, 203–217, <https://doi.org/10.1177/1931243110387092>, 2010.



- Landwehr, P. M. and Carley, K. M.: Social media in disaster relief: Usage patterns, data mining tools, and current research directions, *Data mining and knowledge discovery for big data: Methodologies, Challenge and Opportunities*, pp. 225–257, https://doi.org/10.1007/978-3-642-40837-3_7, 2014.
- 1110 Laylavi, F., Rajabifard, A., and Kalantari, M.: Event relatedness assessment of Twitter messages for emergency response, *Inform Process Manag.*, 53, 266–280, <https://doi.org/10.1016/j.ipm.2016.09.002>, 2017.
- Le Coz, J., Patalano, A., Collins, D., Guillén, N. F., García, C. M., Smart, G. M., Bind, J., Chiaverini, A., Le Boursicaud, R., Dramais, G.,
1115 et al.: Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand, *J Hydrol.*, 541, 766–777, <https://doi.org/10.1016/j.jhydrol.2016.07.036>, 2016.
- Lee, C.-H., Yang, H.-C., Chien, T.-F., and Wen, W.-S.: A novel approach for event detection by mining spatio-temporal information on microblogs, in: *2011 International Conference on Advances in Social Networks Analysis and Mining*, Kaohsiung, Taiwan, July 25–27, 2011, pp. 254–259, IEEE, 2011.
- 1120 Leon, E., Alberoni, C., Wister, M., and Hernández-Nolasco, J. A.: Flood early warning system by Twitter using LoRa, in: *The Twelfth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, Athens, Greece, November 18–22, 2018, vol. 2, p. 1213, MDPI, 2018.
- Li, H., Guevara, N., Herndon, N., Caragea, D., Neppalli, K., Caragea, C., Squicciarini, A. C., and Tapia, A. H.: Twitter Mining for Disaster Response: A Domain Adaptation Approach., in: *The 12th International Conference on Information Systems for Crisis Response and Management*, Norway, May 24–27, 2015, 2015.
- 1125 Li, L., Zhang, Q., Tian, J., and Wang, H.: Characterizing information propagation patterns in emergencies: A case study with Yiliang Earthquake, *Int J Inform Manage.*, 38, 34–41, <https://doi.org/10.1016/j.ijinfomgt.2017.08.008>, 2018a.
- Li, L., Bensi, M., Cui, Q., Baecher, G. B., and Huang, Y.: Social media crowdsourcing for rapid damage assessment following a sudden-onset natural hazard event, *Int J Inform Manage.*, 60, 102–110, <https://doi.org/10.1016/j.ijinfomgt.2021.102378>, 2021.
- 1130 Li, Z., Wang, C., Emrich, C. T., and Guo, D.: A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods, *Cartogr Geogr Inf Sc.*, 45, 97–110, <https://doi.org/10.1080/15230406.2016.1271356>, 2018b.
- Lin, W.-Y., Wu, T.-H., Tsai, M.-H., Hsu, W.-C., Chou, Y.-T., and Kang, S.-C.: Filtering disaster responses using crowdsourcing, *Automat Constr.*, 91, 182–192, <https://doi.org/10.1016/j.autcon.2018.03.016>, 2018.
- Lin, X., Spence, P. R., Sellnow, T. L., and Lachlan, K. A.: Crisis communication, learning and responding: Best practices in social media,
1135 *Comput Hum Behav.*, 65, 601–605, <https://doi.org/10.1016/j.chb.2016.05.080>, 2016.
- Liu, F. and Lee, H. J.: Use of social network information to enhance collaborative filtering performance, *Expert Syst Appl.*, 37, 4772–4778, <https://doi.org/10.1016/j.eswa.2009.12.061>, 2010.
- Liu, L. and Stevenson, M. D.: A cross-cultural analysis of stance in disaster news reports, *Australian Review of Applied Linguistics*, 36, 197–220, <https://doi.org/10.1075/aral.36.2.05liu>, 2013.
- 1140 Liu, W., Lai, C.-H., and Xu, W. W.: Tweeting about emergency: A semantic network analysis of government organizations’ social media messaging during hurricane Harvey, *Public Relat Rev.*, 44, 807–819, <https://doi.org/10.1016/j.pubrev.2018.10.009>, 2018.
- Liu, X., Kar, B., Ishino, F. A. M., Zhang, C., and Williams, F.: Extracting Reliable Twitter Data for Flood Risk Communication using Manual Assessment and Google Vision API from Text and Images, Preprints, 2020.
- LÓPEZ-MARRERO, T.: An integrative approach to study and promote Nat Hazards. adaptive capacity: A case study of two flood-prone
1145 communities in puerto rico, *Geogr J.*, 176, 150–163, <https://doi.org/10.1111/j.1475-4959.2010.00353.x>, 2010.



- Loynes, C., Ouenniche, J., and De Smedt, J.: The detection and location estimation of disasters using Twitter and the identification of Non-Governmental Organisations using crowdsourcing, *Ann Oper Res*, 308, 339–371, <https://doi.org/10.1007/s10479-020-03684-8>, 2022.
- Lu, H. and Yuan, S.: What motivates information sharing about disaster victims on social media? Exploring the role of compassion, sadness, expectancy violation, and enjoyment, *Int J Disast Risk Re.*, 63, 102 431, <https://doi.org/10.1016/j.ijdr.2021.102431>, 2021.
- 1150 Lu, Y. and Yang, D.: Information exchange in virtual communities under extreme disaster conditions, *Decis Support Syst*, 50, 529–538, <https://doi.org/10.1016/j.dss.2010.11.011>, 2011.
- Luna, S. and Pennock, M. J.: Social media applications and emergency management: A literature review and research agenda, *Int J Disast Risk Re.*, 28, 565–577, <https://doi.org/10.1016/j.ijdr.2018.01.006>, 2018.
- Mac Kim, S., Wan, S., Paris, C., Jin, B., and Robinson, B.: The effects of data collection methods in Twitter, in: *Proceedings of the First Workshop on NLP and Computational Social Science*, pp. 86–91, 2016.
- 1155 Madichetty, S.: Classifying informative and non-informative tweets from the twitter by adapting image features during disaster, *Multimed Tools Appl*, 79, 28 901–28 923, <https://doi.org/10.1007/s11042-020-09343-1>, 2020.
- Madichetty, S. and Sridevi, M.: A novel method for identifying the damage assessment tweets during disaster, *Future Generation Computer Systems*, 116, 440–454, <https://doi.org/10.1016/j.future.2020.10.037>, 2021.
- 1160 Mandel, B., Culotta, A., Boulahanis, J., Stark, D., Lewis, B., and Rodrigue, J.: A demographic analysis of online sentiment during hurricane irene, in: *Proceedings of the second workshop on language in social media*, Montréal, Canada, June, 2012, pp. 27–36, 2012.
- Martín, Y., Li, Z., and Cutter, S. L.: Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew, *PLoS one*, 12, e0181 701, <https://doi.org/10.1371/journal.pone.0181701>, 2017.
- Martínez-Rojas, M., del Carmen Pardo-Ferreira, M., and Rubio-Romero, J. C.: Twitter as a tool for the management and analysis of emergency situations: A systematic literature review, *Int J Inform Manage.*, 43, 196–208, <https://doi.org/10.1016/j.ijinfomgt.2018.07.008>, 2018.
- 1165 Mazoyer, B., Cagé, J., Hudelot, C., and Viaud, M.-L.: Real-time collection of reliable and representative tweets datasets related to news events, in: *First International Workshop on Analysis of Broad Dynamic Topics over Social Media (BroDyn 2018)* co-located with the 40th European Conference on Information Retrieval (ECIR 2018), Grenoble, France, March 26, 2018, 2018.
- 1170 McClendon, S. and Robinson, A. C.: Leveraging geospatially-oriented social media communications in disaster response, *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, 5, 22–40, <https://doi.org/10.4018/jiscrm.2013010102>, 2013.
- McCormick, T. H., Lee, H., Cesare, N., Shojaie, A., and Spiro, E. S.: Using Twitter for demographic and social science research: Tools for data collection and processing, *Sociol Method Res.*, 46, 390–421, <https://doi.org/10.1177/0049124115605339>, 2017.
- 1175 Mendoza, M., Poblete, B., and Castillo, C.: Twitter under crisis: Can we trust what we RT?, in: *Proceedings of the first workshop on social media analytics*, Washington D.C. District of Columbia, July 25 - 28, 2010, pp. 71–79, 2010.
- Middleton, S. E., Middleton, L., and Modafferi, S.: Real-time crisis mapping of natural disasters using social media, *IEEE Intell Syst*, 29, 9–17, <https://doi.org/10.1109/MIS.2013.126>, 2013.
- Muhammad, K., Ahmad, J., and Baik, S. W.: Early fire detection using convolutional neural networks during surveillance for effective disaster management, *Neurocomputing*, 288, 30–42, <https://doi.org/10.1016/j.neucom.2017.04.083>, 2018.
- 1180 Mukkamala, A. M. and Beck, R.: Enhancing Disaster Management Through Social Media Analytics To Develop Situation Awareness: What Can Be Learned From Twitter Messages About Hurricane Sandy?, in: *PACIS 2016 Proceedings*, Chiayi, Taiwan, June 27-July 1, 2016, AIS Electronic Library (AISeL), 2016.



- Musaev, A., Sheinidashtegol, P., Conrad, E., and Navathe, S. B.: Mining Twitter Data for Landslide Events Reported Worldwide, *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, 10, 47–64, <https://doi.org/10.4018/ijiscram.2018100103>, 2018.
- Nair, M. R., Ramya, G., and Sivakumar, P. B.: Usage and analysis of Twitter during 2015 Chennai flood towards disaster management, *Procedia Comput Sci*, 115, 350–358, 2017.
- Nazer, T. H., Xue, G., Ji, Y., and Liu, H.: Intelligent disaster response via social media analysis a survey, *ACM SIGKDD Explorations Newsletter*, 19, 46–59, <https://doi.org/10.1145/3137597.3137602>, 2017.
- Neppalli, V. K., Caragea, C., Squicciarini, A., Tapia, A., and Stehle, S.: Sentiment analysis during Hurricane Sandy in emergency response, *Int J Disast Risk Re.*, 21, 213–222, <https://doi.org/10.1016/j.ijdr.2016.12.011>, 2017.
- Neubaum, G., Rösner, L., Rosenthal-von der Pütten, A. M., and Krämer, N. C.: Psychosocial functions of social media usage in a disaster situation: A multi-methodological approach, *Comput Hum Behav*, 34, 28–38, <https://doi.org/10.1016/j.chb.2014.01.021>, 2014.
- 1195 Nguyen, D. T., Ofli, F., Imran, M., and Mitra, P.: Damage assessment from social media imagery data during disasters, in: *Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining*, Sydney, Australia, 31 July - 3 August 2017, pp. 569–576, 2017.
- Niles, M. T., Emery, B. F., Reagan, A. J., Dodds, P. S., and Danforth, C. M.: Social media usage patterns during natural hazards, *PloS one*, 14, e0210484, <https://doi.org/10.1371/journal.pone.0210484>, 2019.
- 1200 Ofli, F., Meier, P., Imran, M., Castillo, C., Tuia, D., Rey, N., Briant, J., Millet, P., Reinhard, F., Parkan, M., et al.: Combining human computing and machine learning to make sense of big (aerial) data for disaster response, *Big data*, 4, 47–59, <https://doi.org/10.1089/big.2014.0064>, 2016.
- Ogie, R. I., Clarke, R. J., Forehead, H., and Perez, P.: Crowdsourced social media data for disaster management: Lessons from the PetaJakarta.org project, *Comput Environ Urban.*, 73, 108–117, <https://doi.org/10.1016/j.compenvurb.2018.09.002>, 2019.
- 1205 Olteanu, A., Castillo, C., Diaz, F., and Vieweg, S.: Crisislex: A lexicon for collecting and filtering microblogged communications in crises, in: *Proceedings of the international AAAI conference on web and social media*, Ann Arbor, Michigan USA, June 1-4, 2014, vol. 8, pp. 376–385, 2014.
- Olteanu, A., Vieweg, S., and Castillo, C.: What to expect when the unexpected happens: Social media communications across crises, in: *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, Vancouver BC Canada, March 14 - 18, 2015, pp. 994–1009, 2015.
- 1210 Oussalah, M., Bhat, F., Challis, K., and Schnier, T.: A software architecture for Twitter collection, search and geolocation services, *Knowl-Based Syst*, 37, 105–120, <https://doi.org/10.1016/j.knosys.2012.07.017>, 2013.
- Palen, L. and Hughes, A. L.: Social media in disaster communication, *Handbook of disaster research*, pp. 497–518, https://doi.org/10.1007/978-3-319-63254-4_24, 2018.
- 1215 Palen, L., Starbird, K., Vieweg, S., and Hughes, A.: Twitter-based information distribution during the 2009 Red River Valley flood threat, *Bulletin of the American Society for Information Science and Technology*, 36, 13–17, <https://doi.org/10.1002/bult.2010.1720360505>, 2010.
- Panagiotopoulos, P., Barnett, J., Bigdeli, A. Z., and Sams, S.: Social media in emergency management: Twitter as a tool for communicating risks to the public, *Technol Forecast Soc*, 111, 86–96, <https://doi.org/10.1016/j.techfore.2016.06.010>, 2016.
- 1220 Pender, B., Currie, G., Delbosc, A., and Shiwakoti, N.: Social media use during unplanned transit network disruptions: A review of literature, *Transport Rev*, 34, 501–521, <https://doi.org/10.1080/01441647.2014.915442>, 2014.



- Phengsuwan, J., Thekkummal, N. B., Shah, T., James, P., Thakker, D., Sun, R., Pullarkatt, D., Hemalatha, T., Ramesh, M. V., and Ranjan, R.: Context-based knowledge discovery and querying for social media data, in: 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), pp. 307–314, IEEE, 2019.
- 1225 Plachouras, V., Stavrakas, Y., and Andreou, A.: Assessing the coverage of data collection campaigns on Twitter: A case study, in: On the Move to Meaningful Internet Systems: OTM 2013 Workshops: Confederated International Workshops: OTM Academy, OTM Industry Case Studies Program, ACM, EI2N, ISDE, META4eS, ORM, SeDeS, SINCOM, SMS, and SOMOCO 2013, Graz, Austria, September 9-13, 2013, Proceedings, pp. 598–607, Springer, 2013.
- Podhoranyi, M.: A comprehensive social media data processing and analytics architecture by using big data platforms: a case study of twitter flood-risk messages, *Earth Sci Inform.*, 14, 913–929, <https://doi.org/10.1007/s12145-021-00601-w>, 2021.
- 1230 Pourebrahim, N., Sultana, S., Edwards, J., Gochanour, A., and Mohanty, S.: Understanding communication dynamics on Twitter during natural disasters: A case study of Hurricane Sandy, *Int J Disast Risk Re.*, 37, 101 176, <https://doi.org/10.1016/j.ijdr.2019.101176>, 2019.
- Purohit, H., Castillo, C., Diaz, F., Sheth, A., and Meier, P.: Emergency-relief coordination on social media: Automatically matching resource requests and offers, *First Monday*, 19, <https://doi.org/10.5210/fm.v19i1.4848>, 2014.
- 1235 Qarabash, N. A. and Qarabash, H. A.: Twitter location-based data: evaluating the methods of data collection provided by twitter API, *International Journal of Computing*, 19, 583–589, <https://doi.org/10.47839/ijc.19.4.1992>, 2018.
- Radianti, J., Hiltz, S. R., and Labaka, L.: An overview of public concerns during the recovery period after a major earthquake: Nepal twitter analysis, in: 2016 49th Hawaii international conference on system sciences (HICSS), Koloa, HI, USA, January 5-8, 2016, pp. 136–145, IEEE, 2016.
- 1240 Ragini, J. R., Anand, P. R., and Bhaskar, V.: Big data analytics for disaster response and recovery through sentiment analysis, *Int J Inform Manage.*, 42, 13–24, <https://doi.org/10.1016/j.ijinfomgt.2018.05.004>, 2018.
- Rahmadan, M. C., Hidayanto, A. N., Ekasari, D. S., Purwandari, B., et al.: Sentiment analysis and topic modelling using the lda method related to the flood disaster in jakarta on twitter, in: 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, November 19-20, 2020, pp. 126–130, IEEE, 2020.
- 1245 Rajput, A. A., Li, Q., Zhang, C., and Mostafavi, A.: Temporal network analysis of inter-organizational communications on social media during disasters: A study of Hurricane Harvey in Houston, *Int J Disast Risk Re.*, 46, 101 622, <https://doi.org/10.1016/j.ijdr.2020.101622>, 2020.
- Ravi Shankar, A., Fernandez-Marquez, J. L., Pernici, B., Scalia, G., Mondardini, M. R., and Di Marzo Serugendo, G.: Crowd4Ems: A crowdsourcing platform for gathering and geolocating social media content in disaster response, *Int Arch Photogramm*, 42, 331–340, <https://doi.org/10.5194/isprs-archives-XLII-3-W8-331-2019>, 2019.
- 1250 Remy, C., Pervin, N., Toriumi, F., and Takeda, H.: Information diffusion on twitter: everyone has its chance, but all chances are not equal, in: 2013 International Conference on Signal-Image Technology & Internet-Based Systems, NW Washington, DC United States, December 2-5, 2013, pp. 483–490, IEEE, 2013.
- Resch, B., Usländer, F., and Havas, C.: Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment, *Cartogr Geogr Inf Sc*, 45, 362–376, <https://doi.org/10.1080/15230406.2017.1356242>, 2018.
- 1255 Reuter, C. and Kaufhold, M.-A.: Fifteen years of social media in emergencies: a retrospective review and future directions for crisis informatics, *J Conting Crisis Man*, 26, 41–57, <https://doi.org/10.1111/1468-5973.12196>, 2018.



- Reuter, C. and Schröter, J.: Microblogging during the European floods 2013: What Twitter may contribute in German emergencies, *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, 7, 22–41, <https://doi.org/10.4018/IJISCRAM.2015010102>, 2015.
- 1260 Reuter, C. and Spielhofer, T.: Towards social resilience: A quantitative and qualitative survey on citizens' perception of social media in emergencies in Europe, *Technol Forecast Soc*, 121, 168–180, <https://doi.org/10.1016/j.techfore.2016.07.038>, 2017.
- Reynard, D. and Shirgaokar, M.: Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster?, *Transport Res D - TR E*, 77, 449–463, <https://doi.org/10.1016/j.trd.2019.03.002>, 2019.
- 1265 Rosser, J. F., Leibovici, D. G., and Jackson, M. J.: Rapid flood inundation mapping using social media, remote sensing and topographic data, *Natural Hazards*, 87, 103–120, <https://doi.org/10.1007/s11069-017-2755-0>, 2017.
- Rossi, C., Acerbo, F. S., Ylinen, K., Juga, I., Nurmi, P., Bosca, A., Tarasconi, F., Cristoforetti, M., and Alikadic, A.: Early detection and information extraction for weather-induced floods using social media streams, *Int J Disast Risk Re.*, 30, 145–157, <https://doi.org/10.1016/j.ijdr.2018.03.002>, 2018.
- 1270 Said, N., Ahmad, K., Riegler, M., Pogorelov, K., Hassan, L., Ahmad, N., and Conci, N.: Natural disasters detection in social media and satellite imagery: a survey, *Multimed Tools Appl*, 78, 31 267–31 302, <https://doi.org/10.1007/s11042-019-07942-1>, 2019.
- Sakaki, T., Okazaki, M., and Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors, in: *Proceedings of the 19th international conference on World wide web*, Raleigh, USA, April 26-30, 2010, pp. 851–860, 2010.
- Sakaki, T., Okazaki, M., and Matsuo, Y.: Tweet analysis for real-time event detection and earthquake reporting system development, *IEEE T Knowl Data En*, 25, 919–931, <https://doi.org/10.1109/TKDE.2012.29>, 2012.
- 1275 Saroj, A. and Pal, S.: Use of social media in crisis management: A survey, *Int J Disast Risk Re.*, 48, 101 584, <https://doi.org/10.1016/j.ijdr.2020.101584>, 2020.
- Schempp, T., Zhang, H., Schmidt, A., Hong, M., and Akerkar, R.: A framework to integrate social media and authoritative data for disaster relief detection and distribution optimization, *Int J Disast Risk Re.*, 39, 101 143, <https://doi.org/10.1016/j.ijdr.2019.101143>, 2019.
- 1280 Shi, Y., Sayama, T., Takara, K., and Ohtake, K.: Detecting flood inundation information through Twitter: The 2015 Kinu River flood disaster in Japan, *Journal of Natural Disaster Science*, 40, 1–13, <https://doi.org/10.2328/jnds.40.1>, 2019.
- Shibuya, Y. and Tanaka, H.: Using social media to detect socio-economic disaster recovery, *IEEE Intell Syst*, 34, 29–37, <https://doi.org/10.1109/MIS.2019.2918245>, 2019.
- Simon, T., Goldberg, A., and Adini, B.: Socializing in emergencies—A review of the use of social media in emergency situations, *Int J Inform Manage.*, 35, 609–619, <https://doi.org/10.1016/j.ijinfomgt.2015.07.001>, 2015.
- 1285 Singh, J. P., Dwivedi, Y. K., Rana, N. P., Kumar, A., and Kapoor, K. K.: Event classification and location prediction from tweets during disasters, *Ann Oper Res.*, 283, 737–757, <https://doi.org/10.1007/s10479-017-2522-3>, 2019.
- Smith, L., Liang, Q., James, P., and Lin, W.: Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework, *J Flood Risk Manag*, 10, 370–380, <https://doi.org/10.1111/jfr3.12154>, 2017.
- 1290 Son, J., Lee, H. K., Jin, S., and Lee, J.: Content features of tweets for effective communication during disasters: A media synchronicity theory perspective, *Int J Inform Manage.*, 45, 56–68, <https://doi.org/10.1016/j.ijinfomgt.2018.10.012>, 2019.
- Spence, P. R., Lachlan, K. A., Lin, X., and Del Greco, M.: Variability in Twitter content across the stages of a natural disaster: Implications for crisis communication, *Communication Quarterly*, 63, 171–186, <https://doi.org/10.1080/01463373.2015.1012219>, 2015.
- Spence, P. R., Lachlan, K. A., and Rainear, A. M.: Social media and crisis research: Data collection and directions, *Comput Hum Behav*, 54, 667–672, <https://doi.org/10.1016/j.chb.2015.08.045>, 2016.
- 1295



- Spinsanti, L. and Ostermann, F.: Automated geographic context analysis for volunteered information, *Appl Geogr.*, 43, 36–44, <https://doi.org/10.1016/j.apgeog.2013.05.005>, 2013.
- Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., and Demirbas, M.: Short text classification in twitter to improve information filtering, in: *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, Geneva, Switzerland, July 19–23, 2010, pp. 841–842, 2010.
- 1300 Srivastava, S., Singh, J. P., and Mangal, D.: Time and Domain Specific Twitter Data Mining for Plastic Ban based on Public Opinion, in: *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, Bengaluru, India, March 5–7, 2020, pp. 755–761, IEEE, 2020.
- St Denis, L., Hughes, A., Diaz, J., Solvik, K., Joseph, M., and Balch, J. K.: What I Need to Know is What I Don't Know!": Filtering Disaster
1305 Twitter Data for Information from Local Individuals, in: *Proceedings of 17th International Conference on Information Systems for Crisis Response and Management*, Virginia, USA, May 24–27, 2020, pp. 730–743, 2020.
- Starbird, K., Palen, L., Hughes, A. L., and Vieweg, S.: Chatter on the red: what hazards threat reveals about the social life of microblogged information, in: *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, Savannah Georgia, USA, February 6 - 10, 2010, pp. 241–250, 2010.
- 1310 Steelman, T. A., McCaffrey, S. M., Velez, A.-L. K., and Briefel, J. A.: What information do people use, trust, and find useful during a disaster? Evidence from five large wildfires, *Nat Hazards.*, 76, 615–634, <https://doi.org/10.1007/s11069-014-1512-x>, 2015.
- Steiger, E., De Albuquerque, J. P., and Zipf, A.: An advanced systematic literature review on spatiotemporal analyses of twitter data, *T GIS*, 19, 809–834, <https://doi.org/10.1111/tgis.12132>, 2015.
- Stephenson, J., Vaganay, M., Coon, D., Cameron, R., and Hewitt, N.: The role of Facebook and Twitter as organisational communication
1315 platforms in relation to flood events in Northern Ireland, *J Flood Risk Manag.*, 11, 339–350, <https://doi.org/10.1111/jfr3.12329>, 2018.
- Stieglitz, S., Mirbabaie, M., Ross, B., and Neuberger, C.: Social media analytics—Challenges in topic discovery, data collection, and data preparation, *Int J Inform Manage.*, 39, 156–168, <https://doi.org/10.1016/j.ijinfomgt.2017.12.002>, 2018.
- Suzuki, Y.: Filtering method for twitter streaming data using human-in-the-loop machine learning, *Journal of Information Processing*, 27, 404–410, <https://doi.org/10.2197/ipsjip.27.404>, 2019.
- 1320 Tandoc Jr, E. C. and Takahashi, B.: Log in if you survived: Collective coping on social media in the aftermath of Typhoon Haiyan in the Philippines, *New Media Soc.*, 19, 1778–1793, <https://doi.org/10.1177/1461444816642755>, 2017.
- Tang, J., Yang, S., and Wang, W.: Social media-based disaster research: Development, trends, and obstacles, *Int J Disast Risk Re.*, 55, 102095, <https://doi.org/10.1016/j.ijdr.2021.102095>, 2021.
- Tang, Z., Zhang, L., Xu, F., and Vo, H.: Examining the role of social media in California's drought risk management in 2014, *Nat Hazards.*,
1325 79, 171–193, <https://doi.org/10.1007/s11069-015-1835-2>, 2015.
- Terpstra, T., Stronkman, R., de Vries, A., and Paradies, G. L.: Towards a realtime Twitter analysis during crises for operational crisis management, in: *9th Proceedings of the International Conference on Information Systems for Crisis Response and Management*, Vancouver, Canada, April 22–25, 2012, 2012.
- Tim, Y., Pan, S. L., Ractham, P., and Kaewkitipong, L.: Digitally enabled disaster response: the emergence of social media as boundary
1330 objects in a flooding disaster, *Inform Syst J*, 27, 197–232, <https://doi.org/10.1111/isj.12114>, 2017.
- Tsao, S.-F., Chen, H., Tisseverasinghe, T., Yang, Y., Li, L., and Butt, Z. A.: What social media told us in the time of COVID-19: a scoping review, *The Lancet Digital Health*, 3, e175–e194, [https://doi.org/10.1016/S2589-7500\(20\)30315-0](https://doi.org/10.1016/S2589-7500(20)30315-0), 2021.



- Uchida, O., Kosugi, M., Endo, G., Funayama, T., Utsu, K., Tajima, S., Tomita, M., Kajita, Y., and Yamamoto, Y.: A real-time information sharing system to support self-, mutual-, and public-help in the aftermath of a disaster utilizing Twitter, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 99, 1551–1554, <https://doi.org/10.1587/transfun.e99.a.1551>, 2016.
- 1335 Valenzuela, S., Puente, S., and Flores, P. M.: Comparing disaster news on Twitter and television: An intermedia agenda setting perspective, *Journal of Broadcasting & Electronic Media*, 61, 615–637, <https://doi.org/10.1080/08838151.2017.1344673>, 2017.
- Veil, S. R., Buehner, T., and Palenchar, M. J.: A work-in-process literature review: Incorporating social media in risk and crisis communication, *J Conting Crisis Man*, 19, 110–122, <https://doi.org/10.1111/j.1468-5973.2011.00639.x>, 2011.
- 1340 Verma, R., Karimi, S., Lee, D., Gnawali, O., and Shakeri, A.: Newswire versus social media for disaster response and recovery, in: 2019 Resilience Week (RWS), Texas, USA, November 4-7, 2019, vol. 1, pp. 132–141, IEEE, 2019.
- Volkova, S.: Twitter data collection: Crawling users, neighbors and their communication for personal attribute prediction in social media, Center for Language and Speech Processing, Johns Hopkins University, 26, 2014.
- Vongkusolkit, J. and Huang, Q.: Situational awareness extraction: a comprehensive review of social media data classification during natural hazards, *Annals of GIS*, 27, 5–28, 2021.
- 1345 Wachowicz, M., Arteaga, M. D., Cha, S., and Bourgeois, Y.: Developing a streaming data processing workflow for querying space–time activities from geotagged tweets, *Comput Environ Urban.*, 59, 256–268, <https://doi.org/10.1016/j.compenvurbsys.2015.12.001>, 2016.
- Wang, R.-Q., Mao, H., Wang, Y., Rae, C., and Shaw, W.: Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data, *Comput Geosci.*, 111, 139–147, <https://doi.org/10.1016/j.cageo.2017.11.008>, 2018.
- 1350 Wang, Y., Hao, H., and Platt, L. S.: Examining risk and crisis communications of government agencies and stakeholders during early-stages of COVID-19 on Twitter, *Comput Hum Behav*, 114, 106 568, <https://doi.org/10.1016/j.chb.2020.106568>, 2021.
- Wang, Z. and Ye, X.: Social media analytics for natural disaster management, *Int J Geogr Inf Sci.*, 32, 49–72, <https://doi.org/10.1080/13658816.2017.1367003>, 2018.
- Wang, Z., Ye, X., and Tsou, M.-H.: Spatial, temporal, and content analysis of Twitter for wildfire hazards, *Nat Hazards.*, 83, 523–540, <https://doi.org/10.1007/s11069-016-2329-6>, 2016.
- 1355 Wang, Z., Lam, N. S., Obradovich, N., and Ye, X.: Are vulnerable communities digitally left behind in social responses to natural disasters? An evidence from Hurricane Sandy with Twitter data, *Appl Geogr.*, 108, 1–8, <https://doi.org/10.1016/j.apgeog.2019.05.001>, 2019.
- Wendt, J. D., Wells, R., Field, R. V., and Soundarajan, S.: On data collection, graph construction, and sampling in twitter, in: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), California, August 18 - 21, 2016, pp. 985–992, IEEE, 2016.
- 1360 Westerman, D., Spence, P. R., and Van Der Heide, B.: Social media as information source: Recency of updates and credibility of information, *J Comput-Mediat Comm*, 19, 171–183, <https://doi.org/10.1111/jcc4.12041>, 2014.
- Williams, B. D., Valero, J. N., and Kim, K.: Social media, trust, and disaster: Does trust in public and nonprofit organizations explain social media use during a disaster?, *Quality & Quantity*, 52, 537–550, <https://doi.org/10.1007/s11135-017-0594-4>, 2018.
- 1365 Win, S. S. M. and Aung, T. N.: Target oriented tweets monitoring system during natural disasters, in: 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, May 24–26, 2017, pp. 143–148, IEEE, 2017.
- Wu, D. and Cui, Y.: Disaster early warning and damage assessment analysis using social media data and geo-location information, *Decis Support Syst*, 111, 48–59, <https://doi.org/10.1016/j.dss.2018.04.005>, 2018.
- Wu, Q., Ma, S., and Liu, Y.: Sub-event discovery and retrieval during natural hazards on social media data, *World Wide Web*, 19, 277–297, <https://doi.org/10.1007/s11280-015-0359-8>, 2016.
- 1370



- Xiao, Y., Huang, Q., and Wu, K.: Understanding social media data for disaster management, *Nat Hazards*, 79, 1663–1679, <https://doi.org/10.1007/s11069-015-1918-0>, 2015.
- Xu, Z., Zhang, H., Sugumaran, V., Choo, K.-K. R., Mei, L., and Zhu, Y.: Participatory sensing-based semantic and spatial analysis of urban emergency events using mobile social media, *EURASIP J Wirel Comm*, 2016, 1–9, <https://doi.org/10.1186/s13638-016-0553-0>, 2016.
- 1375 Yabe, T. and Ukkusuri, S. V.: Integrating information from heterogeneous networks on social media to predict post-disaster returning behavior, *J Comput Sci-Neth*, 32, 12–20, <https://doi.org/10.1016/j.jocs.2019.02.002>, 2019.
- Yan, Y., Eckle, M., Kuo, C.-L., Herfort, B., Fan, H., and Zipf, A.: Monitoring and assessing post-disaster tourism recovery using geotagged social media data, *ISPRS Int Geo-Inf.*, 6, 144, <https://doi.org/10.3390/ijgi6050144>, 2017.
- Yang, J., Yu, M., Qin, H., Lu, M., and Yang, C.: A twitter data credibility framework—Hurricane Harvey as a use case, *ISPRS Int Geo-Inf.*, 8, 111, <https://doi.org/10.3390/ijgi8030111>, 2019a.
- 1380 Yang, T., Xie, J., Li, G., Mou, N., Li, Z., Tian, C., and Zhao, J.: Social media big data mining and spatio-temporal analysis on public emotions for disaster mitigation, *ISPRS Int Geo-Inf.*, 8, 29, <https://doi.org/10.3390/ijgi8010029>, 2019b.
- Yeo, J., Knox, C. C., and Hu, Q.: Disaster recovery communication in the digital era: Social media and the 2016 southern Louisiana flood, *Risk analysis*, 42, 1670–1685, <https://doi.org/10.1111/risa.13652>, 2022.
- 1385 Yigitcanlar, T., Regona, M., Kankanamge, N., Mehmood, R., D’Costa, J., Lindsay, S., Nelson, S., and Brhane, A.: Detecting natural hazard-related disaster impacts with social media analytics: the case of Australian states and territories, *Sustainability*, 14, 810, <https://doi.org/10.3390/su14020810>, 2022.
- Yin, J., Lampert, A., Cameron, M., Robinson, B., and Power, R.: Using social media to enhance emergency situation awareness, *IEEE Intell Syst*, 27, 52–59, <https://doi.org/10.1109/MIS.2012.6>, 2012.
- 1390 Yoo, E., Rand, W., Eftekhari, M., and Rabinovich, E.: Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises, *J Oper Manag*, 45, 123–133, <https://doi.org/10.1016/j.jom.2016.05.007>, 2016.
- Yu, M., Yang, C., and Li, Y.: Big data in natural disaster management: a review, *Geosci J*, 8, 165, <https://doi.org/10.3390/geosciences8050165>, 2018.
- Yuan, F. and Liu, R.: Feasibility study of using crowdsourcing to identify critical affected areas for rapid damage assessment: Hurricane Matthew case study, *Int J Disast Risk Re.*, 28, 758–767, <https://doi.org/10.1016/j.ijdr.2018.02.003>, 2018.
- 1395 Yuan, F., Li, M., Liu, R., Zhai, W., and Qi, B.: Social media for enhanced understanding of disaster resilience during Hurricane Florence, *Int J Inform Manage.*, 57, 102289, <https://doi.org/10.1016/j.ijinfomgt.2020.102289>, 2021.
- Zhang, A. J., Albrecht, L., and Scott, S. D.: Using twitter for data collection with health-care consumers: a scoping review, *Int J Qual Meth*, 17, 1609406917750782, <https://doi.org/10.1177/1609406917750782>, 2018.
- 1400 Zhang, C., Fan, C., Yao, W., Hu, X., and Mostafavi, A.: Social media for intelligent public information and warning in disasters: An interdisciplinary review, *Int J Inform Manage.*, 49, 190–207, <https://doi.org/10.1016/j.ijinfomgt.2019.04.004>, 2019a.
- Zhang, H., Fang, Q., Qian, S., and Xu, C.: Multi-modal knowledge-aware event memory network for social media rumor detection, in: *Proceedings of the 27th ACM international conference on multimedia*, France, October 21 - 25, 2019, pp. 1942–1951, 2019b.
- Zhou, L., Wu, X., Xu, Z., and Fujita, H.: Emergency decision making for natural disasters: An overview, *Int J Disast Risk Re.*, 27, 567–576, <https://doi.org/10.1016/j.ijdr.2017.09.037>, 2018.
- 1405 Zhou, Y., Yang, L., Van de Walle, B., and Han, C.: Classification of microblogs for support emergency responses: Case study Yushu earthquake in China, in: *2013 46th Hawaii International Conference on System Sciences*, Maui, USA, January 7-10, 2013, pp. 1553–1562, IEEE, 2013.