Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2020, 9, 532; doi:10.3390/ijgi9090532

Extracting Reliable Twitter Data for Flood Risk Communication using Manual Assessment and Google Vision API from Text and Images

Xiaohui Liu[1*], Bandana Kar[2], Francisco Alejandro Montiel Ishino[1], Chaoyang Zhang[3] and Faustine Williams[1]

[1*]National Institute on Minority and Health Disparities, National Institutes of Health, Bethesda, MD, U.S.A., 20814
[2]National Security Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, TN, U.S.A., 37831
[3]School of Computing, University of Southern Mississippi, 118 College Drive, Hattiesburg, MS, U.S.A., 39406
***Correspondence:** Xiaohui Liu; Xiaohui.Liu@nih.gov

Abstract

While Twitter has been touted to provide up-to-date information about hazard events, the reliability of tweets is still a concern. Our previous publication extracted relevant tweets containing information about the 2013 Colorado flood event and its impacts. Using the relevant tweets, this research further examined the reliability (accuracy and trueness) of the tweets by examining the text and image content and comparing them to other publicly available data sources. Both manual identification of text information and automated (Google Cloud Vision API) extraction of images were implemented to balance accurate information verification and efficient processing time. The results showed that both the text and images contained useful information about damaged/flooded roads/street networks. This information will help emergency response coordination efforts and informed allocation of resources when enough tweets contain geocoordinates or locations/venue names. This research will help identify reliable crowdsourced risk information to enable near-real time emergency response through better use of crowdsourced risk communication platforms.

Keywords: Twitter; data reliability; risk communication; data mining; Google Cloud Vision API

1. Introduction

Increased frequency and severity of climate-related hazards (e.g., floods, wildfires, hurricanes, and heat weaves) and anthropogenic hazards (e.g., mass shooting, epidemics) have brought unprecedented challenges to nations and individuals worldwide[1]. Risk and crisis communication regarding disasters are paramount in helping the population prepare for and respond to extreme events by providing necessary information to plan and mitigate potential damages to life and property[2,3]. The proliferation of information technology and Web 2.0 have transformed the way individuals and organizations communicate and interact with others across the globe. For instance, according to the Pew Research Center, around 30% of Americans often depend on social media and social networking sites (e.g., Facebook, Twitter etc. for their news or information about specific events[4]). Consequently, traditional mainstream media have adopted new strategies to show their presence, distribute their content as well as engage with their consumers on social media[5]. Similarly, varying online content created by members of the public are being consumed and shared on various social media platforms (e.g., Twitter), thereby, enriching and challenging traditional communication, especially, during emergency management phases[6]. From the socio-psychological perspective, reasons that generally drive people to share information on social media are self-efficacy, self-fulfilment, altruism, social engagement, reciprocity, and reputation[7,8]. Driven by these reasons, numerous scenarios have used social media platforms to warn the public about disasters, report damages, engage with stakeholders, and help organize relief efforts[9-13].

Citizen science-based platforms (e.g., iCoast, Tweet Earthquake Dispatch, CitizenScience.gov) allows citizens to collaborate with scientists in collecting and analyzing data, reporting observations

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2020, 9, 532; doi:10.3390/ijgi9090532

and disseminating results about scientific problems [14]. Crowdsourcing platforms, such as Twitter and Facebook, are social media and social networking sites, that allow non-experts to generate new knowledge and data sets[15,16]. Although, both citizen science and crowdsourcing engage socio-culturally diverse and geographically dispersed citizens for data and knowledge creation/collection, each has subtle differences[17,18]. While crowdsourcing remains an ill-defined approach that uses large networks of people, citizen science solely uses scientists, volunteers, and lay people with interests and knowledge about a specific topic [19]. Because tweets are generated via crowdsourcing and tend to contain rumors and hoaxes, we assumed the tweets to be inaccurate and implemented a hierarchical approach to verify the reliability and relevant of the tweets using scientifically derived and confirmed data.

Despite the importance of social media in risk communication, there are challenges that need to be addressed. First, information overload due to massive amounts of user-generated content can overwhelm users in discerning relevant information[20]. Second, crowdsourced social media data often lack metadata that provide information about the creator, time, date, device used to generate data, purpose, and standard, making it less credible[21-23]. Third, robot-controlled social media accounts, commercial spam, and collective attention spam/misinformation advent with social media prevalence [24,25], could also impede the quality of crowdsourced data. Finally, heuristic plays a significant role in deciding what or whether to share information on social media. This has become influential during complicated and unanticipated crisis situations, thereby contributing to the possibility of introducing errors and biased judgements to shared risk information [26]. These challenges are more pronounced in case of crowdsourced sites.

Even when the above issues are controlled, information relevance determines the usability of social media crisis information. Thus, evaluating relevance of social media content is critical[10], and hence, it is paramount to assess the quality and trustworthiness of data to ensure the information shared is accurate and true for decision making and public consumption during crisis. The goal of this research is to extract risk information from tweets during the 2013 Colorado flood and assess the reliability (accuracy and trueness) of this information. This was done by examining the text and image content and comparing the content to publicly available information from federal, state and local governments and emergency management agencies.

2. Literature Review

Risk communication, a principal element of emergency management, is defined as "the process of exchanging information among interested parties about the nature, magnitude, significance, or control of a risk"[27]. Risk communication is paramount to governments, organizations, businesses, and individuals because it provides information about potential disasters/crisis, possible impacts and/or damages, and countermeasures. Social media-based approaches are characterized by collaborative, participatory, and multidirectional communications that allow both impacted and interested populations to share unlimited information about a hazard, irrespective of its geographic location and time [28]. For instance, social networking sites (e.g. Facebook) and short-blog services (e.g. Twitter) were extensively used during 2017 Hurricane Harvey [29], 2017 Hurricane Maria [30], and 2018 California wildfire [12,13] and even during COVID-19 pandemic [31,32].

Data reliability can be defined as "the accuracy and completeness of data, given the uses they are intended for" [33]. Existing research assessing reliability of crowdsourced data tends to focus on evaluating quality of content (e.g., presence of metadata [21], detection of rumors [24]), and developing machine learning algorithms or models to asses data reliability [34-36]. Citizen scientists, subject matter experts are also used in reliability validation to differentiate and justify perceived "true incidents" [37,38]. Based on this need, Amazon Mechanical Turk has increasingly been adopted by researchers to verify the effectiveness and reliability of crowdsourced data in addition to other manual identification approaches [39-43].

Despite the abundance of existing evaluation methods, some algorithm-based studies rarely incorporate potentially relevant external data sources to the research context, such as meteorological and geospatial data in flood studies [44] and digital elevation models (DEM) in earthquake or landslide

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2020, 9, 532; doi:10.3390/ijgi9090532

studies [45]. As a result, these studies may fail to capture all the necessary information for reliability validation. Therefore, this research designed a workflow to work closely with reference documents to extract reliable risk information.

Reliability in this research refers to "accuracy of information and the extent to which the data reflects actuality". Using this definition, a workflow was developed to assess reliability of extracted risk information from relevant tweets that were obtained for the 2013 Colorado flood event. Using the workflow, we examined the tweet text and images leveraging human intelligence and Google Cloud Vision API (GCV API). The relevant tweets were extracted via several data mining techniques and can be found from previous publication[10]. GCV API allowed automatic identification of image content, labeling of images, matching other online information by leveraging pre-trained machine learning models, and has been widely used by other research [46,47].

3. Materials and methods

3.1. Study site

The 2013 Colorado flood severely affected Front Range, EL Paso County, Boulder County, and part of the Denver metropolitan area. The severe flash flooding caused by days of heavy precipitation that spanned from September 9th to 18th brought considerable damages to the region. Boulder County, the study site, received 9.4 inches of precipitation on September 12th alone, which was equivalent to the county's average annual precipitation[48]. Other counties had relatively less but increasing precipitation from September 9th until September 15th.

3.2. Datasets and processing

The datasets used in this study include historical tweets, geospatial data sets corresponding to the flood event and the study site (e.g., Boulder flood extent map, Boulder street map), and reference documents including news articles and agency reports from the National Weather Services, state and local government agencies. A discussion of the data processing steps and analytical approaches is presented below.

3.2.1. Tweets of 2013 Colorado floods

Historical tweets were purchased from Twitter Inc. using two types of keywords: 1) location names (Colorado, Boulder, Front Range, El Paso County and Boulder County, Denver metro), and 2) hazard event/impacts (flash flooding, flooding, rain 2013, emergency, impact, damaged bridges and roads, damaged houses, financial losses, evacuate, and evacuation). Any tweet that contained either the location name or hazard event/impact was included in the analysis. The tweets covered a 10-day duration from September 9th to 18th and captured all flooding event. From the 1 million tweets, 5202 (0.44 %) tweets that were in English language and geo-tagged to Colorado were extracted. Our previous study mined the tweets and extracted 720 (14% of the geo-tagged tweets and 0.31% of raw tweets) relevant tweets using six different computational and spatiotemporal analytical approaches[10,49]. The relevant tweets contained considerable flooding related information with a threshold relevance score of 1.3 [10].

3.2.2. GIS data

To understand the spatial distribution of tweets with respect to the flood impacted area, flood extent dataset was obtained from City of Boulder [50]. This dataset was generated using field surveys, Digital Globe Worldview satellite imagery, and public input from Boulder crowdsourced online apps. Street network data from City of Boulder was used to evaluate reliability of tweets about damages to flooded roads and streets [51].

3.2.3. NOAA Warning/alert messages.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2020 doi:10.20944/preprints202008.0487.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2020, 9, 532; doi:10.3390/ijgi9090532

Warning/alert messages sent by the National Weather Services during the 2013 Colorado flooding event were obtained from the NOAA Weather Forecast Office at Boulder [52]. The messages contained meteorological forecasts, observations, public watches, warnings, advisories, and areas that may be impacted during the flooding event. These alert/warning messages were used as official reference information in evaluating reliability of tweets.

3.2.4. Reference documents

Damage assessment reports from federal, state and local governments as well as from emergency management agencies were obtained about the Colorado flooding event from their respective websites. The documents include "situational awareness report" [53], rainfall assessment report [54], damage assessment report [55]. These reports provided situational awareness about cause of flooding, flooding extent, severity, as well as damages to properties and infrastructures in affected regions. Additionally, newspaper articles that validated incidents and/or facts (i.e., damage to specific roads) were also used as reference documents [56,57].

3.3. Analytics and techniques

This section presents the steps used to assess the reliability of relevant tweets. In the context of risk communication, relevant information may not be reliable, e.g., mention of the time and/or location of the event cannot be deemed as reliable unless the relevant information is verified to be accurate and true. Based on this rationale, this research sequentially extracted relevant tweets first and then evaluated their reliability (Figure 1). Specifically, the bag-of-words model was applied to geo-tagged tweets to extract assumed relevant tweets. The bag-of-words extraction used topic-specific search terms, top frequency words and high-frequency hashtags, to measure the relevance of a document (i.e., tweets) to the search terms and extract the assumed relevant documents. The relevance of these tweets was determined first following which their reliability was evaluated.

The text analysis involves a few consecutive steps. The first and foremost step is to search for evidence information from reference documents, especially weather warning/alert messages from the National Weather Services and the state and local emergency management agencies. Events, names of damaged roads, streets, and the posted time of each tweet were manually identified[58] from relevant tweets and then used as keywords to search for related information in reference documents. If no such information can be found, the topic, posted time, location of tweets can be documented for further use. The next step is to holistically assess if the documented unverified tweets have any association with other tweets based on topic, posted time, or location. Finally, news information may also be a complementary reference source if available. If evidence can be found from reference documents or enough tweets from multiple Twitter users presented facts that fit the hazard context in the relevant tweets, the studied tweets can be considered reliable.

In the image analysis process, 308 images were downloaded from 720 relevant tweets. The images were considered reliable if they met either of the following two conditions: (1) gain evidence from credible sources, or (2) mutually prove each other. Both manual and automatic evaluation approaches were implemented to analyze the 308 images. In the manual approach, images were manually examined for damages to roads/streets, properties as well as their corresponding tweet text content. Next, the image content, geographical locations, and text content were compared to reference documents. In the automatic/Artificial Intelligence (AI) approach, Google Cloud Vision API 308 images were uploaded to Google Cloud's Vision API (application programming interface) [59], which were classified and assigned categorical labels using Google's pre-trained machine learning models. This approach aims to leverage existing AI (artificial intelligence) tool to improve the efficiency of extracting flood related features to facilitate the tweets reliability evaluation process.

doi:10.20944/preprints202008.0487.v1

Peer-reviewed version available at *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 532; doi:10.3390/ijgi9090532

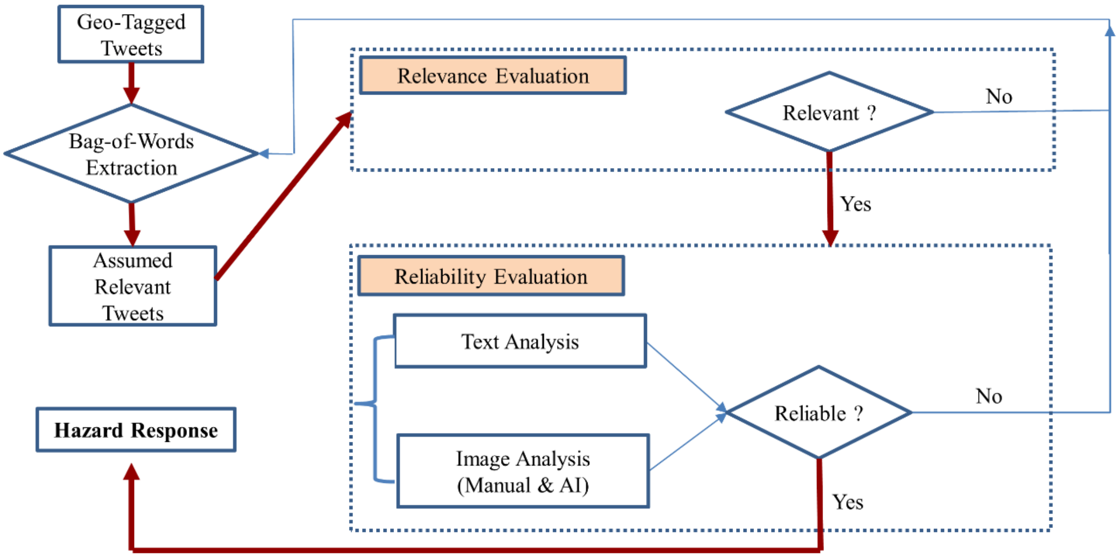

Figure 1. Reliability evaluation workflow.

4. Result and discussion

4.1. Evaluation of text content

Three authors of this paper worked on manual evaluation of tweet text, and each tweet was evaluated by at least two authors to minimize human error or bias. As a result, 584 out of 720 relevant tweets were verified to have reliable information. Examples of unverified tweets include tweets solely about emotions or contained information that cannot be verified based on our evaluation criteria. Table 1 shows how the names of damaged roads/streets, tweet post time, and detailed damage/impact were extracted manually. The location of the tweets shown in Table 1 were marked in Figure 2 using their ID number. A detailed description of assessing reliability of each tweet is presented below.

Table 1. Example of Identified Roads/Streets.

ID	Roads/streets	Posted Time	Associated Risk Information
1	West of Broadway	09/12 03:02	Boulder Creek is about to spill its bank.
2	Broadway & Arapahoe Avenue	09/12 05:30	Water at Boulder Creek has come up 2.5 feet in 10 mins.
3	8th Street & Marine Street	09/12 05:52	Gregory canyon drainage overtopping the underground culvert, flowing onto 8th St. near Marine.
4	28th Street & Colorado Avenue	09/12 06:09	Knee deep water at 28th St & Colorado Ave.
5	15th Street	09/12 08:39	River taking back Boulder neighborhood street.
6	Highway 36 underpass	09/12 22:23	It's raining! It's pouring!
7	8th Street between University of Colorado and Marine	09/13 03:22	…basically, a raging torrent.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2020 doi:10.20944/preprints202008.0487.v1

Peer-reviewed version available at *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 532; doi:10.3390/ijgi9090532

8	30th Street & Foothills	09/13 00:49	Colorado Avenue is closed between 30th and Foothill.
9	30th Street	09/13 01:08	Water is coming up through drains on 30th and Colorado Ave…this could get ugly.
10	Highway 36	09/13 01:30	Barely make it out of Boulder. Couldn't get to hwy 36.
11	Highway 36	09/13 02:33	Highway 36 is flooded, not way out.
12	Highway 36 & Foothills	09/13 05:32	Over 3 feet of water flooding.

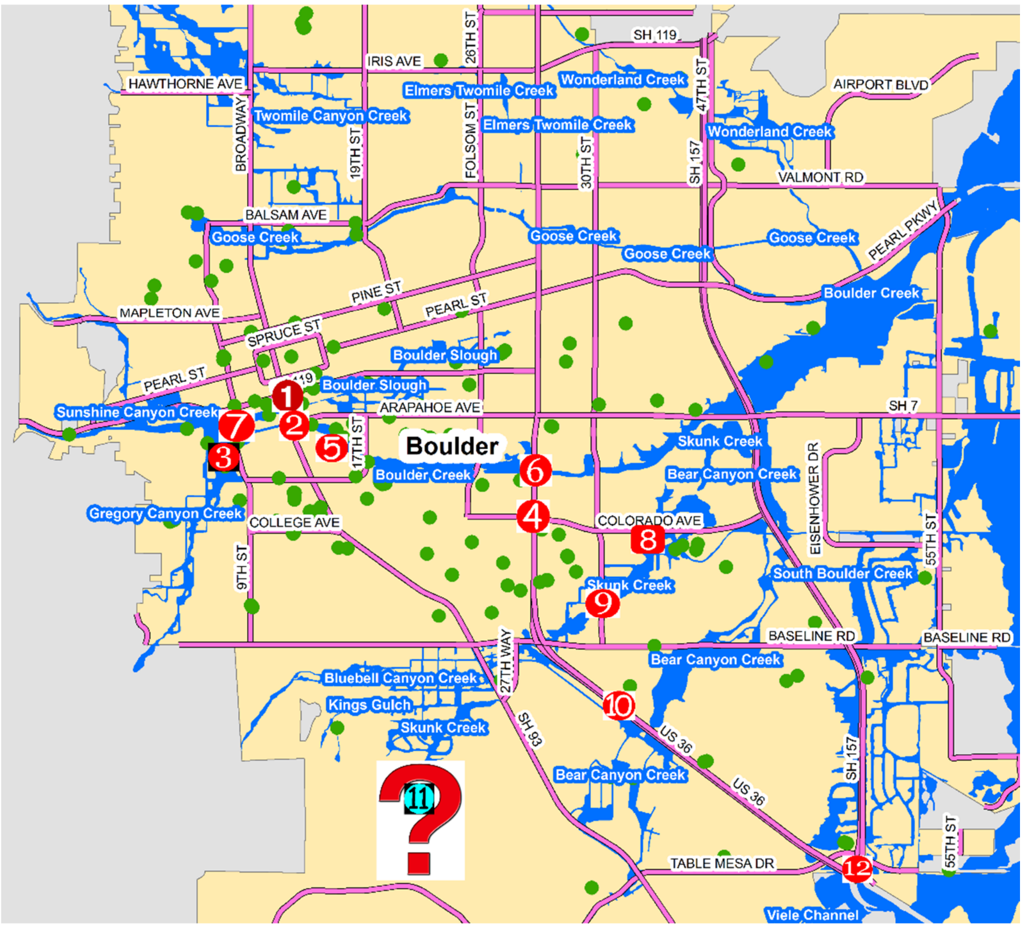

Figure 2. Example of identified roads/streets.

Using the key phrase "west of Broadway", a related NOAA warning/alert message was found from tweet #1 in Table 1: "Hourly rainfall intensity at the Sugarloaf RAWS station 6 mi. west of Boulder compared with gage height on Boulder Creek at Boulder (**west of Broadway**). The first flood peak closely followed the **heavy rainfall before midnight on 9/11-12**, when 3.5" fell in 6 hours. (Data: rainfall: RAWS via WRCC; and streamflow: Colorado DWR; plotted by Jeff Lukas, WWA)".

The above message mentioned the gauge height on Boulder Creek at west of Broadway following the flood peak that resulted from heavy rainfall before midnight on September 11th. This corresponds to tweet #1 and explains why "Boulder Creek is about to spill its bank at west of Broadway" at 3:02 am on September 12th. Therefore, tweet #1 in Table 1 was considered reliable in terms of its location, time, and content.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2020 doi:10.20944/preprints202008.0487.v1

Peer-reviewed version available at *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 532; doi:10.3390/ijgi9090532

When searching for "*Broadway*" and "*Arapahoe Avenue*", no direct evidence was found in tweet #2, which may be because *Arapahoe Avenue* is a county road and is generally too specific to be mentioned in official warning or damage assessment reports. However, as shown in symbol #2 in Figure 2, the Boulder Creek flooded the crossing of Broadway and *Arapahoe Avenue*, allowing the observer to detect increased water level of 2.5 feet within 10 minutes. Additionally, the tweet posting time (5:30 am) was within the period when Boulder Creek was officially identified to have experienced a rapid accumulation of precipitation (see section 3.1).

The crossing of *8th Street* and *Marine Street* (symbol # 3 in Figure 2) was adjacent to and flooded by Gregory Canyon Creek, which corresponded to tweet #3 (see Table 1) indicating that the drainage at Gregory Canyon overflooded 8th street. Based on the time, tweet #2 identified a rapid increase of water level on Boulder Creek at 5:30am, and 20 minutes later, this tweet reported inundation of roads due to flooding of Gregory Canyon Creek that is close to Boulder Creek. This confirmed that risk information in tweet #3 is reliable based on content, time, and geographical locations.

The intersection of *28th Street* and Colorado Avenue (symbol #4 in Figure 2) is between Boulder Creek and Skunk Creek, and tweet #4 was posted at the peak of the flooding when water overflowed from the creeks. The multi-day continuous rainfall flooded most tributaries and thereafter inundated most roads in Boulder City. An estimation of road damage was found in an official damage assessment report by [55]: "*Authorities estimate the flooding damaged or destroyed almost 485 miles of roads and 50 bridges in the impacted counties"*. This tweet reported flooded roads with "knee deep water" and was posted right after continuous heavy rainfall. Therefore, it can be treated as a reliable tweet.

Tweet #5 in Table 1 was posted in a similar context as tweet #4, and the user seemed to have witnessed the flooded neighborhood streets. Since this tweet was reliable, *15th street* (where the tweet was posted) could be marked as inundated so that others can avoid this road.

State *Highway 36* was mentioned several times in tweets #6, #10, #11, and #12 (Table 1). The earliest mention was on September 12th when excessive rainfall continued to intensify the flooding situation. Those tweets also disclosed other details about *Highway 36*, such as "raining and pouring", "flooded by over 3 feet of water", and "its subsequent closure". Evidence of this was also found in an official damage assessment report [55]: "*Based on FEMA information, the flooding destroyed more than 350 homes with over 19,000 homes and commercial buildings damaged, many of which were impossible to reach except on foot. Flooding resulted in a total of 485 miles of damaged roadway, destroyed 30 state highway bridges, and severely damaged another 20 bridges. During the height of the flooding, authorities were forced to close 36 state highways. Some highways could not be repaired for weeks or even months"*. These assessments confirmed the reliability of the tweets.

Tweets #6, #10, and #12 were also geo-located along Highway 36, but tweet #11 was posted beyond the city limits of Boulder. Because this tweet was posted from a place that is farther from the impacted location, it was hard to prove its reliability without referring to other tweets that also mentioned Highway 36. However, because the content mentioned in this tweet was also mentioned in other tweets, this tweet was considered reliable. Consequently, keywords that were verified to be related to important incidents/places, such as Highway 36, could be used to extract tweets that were beyond the spatial limit of the study area or even do not possess any geo-location information. This approach would yield a larger volume of relevant tweets.

Tweet #7 posted at 3:22 am on 9/13 mentioned that a portion of *8th Street* between University of Colorado Boulder and Marine Street (symbol # 7 in Figure 2) experienced severe rainfall. Given the site was located near Boulder Creek, Sunshine Canyon Creek, and Gregory Canyon Creek junction, 8th street was highly likely to have been flooded at that time. A piece of news by Huffing Post reported that, "*around 80 buildings on campus were damaged in some form, CU Boulder police tweeted, and raw sewage was flowing from a pipe in one area.*" [56] confirming this tweet. A campus damage assessment report [57] also mentioned that "*80 of 300 structures on the Boulder campus sustained some damage. The damage is described as "widespread" but not severe*." These two news articles confirmed the reliability of the tweet.

Tweets #8 and #9 were geo-located along the flooded Skunk Creek (symbols #8 and #9 in Figure 2). While 30th Street was flooded, the adjacent *Colorado Avenue* was already closed. Both streets are

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2020 doi:10.20944/preprints202008.0487.v1

Peer-reviewed version available at *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 532; doi:10.3390/ijgi9090532

in the Foothills area, which was reported to have been seriously impacted by flood in a damage assessment report summary: "*Foothills around Boulder also saw severe flooding and debris flows*" [54].

4.2. Evaluation of image content

Interactive delivery of information (stories) through images is more engaging because it is an effective way to visualize information that enables brain to process and organize the information. Through images people can develop a deep understanding about the severity and significance of issues associated with a disaster [60]. However, previous studies have shown that around 4% of tweets are spams [61], and fake images tend to be propagated via web especially during crises [62]. Despite abundant research on filtering out spam or phishing tweets [63], studies focusing on diffusion of fake images are sparse [62]. Given this limitation, 308 images were downloaded from relevant tweets and two strategies were implemented to evaluate the reliability of those images. Results showed that manual approach identified 60 (19%) reliable images compared to AI approach which detected only 34 (11%). The following section presents the results of both approaches.

4.2.1. Manual approach

This section illustrates the method of organizing images based on locations, time, and the photographer, and other categories into which the images can be grouped into so that images within a group could be compared with each other to elucidate the themes or topics of those categories. Figures 3, 4, 5 and 6 show 24 most representative images out of 308 that were identified and grouped into different themes. All images in Figure 3 depict the flood conditions of Boulder Creek at different time, across the creek, and at different angles. Figure 4 includes images about submerged ground on different streets & intersections, some streets were mentioned in 4.1.2 and 4.1.4.

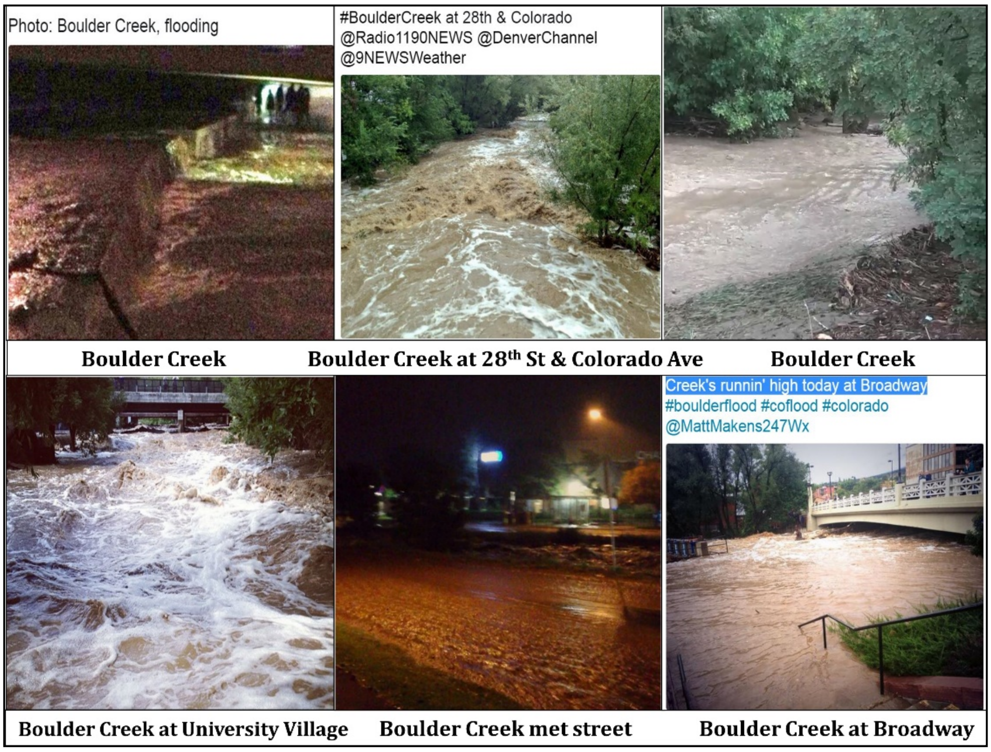

Figure 3. Images of Boulder Creek.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 August 2020 doi:10.20944/preprints202008.0487.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2020, 9, 532; doi:10.3390/ijgi9090532

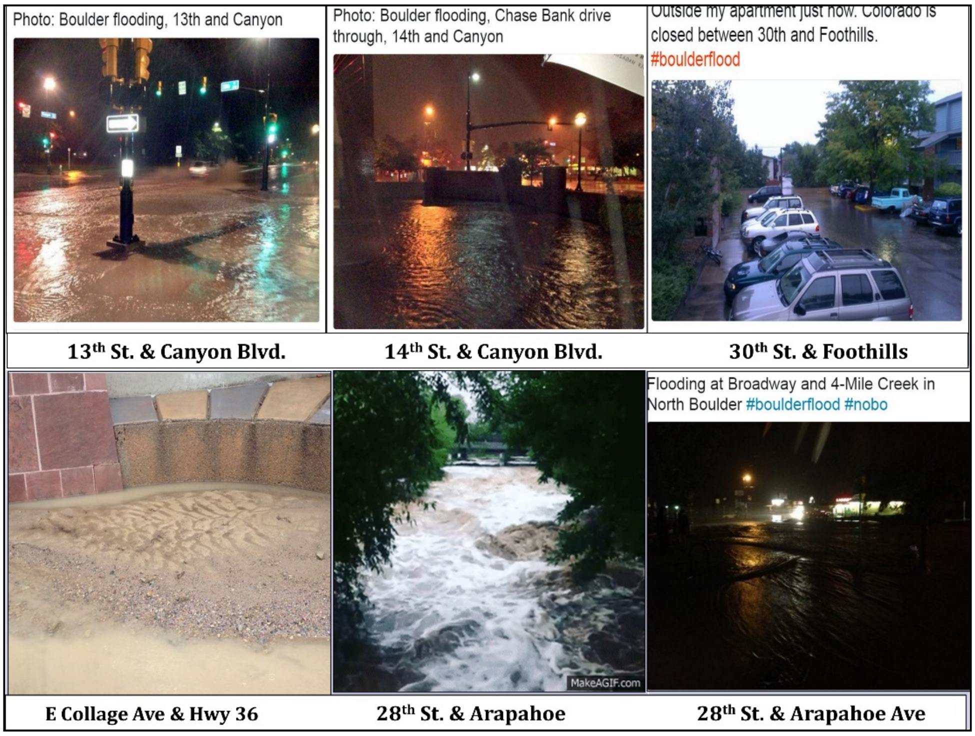

Figure 4. Images of flooded streets.

The images shown in Figure 5 were taken at the same location by different people, at different time, and from different angles. The flood water falling from the bridge created unusual waterfall and attracted people to take pictures to report the severity and rarity of the flood. The bottom three images in Figure 5 recorded the increased water level at Boulder Creek under Broadway Bridge, which clearly displays the temporal change in flood severity. This finding is critical for crowdsourcing-based risk communication because massive images could mutually verify each other despite lack of external information.

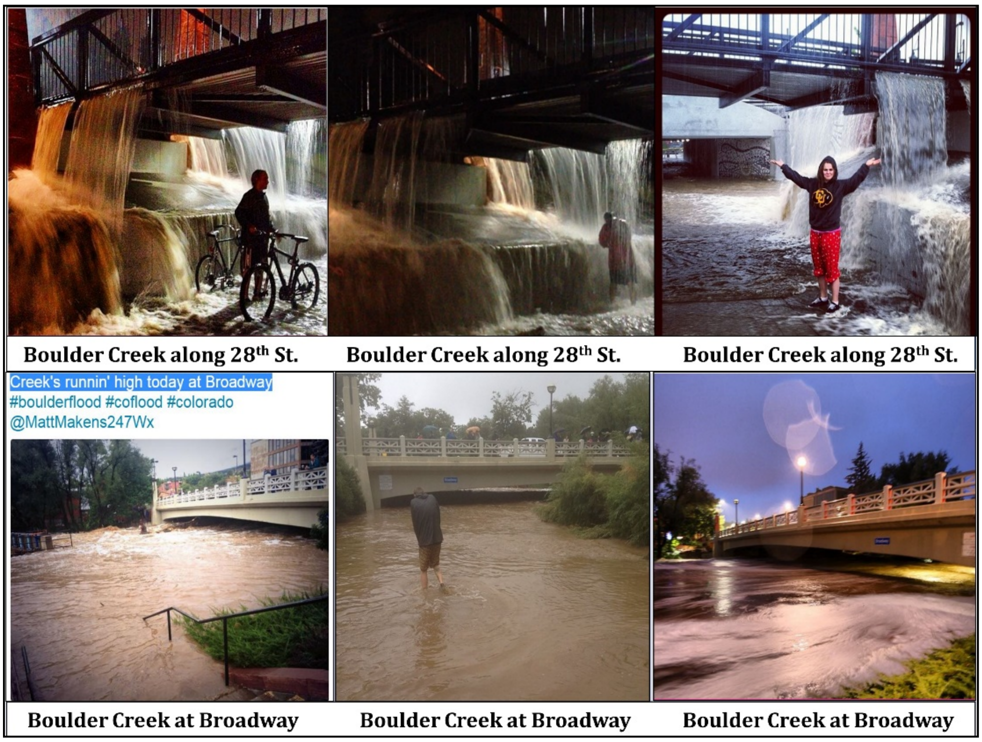

Figure 5. Images mutually prove each other.

Images in Figure 6 were posted by a local news reporter, who continuously reported flood situations in several locations along with pictures on Twitter. The locations that were mentioned by the reporter were: *Colorado Avenue*, the backyard of *Boulder High School*, *Folsom Field Stadium*, and 28th *Street* & *Arapahoe Ave*. The text and images posted by the reporter could be regarded as reliable.

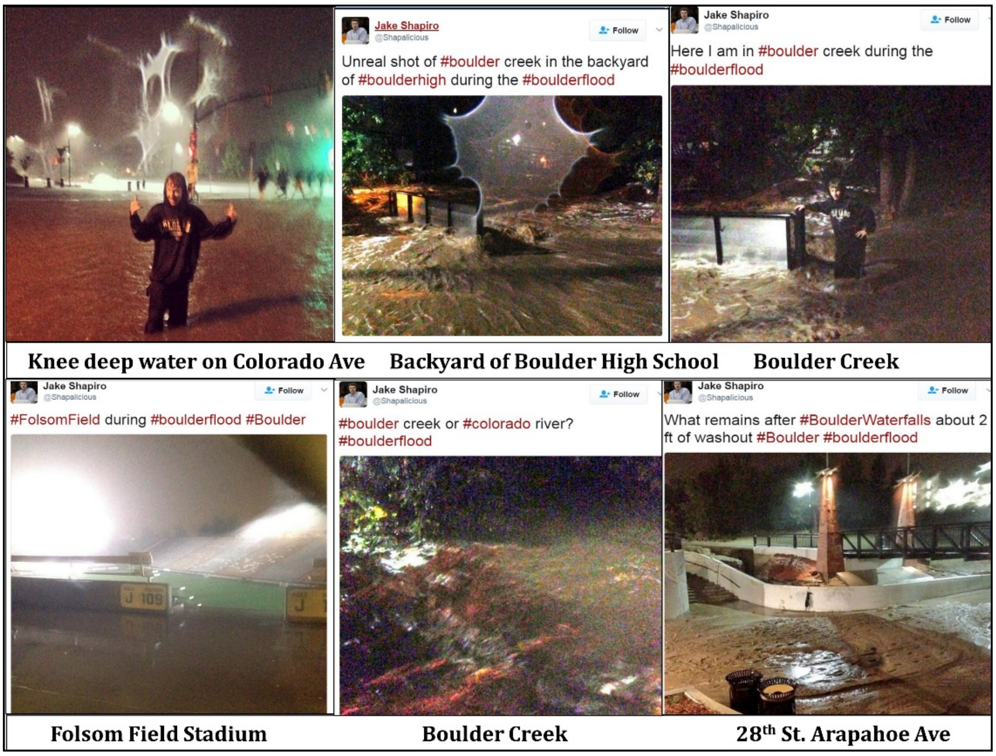

Figure 6. Images took by a local news reporter.

4.2.2. AI approach

This section illustrates an AI approach to detecting flood related features using GCV API and presents the result of the automatic detection. GCV API provided two types of automatic detection: image and web detection [59]. For each image fed to the API, its pre-trained machine learning models generated image detection and web detection results. Image detection results include image annotations by detecting the features within images, and web detection uses the image content and its metadata to crawl the web and detect relevant information from the internet. Accuracy of image detection is based on the availability of training data and the detection algorithm. Accuracy of web detection is based on image content, metadata, and the availability of related information on the web.

Among the 34 images detected by GCV API to be relevant to 2013 Colorado flood, web detection outperformed the results from image detection, one example of which could be found in Figure 7. According to figure 7, web detection accurately detected the scene as the 2013 Colorado flood while image detection only recognized the water feature in the image. On the other hand, figure 8 shows that both modes of detection failed to identify the inundated condition of the parking lot with most cars only visible from the top. GCV API can only detect the type or size of cars, e.g., "family car" or "luxury vehicle", which was largely because the training sample images used to train the underlying models did not include scenes of car parking lot flooding. This also reveals that even industry leading image detection technology has limitation in identifying flood related content from images and suggests the current limitations of AI based image processing approaches.

Image Detection	Web Detection
water	University of Colorado Boulder
water feature	Salina, Colorado
water resources	2013 Colorado Floods
	flash flood
	emergency evacuation

Figure 7. Web detection performed better.

Image Detection	Web Detection
car	car
luxury vehicle	luxury vehicle
family car	mid-size car
motor vehicle	compact car
vehicle	family car

Figure 8. Both image and web detection were inaccurate.

4.3. Extracting added tweets using verified keywords

Sections 4.1 and 4.2 identified 584 reliable tweets and 60 reliable images, which accounts for 11% and 1% of 5202 geo-tagged tweets, respectively, and 0.05% and 0.01% of all 1,195,183 purchased tweets, respectively. To make better use of this data source, we selected a group of keywords/locations (e.g., Highway 36) from the verified reliable tweets discussed in sections 4.1 and 4.2 and used the keywords to extract more tweets that do not possess any geo-location information. We believe that doing this would yield a larger volume of relevant tweets that were discarded due to lack of geoinformation. Without geolocation, it is possible that those tweets may be sent from outside the study area, but the time frame (September 9th to 18th, 2013) and keywords (a. location names: Colorado, Boulder, etc., and b. hazard event/impacts: flooding, rain etc.) used to download those tweets from Twitter database significantly decreased this possibility.

The keywords we used were from Table 1, which included: West of Broadway, Broadway, Arapahoe Ave, Marine St, 28th St, Colorado Ave, Boulder Creek, Highway 36/US-36, and Skunk Creek. Using these keywords, we found 2472 additional non-repetitive relevant and reliable tweets and 752 reliable images, which account for 0.2% and 0.06% of all 1,195,183 raw tweets, respectively. This is a big improvement than using geo-tagged tweets alone for this research workflow.

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2020, 9, 532; doi:10.3390/ijgi9090532

5. Discussion, implications for risk communication, and future research

The goal of this research was to apply an integrated workflow to extract and evaluate reliable risk information to facilitate risk communication, increase situational awareness, and prompt public response to natural hazards. Crowdsourced risk communication could provide valuable risk information if relevance and reliability evaluations are done properly to alleviate or eliminate data quality issues. This research integrates relevant references with prevalent approaches to design an "in context" research workflow that sheds light on crowdsourced risk information evaluation to make better use of this increasingly popular risk communication channel.

In this study, we implemented text and image content analysis to extract and evaluate tweet reliability because research on using both image and text analysis was relatively rare in Twitter based flood research while majority research focus only on Twitter text. Another reason for using this approach is that information extracted from text and images is oftentimes complementary; so including both would extract more information than only using only text or images.

The strengths of this research are: 1) precipitation data was used to account for the cause of flood, 2) geospatial data were used to understand the spatial extent, 3) relevant official documents were closely referenced, and 4) both manual and AI approaches were implemented in image content analysis to ensure accuracy and efficient processing time. Manual and AI approaches combined the advantages of human intelligence and computing efficiency. While leveraging human intelligence to validate textual content of tweets is not novel in Twitter text mining research, it brought a unique contribution to the flood research. Specifically, it allowed identification of different scenarios and process information beyond plain text (e.g., associate events in different images or associate events based on their proximity to events in the surrounding areas by pinpointing them on maps), which is impossible for current AI approaches to achieve. Given that the current neural networks (E.g., ResNet, UNet) used for disaster situations require human intelligence to collect and label significant amount of training images, our manual approach complemented the AI approach. The GCV API could be replaced with other AI algorithms. However, our research workflow can be repurposed to be used by researchers interested in designing automatic or semi-automatic systems to extract reliable and relevant data and information from social media streams searching for disaster response.

Despite their advantages, both the manual and AI approaches have certain limitations in terms of their usability and implementation. First, given the time-consuming and expensive nature of the proposed manual approach, its implementation may require a team of specialists to dedicate huge efforts to extract relevant and reliable risk information in an emergency setting. One promising phenomenon to counterbalance this limitation, though, is the emergence of volunteered citizen scientists who involved themselves in disaster response activities by voluntarily providing technical support or processed information to facilitate humanitarian efforts in recent disasters [64,65]. For instance, CitizenScience.gov, citizen science efforts by the United States Geological Survey (https://www.usgs.gov/topic/citizen-science) and FEMA's crowdsourcing and citizen science efforts have allowed citizens to participate during emergency management and response efforts to complement the activities underway by the decision-makers. With the involvement of these digital humanitarians, we believe that the workflow outlined in our research can partially or fully be adopted in disaster responses. Further, the AI approach was able to detect reliable information for 11% of the images, which is less than the percentage achieved in manual approach (19%) and most of the images identified by AI approach were also identified by manual approach. AI has low accuracy because it was developed for general purpose image detection and understanding, but not tailored for flood / disaster learning. If more images are used to train the AI model, it has the potential to significantly improve the accuracy. This approach requires less human labor investment; so, it is complementary to the manual approach and is advantageous when significant number of tweets are available. Finally, human errors and heuristic bias may be introduced in manual approaches, even though multiple authors cross-checked the results.

Considering the limitation of this research workflow, future research would focus on streamlining the process and automating the entire workflow of assessing relevance and reliability of Twitter data. Moreover, integration of citizen-led reliability evaluation efforts following well-informed protocols will greatly boost the usefulness of this research workflow. Space and air-borne images can also be used to assess the reliability of tweets. While researchers are working to maximize the amount of risk information from Twitter, it is essential for emergency management agencies to develop easy-to-follow standards tailored for Twitter users to encourage dissemination of relevant and reliable crisis information to facilitate their use for response activities.

Funding

This research is part of a broader project that was funded by the National Science Foundation (Grant # CMMI-1335187).

Author contribution

Xiaohui Liu: Methodology, Analysis, Writing, Original draft preparation, Revision, **Bandana Kar**: Conceptualization, Reviewing, Editing, Revision **Francisco Alejandro Montiel Ishino**: Reviewing and Editing, **Chaoyang Zhang**: Methodology, **Faustine Williams**: Reviewing and Editing

Acknowledgments

Drs. Liu, Montiel Ishino, and Williams' efforts were supported by the Division of Intramural Research, National Institute on Minority Health and Health Disparities, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily reflect the views of the National Institutes of Health.

Dr. Bandana Kar has participated in this project in her own independent capacity and not on behalf of UT-Battelle, LLC, or its affiliates or successors. The views and conclusions expressed in this article are those of the authors and do not reflect the policies or opinions of the funding agency, Oak Ridge National Laboratory, UT-Battelle, the Department of Energy, or the US Government.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Xiaohui Liu http://orcid.org/0000-0002-4161-0388
Bandana Kar http://orcid.org/0000-0002-0510-658X
Francisco Alejandro Montiel Ishino https://orcid.org/0000-0002-2837-726X
Chaoyang Zhang https://orcid.org/0000-0001-6873-1780
Faustine Williams https://orcid.org/0000-0001-5810-1291

References

1. Hannah R, Max R. Natural Disasters. https://ourworldindata.org/natural-disasters. Published 2020. Accessed Feb 2020.
2. Newell BR, Rakow T, Yechiam E, Sambur M. Rare disaster information can increase risk-taking. *Nature Climate Change.* 2016;6(2):158-161.
3. Bradley DT, McFarland M, Clarke M. The effectiveness of disaster risk communication: a systematic review of intervention studies. In: *Effective Communication During Disasters.* Apple Academic Press; 2016:81-120.
4. Elisa S, Grieco E. Americans Are Wary of the Role Social Media Sites Play in Delivering the News. In:2019.
5. Richards D. Mainstream media vs. social media: what does the future hold? In. https://tuckerhall.com/mainstream-media-vs-social-media-future-hold/2017.
6. Amber S. The use of social media in crisis communication. In: *Risk Communication and Community Resilience.* Routledge; 2019:267-282.

7. Moon G. Why People Share: The Psychology of Social Sharing. https://coschedule.com/blog/why-people-share/. Published 2014. Accessed.
8. Oh S, Syn SY. Motivations for sharing information and social support in social media: A comparative analysis of Facebook, Twitter, Delicious, YouTube, and Flickr. *Journal of the Association for Information Science and Technology.* 2015;66(10):2045-2060.
9. Starbird K. Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. Paper presented at: Eleventh International AAAI Conference on Web and Social Media2017.
10. Liu X, Kar B, Zhang C, Cochran DM. Assessing relevance of tweets for risk communication. *International Journal of Digital Earth.* 2018;12(7):781-801.
11. Boulton CA, Shotton H, Williams HT. Using social media to detect and locate wildfires. Paper presented at: Tenth International AAAI Conference on Web and Social Media2016.
12. Brengarth LB, Mujkic E. WEB 2.0: How social media applications leverage nonprofit responses during a wildfire crisis. *Computers in Human Behavior.* 2016;54:589-596.
13. Sachdeva S, McCaffrey S. Using social media to predict air pollution during California wildfires. Paper presented at: Proceedings of the 9th International Conference on Social Media and Society2018.
14. Bonney R. 'Citizen Science: A Lab Tradition' [in] Living Bird: For the Study and Conservation of Birds. *Living Bird: For the Study and Conservation of Birds.* 1996;15(4):7-15.
15. Greengard S. Following the crowd. *Communications of the ACM.* 2011;54(2):20-22.
16. Hetmank L. Components and Functions of Crowdsourcing Systems-A Systematic Literature Review. *Wirtschaftsinformatik.* 2013;4:2013.
17. Onsrud H, Camara G, Campbell J, Chakravarthy NS. Public commons of geographic data: Research and development challenges. Paper presented at: International Conference on Geographic Information Science2004.
18. Onsrud H, Campbell J. Big opportunities in access to" Small Science" data. *Data Science Journal.* 2007;6:OD58-OD66.
19. Wiggins A, Crowston K. From conservation to crowdsourcing: A typology of citizen science. Paper presented at: 2011 44th Hawaii international conference on system sciences2011.
20. Sutton JN, Palen L, Shklovski I, Fifth International IC. Backchannels on the front lines : emergency uses of social media in the 2007 Southern California wildfires. 2008, 2008; Boulder, CO.
21. Meek S, Jackson MJ, Leibovici DG. A flexible framework for assessing the quality of crowdsourced data. 2014.
22. Castillo C, Mendoza M, Poblete B. Information credibility on twitter. Paper presented at: Proceedings of the 20th international conference on World wide web2011.
23. Liu J, Li J, Li W, Wu J. Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing.* 2016;115:134-142.
24. Lee K, Caverlee J, Pu C. Social spam, campaigns, misinformation and crowdturfing. Paper presented at: Proceedings of the 23rd International Conference on World Wide Web2014.
25. Starbird K, Maddock J, Orand M, Achterman P, Mason RM. Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. *IConference 2014 Proceedings.* 2014.

26. Dale S. Heuristics and biases: The science of decision-making. *Business Information Review.* 2015;32(2):93-99.
27. Covello VT. Risk communication: An emerging area of health communication research. *Annals of the International Communication Association.* 1992;15(1):359-373.
28. Kar B. Citizen science in risk communication in the era of ICT. *Concurrency and Computation: Practice and Experience.* 2016;28(7):2005-2013.
29. Greenhalgh A. Social Media Flooded with Rescue Requests During Hurricane Harvey. 2018.
30. Kishore N, Marqués D, Mahmud A, et al. Mortality in puerto rico after hurricane maria. *New England journal of medicine.* 2018;379(2):162-170.
31. Jahanbin K, Rahmanian V. Using Twitter and web news mining to predict COVID-19 outbreak. *Asian Pacific Journal of Tropical Medicine.* 2020:13.
32. Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: the critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. *Canadian Journal of Emergency Medicine.* 2020:1-7.
33. Government Accountability Office. Applied Research and Methods: Assessing the Reliability of Computer-Processed Data (GAO-09-680G). https://www.gao.gov/assets/80/77213.pdf. Published 2009. Accessed 2020.
34. Resch B, Usländer F, Havas C. Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. *Cartography and Geographic Information Science.* 2018;45(4):362-376.
35. Viviani M, Pasi G. Credibility in social media: opinions, news, and health information—a survey. *WIREs Data Mining and Knowledge Discovery.* 2017;7(5):e1209.
36. Xiang Z, Du Q, Ma Y, Fan W. Assessing Reliability of Social Media Data: Lessons from Mining TripAdvisor Hotel Reviews. Paper presented at: Information and Communication Technologies in Tourism 2017; 2017//, 2017; Cham.
37. Alonso O, Marshall CC, Najork MA. A human-centered framework for ensuring reliability on crowdsourced labeling tasks. Paper presented at: First AAAI Conference on Human Computation and Crowdsourcing2013.
38. Oh O, Agrawal M, Rao HR. Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *Mis Quarterly.* 2013:407-426.
39. Guberman J, Schmitz C, Hemphill L. Quantifying toxicity and verbal violence on Twitter. Paper presented at: Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion2016.
40. McCormick TH, Lee H, Cesare N, Shojaie A, Spiro ES. Using Twitter for demographic and social science research: Tools for data collection and processing. *Sociological methods & research.* 2017;46(3):390-421.
41. Williams J, Dagli C. Twitter language identification of similar languages and dialects without ground truth. Paper presented at: Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)2017.
42. Liu W, Ruths D. What's in a name? using first names as features for gender inference in twitter. Paper presented at: 2013 AAAI Spring Symposium Series2013.
43. Burger JD, Henderson J, Kim G, Zarrella G. Discriminating gender on Twitter. Paper presented at: Proceedings of the conference on empirical methods in natural language processing2011.
44. Denis LAS, Palen L, Anderson KM. Mastering Social Media: An Analysis of Jefferson County's Communications during the 2013 Colorado Floods. 2014.

45. Caragea C, McNeese N, Jaiswal A, et al. Classifying text messages for the Haiti earthquake. 2011.
46. Chen S-H, Chen Y-H. A content-based image retrieval method based on the google cloud vision api and wordnet. Paper presented at: Asian conference on intelligent information and database systems2017.
47. d'Andrea C, Mintz A. Studying the Live Cross-Platform Circulation of Images With Computer Vision API: An Experiment Based on a Sports Media Event. *International Journal of Communication.* 2019;13:21.
48. Colorado Climate Center. The 2013 Colorado Floods accumulated precipitation. https://www.ncdc.noaa.gov/news/visualizing-september-2013-colorado-flood. Published 2013. Accessed.
49. Liu X. Evaluating Relevance and Reliability of Twitter Data for Risk Communication. 2017.
50. City of Boulder. The 2013 Boulder Flood Extent Map. https://bouldercolorado.gov/open-data/city-of-boulder-september-2013-flood-extents. Published 2014. Accessed2016.
51. City of Boulder. Main Roads and Streets map. https://bouldercolorado.gov/maps. Published 2014. Accessed2016.
52. NOAA NOaAA. NOAA Warning/alert. In:2013.
53. FEMA. Reducing Losses through Higher Regulatory Standards. https://www.fema.gov/media-library-data/1429759760513-f96124536d2c3ccc07b3db4a4f8c35b5/FEMA_CO_RegulatoryLAS.pdf. Published 2015. Accessed 2016.
54. Front Range Rain Event and Floods - September 2013 [press release]. 2014.
55. NOAA. The record Front Range and eastern Colorado Floods of September 11-17. In:2014.
56. Kingkade T. University Of Colorado Students Post Scenes From Flooded Boulder Campus. 2013.
57. Department of Higher Education. 2013 Colorado flood: Campus impacts. In:2013.
58. De Choudhury M, Diakopoulos N, Naaman M. Unfolding the event landscape on twitter: classification and exploration of user categories. Paper presented at: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work2012.
59. Google. Google Vision API. In:2020.
60. Vis F, Faulkner S, Parry K, Manyukhina Y, Evans L. Twitpic-ing the riots: Analysing images shared on Twitter during the 2011 UK riots. 2013.
61. Kelly R. Twitter study reveals interesting results about usage. *PearAnalytics August 12th.* 2009.
62. Gupta A, Lamba H, Kumaraguru P, Joshi A. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. Paper presented at: Proceedings of the 22nd international conference on World Wide Web2013.
63. Benevenuto F, Magno G, Rodrigues T, Almeida V. Detecting spammers on twitter. Paper presented at: Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)2010.
64. Horita FEA, Degrossi LC, de Assis LFG, Zipf A, de Albuquerque JP. The use of volunteered geographic information (VGI) and crowdsourcing in disaster management: a systematic literature review. 2013.
65. Klonner C, Marx S, Usón T, Porto de Albuquerque J, Höfle B. Volunteered Geographic Information in Natural Hazard Analysis: A Systematic Literature Review of Current Approaches with a Focus on Preparedness and Mitigation. *ISPRS International Journal of Geo-Information.* 2016;5(7):103.