ORIGINAL PAPER

A multi-level analytic framework for disaster situational awareness using Twitter data

Wei Zhai*

Abstract

During a natural disaster, mining messages from social media platforms can facilitate local agencies, rescue teams, humanitarian aid organizations, etc., to track the situational awareness of the public. However, for different stakeholders, the concerns about people's situational awareness in a natural disaster event are different. Therefore, I developed a Twitter-based analytic framework to take perception-level situational awareness, humanitarian-level situational awareness, and action-level situational awareness into consideration. Specifically, perception-level situational awareness mainly reflects people's perception of the ongoing natural disaster event (i.e., if people are discussing the disaster event). Decision-makers can rapidly have a big picture of severely impacted regions. Humanitarian-level situational awareness represents tweets that are associated with the humanitarian categories based on the definition from the United Nations Office for the Coordination of Humanitarian Affairs. The detection of humanitarian-level situational awareness can help response teams understand the specific situations and needs of local communities. In terms of the action-level situational awareness, I extracted noun-verb pairs in each tweet to explicitly represent the specific event described in a given tweet, so that the response teams can quickly act on the situation case by case. Moreover, to shed light on disaster resilience and social vulnerability, I further examined the demographic characteristics of three levels of situational awareness. I empirically demonstrated the analytic framework using geo-tagged tweets during 2018 Hurricane Michael.

Keywords: Disaster management, Situational awareness, Twitter data, Natural language processing, Machine learning

1 Introduction

Disaster managers and responders previously obtained situational awareness by conventional methods such as on-site observations, and interviews. However, it takes at least several months and dedicated investments to complete a traditional social survey across the city (Savage et al., 2013). Given that management during a disaster is time-constrained, the social survey method is expected to be improved in practice. Admittedly, knowing the real-time situation is very difficult for any person, especially during a high-stress survival situation. Emerging social media platforms could provide near-real-time information for decision-makers to track the latest situations.

Even for preliminary level investigations, social media can present rich snapshots of the general public's perception of the disaster at a macro scale, which is difficult to be accomplished by traditional surveys (Huang & Xu, 2014). Considering that social media messages are generated by people who are being exposed to the disaster, it also provides a bottom-up perspective in terms of understanding the disaster event (Ford et al., 2016). Among the existing social media apps, Twitter has been most widely deployed as an effective communication channel in the face of natural disasters, regarding its geo-enabled function and wide usage by people (Hughes & Palen, 2009; Imran et al., 2015; Kryvasheyeu et al., 2016; Yuan et al., 2020; Yuan et al., 2021).

In the literature, the sentiment analysis of Twitter data is often informative enough in providing

*Correspondence: wei.zhai@utsa.edu
The University of Texas at San Antonio, San Antonio, USA

actionable knowledge to disaster response teams (e.g., Huang & Xiao, 2015; Cervone et al., 2016; De Albuquerque et al., 2015; Verma et al., 2011). To this end, some emerging methods in terms of disaster-specific summarization and detection on Twitter have been proposed recently (Kedzie et al., 2015; Nguyen et al., 2015; Rudra et al., 2015; Vieweg et al., 2014; Zhai & Peng, 2020). Even though these approaches are demonstrated to be effective, few of them can be used by different stakeholders, particularly local rescue and response teams. The major challenge is that different stakeholders have different needs during the disaster. Hence, a research question is: how to extract different levels of situational awareness from tweets so that the information can be employed by multiple types of end-users? In specific, the state-level decision-makers are more interested in the macro-level disaster situations so that they could estimate the statewide or regional damage to have a big picture of the disaster impact quickly. By contrast, local rescue and response teams would be more concerned with the specific needs of affected people (e.g., shelter need, medical assistance, donation need) or specific infrastructure damages (e.g., road damage, power system failure).

Hence, the motivation of this study is to detect multi-level situational awareness using a Twitter-based analytic framework so that it can be beneficial to different stakeholders in practice. Specifically, I summarized three types of situational awareness during a natural disaster event, namely perception-level situational awareness, humanitarian-level situational awareness, and action-level situational awareness (Fig. 1). Perception-level situational awareness represents if the individual is concerned about the disaster so that the state-level decision-makers can quickly understand the impacts of the disaster on a regional scale. Humanitarian-level situational awareness summarizes the response of individuals associated with humanitarian topics, which would help state-level and local decision-makers understand the specific humanitarian concerns of each city. Action-level situational awareness reflects the specific needs and expressed feelings of each person, which can help response teams quickly track the situations of individuals in an explicit manner. Moreover, to shed new light on the emerging theory of resilience and social vulnerability (Zhai et al., 2021; Fu et al., 2022), it is imperative to implement the framework across space and time. To show the feasibility of the developed framework, I conducted a case study using Twitter data generated during 2018 Hurricane Michael.
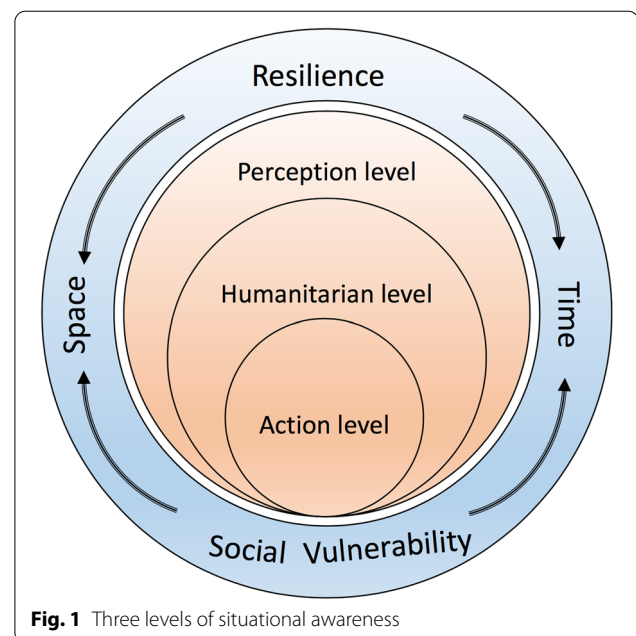
Fig. 1 Three levels of situational awareness

2 Literature review
2.1 The role of twitter in disaster management
The first typical usage of Twitter in disaster management is estimating the damage. Rapid detection of damage using social media platforms can help rescue people in danger, determine evacuation orders and prepare for the subsequent disaster hit. For example, Zhou et al. (2022) developed new algorithms for identifying rescue request tweets during 2017 Hurricane Harvey. Ganz et al. (2015) developed an intelligent system that provides situational awareness to support victim searches and rescue operations. Deng et al. (2016) introduced a new approach using crowdsourcing data and examined the significant correlation between Typhoon damages and social media activities. Guan and Chen (2014) developed a Twitter-based measure to quantify the evolution of the disaster event, and then demonstrated the close relationship between the disaster damage and Twitter activities. Kryvasheyeu et al. (2016) demonstrated a significant relationship between proximity to the hurricane's path and hurricane-related tweets, by analyzing Twitter data in 50 metropolitan areas. Most recently, Feng et al. (2022) grouped and reviewed the studies on the extraction and analysis of natural disaster-related volunteered geographic information from social media.

Second, a growing number of researchers have found that Twitter shows advantages in terms of information dissemination and crisis communication (e.g., Chatfield & Reddick, 2015; Hughes & Palen, 2009; Ghahremanlou et al., 2015). These characteristics make it possible

to support the traditional alter and warning system. For instance, Landwehr et al. (2016) developed a Tsunami Warning and Response Social Media System based on the Twitter platform. The system is developed to support disaster management throughout all phases of a Tsunami. Sakaki et al. (2010) found the trajectory of the event location in an earthquake using Twitter. Kent and Capello Jr. (2013) collected social media data to examine the underlying demographic characteristics during a fire. Provided the advantage in information dissemination, Twitter data has also been widely applied for disaster response and relief (e.g., Ashktorab et al., 2014; Kumar et al., 2011; Spinsanti & Ostermann, 2013; Zhai et al., 2020). Gao et al. (2011) listed the pros and cons of social media when Twitter was applied in disaster relief coordination. Vieweg (2012) demonstrated that Humanitarian Aid and Disaster Relief (HADR) responders can obtain tactical and actionable information from Twitter.

2.2 Situational awareness in natural disasters

Situational awareness is not a new concept, although the term itself is fairly recent. The concept of situational awareness has been widely applied in air traffic control (e.g., Endsley & Rodgers, 1994), nuclear power plant operation (e.g., Hogg et al. 1995), and anesthesiology (e.g. Gaba et al., 1995). To provide situational awareness information for emergency response, Kim et al. (2007) employed mobile devices embedded with map-based visual analytic functions.

During a disaster event, the individual's spatial information is essentially important, which could reflect the specific locations of activities and events. However, researchers often neglected the spatial characteristics of tweets (e.g., Cameron et al., 2012; Imran et al., 2015). To redefine the traditional situational awareness concept from a geographic perspective, Huang and Xiao (2015) first proposed the concept of geographic situational awareness, which is defined as knowing what is happening in space. Wang and Ye (2019) concluded a theoretical framework to integrate space, time, and content dimensions of social media data. Rudra et al. (2015) proposed a novel approach to extracting situational awareness information with a classification-summarization approach. Albuquerque et al. (2015) improved the extraction of disaster-related tweets using the relation between geotagged tweets and spatial characteristics of the disaster. Wang et al. (2016) integrated wildfire-related Twitter activities with the space and time of tweets to reveal the situational awareness of the wildfire event. Mandel et al. (2012) demonstrated that the quantity of tweets sent out is highly related to the peaks of a disaster event and the users' concerns for Hurricane Irene are highly associated with their locations and genders.

Apart from the spatial component, situational awareness also involves a temporal component in natural disasters. More specifically, situational awareness is a dynamic construct, depending on the perceptions of individuals, emergency characteristics, and surrounding damages. The content categories of situational awareness defined in previous studies (e.g., Vieweg, 2012; Vieweg et al., 2010), such as casualties, damage, donation efforts, alerts, etc., largely focus on extracting information from the content of tweets, but the useful information posted before and after a disaster event has not been fully explored (Huang & Xu, 2014).

Previous studies sifted disaster-relevant tweets by using keywords (e.g., Huang & Xiao, 2015; Wang et al., 2016; Zou et al., 2018), which may miss many disaster-relevant tweets that do not include keywords in the predefined corpus. Moreover, most of these studies mainly rely on tweets on a macro level. In other words, researchers are more interested in understanding if the tweets are related to the disaster, instead of exploring further practical information. In this sense, local agencies and rescue teams may have difficulties addressing specific problems that are urgent during the disaster. As a result, beyond identifying macro-level situational awareness using tweets, this research also aims to extract more specific and actionable information from tweets to support local response teams in practice.

2.3 Techniques for situational awareness mining on twitter

Identifying disaster-relevant tweets is essential for disaster management, even though it is still challenging in terms of techniques. Sakaki et al. (2010) introduced a probabilistic spatiotemporal model to explore the interaction between disaster events and tweets. Kent and Capello (2012) employed a regression model to identify demographic characteristics so as to reflect the voluntary participation of the impacted neighborhood based on fire-related tweets. Natural language processing (NLP) techniques and statistical models are the fundamental methods to deal with traditional Twitter-related tasks (e.g., Corvey et al., 2010). However, most NLP techniques are developed based on formal texts. In addition, due to the scarcity of human-annotated data, crisis informatics researchers often failed to fully leverage state-of-the-art methods in their research.

To achieve a better performance in analyzing disaster data, a few machine learning approaches have been developed and applied. In order to build a platform for situational awareness detection, Cameron et al. (2012) employed an SVM classifier to classify related tweets.

Verma et al. (2011) adopted Naïve Bayes and MaxEnt classifiers to detect situational awareness from tweets. Imran et al. (2014) introduced a platform called AIDR, supported by a random forest classifier, to identify the tweets in a disaster event. Imran, Mitra, and Castillo (2016) compared the performance of traditional machine learning-based classifiers based on labeled tweets. With the emerging of Word2vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014), the combination of word embeddings and DNNs becomes possible. For instance, Kim (2014) deployed Convolutional Neural Networks (CNN) to solve sentence-level classification problems. Kalchbrenner et al. (2014) used Dynamic Convolutional Neural Network (DCNN) to deal with the semantic modeling of sentences. Although studies show that CNN models outperform the traditional methods, more specific and actionable information, such as 'airport shut' or 'building collapse', is difficult to be extracted from a tweet. While Latent Dirichlet Allocation (LDA) could generate more disaster-related topics, the topics are often too general for decision-makers to act on (Vieweg et al., 2014).

3 Research design

3.1 Approach for multi-level situational awareness detection

The flowchart of the developed approach is illustrated in Fig. 2. The procedure includes the following parts. First, historical disaster-related tweets would be collected and labeled manually. Two-level categories should be labeled for each tweet. The first level is to determine if the tweet is related to the disaster event (perception-level situational awareness), i.e., binary classification. The second level is to categorize the disaster-related tweets into six humanitarian categories (humanitarian-level situational awareness). Second, historical tweets should be preprocessed and pre-trained by Word2vec so that the text can be converted into a numerical format. In order to build a deep learning classifier, the preprocessed data can be split into the training set (70%), the validation set (10%), and the test set (20%). Third, when a hurricane is approaching, the geotagged tweets can be crawled based on the Twitter Streaming API. The classifier would be used to detect perception-level situational awareness (binary classification) and humanitarian-level situational awareness (six-category classification) for
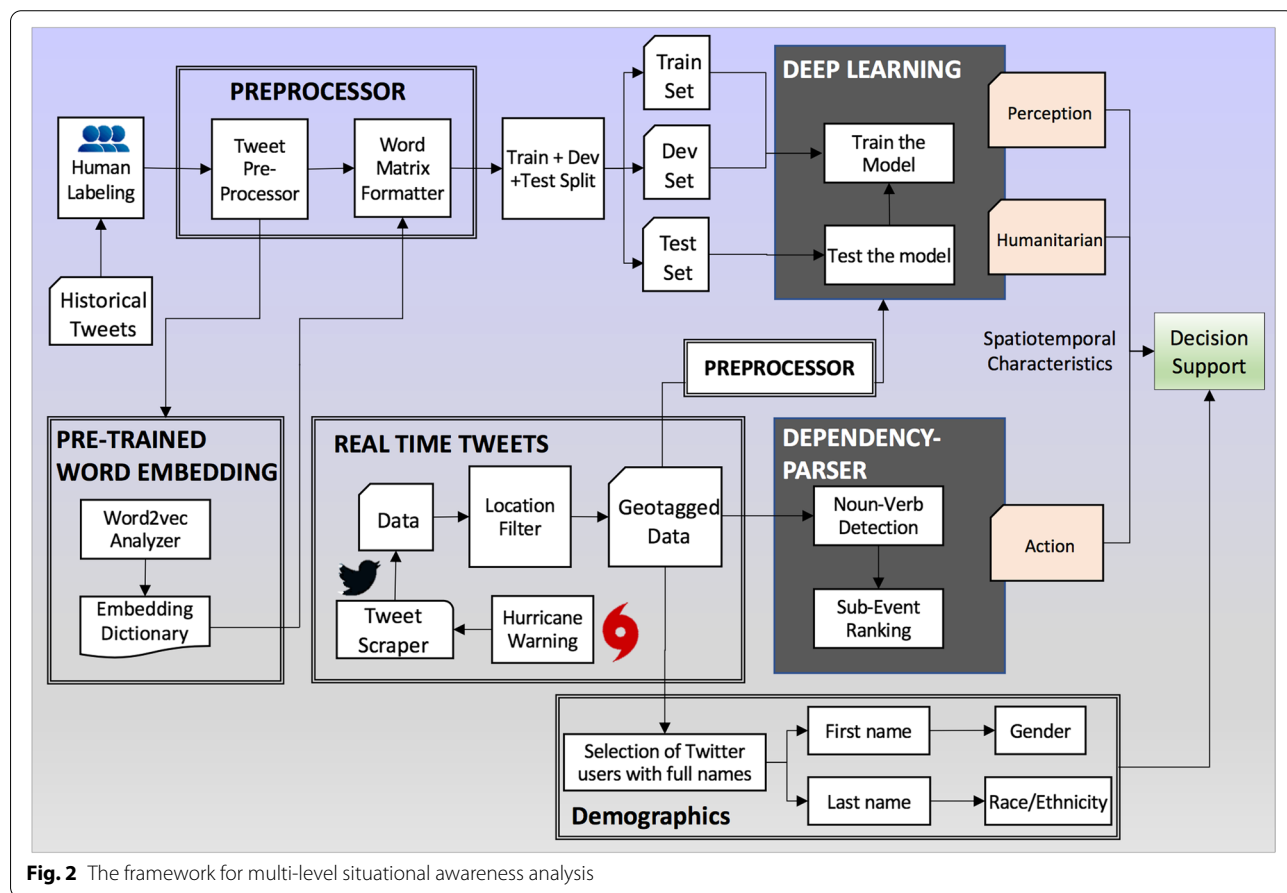

Fig. 2 The framework for multi-level situational awareness analysis

each tweet, respectively. Simultaneously, to understand action-level situational awareness, the dependency parser would be employed to extract all noun-verb pairs in each tweet so as to extract the up-to-date topics. Finally, to shed light on the resilience and social vulnerability of the community, it is necessary to estimate the demographics of Twitter users.

3.2 Deep learning model

In order to identify perception-level situational awareness and humanitarian-level situational awareness, the deep learning model can be used for the classification tasks. However, insufficient labeled data in the early phase of a disaster hinders machine learning tasks, which delays disaster response. To enrich the labeled data, I applied a graph-based deep learning framework that could be employed to learn an inductive semi-supervised model. The specific approach of graph-based semi-supervised learning with convolution neural networks can be found in Alam et al. (2018). The graph-based network learns internal representations of the input by predicting contextual nodes in a graph that encodes similarity between labeled and unlabeled training data. To accomplish the task, I performed the training using the function of *semi_supervised.LabelPropagation* from the package *sklearn* in Python.

Technically, calculating distance between $n(n-1)/2$ pairs of tweets to construct the graph is extremely expensive. Hence, I adopted the k-nearest neighbor-based method to find the nearest neighbors of tweets (k=10 in this study). The nearest neighbor graph consists of n vertices. Moreover, for each vertex, there is an edge set consisting of a subset of n tweets. The edge is defined by the distance measure $d(i, j)$ between tweets t_i and t_j, representing the similarity between the two tweets. Then I used pre-trained word embeddings to initialize the embedding matrix E in the network and trained a continuous bag-of-words (CBOW) Word2vec model (vector dimensions=300, window size=5) (Mikolov et al., 2013). The word embeddings were generated using the *Gensim* package in Python. Before training the deep learning model, I preprocessed all tweets by normalizing all characters to the lower-cased forms, removing special characters (i.e., '#$%^&*'), spelling out every digit, truncating elongations to two characters, converting all usernames to "userID", and converting all URLs to "HTTP". It is also essential to remove all punctuation marks except for periods, semicolons, questions, and exclamation marks.

3.3 Twitter dependency-parser

To perform the action-level analysis of situational awareness detection, practitioners are quite interested

in a specific topic discussed in each tweet. In this sense, the dependency parser can be employed to explore the representative part-of-speech (POS) taggers in each tweet. However, explicitly matching all nouns and verbs in a sentence is still challenging, not to mention that the tweet is informal in many cases. For instance, in the tweet: *"New construction just collapsed in front of me in Panama City Beach from #Hurricane Michael!!! Now pray for the people affected"*, the words 'collapsed' and 'pray' could be actionable verbs for decision-makers. The challenge is that the noun 'construction' is only supposed to be related to the term 'collapsed', instead of 'pray'. In other words, (*construction, collapsed*) is an actionable event, whereas (*construction, pray*) is not. In addition, expected nouns do not necessarily appear adjacent to the verbs. For example, in the tweet, *"Devastation in northwest Florida: strong bursts of Hurricane caused serious damage in residential areas. This is the panorama in Panama City. #HurricaneMichael"*, (damage, caused) is expected to be extracted as an actionable event for the response teams. However, the noun 'Hurricane' stays closer to the verb 'caused' than the noun 'damage'. In this sense, (*hurricane, caused*) would be identified automatically based on traditional POS taggers.

To address the aforementioned challenges, some new methods to construct the dependency relationship of tokens in tweets in the text-mining domain have been developed in the literature (e.g., Cai et al., 2009; Kong et al., 2014). Based on the approaches from Cai et al. (2009) and Kong et al. (2014), a noun can be matched with a verb based on the Twitter dependency parser. The representative noun-verb pairs in a tweet can rapidly reveal actionable information expressed by a user so as to help response teams understand the individual's need rapidly. To accomplish this challenge, this paper employed *TWEEBOPARSER* (https://github.com/ikekonglp/TweeboParser), the first syntactic dependency parser designed explicitly for English tweets (Kong et al., 2014). But if we aim to know the most representative noun-verb pairs within a region, further analysis is needed. Inspired by Rudra et al. (2018), we can rank the identified noun-verb pairs based on different factors. First, we can calculate the Szymkiewicz-Simpson overlap score of a pair P (N, V) using Equation 1:

$$\text{Score(P)} = \frac{|X \cap Y|}{\min(|X|, |Y|)} \tag{1}$$

where X, Y represents the set of tweets containing N and V, respectively. However, Equation 1 alone does not discriminate between frequent and infrequent noun-verb pairs. Then, a discounting factor δ proposed by Pantel and Lin (2002) is also calculated to take the frequency of

the noun-verb pair into consideration. The factor δ can be represented by the Equation 2:

$$\delta(P) = \frac{|X \cap Y|}{1 + (|X|, |Y|)} * \frac{\min(|X|, |Y|)}{1 + \min(|X|, |Y|)} \tag{2}$$

Then, the weight score of a noun-verb pair is calculated using Equation 3, which reflects the importance of the sub-event:

$$Weight(P) = Score(P) * \delta(P) \tag{3}$$

3.4 Demographic inference

To protect the privacy of users, the demographic information of Twitter users is not displayed in the user profile. To enrich the findings of this research, we estimated the Twitter user's gender and race using their usernames using the approach adopted by Yuan et al. (2021). First, it is to preprocess Twitter users' initial usernames, including removing punctuations in Twitter users' names and dividing usernames into tokens by white space. After that, it is essential to remove the titles (e.g., professor and president), suffix acronyms (e.g., M.D. and M.A.), and prefix acronyms (Mr. and Ms.) in usernames. Next, two databases can be used to infer gender and race, respectively. The last names database contains 162,254 last names from the 2017 Census and the ratio of users of different races. According to the highest ratio of race corresponding to the first name and surname in the United States Census Bureau data, the user's race can be estimated. To infer users' gender, I employed the first name database with 32,469 first names, including 18,309 and 14,160 first names for females and males, respectively. Thereafter, the gender of every user can be classified through the user's first name.
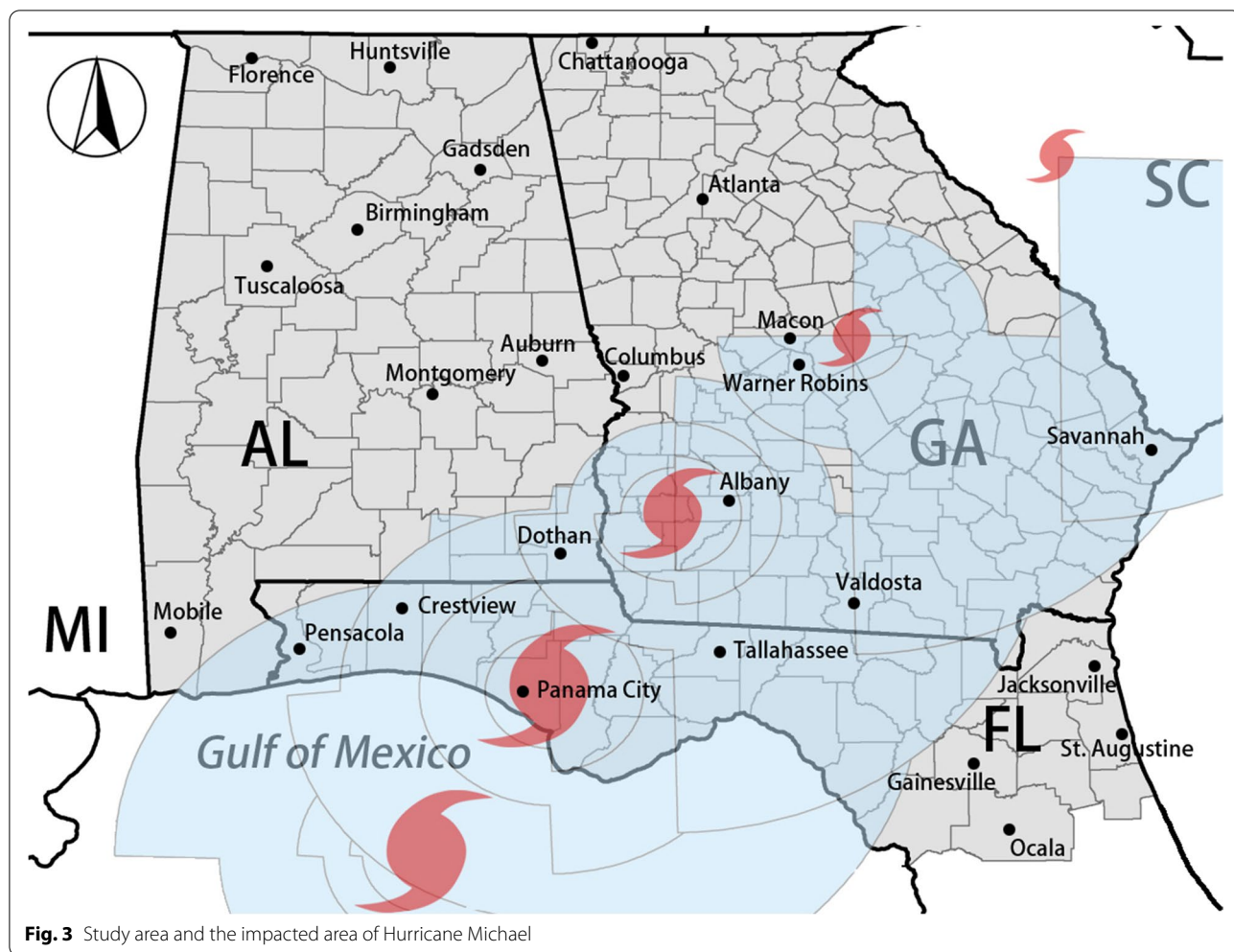

Fig. 3 Study area and the impacted area of Hurricane Michael

4 Case study and data

4.1 Hurricane Michael and study area

To show the feasibility of the developed framework, I applied the multi-level analytic framework to investigate the impacts of Hurricane Michael. Hurricane Michael was the fourth-strongest landfalling hurricane in the United States on record. Once making landfall on Oct. 10, 2018, Hurricane Michael began to weaken, as it moved into South Carolina early on Oct. 11. Fig. 3 indicates that the main impacted areas include northwest Florida and southern Georgia. Then, I captured geotagged tweets using the Twitter Streaming API from Oct. 8, 12:00 am to Oct. 19, 11:59 pm. Specifically, each tweet contains the tweet ID, coordinates, registered city of the user, timestamp, and content. Note that we not only considered the tweets with coordinates, but also took the tweets with city information into account. The major reason is that the tweets with coordinates only account for less than 2% of the total tweets. Nevertheless, if the Twitter user's city information is also considered, there would be over 20% of tweets that contain location information. Thereafter, this study would aggregate the detected results on a county level. To analyze the situational awareness changes over time, I divided the collected tweets into

Table 1 Class distribution of annotated disaster events

Humanitarian-level Category	Affected Individual	Donations and Volunteering	Infrastructure and Utility	Sympathy and Support	Other Useful Information	Not Related or Irrelevant	Total
2013 Pakistan Earthquake	261	228	21	146	554	247	1,457
2014 California Earthquake	167	58	312	125	80	764	1,506
2014 Chile Earthquake	249	12	43	931	840	363	2,438
2014 India Floods	995	46	67	72	312	483	1,975
2014 Mexico Hurricane Odile	191	181	675	236	382	42	1,707
2015 Nepal Earthquake	756	1,001	50	1,017	1,504	6,658	11,284
2015 Vanuatu Cyclone Pam	254	383	231	320	668	713	2,569
2014 Philippine Typhoon Hagupit	127	388	301	424	9,103	159	10,502
Other disasters	943	636	492	1467	1,742	929	6,209
Total	3,943	2,933	2,192	4,738	15,185	10,358	39,647

Table 2 Binary classification results of tweets (perception-level situational awareness)

Natural Disaster Events	Models	Area Under the Curve (AUC)			F1-score		
		SVM	RF	Graph-based CNN	SVM	RF	Graph-based CNN
2013 Pakistan Earthquake	M_{event}	77.43	79.17	82.16	0.771	0.782	0.813
	M_{cross}	70.36	71.24	74.92	0.692	0.715	0.751
2014 California Earthquake	M_{event}	77.54	78.97	80.23	0.756	0.772	0.791
	M_{cross}	52.89	55.89	63.87	0.501	0.534	0.615
2014 Chile Earthquake	M_{event}	89.36	89.23	92.95	0.882	0.890	0.928
	M_{cross}	80.12	80.06	83.92	0.802	0.801	0.828
2014 Mexico Hurricane Odile	M_{event}	92.13	93.54	95.36	0.919	0.936	0.965
	M_{cross}	85.77	87.17	88.26	0.842	0.867	0.8813
2014 India Floods	M_{event}	90.07	91.12	93.23	0.901	0.908	0.936
	M_{cross}	83.92	84.78	85.89	0.839	0.849	0.860
2015 Nepal Earthquake	M_{event}	82.03	83.25	85.97	0.809	0.812	0.831
	M_{cross}	79.12	80.21	81.12	0.769	0.788	0.783
2015 Vanuatu Cyclone Pam	M_{event}	87.12	88.67	91.23	0.855	0.867	0.893
	M_{cross}	78.23	79.12	84.36	0.761	0.782	0.823
2014 Philippines Typhoon Hagupit	M_{event}	81.09	82.95	87.01	0.791	0.803	0.852
	M_{cross}	71.79	72.98	81.78	0.684	0.693	0.796
Average difference between M_{event} and M_{out}		9.32 (6.12)	9.43 (5.58)	8.00 (3.38)	0.10 (0.06)	0.09 (0.06)	0.08 (0.04)

three groups: before the hurricane made landfall (Oct. 8, 12:00 am - Oct. 10, 2:00 pm), during the hurricane event (Oct. 10, 2:00 pm- Oct. 12, 11:59 pm) and after the hurricane dissipated (Oct. 13, 12:00 am – Oct. 19, 11:59 pm).

4.2 Labeled disaster-relevant dataset

A major reason that not many studies adopt the supervised learning method is that creating a large and labeled Twitter corpus is time-and-money consuming. Imran, Mitra, and Srivastava (2016) first opened their labeled disaster-relevant Twitter datasets, which contain nearly 50,000 labeled tweets. In this research, regarding the perception-level situational awareness, tweets can be categorized into two classes, *Disaster-relevant* and *Not related.* In terms of humanitarian-level situational awareness, six categories are specified. The dataset used in this study contains eight disaster events. The distribution of categorized tweets can be seen in Table 1. Then, I built a deep learning classifier using graph-based semi-supervised learning and trained the model based on the categorized tweets.

5 Results and findings

5.1 Accuracies of the deep learning model

Table 2 indicates the results of the accuracy of the deep learning model for detecting perception-level situational awareness. M_{event} represents that the model of a disaster

event is only trained and tested based on the tweets generated in this event (defined as *event-based data*). M_{cross} represents that the model is trained and tested based on tweets from other events. To be consistent with the existing work (e.g., Nguyen et al., 2017; Yu et al., 2019), I compared the performance of the deep learning model with traditional machine learning models, namely Random Forest (RF) and Support Vector Machine (SVM). Unigram, bigram, and trigram features are generated from the tweets as features for training RF and SVM models. Table 2 further shows that the graph-based CNN model outperforms the traditional machine learning models irrespective of if the model is trained with event-based data or cross-event data. In addition, Table 2 indicates that the AUC derived from cross-event data is lower than that derived from the event-based data. An interesting finding is that the average difference number between M_{event} and M_{out} in Table 3 is less than that in Table 2. Therefore, even though the accuracy of binary classification would be higher than that of the multiclass task, the multiclass classifier shows higher robustness with cross-event data.

5.2 Accuracy of dependency-parser

To explore the performance of the action-level situational awareness detection, I calculated the accuracy of extracted noun-verb pairs. In specific, two paid

Table 3 Multi-class classification results of tweets (humanitarian-level situational awareness)

Natural Disaster Events	Models	Accuracy			Macro F1		
		SVM	RF	Graph-based CNN	SVM	RF	Graph-based CNN
2013 Pakistan Earthquake	M_{event}	72.15	73.01	74.13	0.684	0.692	0.713
	M_{cross}	67.85	69.77	70.01	0.663	0.672	0.688
2014 California Earthquake	M_{event}	69.87	70.17	72.23	0.678	0.682	0.709
	M_{cross}	65.71	66.21	68.98	0.645	0.652	0.671
2014 Chile Earthquake	M_{event}	71.98	73.21	75.12	0.725	0.733	0.752
	M_{cross}	68.98	70.12	72.23	0.691	0.701	0.712
2014 Mexico Hurricane Odile	M_{event}	67.54	68.43	70.78	0.640	0.696	0.716
	M_{cross}	64.47	66.02	68.87	0.621	0.672	0.702
2014 India Floods	M_{event}	81.19	82.73	85.86	0.810	0.832	0.863
	M_{cross}	77.98	78.77	81.23	0.792	0.811	0.832
2015 Nepal Earthquake	M_{event}	67.28	68.46	70.12	0.793	0.707	0.712
	M_{cross}	66.12	67.34	69.68	0.771	0.683	0.692
2015 Vanuatu Cyclone Pam	M_{event}	70.98	71.26	72.23	0.702	0.712	0.709
	M_{cross}	67.32	68.25	70.19	0.687	0.692	0.676
2014 Philippines Typhoon Hagupit	M_{event}	81.79	82.15	83.41	0.791	0.818	0.859
	M_{cross}	77.12	78.23	80.09	0.765	0.784	0.823
Average difference between M_{event} and M_{cross}		3.40 (1.02)	3.09 (0.91)	2.83 (1.25)	0.02 (0.01)	0.03 (0.01)	0.03 (0.01)

Table 4 Accuracy of noun-verb pair extraction in each disaster event

	2013 Pakistan Earthquake	2014 California Earthquake	2014 Chile Earthquake	2014 Mexico Hurricane Odile	2014 India Floods	2015 Nepal Earthquake	2015 Vanuatu Cyclone Pam	2014 Philippines Typhoon Hagupit	Total
Accuracy	0.701	0.745	0.667	0.712	0.692	0.656	0.668	0.693	0.696
SD	0.121	0.132	0.112	0.096	0.104	0.143	0.154	0.124	0.127
N	100	100	100	100	100	100	100	100	800

SD Standard Deviation

volunteers manually identified the noun-verb pairs in a tweet, which is seen as the ground truth of actionable information for the tweet. For each disaster event, 100 tweets are randomly selected. Note that many tweets do not include a valid noun-verb pair because of the informal grammar. I only selected tweets that contain at least one noun-verb pair. Assume that there are n noun-verb pairs in a tweet, and m noun-verb pairs are correctly automatically extracted. In this sense, the accuracy of the noun-verb pair identification for one tweet is m/n. By averaging the accuracies of sampled tweets in each disaster event, the overall accuracy of the noun-verb pair extractions could be evaluated.
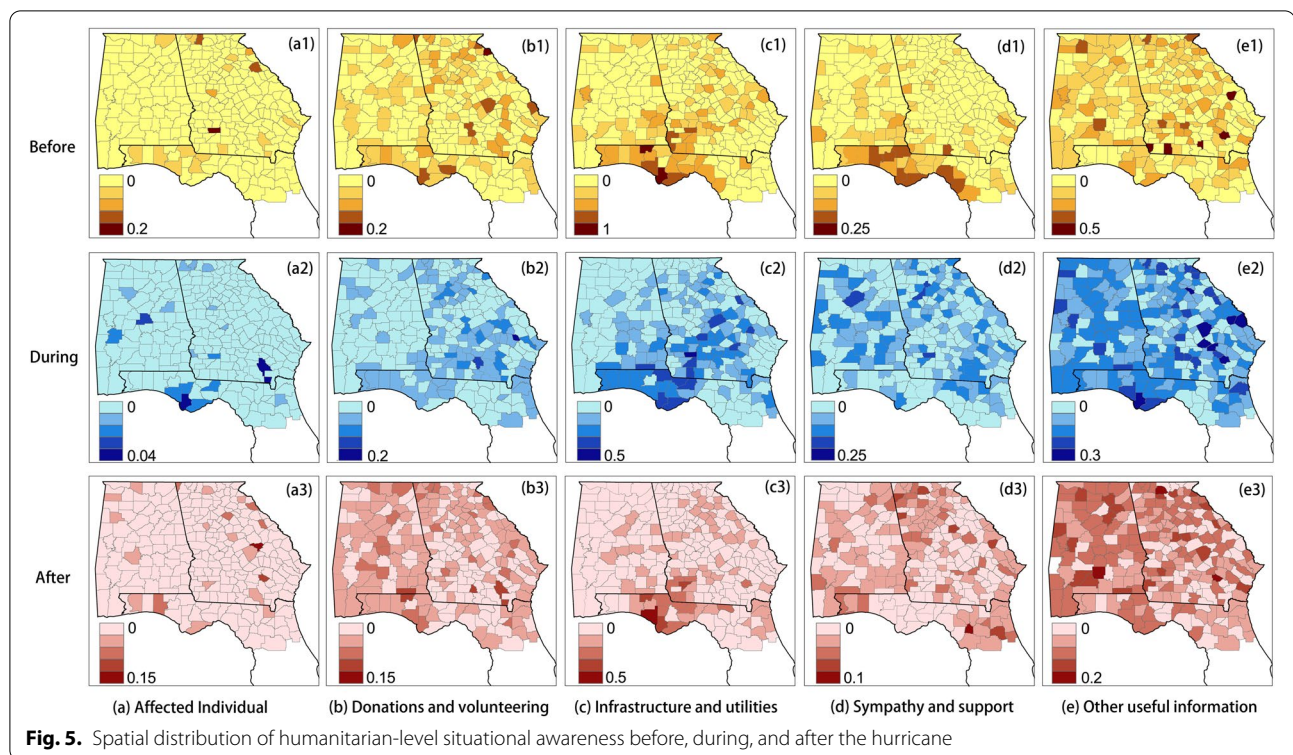
Table 4 indicates the accuracy of extracting noun-verb pairs in each disaster event. Clearly, the average accuracy value for all examined tweets is 0.686. The main type of error in the extracted noun-verb pairs is the mismatch of nouns and verbs in a tweet. For instance, in the tweet "#HurrincaeMichael cuts power lines in Mexico Beach.", the extracted noun-verb pair is (lines, cuts) whereas the accurate pair is (power, cuts), which represents more actionable information. In addition, the informal usage of English grammar may also lead to mismatches. It is worth noting that the accuracy of 2014 California data is significantly higher than that of others. Part of the reason is that the disaster event occurred in the US and the users

may have a good command of English grammar. Overall, even though the accuracy varies with the disaster event, the results reveal nevertheless that the dependency-parser approach can automatically extract the actionable information of a tweet with an accuracy of around 0.7. Considering that ambiguity remains a major challenge in the NLP domain, the accuracy of the method is acceptable in practice.

5.3 Perception-level situational awareness
Based on the pre-trained deep learning model, decision-makers can predict the perception-level situational awareness of captured tweets to have a big picture of the disaster impact. Fig. 4 indicates the spatial distribution of the percentage of disaster-relevant tweets. It is not surprising that before Hurricane Michael, the coastal counties would exhibit a higher level of perception of the upcoming disaster (see Fig. 4a). More specifically, Hurricane Michael first made landfall in Panama City and Mexico Beach City based on Fig. 4, so that residents would be more concerned about the disaster before the hurricane. Even though some coastal counties are not indeed affected by the hurricane, the perception of coastal counties before the hurricane is also higher than inland counties. It could be because the predicted hurricane trajectory is unpredictable so that

Fig. 4 Spatial distribution of perception-level GSA before, during and after Hurricane Michael

Fig. 5. Spatial distribution of humanitarian-level situational awareness before, during, and after the hurricane
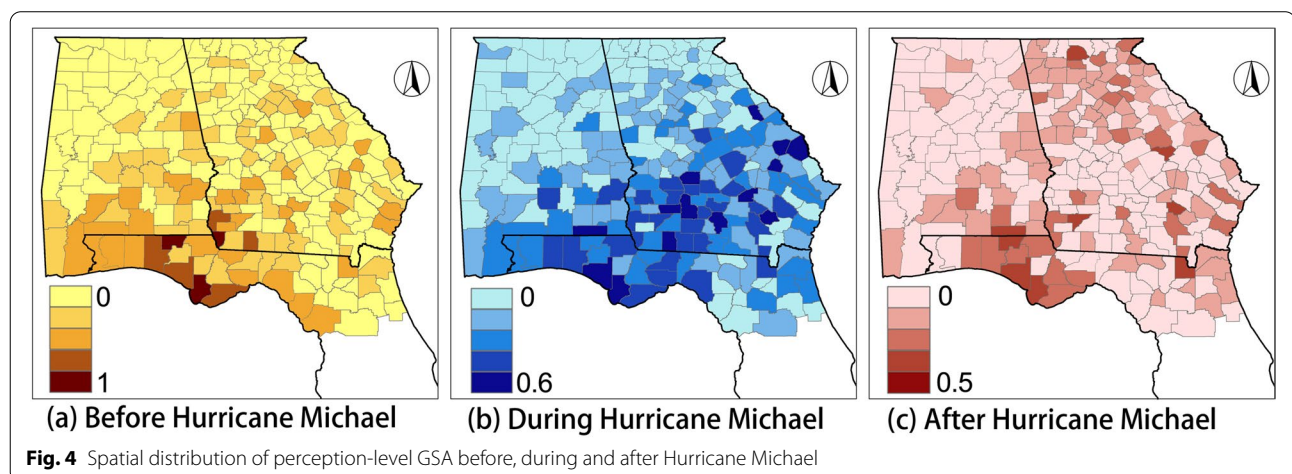
the uncertainty makes the residents in coastal counties also concerned. Fig. 4b also shows that counties along the path of Hurricane Michael have a higher percentage of disaster-related tweets during Hurricane Michael. Intuitively, people would have a higher level of perception when they are in the disaster-affected area, meaning that perception-level situational awareness could quickly help decision-makers have a big picture of the mainly impacted regions. An interesting finding after the hurricane is that the less impacted regions, such as counties in northern Georgia, have more concerns about the hurricane (Fig. 4c).

It is worth noting that residents in the most damaged coastal cities (i.e., Panama City and Mexico Beach City) were concerned about the hurricane all the time. However, the percentages of perception-level situational awareness in coastal counties are the highest before the hurricane (Fig. 4a). Relatively, the percentages of perception-level situational awareness in Panama City and Mexico Beach City are in the range of 30%-50% during and after the hurricane. The result shows that in regions that would be severely impacted, the expressed perception of the disaster is most significant before the hurricane. One possible explanation for this finding is that residents may be more concerned about their own safety during and after the hurricane, instead of expressing their concerns about the disaster on social

media platforms. Another possible reason could be that local communication infrastructures were damaged due to the hurricane so that few people would get access to the Internet during and shortly after the hurricane.

5.4 Humanitarian-level situational awareness
Figure 5 indicates the spatial distribution of tweets categorized by the humanitarian topics before, during, and after the hurricane. Specific findings of each category can be seen below.

5.4.1 *Affected individual*
Tweets in this category account for the least percentage. Only during the hurricane, *Affected Individual* tweets are mainly located in the most impacted regions (Fig. 5a2). Intuitively, the percentages before the hurricane are less than that during the hurricane, considering that the hurricane has not made landfall yet and no damage occurred. The identified tweets associated with *Affected Individual* can help disaster management organizations quickly estimate the fatalities and injuries during and shortly after the disaster event. Admittedly, the estimated numbers may not be accurate. But the magnitudes of the numbers nevertheless help decision-makers understand the general severity of the disaster in terms of affecting local residents.

5.4.2 Donations and volunteering

Figure 5b1 shows that the tweets in the category *Donations and Volunteering* are randomly distributed before the hurricane. The reason is that people are not sure which region would be severely damaged by the hurricane. Figure 5b2 shows that Twitter users in the impacted regions would be more concerned about donations during the hurricane so that they express their needs via Twitter. Figure 5b3 indicates that after Hurricane Michael, the percentage of tweets associated with *Donations and volunteering* increases significantly, meaning that people are more willing to help the impacted cities after the disaster event. The result could be used to help the state government optimize the distribution of relief materials before, during, and after the hurricane.

5.4.3 Infrastructure and utility

Regarding the tweets associated with *Infrastructure and Utility*, the spatial distribution is highly correlated with the geographic path of the hurricane. People would respond to and spread the incident shortly after it happens. For example, it is very common that drivers would report a road segment damage or block when they are on site. Specifically, before the hurricane made landfall, some Twitter users had been concerned about the infrastructure and utility damages (Fig. 5c1). Figure 5c2 indicates that the infrastructure and utility damages in the mainly impacted regions are more frequently mentioned on Twitter during the hurricane. The spatiotemporal pattern of this humanitarian category is able to help decision-makers determine the priority in terms of repairing the infrastructures and utilities in the disaster event.

5.4.4 Sympathy and support

Figure 5d1 indicates that Twitter users in coastal counties would like to express their sympathy and support before the hurricane made landfall. An interesting finding is that users in the most impacted areas are not likely to express sympathy and support during and after the hurricane (Fig. 5d2 and Fig. 5d3). By contrast, Twitter users from counties closer to the impacted area show more sympathy and willingness to emotionally support affected people online. It is reasonable because people in safe areas would prefer to pray for the impacted cities and encourage more people to provide essential support.

5.4.5 Other useful information

Compared to other categories, the tweets classified as *Other Useful Information* account for the most percentage, whereas Tweets in this category are relatively randomly distributed. A surprising finding is that, after the hurricane, Twitter users in many counties from Alabama show more concern for the hurricane even though they are just slightly influenced by the hurricane.

5.5 Action-level situational awareness

Based on the approach introduced in Section 3.3, I extracted the noun-verb pairs of each tweet. Then, I calculated the importance of extracted pairs based on the Equation. 3. However, there are many disaster-irrelevant tweets that may impact the identification of specific noun-verbs associated with disasters. I therefore only listed the five most important noun-verb pairs identified in each humanitarian category (see Table 4).

In Table 5, each extracted noun-verb pair is followed by a number, which represents the weight (or the importance) of this pair calculated based on Equation.3. Generally speaking, the noun-verb pairs can allow response
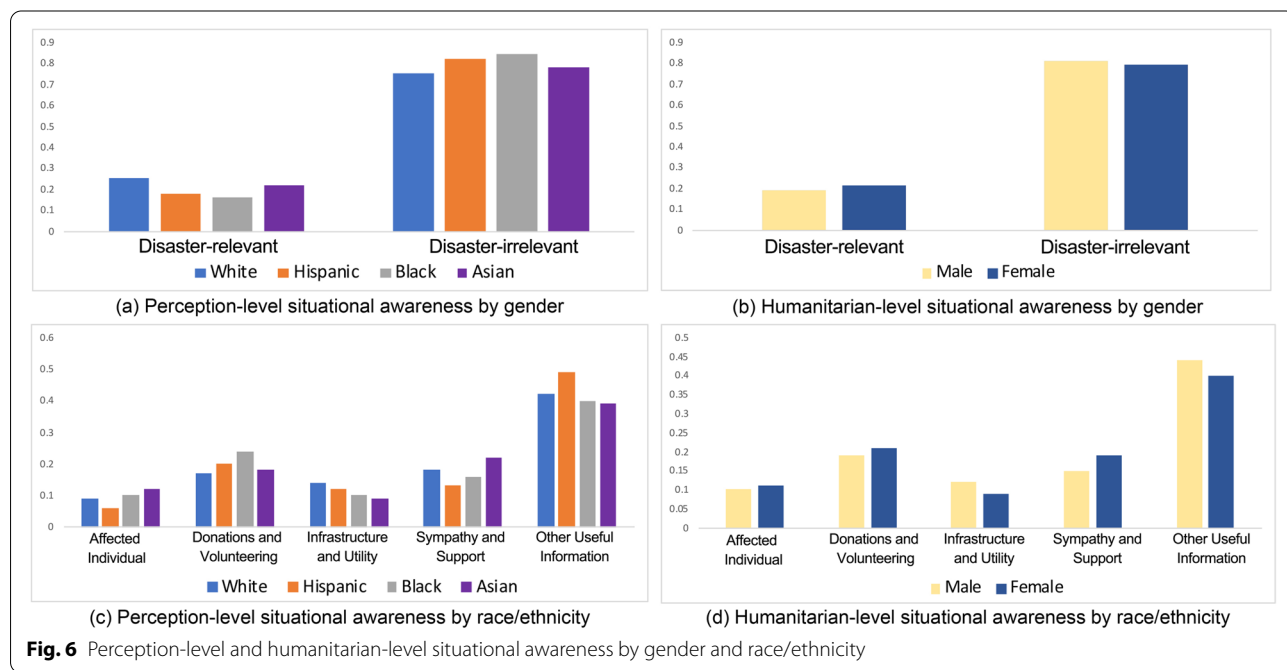
Table 5 Action-level situational awareness over different phases (top 5)

Phase	Affected Individual	Infrastructure and Utility	Donations and Volunteering	Sympathy and Emotional Support	Other Useful Information
Before	(we, close), 0.31 (we, leave), 0.19 (hurricane, attack), 0.18 (mom, worry),0.15 (people, evacuate), 0.13	(flight, make), 0.25 (traffic, congest), 0.20 (airport, close), 0.17 (road, block), 0.14 (airport, shut), 0.13	(hurricane, hit), 0.23 (money, donate), 0.22 (friend, help), 0.17 (volunteers, help), 0.15 (hurricane, get), 0.12	(hurricane, pray), 0.21 (safe, stay), 0.16 (god, bless), 0.15 (Florida, pray), 0.09 (Florida, help), 0.07	(storm, come), 0.33 (storm, report), 0.15 (rain, start), 0.13 (rain, come), 0.11 (landfall, make), 0.06
During	(we, evacuate), 0.24 (people, evacuate), 0.22 (I, leave), 0.18 (we, leave), 0.14 (nobody, hurt), 0.12	(road, block), 0.31 (airport, shut), 0.29 (closure, force), 18 (electricity, lost), 13 (power, lost), 13	(supplies, get), 0.21 (help, need), 0.19 (money, support), 0.18 (volunteer, need), 0.14 (donation, support), 0.12	(family, pray), 0.44 (Florida, pray), 0.32 (we, pray), 0.28 (Florida, pray), 0.16 (boats, rescue), 0.15	(category, downgrade), 0.21 (hurricane, rain), 0.17 (place, ruin), 0.12 (landfall, make), 0.10 (we, reopen), 0.08
After	(hurricane, beat), 0.20 (we, survive), 0.18 (survivor, be), 0.15 (citizens, trapped), 0.12 (crew, thank), 0.12	(power, lost), 18 (house, inundate), 18 (house, clean), 16 (water, go), 14 (airport, open), 13	(we, donate), 0.20 (Florida, help), 0.12 (money, prepare), 0.10 (money, need), 0.09 (we, rebuild), 0.05	(guys, love), 0.30 (prayer, thank), 0.28 (staff, beg), 0.17 (safe, stay), 0.15 (god, bless), 0.12	(today, open), 0.20 (clouds, come), 0.17 (hurricane, stop), 0.16 (hurricane, left), 0.13 (storm, left), 0.10

teams to rapidly know the specific topic, emergency, or needs in each tweet. For instance, Table 5 shows that the pairs (*traffic congested*) and (*road, block*) are highly important in the humanitarian category *Infrastructure and Utility*, meaning that the specific road has been blocked or congested. Thereafter, the decision-makers can extract tweets containing related pairs and check the location of these tweets, so that the evacuation route and traffic could be rearranged. Similarly, the pair (*house, inundate*) appears frequently after the hurricane dissipated, which would remind rescue teams that people may be trapped in the houses.

Apart from specific information that is beneficial to rescue teams, Table 5 indicates some interesting patterns in each humanitarian category. For the category *Affected Individual*, people are more concerned about the evacuation before and during the hurricane. An interesting finding is that before the hurricane, the pair (*we, close*) is the most important within the study area. By checking the original tweets, I found that many Twitter users sending this topic were restaurant and store owners. They closed their places and spread the shutdown information on Twitter. For the category *Infrastructure and Utility*, transportation infrastructures and power infrastructures are more frequent before and during the hurricane. After the hurricane, people would be more concerned about their properties. An unexpected finding in the category *Donations and Volunteering* is that (*hurricane, hit*) is the most frequent pair before the hurricane. It is because many humanitarian and aid organizations prefer to warn about the hurricane landfall in advance, and then encourage people to donate money. In this sense, the tweet containing (*hurricane, hit*) often includes more than one noun-verb pair. In specific, most of the tweets containing the (*hurricane, hit*) pair also contain the pair (*money, donate*) or (*we, donate*). In terms of the category *Other*

Fig. 6 Perception-level and humanitarian-level situational awareness by gender and race/ethnicity

Table 6 Action-level situational awareness by gender and race/ethnicity (top 10)

White	Hispanic	Black	Asian	Female	Male
(storm, come)	(hurricane, rain)	(hurricane, left)	(we, leave)	(hurricane, rain)	(storm, come)
(safe, stay)	(god, bless)	(volunteers, help)	(hurricane, left)	(we, leave)	(safe, stay)
(hurricane, hit)	(rain, come)	(prayer, thank)	(traffic, congest)	(volunteers, help)	(hurricane, hit)
(money, donate)	(landfall, make)	(we, pray)	(we, rebuild)	(family, pray),	(money, donate)
(traffic, congest)	(money, need)	(people, evacuate)	(power, lost)	(volunteer, need)	(hurricane, left)
(power, lost)	(staff, beg)	(safe, stay)	(airport, shut)	(house, inundate)	(storm, report)
(house, clean)	(we, leave)	(god, bless)	(money, support)	(house, clean)	(money, donate)
(hurricane, pray)	(mom, worry)	(we, reopen)	(storm, report)	(rain, come)	(traffic, congest)
(rain, come)	(we, survive)	(donation, support)	(storm, left)	(people evacuate)	(staff, beg)
(landfall, make)	(volunteer, need)	(money, prepare)	(clouds, come)	(friend, help)	(we, leave)

Useful Information, the extracted noun-verb pairs are mainly related to the landfall and dissipation of the hurricane, and the weather conditions.

5.6 Disparities of situational awareness by demographics
The identification of demographic information allows me to explore the disparities of situational awareness by race/ethnicity and gender (Fig. 6). First, regarding race/ethnicity, Whites are more likely to post disaster-related tweets during the hurricane compared to other groups at the perception level (Fig. 6a). Moreover, at the humanitarian level, I found that Blacks and Hispanics show more interest in donation and volunteering (Fig. 6c), which is caused by the more severe disaster impacts on the minority group, due to their insufficient preparations against the hurricane. In the meantime, Whites could have more concerns about the infrastructure and utility than other groups. For example, Whites are more likely to face traffic congestion since they could have a higher rate of car ownership than minority groups. Second, considering the gender, I found that females are more concerned about the hurricane than males at the perception level (Fig. 6b). At the humanitarian level, it is clear that females show more interest in sympathy and donations than males (Fig. 6d).

I also summarized the top five topics at the action level across races/ethnicities and genders (Table 6). The results indicate that Whites show more interest in the hurricane before the event such as the noun-verb pair (*storm, come*) whereas Blacks and Asians are more concerned about the post-disaster impact such as the pairs (*hurricane, left*), (*people, evacuate*), (*we, leave*), etc. Furthermore, I found that Whites and Asians are more interested in the impact on infrastructure such as the pairs (*traffic, congest*) and (*power, lost*). With regard to the gender difference, it is clear that females show more sympathy such as the pairs (*volunteers, help*), (*volunteer, need*), etc., and infrastructures such as (*house, inundate*).

6 Conclusion
6.1 Limitations
So far, I have not discussed the limitations in terms of using Twitter to monitor situational awareness in a disaster event.

First, Twitter users cannot represent all local residents because not everyone uses Twitter. It has been widely acknowledged that some disadvantaged groups (i.e., low income, low education, and elderly) could lack the devices and motivations to access social media apps; thus, they may be less likely to publish or receive disaster-relevant information. In addition, based on

the privacy policy of Twitter, only a small percentage of geotagged tweets with coordinates are accessible for public use (less than 2%). Even though I also consider the tweets with city information in this study, the geolocation tweets are still less than 10%. Furthermore, uncertainty and representativeness have been found in previous studies (e.g., Starbird et al., 2016; Pavalanathan & Eisenstein, 2015).

Second, the quality and amount of training data determine the effectiveness and accuracy of the proposed approach. Specifically, for detecting perception-level and humanitarian-level situational awareness, the deep learning model heavily depends on labeled tweets. One concern of labeled tweets is that some tweets may belong to more multi-categories on the humanitarian level. For instance, the tweet "*The airport passenger terminal will close tomorrow# Pray for Florida.*" can belong to both *Infrastructure and utility damage* and *Sympathy and emotional support*. This common and inevitable problem may decrease the performance of the proposed approach. However, given these disadvantages, Goodchild and Glennon (2010) demonstrated that the benefits of using the geotagged social media data outweigh its limitations.

Third, this study filtered the Twitter users using their usernames from disaster-related tweets in order to infer their demographic characteristics. However, the filtered tweets' representativeness in the overall disaster-related tweets remains to be enhanced because not all Twitter users use real names as their usernames. While this study ignored this representativeness issue, future work could further examine the accuracy of this method.

7 Conclusion
This study contributes to the literature from two aspects. First, to meet the demand from different stakeholders, I developed a multi-level framework for detecting situational awareness using Twitter data in disaster management. This framework is not only feasible for analyzing Twitter data, but also for conventional approaches (e.g., survey or interview) in terms of the investigation of situational awareness. Second, I examined the performance of using cross-event tweets in perception-level and humanitarian-level situational awareness detections, regarding those previous methods are mostly based on event datasets that are not practical when the event data is not available. In addition, to demonstrate the reliability of the action-level situational awareness detection, I examined the accuracy of the dependency parser in terms of extracting actionable noun-verb pairs.

In my future work, I will integrate more open data, as well as traditional survey data, to monitor the residents' situational awareness before, during and after a

hurricane. For instance, instead of only using Twitter as a data source, the authoritative datasets (e.g., remote sensing data) are expected to be combined to improve the identification of relevant messages from social media platforms. Moreover, to improve the accuracy and effectiveness of the analytic framework, I would annotate more tweets in more types of disaster events, so that the framework can be generalized to more types of disasters.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
Not applicable.

References
Alam, F., Joty, S., & Imran, M. (2018). Graph based semi-supervised learning with convolution neural networks to classify crisis related tweets. In *Twelfth international AAAI conference on web and social media*.
Ashktorab, Z., Brown, C., Nandi, M., & Culotta, A. (2014). Tweedr: Mining twitter to inform disaster response. In *ISCRAM* (pp. 269–272).
Cai, D., Hu, Y., Miao, X., & Song, Y. (2009). Dependency grammar based English subject-verb agreement evaluation. In *Proceedings of the 23rd Pacific Asia conference on language, information and computation, volume 1 (Vol. 1)*.
Cameron, M. A., Power, R., Robinson, B., & Yin, J. (2012). Emergency situation awareness from twitter for crisis management. In *Proceedings of the 21st international conference on world wide web (pp. 695-698)*. ACM.
Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J., & Waters, N. (2016). Using twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. *International Journal of Remote Sensing, 37*(1), 100–124.
Chatfield, A. T., & Reddick, C. G. (2015). Understanding risk communication gaps through e-government website and twitter hashtag content analyses: The case of Indonesia's Mt. Sinabung eruption. *Journal of Homeland Security and Emergency Management, 12*(2), 351–385.
Corvey, W. J., Vieweg, S., Rood, T., & Palmer, M. (2010). Twitter in mass emergency: What NLP can contribute. In *Proceedings of the NAACL HLT 2010 workshop on computational linguistics in a world of social media* (pp. 23–24).

De Albuquerque, J. P., Herfort, B., Brenning, A., & Zipf, A. (2015). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science, 29*(4), 667–689.
Deng, Q., Liu, Y., Zhang, H., Deng, X., & Ma, Y. (2016). A new crowdsourcing model to assess disaster using microblog data in typhoon Haiyan. *Natural Hazards, 84*(2), 1241–1256.
Endsley, M. R., & Rodgers, M. D. (1994). Situation awareness information requirements analysis for en route air traffic control. In *Proceedings of the human factors and ergonomics society annual meeting (Vol. 38, no. 1, pp. 71-75)*. SAGE Publications.
Feng, Y., Huang, X., & Sester, M. (2022). Extraction and analysis of natural disaster-related VGI from social media: Review, opportunities and challenges. *International Journal of Geographical Information Science, 36*(7), 1275–1316.
Ford, J. D., Tilleard, S. E., Berrang-Ford, L., Araos, M., Biesbroek, R., Lesnikowski, A. C., et al. (2016). Opinion: Big data has big potential for applications to climate change adaptation. *Proceedings of the National Academy of Sciences, 113*(39), 10729–10732.
Fu, X., Li, C., & Zhai, W. (2022). Using Natural Language Processing to Read Plans: A Study of 78 Resilience Plans From the 100 Resilient Cities Network. *Journal of the American Planning Association*, 1–12.
Gaba, D. M., Howard, S. K., & Small, S. D. (1995). Situation awareness in anesthesiology. *Human Factors, 37*(1), 20–31.
Ganz, A., Schafer, J. M., Tang, J., Yang, Z., Yi, J., & Ciottone, G. (2015). Urban search and rescue situational awareness using DIORAMA disaster management system. *Procedia Engineering, 107*, 349–356.
Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems, 26*(3), 10–14.
Ghahremanlou, L., Sherchan, W., & Thom, J. A. (2015). Geotagging twitter messages in crisis management. *The Computer Journal, 58*(9), 1937–1954.
Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: A research frontier. *International Journal of Digital Earth, 3*(3), 231–241.
Guan, X., & Chen, C. (2014). Using social media data to understand and assess disasters. *Natural Hazards, 74*(2), 837–850.
Hogg, D. N., & FOLLES⊘, K. N. U. T., Strand-Volden, F., & Torralba, B. (1995). Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. *Ergonomics, 38*(11), 2394–2413.
Huang, Q., & Xiao, Y. (2015). Geographic situational awareness: Mining tweets for disaster preparedness, emergency response, impact, and recovery. *ISPRS International Journal of Geo-Information, 4*(3), 1549–1568.
Huang, Q., & Xu, C. (2014). A data-driven framework for archiving and exploring social media data. *Annals of GIS, 20*(4), 265–277.
Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management, 6*(3-4), 248–260.
Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. *ACM Computing Surveys (CSUR), 47*(4), 67.
Imran, M., Castillo, C., Lucas, J., Meier, P., & Vieweg, S. (2014). AIDR: Artificial intelligence for disasterresponse. In *Proceedings of the 23rd international conference on world wide web* (pp. 159–162).
Imran, M., Mitra, P., & Castillo, C. (2016). Twitter as a lifeline: Human-annotated twitter corpora for NLP of crisis-related messages. arXiv preprint arXiv:1605.05894.
Imran, M., Mitra, P., & Srivastava, J. (2016). Cross-language domain adaptation for classifying crisis-related short messages. arXiv preprint arXiv:1602.05388.
Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. arXiv preprint arXiv:*1404.2188*.
Kedzie, C., McKeown, K., & Diaz, F. (2015). Predicting salient updates for disaster summarization. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (Vol. 1, pp. 1608-1617).
Kent, J. D., & Capello Jr., H. T. (2013). Spatial patterns and demographic indicators of effective social media content during theHorsethief canyon fire of 2012. *Cartography and Geographic Information Science, 40*(2), 78–89.

Kim, S. Y., Jang, Y., Mellema, A., Ebert, D. S., & Collinss, T. (2007). Visual analytics on mobile devices for emergency response. In *2007 IEEE Symposium on Visual Analytics Science and Technology* (pp. 35-42). IEEE.

Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

Kong, L., et al. (2014). *"a dependency parser for tweets." proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science Advances, 2*(3), e1500779.

Kumar, S., Barbier, G., Abbasi, M. A., & Liu, H. (2011). Tweettracker: An analysis tool for humanitarian and disaster relief. In *Fifth international AAAI conference on weblogs and social media.* (Vol. 5, No. 1, pp. 661–662).

Landwehr, P. M., Wei, W., Kowalchuck, M., & Carley, K. M. (2016). Using tweets to support disaster planning, warning and response. *Safety Science, 90*, 33–47.

Mandel, B., Culotta, A., Boulahanis, J., Stark, D., Lewis, B., & Rodrigue, J. (2012). A demographic analysis of online sentiment during hurricane irene. In *Proceedings of the second workshop on language in social media* (pp. 27–36). Association for Computational Linguistics.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Nguyen, D. T., Al Mannai, K. A., Joty, S., Sajjad, H., Imran, M., & Mitra, P. (2017). Robust classification of crisis-related data on social networks using convolutional neural networks. In *Eleventh international AAAI conference on web and social media*.

Nguyen, M. T., Kitamoto, A., & Nguyen, T. T. (2015). Tsum4act: A framework for retrieving and summarizing actionable tweets during a disaster for reaction. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 64–75). Springer.

Pantel, P., & Lin, D. (2002). Discovering word senses from text. In *Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 613–619).

Pavalanathan, U., & Eisenstein, J. (2015). Confounds and consequences in geotagged Twitter data. *arXiv preprint arXiv:1506.02275*.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543).

Rudra, K., Ghosh, S., Ganguly, N., Goyal, P., & Ghosh, S. (2015). Extracting situational information from microblogs during disaster events: A classification-summarization approach. In *Proceedings of the 24th ACM international on conference on information and knowledge management* (pp. 583–592). ACM.

Rudra, K., Goyal, P., Ganguly, N., Mitra, P., & Imran, M. (2018). Identifying sub-events and summarizingdisaster-related information from microblogs. In *The 41st International ACM SIGIR Conference on Research &Development in Information Retrieval* (pp. 265–274).

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th international conference on world wide web* (pp. 851–860).

Savage, M., Devine, F., Cunningham, N., Taylor, M., Li, Y., Hjellbrekke, J., et al. (2013). A new model of social class? Findings from the BBC's great British class survey experiment. *Sociology, 47*(2), 219–250.

Spinsanti, L., & Ostermann, F. (2013). Automated geographic context analysis for volunteered information. *Applied Geography, 43*, 36–44.

Starbird, K., Spiro, E., Edwards, I., Zhou, K., Maddock, J., & Narasimhan, S. (2016). Could this be true?: I think so! Expressed uncertainty in online rumoring. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 360–371). ACM.

Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., et al. (2011). Natural language processing to the rescue? Extracting" situational awareness" tweets during mass emergency. In *Fifth international AAAI conference on weblogs and social media*.

Vieweg, S., Castillo, C., & Imran, M. (2014). Integrating social media communications into the rapid assessment of sudden onset disasters. In *International conference on social informatics* (pp. 444–461). Springer.

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: What twitter may contribute to situational awareness. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1079–1088). ACM.

Vieweg, S. E. (2012). *Situational awareness in mass emergency: A behavioral and linguistic analysis of microblogged communications (doctoral dissertation, University of Colorado at Boulder)*.

Wang, Z., & Ye, X. (2019). Space, time, and situational awareness in natural hazards: A case study of hurricane Sandy with social media data. *Cartography and Geographic Information Science, 46*(4), 334–346.

Wang, Z., Ye, X., & Tsou, M. H. (2016). Spatial, temporal, and content analysis of twitter for wildfire hazards. *Natural Hazards, 83*(1), 523–540.

Yu, M., Huang, Q., Qin, H., Scheele, C., & Yang, C. (2019). Deep learning for real-time social media text classification for situation awareness–using hurricanes Sandy, Harvey, and Irma as case studies. *International Journal of Digital Earth, 12*(11), 1230–1247.

Yuan, F., Li, M., Liu, R., Zhai, W., & Qi, B. (2021). Social media for enhanced understanding of disaster resilience during hurricane Florence. *International Journal of Information Management, 57*, 102289.

Yuan, F., Li, M., Zhai, W., Qi, B., & Liu, R. (2020). Social media based demographics analysis for understanding disaster response disparity. In *Construction research congress 2020: Computer applications* (pp. 1020–1028). American Society of Civil Engineers.

Zhai, W., & Peng, Z. R. (2020). Damage assessment using google street view: Evidence from hurricane michael in Mexico Beach, Florida. *Applied Geography, 123*, 102252.

Zhai, W., Peng, Z. R., & Yuan, F. (2020). Examine the effects of neighborhood equity on disaster situational awareness: Harness machine learning and geotagged twitter data. *International Journal of Disaster Risk Reduction, 48*, 101611.

Zhai, W., Peng, Z. R., & Bai, X. (2021). Prototypical Resilience Projects for Post-disaster Recovery Planning: From Theory to Action. *Journal of Planning Education and Research*, 0739456X211048928.

Zhou, B., Zou, L., Mostafavi, A., Lin, B., Yang, M., Gharaibeh, N., & Mandal, D. (2022). VictimFinder: Harvesting rescue requests in disaster response from social media with BERT. *Computers, Environment and Urban Systems, 95*, 101824.

Zou, L., Lam, N. S., Cai, H., & Qiang, Y. (2018). Mining twitter data for improved understanding of disaster resilience. *Annals of the American Association of Geographers, 108*(5), 1422–1441.

Publisher's Note