A Multi-label Classification Approach to Identify Hurricane-induced Infrastructure Disruptions Using Social Media Data

Kamol Chandra Roy, Samiul Hasan*

Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, Florida

&

Pallab Mozumder

Department of Earth & Environment and Department of Economics, Florida International University, Miami, Florida.

Abstract: *Rapid identification of infrastructure disruptions during a disaster plays an important role in restoration and recovery operations. Due to the limitations of using physical sensing technologies, such as the requirement to cover a large area in a short period of time, studies have investigated the potential of social sensing for damage/disruption assessment following a disaster. However, previous studies focused on identifying whether a social media post is damage related or not. Hence, advanced methods are needed to infer actual infrastructure disruptions and their locations from such data. In this paper, we present a multi-label classification approach to identify the co-occurrence of multiple types of infrastructure disruptions considering the sentiment towards a disruption—whether a post is reporting an actual disruption (negative), or a disruption in general (neutral), or not affected by a disruption (positive). In addition, we propose a dynamic mapping framework for visualizing infrastructure disruptions. We use a geo-parsing method that extracts location from the texts of a social media post. We test the proposed approach using Twitter data collected during hurricanes Irma and Michael. The proposed multi-label classification approach performs better than a baseline method (using simple keyword search and sentiment analysis). We also find that disruption related tweets, based on specific keywords, do not necessarily indicate an actual disruption. Many tweets represent general conversations, concerns about a potential disruption, and positive emotion for not being affected by any disruption. In addition, a dynamic disruption map has potential in showing county and point/coordinate level disruptions. Identifying disruption types and their locations are vital for disaster recovery, response, and relief actions. By inferring the co-occurrence of multiple disruptions, the proposed approach may help coordinate among infrastructure service providers and disaster management organizations.*

1 INTRODUCTION

Cities and communities all over the world largely depend on critical infrastructure systems/services such as electrical power, water distribution, communication services, and transportation networks. The growing interconnectedness and interdependency among these systems have changed the organizational and operational factors and increased the vulnerability in the face of unwanted disruptions. These systems provide critical services to a large population, and thus when disrupted they affect our quality of life, local and regional economy, and the overall community well-being. The need to quickly identify disaster-induced infrastructure disruptions is growing because of the increasing number of natural disasters such as hurricane Michael, Irma, Harvey, and Florence and their enormous impacts to affected communities.

For instance, hurricane Irma caused a substantial number of power outages in addition to transportation, communication, drinking water, and wastewater related disruptions. More than six million customers faced power outages during Irma. Storm related high winds and sustained storm surges cost approximately 3,300 megawatts of power generation (NERC, 2018). Around 27.4% of cell phone towers in Florida were damaged due to hurricane Irma as reported by the Federal Communications Commission (FCC) (FCC, 2017). Irma caused flooding to several areas throughout Florida, forcing health officials to issue unsafe

[This is the accepted version of the following article: Roy, KC, Hasan, S, Mozumder, P. A multilabel classification approach to identify hurricane-induced infrastructure disruptions using social media data. Comput Aided Civ Inf. 2020; 35: 1387– 1402, which has been published in final form at https://doi.org/10.1111/mice.12573. This article may be used for non-commercial purposes in accordance with the Wiley Self-Archiving Policy http://www.wileyauthors.com/self-archiving].

drinking water and boiling water notices ("The Effect of Hurricane Irma on Water Supply," 2017; "Unsafe Drinking Water After Hurricane Irma," 2019). Moreover, dozens of sewage systems were overflowed after the power went out, which further exacerbated the drinking water condition ("Unsafe Drinking Water After Hurricane Irma," 2019).

To ensure efficient operation and maintenance, it is important to gather real-time information about the performance and integrity of engineering systems. This is typically performed through a computational monitoring process that involves observation of a system, analysis of the obtained data, and prediction of future performance (Dyskin et al., 2018). During a disaster, due to disruptions, the performance of critical infrastructures degrades rapidly—leading to cascading failures (Du, Cai, Sun, & Minsker, 2017; Kadri, Birregah, & Châtelet, 2014). In such extreme events, computational monitoring is required to assess the quickly changing condition of infrastructure systems and warn about an approaching failure or even a catastrophic event.

For effective disaster response and recovery operations, coordinated actions are required from the responsible organizations. Disruptions to infrastructure systems such as electricity/power, cell phone, internet, water, waste water, and other systems significantly affect the recovery time of a community (Mitsova, Escaleras, Sapat, Esnard, & Lamadrid, 2019). Due to the interdependence among infrastructure systems, multiple types of disruptions (e.g., power outages, internet/cell phones, water service) are likely to co-occur during a disaster. To ensure an expedited recovery of the systems, rapid identification of the co-occurrence of disruptions is necessary so that coordinated actions can be taken by multiple agencies.

Although infrastructure performance data can be collected through physical sensing technologies such as drones, satellite, UAV etc. (Jongman, Wagemaker, Romero, & de Perez, 2015; NERC, 2018), they might not be feasible due to the rapidly evolving nature of a disaster spreading over a large area (Fan, Mostafavi, Gupta, & Zhang, 2018). Social media users have been used as sensors during disasters and several studies have found its potential for understanding situational awareness (Huang & Xiao, 2015; Kryvasheyeu Y, Chen H, Moro E, Van Hentenryck P, 2015). Previous studies investigated social media sensing for damage assessment (Kryvasheyeu et al., 2016), recovery (Z. Li, Wang, Emrich, & Guo, 2018), and inundation mapping (Jongman et al., 2015). Studies have also proposed query based approaches to identify topics related to critical infrastructure disruptions (Fan & Mostafavi, 2019; Fan et al., 2018). However, these studies have not considered the co-occurrences of the types and extent of infrastructure disruptions.

During an unfolding disaster, people from the affected regions share their opinions, views, concerns, and eye witnessed events in social media platforms. Such user-generated content can provide valuable information to extract disruption-related information. However, during a disaster, emergency managers face challenges to monitor the massive volume of social media posts in real time (Oyeniyi, 2017). Thus, to get actionable information, it is important to identify whether a post indicates an actual disruption or simply expresses user views or opinions about a disruption. Recent studies have mainly focused on identifying whether a particular social media post is damage related or not (Yuan & Liu, 2018, 2019). However, since infrastructure systems are more interconnected, co-occurrences of disruptions in multiple infrastructures are more likely.

In this study, we develop a multi-label classification approach to identify the co-occurrence and extent of multiple types of infrastructure disruptions. We also present a framework to create dynamic disruption maps and case studies showing the developed approach based on Twitter data collected during hurricanes Irma and Michael. This study has the following contributions:

- We consider multiple types of infrastructure disruptions (e.g., power, transportation, water, wastewater, and other disruption) and their co-occurrences in a social media post, instead of considering a simple binary classification problem (i.e., whether a post is disruption related or not).
- To identify if a disruption related post reflects an actual disruption, we associate sentiments with disruption status—whether a post is reporting an actual disruption (*negative*), or disruption in general (*neutral*), or not affected by a disruption (*positive*).
- We propose a dynamic mapping framework for visualizing infrastructure disruptions by adopting a geo-parsing method that extracts location from tweet texts.

Instead of identifying disruption types and status in a single label, we identify disruption types and disruption status (through sentiment) separately. We adopt this approach since the neutral and positive sentiment about a disruption may also provide valuable information on the level of situational awareness about disruptions during a disaster.

2 LITERATURE REVIEW

According to the Department of Homeland Security, there are 16 critical infrastructure sectors (Homeland Security, 2019). Among these sectors, energy, communication, transportation, water/wastewater systems are the most vulnerable ones to a natural disaster. It is important to identify, characterize, and model infrastructure disruptions for a faster restoration and recovery operation (Fang & Sansavini, 2019; Sriram, Ulak, Ozguven, & Arghandeh, 2019). Studies have focused on the recovery plans and damages due to extreme weather events (Bryson, Millar, Joseph, & Mobolurin, 2002; Fang & Sansavini, 2019; Lambert & Patterson, 2002; Rosenzweig & Solecki, 2014; Sörensen, Webster, & Roggman, 2002). Several studies have proposed approaches to assess the reliability, resilience, vulnerability and failure process of power, transportation, and water supply networks individually (Barabási & Albert, 1999; Buldyrev, Parshani, Paul, Stanley, & Havlin, 2010; Jenelius & Mattsson, 2012; Ouyang & Fang, 2017; Pietrucha-Urbanik & Tchórzewska-Cieślak, 2018; Sumalee & Kurauchi, 2006; Ulak, Kocatepe, Sriram, Ozguven, & Arghandeh, 2018). However, these critical infrastructures are inter-connected and inter-dependent (Alinizzi, Chen, Labi, & Kandil, 2018; Homeland Security, 2019; Martani, Jin, Soga, & Scholtes, 2016). Considering the increased connectedness and interdependencies among infrastructure systems, studies have proposed a holistic approach to assess the resilience to disruptions (Hasan & Foliente, 2015; Lu et al., 2018; Pant, Thacker, Hall, Alderson, & Barr, 2018; Rinaldi, Peerenboom, & Kelly, 2001; Sriram et al., 2019). However, most of these studies are based on synthetic data or post-event data. Thus, they are not suitable for real-time decision-making.

Recently, real-time condition monitoring is becoming very popular in manufacturing, maintenance, and usage of many engineering systems (Dyskin et al., 2018) and civil engineering infrastructures (Chang, Flatau, & Liu, 2003). Computational models have been developed for estimating the properties of constructional materials (Rafiei, Khushefati, Demirboga, & Adeli, 2017), detecting damages to building structures (Rafiei & Adeli, 2017b, 2018a), predicting construction costs (Rafiei & Adeli, 2018b) etc. Another potential approach for monitoring infrastructures is by collecting real-time data using smartphones, leading to citizen-centered and scalable monitoring systems in a disaster context (Alavi & Buttlar, 2019).

During an ongoing disaster and post-disaster period, it is important to collect disruption data to take necessary actions as fast as possible. Due to the intensity and spread of a disaster, physical sensing techniques such as satellite, UAV (Unmanned Aerial Vehicle) etc. (Jongman et al., 2015; NERC, 2018) are not suitable. For example, after hurricane Irma, unmanned aerial drones, amphibious vehicles, airboats are used to perform damage assessment on inaccessible transmission and distribution lines (NERC, 2018). A crowd-sourcing app that allows damage reporting might not be useful because of fewer participants. On the other hand, the ubiquitous use of social media on GPS enabled smartphone device, allows us to collect large-scale user generated data containing live and in situ events during a disaster (Middleton, Middleton, & Modafferi, 2013). Studies have already used social media data for crisis mapping (Birregah et al., 2012; Gao, Barbier, & Goolsby, 2011; Middleton et al., 2013). However, real-time crisis mapping requires location information, but only around 1% to 4% of social media (e.g., Twitter) data posts are geo-tagged (Cheng, Caverlee, & Lee, 2010; C. Li & Sun, 2014; Middleton et al., 2013). Studies have proposed several location-extraction methods from content/textual data (Cheng et al., 2010; C. Li & Sun, 2014; Middleton et al., 2013). In addition, the power of social media to connect a large group of population has drawn significant attention towards using social media platforms for disaster management (Keim & Noji, 2010; Sadri, Hasan, & Ukkusuri, 2019; Tang, Zhang, Xu, & Vo, 2015). Studies have analyzed social media data for understanding human mobility and resilience during a disaster (Roy, Cebrian, & Hasan, 2019; Wang & Taylor, 2014). Kryvasheyeu et al. proposed that social media users can be considered as early warning sensors in detecting and locating disasters (Kryvasheyeu Y, Chen H, Moro E, Van Hentenryck P, 2015). Studies have also explored social media data to understand evacuation behavior (Fry & Binner, 2015; Martín, Li, & Cutter, 2017) and damage assessment (Deng, Liu, Zhang, Deng, & Ma, 2016; Guan & Chen, 2014; Kryvasheyeu et al., 2016; Yuan & Liu, 2018).

Damage assessment plays a vital role in resource allocation and coordination in disaster response and recovery efforts. Previous studies found that affected people provide damage related situational updates in social media (Deng et al., 2016; Guan & Chen, 2014; Kryvasheyeu et al., 2016; Yuan & Liu, 2018). However, these studies do not consider the types of disruptions and are mainly suitable for post-disaster overall damage assessment. Most of these studies adopted simpler indicators of damage assessment such as frequency of disaster related tweets (based on keywords such as 'sandy', 'hurricane sandy', 'damage'). The limitation of using pre-defined keywords is that a large number of such tweets/texts may not contain any damage related information. Some studies (Kotsiantis, Zaharakis, & Pintelas, 2007) adopted supervised machine learning based classification approaches to resolve this limitation. These studies (Cresci, Cimino, Dell'Orletta, & Tesconi, 2015; Yuan & Liu, 2019) adopted support vector machine, naïve Bayes, decision tree classification algorithms to analyze damage related social media posts. However, these studies considered damage identification as a binary (damage related or not) classification problem, which may include posts that are not reporting an actual damage/disruption. In addition, deep learning models were used for image and text data (Mouzannar, Rizk, & Awad, 2018; D. T. Nguyen, Ofli, Imran, & Mitra, 2017). Image data are limited, computationally expensive, and cannot report disruptions in functionality such as power outage, communication disruptions etc.

The most relevant studies towards identifying an infrastructure disruption using social media posts are proposed by Fan et al. (Fan & Mostafavi, 2019; Fan et al., 2018). The first study (Fan et al., 2018) has focused on summarizing the overall topics during a disaster given some predefined keywords, not suitable to identify disruptions from real-time data. In the second study (Fan & Mostafavi, 2019), the authors have developed a graph-based method to identify situational information related to infrastructure disruptions by detecting time slices based on a threshold number of tweets. They compute content similarity within the detected time slices to get credible information. Some limitations of this approach include: it depends on keyword based filtering, which can miss out important information if appropriate keywords are not chosen; it requires the whole dataset as an input, which is not suitable for a real-time prediction; it considers the content posted only on the burst timeframe that might miss out some actual disruption related posts. Moreover, this study does not consider that a single post may have information about multiple types of disruptions and cannot distinguish if a particular post is reporting an actual disruption or not.

In summary, to the best of our knowledge, currently no study exists to identify the co-occurrence of multiple types of infrastructure disruptions using social media data. For this task, a multi-label classification approach (Sorower, 2010) identifying multiple labels from a single input, can be useful.

In this study, we use a multi-label text classification approach to identify multiple disruption types and their status using social media data. To develop our multi-label disruption classification approach, we use eight well-known models on text classification. We present two case studies to identify disruptions using Twitter data from hurricanes Irma and Michael. Finally, we visualize the spatio-temporal dynamics of infrastructure disruptions in a map of the affected regions.

3 DATA PREPARATION

In this study, we use Twitter data collected during hurricanes Irma and Michael for creating a dynamic disruption map of critical infrastructure disruptions. We use two different methods (Twitter streaming API and rest API) for data collection. A brief description of the data is provided in Table 1.

Table 1 Data Description

Hurricane Name	Regions (USA)	No. of Tweets	No. of Users
Irma (Streaming API)	FL, GA, SC	1,810,000	248,763
Irma (Rest API)		2,478,383	16,399
Michael (Streaming API)	FL, GA, SC, NC	3,534,524	1,289,204

Using the streaming API, we collected about 1.81 million tweets posted by 248,763 users between September 5, 2017 and September 14, 2017 during hurricane Irma. We collected the tweets using a bounding box covering Florida, Georgia, and South Carolina. To collect data for more time span and to fill some missing values contained in the steaming API data, we used Twitter's rest API to gather user-specific historical data. Twitter's rest API allows collecting the most recent 3,200 tweets of a given user. We collected user-specific data for 19,000 users, who were active for at least three days within the streaming data collection period. Similarly, we collected data for hurricane Michael using a bounding box covering Florida, Georgia, South Carolina, and North Carolina, containing 3.53 million tweets posted by 1.29 million users covering from October 8, 2018 to October 18, 2018.

To create an annotated disruption dataset, we manually labeled 1,127 tweets from hurricane Irma and 338 tweets (for testing purpose only) from hurricane Michael. The tweets were labeled by 5 human annotators. To ensure that we retrieve the right labels of the disruption types and sentiments, we only considered the labels when all 5 annotators agreed on it. Each tweet can have one or more labels out of the ten possible labels including: *not hurricane related, power/electricity disruption, communication disruption, road/transportation disruption, drinking water disruption, wastewater related disruption, other disruption, positive, negative, and neutral*. The first label indicates whether a tweet is hurricane related or not. The next five labels indicate five types of infrastructure disruptions. The label, *other disruption,* indicates a disruption that does not fall into the five types of infrastructure disruptions considered here. The last three labels indicate the possible sentiment towards a disruption. We give below three examples of disruption related tweets:

- This tweet -"*Update I'm the only community in my area with power I feel really lucky right now but I hope everyone else is safe*"- mentions about *power/electricity disruption* but in a *positive* way. We would label such a tweet as *(power/electricity disruption, positive).*
- This tweet- "*we are in Clermont on Lake Minnehaha. We have no cable or power & cell service is spotty. When will be the worst here*"- mentions about both *power/electricity* and *communication disruptions*. We would label this tweet as *(power/electricity disruption, communication disruption, negative).*
- This tweet -"*im trying to eat and watch as much netflix as i can just incase my power go out*"- mentions about power/electricity disruption but does not indicate an actual disruption. We would label it as *(power/electricity disruption, neutral).*

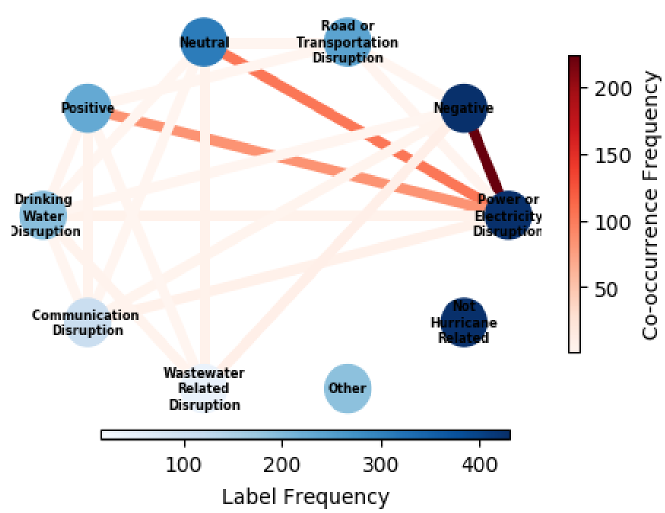


Figure 1 Distribution of Label Frequency and Label Co-occurrence Frequency

Figure 1 shows the frequencies and co-occurrences of the labels in the annotated dataset. It shows that the annotated data contain many "*not hurricane related"* tweets. Among the tweets related to different types of disruptions, power/electricity related disruptions have the highest frequency. Among the sentiment related labels, *negative* sentiment has the highest frequency. On the other hand, power/electricity disruption and negative sentiment are the most frequently co-occurred labels in the annotated dataset.

4 METHODOLOGICAL APPROACH

The methodological approach adopted in this study has three main parts. The first part takes tweet texts as input and identifies disruptions and the sentiment towards the disruption. The second part extracts the geo-location from the tweet's metadata or text. The third part visualizes the disruptions geographically in a dynamic map of disruptions. Figure 2 shows the steps and information flow among those steps. Each part of the framework is described below:

4.1 Disruption Identification

The objective of this step is to identify infrastructure disruptions and sentiments from a given text input, where more than one disruption type might be present. We use a supervised multi-label classification approach. The input texts collected from Twitter posts contain many noises, which may degrade classification performance. Therefore, we process the data before feeding it into the model. The sequential steps are shown in Figure 2 (left side).

Data Pre-processing

In this step, we discard the unnecessary tweets and remove noise from a tweet. Since a retweet (starting with RT in the texts) does not provide any new information in the dynamics of disruption, we discard retweets from the data to avoid false spike in the disruption count. To clean the tweet texts, we remove the stop words (e.g., 'a', 'an' and 'the'), short URLs, emoticons, user mentions (@Twitter user name), punctuations, and special characters (\@/#$ etc.). Finally, we tokenize (splitting texts into words) the texts and apply lemmatization (converting the words into noun, verb etc.) and stemming (converting words into root form) to the tokens.

Data Processing

In this step, we process the data for training models and predicting disruptions. In machine learning, training of a model refers to providing it with training data, which contains both inputs and correct answers, so that the algorithm can find the pattern to map the input features to the target/output features. We convert the preprocessed tokens as TF-IDF (Term Frequency-Inverse Document Frequency), which measures the importance of a word in a document of a corpus (collection of documents). The details on TF-IDF can be found in this study (Ramos & others, 2003). The TF-IDF of a term/word (w) is calculated as follows:

$$TF\text{-}IDF(w) = TF(w) \times IDF(w) \quad (1)$$

where,

$$TF(w) = \frac{\begin{array}{c}the\ number\ of\ times\ term\ w\\ appears\ in\ a\ document\end{array}}{\begin{array}{c}the\ total\ number\ of\ terms\\ in\ the\ document\end{array}}$$

$$IDF(w) = log\frac{\begin{array}{c}the\ total\ number\ of\\ documents\end{array}}{\begin{array}{c}the\ number\ of\\ documents\ with\ term\ w\end{array}}$$

We create the TF-IDF using both unigram and bigram of words. We remove the features that appear in less than 2 documents. To remove the effect of total word counts in a document, we apply $l2$ normalization (sum of the squared value of TF-IDF =1 for a document). To prevent data leakage, we calculate the TF-IDF considering the tweets available in the training dataset. The output of the model may contain multiple disruptions; thus we convert the annotated labels into multi-label formats. We represent the multi-label output as a binary/one hot encoded matrix indicating the presence of disruption type and the sentiment label. In our study, we have 10 possible labels, so, each converted label is represented as 1×10 binary matrix where the value 1 represents the presence and the value 0 represents the absence of a particular label.

Model Selection

The objective of this step is to find the best model that maps an input tweet text to the binary matrix representing one or more types of infrastructure disruptions and sentiment. In our study, we choose a multi-label classification approach for identifying disruptions and sentiments. This approach generalizes the multiclass classification, where a single input/tweet can be assigned to multiple types of disruptions. Let $L = \{\lambda_i\}$ be the set of labels containing disruption types and sentiment, where, $i = 1 \dots \dots |L|$. In our case, $|L| = 10$. The objective of our disruption identification model, h is that: given the input tweet, X the model has to predict the disruption types and sentiments, $Y \subseteq L$.

$$h: X \rightarrow Y \quad (2)$$

We apply three methods that allow using the multiclass classification models for a multi-label classification task. The first method transforms a multi-label classification into multiple binary classification problems. This method is also known as binary relevance (BR) (Sorower, 2010) that trains one binary classifier for each label independently. The equation for a binary classifier, h_{λ_i} for a label λ_i can be expressed as below:

$$h_{\lambda_i}: X \rightarrow \{\neg\lambda_i, \lambda_i\} \quad (3)$$

The BR method transforms the training data into $|L|$ datasets. The dataset D_{λ_i} for label λ_i contains all the original dataset labeled as λ_i if the original example contains λ_i, otherwise, as $\neg\lambda_i$. For an unseen sample, the combined model predicts all labels using the respective classifier. One of the disadvantages of the BR method is that it does not consider the correlation between labels.

The second method transforms the multi-label classification problem into a multi-class classification problem. This method is known as label powerset (LP) that considers each subset of L as a single label. Let, $P(L)$ be the powerset of L, which contains all possible subset of L. LP method considers each element of $P(L)$ as a single label. Now, in training LP learns one single label classifier h, where:

$$h: X \rightarrow P(L) \quad (4)$$

The LP method has advantages over the BR method, because it takes the label correlations into account. However, it requires high computation time if the size of $P(L)$ is very big and majority of the subsets have very few members. Also, the LP method tends to overfit (performs well on training data but performs poorly on test data), when the number of labeled samples of the generated subsets is low.

As the third method, we apply an ensemble technique, known as Random k-Labelsets (RAKEL) adopted from the study (Tsoumakas & Vlahavas, 2007). This method constructs an ensemble of LP classifiers, where each LP classifier is trained on a small random subset of labels. Instead of using $P(L)$, it creates k-labelset $Y \subseteq L$, where $k = |Y|$. If the set of all distinct k-labelset is L^k, then $|L^k| = \binom{|L|}{k}$. Given a user specified integer value for k and m, where, $1 \leq k \leq |L|$ and $1 \leq m \leq |L^k|$, the RAKEL algorithm iteratively constructs an ensemble of m numbers of LP classifiers. However, for $k = 1$ and $m = |k|$, RAKEL method becomes a binary classifier ensemble of BR method. On the other hand, for $k = |L|$, m becomes 1, and consequently, RAKEL method becomes a single label classifier of the LP method. Given a meaningful parameter of k $(2\ to\ |L| - 1)$, at each iteration, $i = 1 \ldots.. m,$ without replacement it randomly selects a k-labelset, Y_i from L^k and learns an LP classifier, h_i. Where,

$$h_i : X \rightarrow P(Y_i) \tag{5}$$

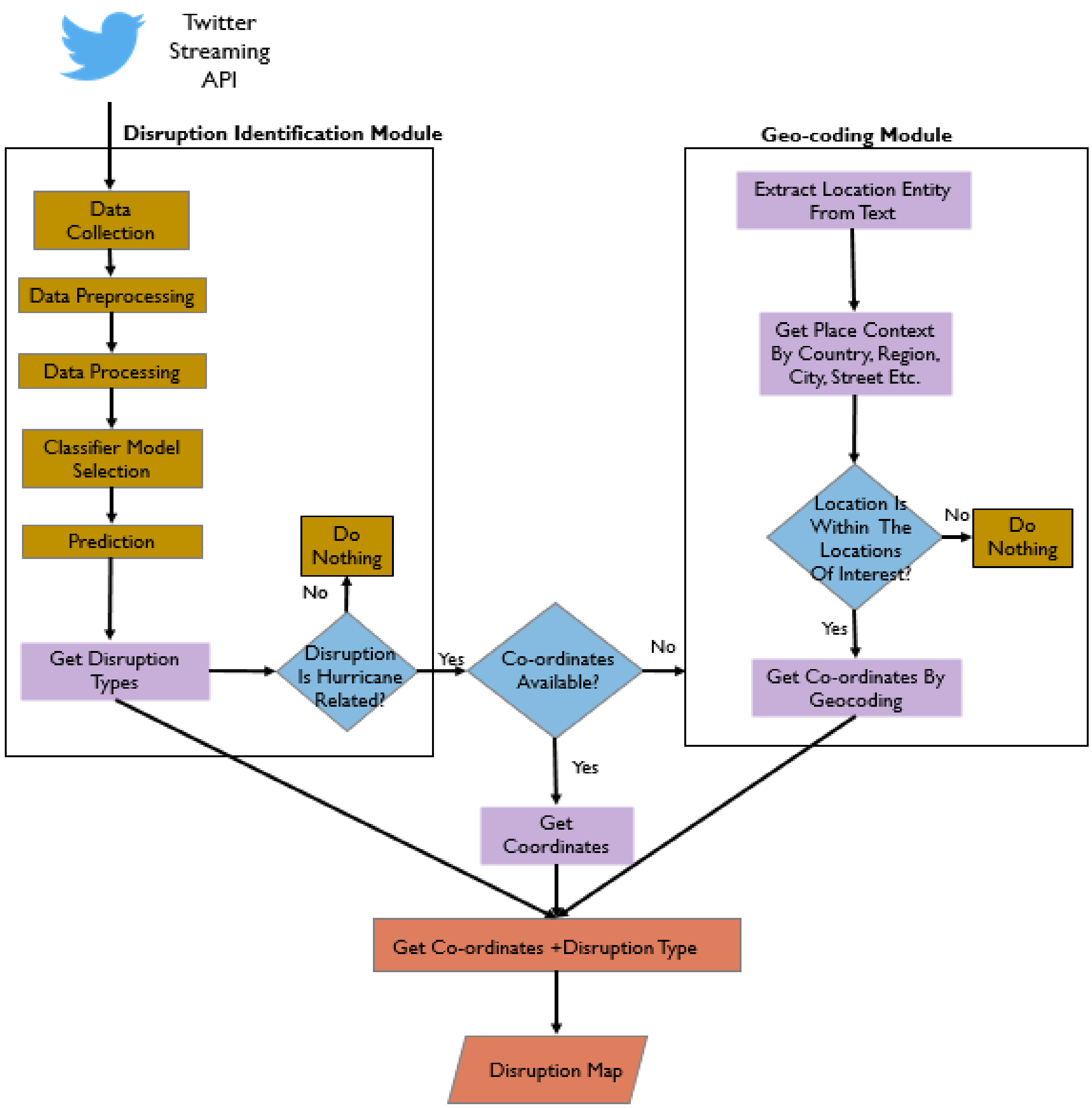


Figure 2 Methodological framework: disruption identification module (left); geo-parsing module (right); and visualization module (middle)

For a given input, the label prediction is accomplished by a voting scheme from the ensemble combination. The RAKEL method solves the overfitting problem of the LP method but loses some correlations as it considers a random subset of the labels (LP method considers all possible subsets). The full description of the RAKEL method can be found in this study (Tsoumakas & Vlahavas, 2007).

In multi-label classification, a prediction cannot be assigned as a hard right or wrong value, because a prediction containing a subset of the actual classes should be considered better than a prediction that contains none of them. Thus, traditional performance metrics (e.g., precision, recall) are not suitable for evaluating our disruption identification model. We choose the best model based on three generally used performance metrics in multi-label classification: subset accuracy, micro F1 score, and hamming loss. Here, subset accuracy and hamming loss are example-based metrics and micro F1 measure is a label-based metric. For each test sample, an example-based metric computes the difference between true and predicted class labels and then calculate the average over all test samples. Whereas, a label-based metric first computes the performance for each class label, and then calculates the average over all class labels. Assuming y as the set of true class labels, $\hat{y}$ as the predicted set of labels, L as the set of labels, y_l as the subset of y with label l, $\hat{y}_l$ the subset of $\hat{y}$ with label l, $n_{samples}$ the number of samples, the equations of these metrics are given below:

$$Subset\ Accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i \neq y_i) \tag{6}$$

$$Micro\ F_1\ Measure(y, \hat{y}) = 2 \times \frac{\frac{|y \cap \hat{y}|}{|y|} \times \frac{|y \cap \hat{y}|}{|\hat{y}|}}{\frac{|y \cap \hat{y}|}{|y|} + \frac{|y \cap \hat{y}|}{|\hat{y}|}} \tag{7}$$

$$Hamming\ Loss(y, \hat{y}) = \frac{1}{n_{samples} \times |L|} \sum_{i=0}^{n_{samples}-1} \sum_{l=0}^{L} 1(\hat{y}_l \neq y_l) \tag{8}$$

We further check the predictive performance of the model computing a confusion matrix for each label (representing disruption types and sentiment). Table 2 shows the components of a confusion matrix. The rows represent the actual labels and the columns represent the predicted labels where positive means the existence of a particular label and negative means the absence of a particular label. For a particular sample, if the actual label is negative, a negative prediction by the model is assigned as true negative and a positive prediction is assigned as false positive. Similarly, if the actual label is positive, a positive prediction is assigned as true positive and a negative prediction is assigned as false negative.

Table 2 Confusion matrix

		Predicted Label	
		Negative (0)	Positive (1)
Actual Label	Negative (0)	True Negative (TN)	False Positive (FP)
	Positive (1)	False Negative (FN)	True Positive (TP)

4.2 Disruption Location Extraction

The objective of this step is to extract the location of the disruptions that are identified by the previous step. Geo-tagged tweets provide location information either as a point type (exact latitude-longitude) or as a polygon type (bounding box). We use this location to indicate the location of a disruption either at a point resolution or a city/county resolution. However, geo-tagged tweets are only a few percentages (1% to 4%) of the total number of tweets. To address this limitation, we implement a location extraction method from tweet texts. This approach has several steps within it. Given a tweet text, the first step is to label each word (e.g., person's name, location, organization etc.), which is known as Named Entity Recognition (NER). We implement our NER model using the Natural Language Toolkit (NLTK), developed by (Bird, Klein, & Loper, 2009). The second step is to extract the location entity, words that are tagged as location, from the labeled words. In the third step, we match the extracted location with the county/city names of the affected regions. Finally, if the extracted locations are matched, we collect the coordinates using the geo-coding API provided by Google Maps. The process of location extraction is shown in Figure 2.

4.3 Dynamic Disruption Mapping

This part of the methodology enables the visualization of the locations of disruptions with disruption types in a dynamic way. We visualize the exact disruption location, only if the location has the exact co-ordinate (location type: point or latitude-longitude). We choose a time interval (t) to count the number of disruptions within a geographical boundary (e.g., county) and then visualize the disruption intensity as a geographical heat map. We did not consider disruption severity in this study. But severity can be assumed to be correlated with the frequency of disruption related tweets from a given area; the higher the frequency of disruption related tweets the higher will be the severity level of disruptions. Hence, a dynamic disruption map can provide insights about the severity of infrastructure disruptions of an area based on the frequency of a specific or all disruption related posts generated from that area.

5 RESULTS

Using Twitter data from real-world hurricanes, we present our results to identify infrastructure disruptions and visualize those disruptions in a dynamic map. To identify disruptions types and sentiment from text data, we use Binary Relevance, Label Powerset, and ensemble based multi-label classification approaches. We compare the performance of these approaches using eight existing models namely: Multinomial Naïve Bayes (MNB), Logistic Regression (LR), *K*-Nearest Neighborhood (KNN), Support Vector Machine (SVC), Random Forest (RF), Decision Tree (DT), Multilayer Perceptron (MLP), and Deep Neural Network (DNN) methods. The details of these well-known methods can be found in these studies (Binkhonain & Zhao, 2019; Khan, Baharudin, Lee, & Khan, 2010). We convert the annotated tweet text as TF-IDF and annotated label as binary matrix (multi-label format) by following the steps described in the data processing section. We use the TF-IDF as input and the binary matrix as output. For each model, we use 70% (788 tweets) of the annotated samples as training and the rest 30% (339 tweets) as test samples. We further validate our best model over 338 tweets from hurricane Michael to test model performance on the data from an unseen hurricane (i.e., for hurricane data which were never used for training the model).

We implement all the models in a personal computer using Python programming language and model parameters are selected using a grid search approach (Pedregosa et al., 2011). Moreover, we implement a baseline method that uses keyword matching and sentiment analysis to identify disruptions and sentiment characteristics, respectively. Currently no benchmark method exists that can identify the co-occurrence of multiple types of disruptions from social media posts. Since a keyword based approach has been used in similar studies (Fan & Mostafavi, 2019; Yuan & Liu, 2018), we choose to use this as a baseline method. The keywords used are listed in Table 3.

For sentiment identification, we use a pre-trained model adopted from this study (Hutto & Gilbert, 2014); this model has been trained on social media texts. We consider this combined (keyword matching and sentiment identification) approach as a baseline method to evaluate if the trained models perform better than this baseline method. Table 4 presents the performance of each model on hurricane Irma test dataset with respect to the selected performance metrics: subset accuracy, micro F1 measure, and hamming loss.

From the results, we can see that Logistic Regression classifier (LP method) has the best subset accuracy and micro F1 scores and Support Vector classifier (RAKEL method) has the best hamming loss score. The models (LR, KNN, SVC, MLP, and Deep DNN) perform better than the baseline method in all approaches (BR, LP, and Ensemble) (see Table 4). Among the three multi-label approaches, LP has the best performance; RAKEL is second; and BR method is the last in terms of the considered performance metrics. The reasons for this result are the following: (i) BR method considers the labels as mutually exclusive or the correlation between the disruptions is ignored; (ii) LP method considers the correlations between the labels/disruptions by considering all label combination; and (iii) RAKEL method falls between the BR and LP methods with respect to label correlations as it considers a random small subset of labels.

Table 3 Keywords for Identifying Disruption Related Tweets

Disruption Types	Keywords
Power/Electricity Disruption	power, electricity, outage, (power, outage), (without, power)
Communication Disruption	internet, wi-fi, cell, (no, internet), (no, network)
Transportation Disruption	road, roads, traffic, transportation, turnpike, i-4, i-95, jam, closed, (traffic, signal), (road, closed)
Drinking Water Disruption	drinkingwater, drinking_water, bottledwater, bottled_water, (drinking, water), (bottled, water)
Wastewater Related Disruption	wastewater, waste_water, drainage, drainagewater, (waste, water), (drainage, water)

To select the best model, we further check the confusion matrix and choose Logistic Regression (LP method) classifier. Figure 3 shows the confusion matrix for the LR (LP) on the test samples from hurricane Irma. The selected best model (LR-LP) shows 74.93% increase (0.351 to 0.614) in subset accuracy, 30.73% increase (0.550 to 0.719) in micro F1 measure, and 44.65% decrease (0.159 to 0.088) in hamming loss compared to the baseline method.

We also check the performance of our best model (LR-LP) for disruption and sentiment identification separately. We validate for hurricanes Irma and Michael, using 339 test data from hurricane Irma and 338 test data from hurricane Michael. Table 5 shows the performance on disruption identification.

Table 4 Model Performance Values (Accuracy, Micro F1-measure, Hamming-loss) (A higher score of subset accuracy or micro F1 measure indicates better performance and a lower score of hamming loss indicates better performance)

Model Name	Binary Relevance (BR)	Label Power set (LP)	Ensemble (RAKEL)
Baseline (keyword search + sentiment)	*0.351, 0.55, 0.159*		
Multinomial Naïve Bayes (MNB)	0.218, 0.519, 0.145	0.472, 0.615, 0.14	0.268, 0.527, 0.151
Logistic Regression (LR)	0.463, 0.709, 0.090	**0.614**, **0.719**, 0.092	0.525, 0.702, 0.094
K-nearest Neighborhood (KNN)	0.490, 0.613, 0.130	0.525, 0.612, 0.125	0.510, 0.598, 0.126
Support Vector Classifier (SVC)	0.472, 0.707, 0.089	0.608, 0.699, 0.096	0.519, 0.709, **0.088**
Random Forest (RF)	0.124, 0.471, 0.170	0.54, 0.635, 0.116	0.357, 0.588, 0.109
Decision Tree (DT)	0.292, 0.628, 0.129	0.522, 0.634, 0.119	0.366, 0.617, 0.124
Multilayer Perceptron (MLP)	0.440, 0.662, 0.099	0.540, 0.615, 0.119	0.507, 0.635, 0.11
Deep Neural Network (DNN)	0.466, 0.342, 0.138	0.569, 0.684, 0.103	-

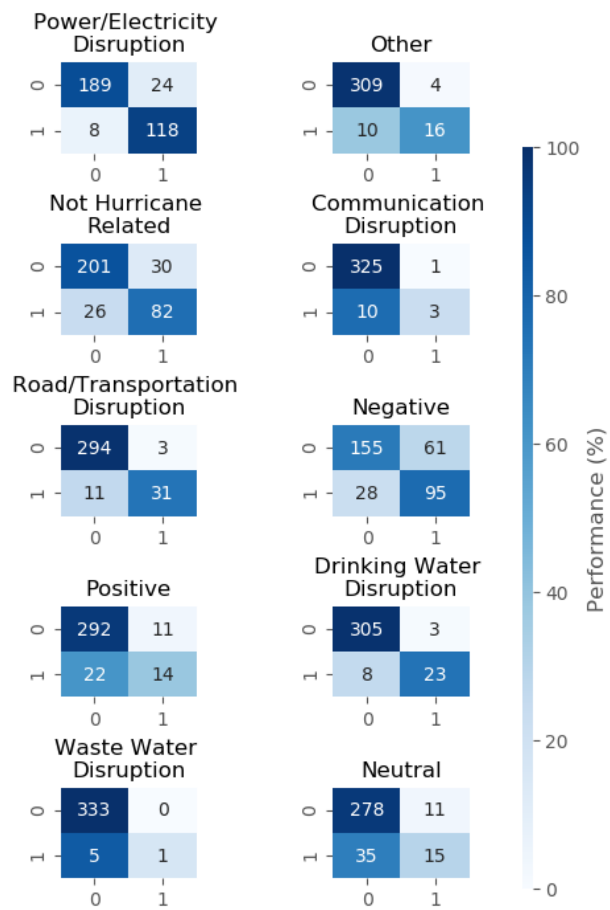


Figure 3 Confusion Matrix (In each panel, the x axis represents the predicted label and the y axis represents the actual label in the test set of hurricane Irma. For a particular label, the value 1 means the presence of this label whereas 0 means the absence of the label. The value within a cell represents the number of times a predicted label matched or mismatched with the actual label)

Except hamming loss for hurricane Michael, our model performed better for both hurricanes with respect to accuracy,

Table 5: Performance Comparison of disruption identification

	Baseline	Model (LR-LP)
Hurricane	Accuracy, Micro F1-measure, Hamming-loss	
Irma	0.351, 0.55, 0.159	0.614, 0.719, 0.092
Michael	0.476, 0.656, 0.115	0.515, 0.658, 0.119

Table 6: Performance comparison of sentiment model

	Baseline	Model (LR-LP)
Hurricane	Accuracy, Micro F1-measure, Hamming-loss	
Irma	0.383, 0.368 0.311	0.673, 0.596, 0.165
Michael	0.571, 0.501, 0.226	0.609, 0.656, 0.175

micro F1, and hamming loss. The baseline method performed better in hurricane Michael test set than the Irma test data set. On the other hand, LR-LP model performed better in Irma data than the Michael data since the model is trained on Irma dataset.

Table 6 shows the performance of LR-LP model against the baseline sentiment model (adopted from (Hutto & Gilbert, 2014)). The (LR-LP) model performed better than the baseline for both Irma and Michael datasets. The baseline method also performed better for Michael data than Irma data for sentiment classification. In summary, our developed model (LR-LP) performed better than the baseline for both hurricanes Irma (hurricane data used to train the model) and Michael (unseen hurricane data representing a future hurricane).

To understand the features that help to correctly identify a disruption, we analyze the training samples that our model correctly predicted (i.e., true positive samples in Table 2). For each disruption type, we rank the words based on their average TF-IDF score. A higher score represents more importance of a word for a disruption type. Figure 4 shows the TF-IDF scores of the top ten words of each disruption type (shown as horizontal bars) and the TF-IDF scores of the same words calculated over all disruption types in the training set (shown as color intensity). We can see that overall words such as 'power', 'water', 'wifi', 'internet', 'traffic', 'drainage' etc. have higher TF-IDF scores (see the color intensity of the corresponding bars in Figure 4). It means that these words are highly important in the overall classification performance. On the other hand, 'power', 'cell', 'stop', 'water', 'drainage', 'close' are the highest ranked words for power/electricity disruption, communication disruption, road/transportation disruption, drinking water disruption, waste water related disruption, and other disruption, respectively. Some words (e.g., 'power', 'water', 'cell') are present in multiple disruption types, indicating that these words would help identify the co-occurrence of multiple disruption types. For example, the presence of 'cell' and 'signal' in the top 3 words of power/electricity and communication disruptions indicates the co-occurrence of these two types of disruptions. Regarding sentiment features, the word 'power' is common in all the three sentiments. The differences among the words present in these three sentiment classes are: (i) the negative (actual disruption) contains the words that are mostly present in the disruption types, (ii) the positive sentiment contains slang words such as 'hell yeah', 'yeah', 'ac loll', (iii) neutral sentiment contains situation and forecast related words such as 'update', 'best update', 'situation', 'chance wont', 'good chance' (see Figure 4).

6 CASE STUDIES: HURRICANES IRMA AND MICHAEL

In this section, we present two case studies of our proposed approach, one for hurricane Irma and another for hurricane Michael. Our best model (LR-LP) predicts the disruption types and status over the input data described in Table 1. As shown in Figure 2, for a geotagged tweet, we obtain the disruption location from the tweet geo-location information. Otherwise, we extract the location from the tweet texts using the geocoding module. We match the extracted location with the city/county of a state and then obtain the coordinate using Google Maps API.

Finally, we plot the disruption types and status in a disruption map. To understand the hurricane context, we also

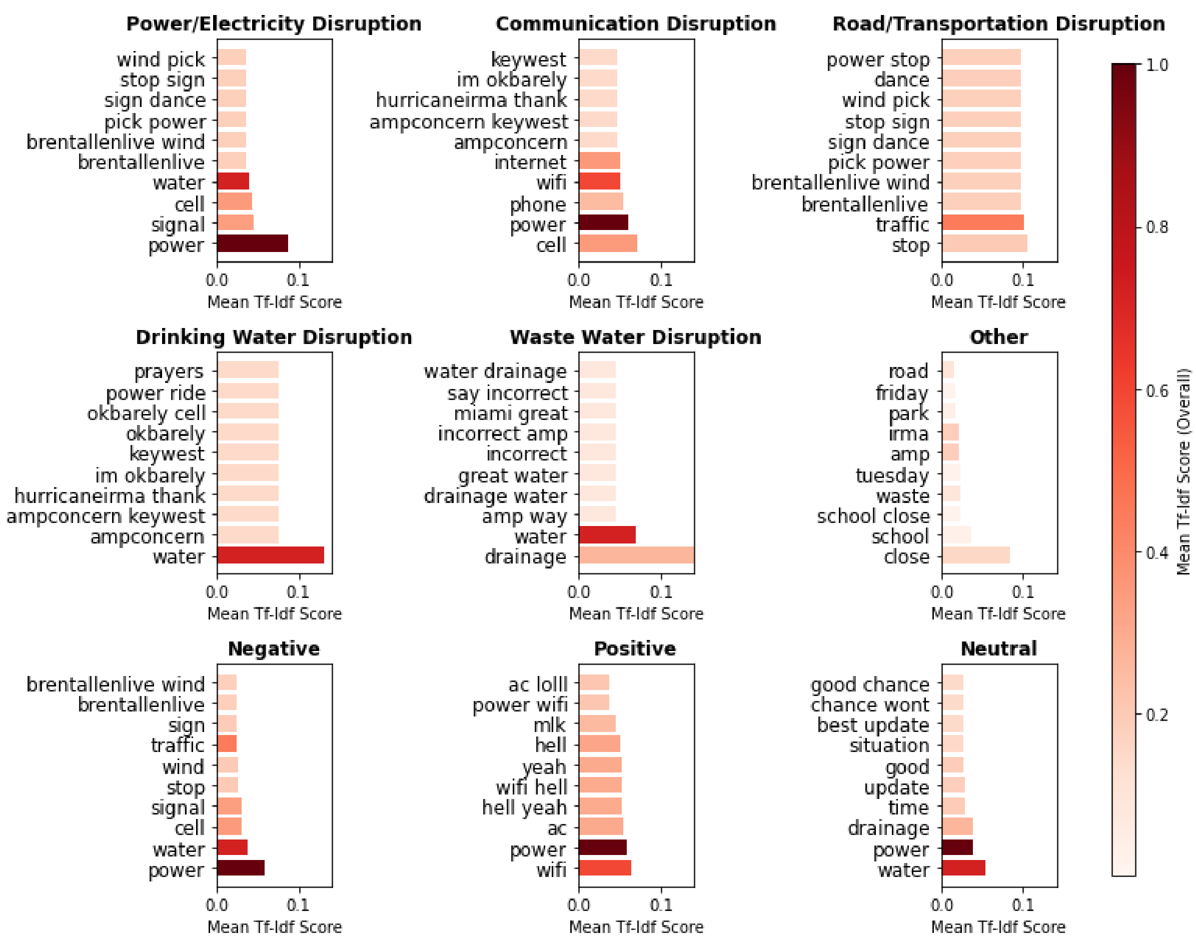
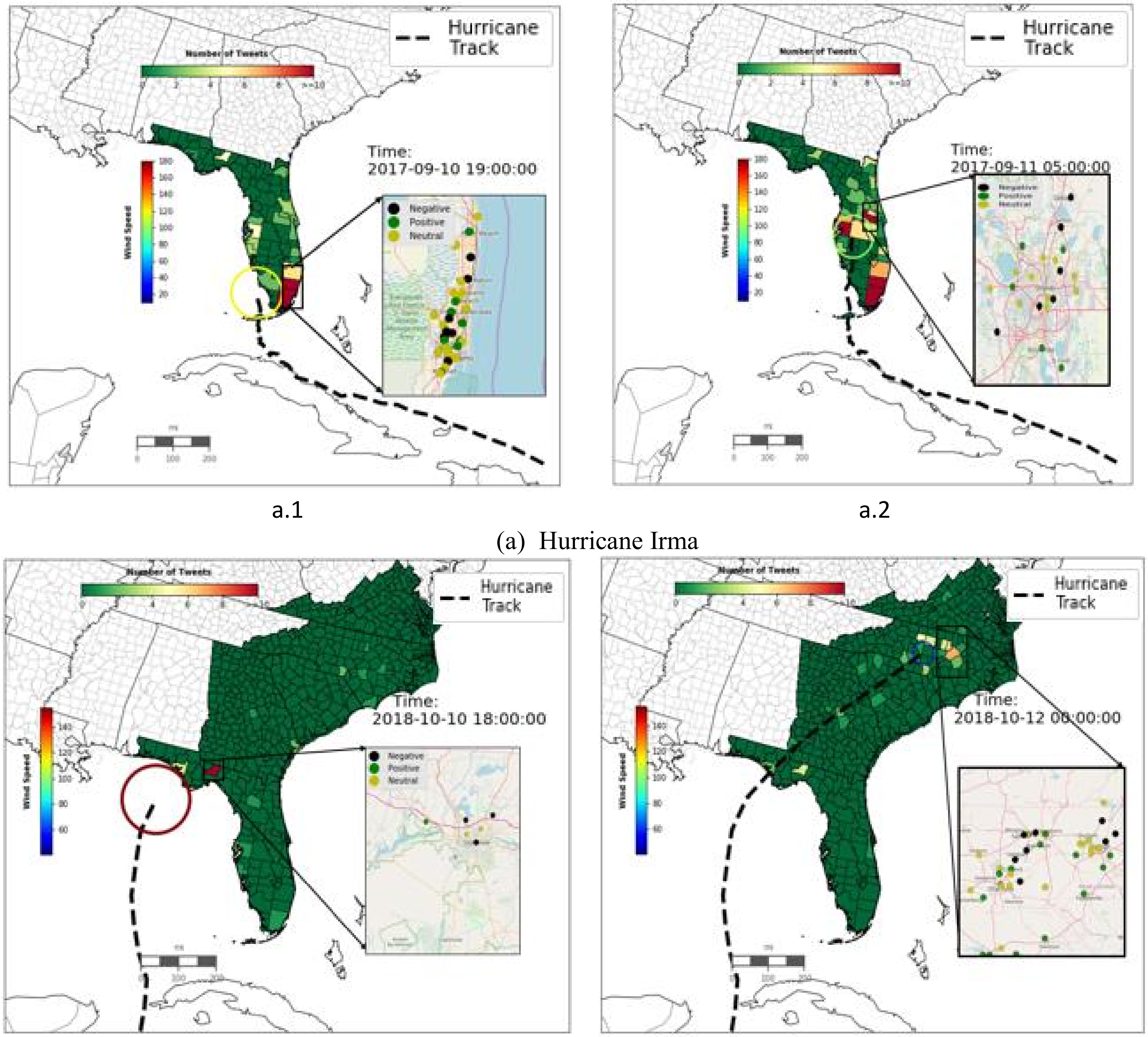


Figure 4 Important features for different disruption types. The X axis shows the mean TF-IDF score (calculated over individual disruption type) and the Y axis shows the words/features. The color of the bar indicates the mean TF-IDF score (calculated over all disruption types). The calculated scores and important features are based on the training dataset.



a.1 a.2

(a) Hurricane Irma

b.1 b.2

(b) Hurricane Michael

Figure 5 Dynamic Disruption Map for Power/Electricity Disruption: (a) Hurricane Irma, (b) Hurricane Michael.

present the hurricane track and wind speed data collected from the National Hurricane Center (NOAA, 2019). Two snapshots of the power/electricity disruption map from each hurricane are shown in Figure 5 (5a.1 and 5a.2 for hurricane Irma, 5b.1 and 5b.2 for hurricane Michael). We use a 3-hour time-interval for aggregating the tweets to create the county-level disruption heat map. The inset plot shows the locations of power/electricity disruptions. We show the location of hurricane center (shown as a circle at the beginning of the hurricane track line), wind speed (through the color of the circle), and disruption related tweets (geographic heat map) which will be updated dynamically as we receive data from Twitter stream. Figure 5a.1 shows a snapshot of Irma at around 7 PM on Sept. 10, 2017. It shows that majority of the power/electricity related posts were generated from Miami-Dade, Broward, Palm-Beach counties when Irma's center was near Collier county with a wind speed of around 120 mph. However, not all the posts are about the actual power outage incident (disruptions are represented by black circles in the inset plot of Figure 5a.1), and a substantial number of these posts are expressing concerns about power outage or expressing that they still have power. The second snapshot (Figure 5a.2) shows that when the center of Irma was near Tampa, most of the disruption related tweets were posted from Orlando, Tampa, and Miami-Dade counties. A dynamic disruption map of Michael shows similar results. On October

10, 2018 around 6 PM (Figure 5b.1), when Michael was about to make its landfall near Tallahassee, most of the power/electricity disruption related tweets were coming from Tallahassee area. Figure 5b.2 shows the second snapshot of Michael around midnight of October 12, 2018 when the center of Michael was over North Carolina. It shows that most of power/electricity related disruptions were coming from Wake, Johnston, Durham and Orange counties of North Carolina.

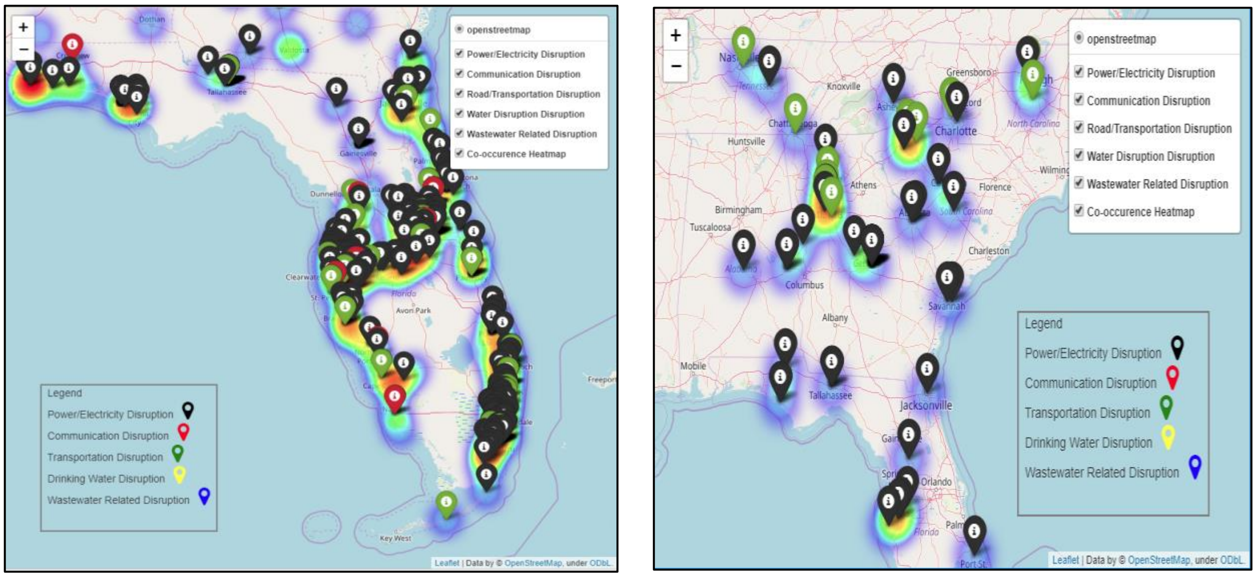
Finally, we visualize the co-occurrence of multiple disruption types in an interactive map. Figure 6 shows a snapshot of the co-occurrence map for hurricane Irma (Figure 6a) and Michael (Figure 6b). We plot this map using only the actual disruption samples (negative sentiment) aggregated over a 1-hour interval. This interactive map allows to explore the disruptions type separately as well as a combination of them. The co-occurrence heat map shows a relative intensity of the disruptions based on the co-occurrences of all the disruption types. For Irma, mostly co-occurred disruptions are power, communication, and transportation disruptions. On the other hand, for hurricane Michael (see Figure 6b) the most co-occurred disruptions are power and transportation disruptions.

In summary, we find that during hurricanes Irma and Michael affected people posted infrastructure related tweets. Those posts may represent actual infrastructure disruptions. A multi-label classification approach (a logistic regression model adopted over a label powerset) has been developed to predict both the disruption types and disruption status from such data. After locating the disruptions using a geocoding approach, a map can visualize the disruptions spatially and temporally. The training time of the model is about 7 sec, and it takes about 1 sec to process, predict, and visualize the data collected over one hour. Thus, this approach can be easily applied in a real-time setting.

7 LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Our study has some limitations. For instance, the annotated dataset is small in comparison to the entire dataset. More annotated samples are likely to increase the accuracy of the model. Although the co-occurrence of multiple disruptions is considered, the approach cannot infer if a disruption is caused by another disruption. Incorporating causality as an input to the model may improve its performance. Another limitation of our approach is that we have checked the accuracy of the method based on human-annotated tweets, which may not represent the total number of disruptions observed in the ground. To check the extent to which the reported disruptions match actual ones, ground truth data on disruptions occurring in different infrastructure systems are required. These datasets, often collected by infrastructure service providers including private companies and public agencies, may contain sensitive information. Collecting ground truth data on infrastructure disruptions from a variety of sources covering multiple states will be a very challenging task. Further studies are needed to verify what percentage of actual disruptions is reported in social media and to what extent these disruptions can be identified using the method developed in this study. In addition, our data cover hurricanes only. Future studies can transfer and validate our approach across other disasters such as wildfire, earthquake, snowfall, and thunderstorms.

In this study, we assume that a post with a negative sentiment is associated with an actual disruption, and a post with a neutral or positive sentiment is associated with no disruption. However, there could be a post with a positive



(a) Hurricane Irma (Time 2017-09-11 05:00:00)

(b) Hurricane Michael (Time 2018-10-12 06:00:00)

Figure 6 Disruption Co-occurrence Map (a) Hurricane Irma (b) Hurricane Michael sentiment, but associated with an actual disruption. These tweets are likely to be a small portion of the entire dataset. In our annotated dataset, we did not find such tweets. Future studies, based on natural language processing, can develop more advanced methods to capture the situations where even a positive tweet could be associated with a disruption.

When the communication network is disrupted, affected people may not have access to social media platforms. In such situations, our model cannot detect disruptions. In the geo-parsing method, we use exact matching process between the extracted location and county/city of the affected regions. Since our approach finds city/county names only, it cannot extract location if street or any finer level location is mentioned in the text. In future studies, text-based location matching can be developed with finer resolution (e.g., street name), which may help in locating disruptions with more specific location information.

For training our models, we adopt a batch learning approach which requires retraining the model to incorporate new data from the data stream. Future studies can explore an incremental learning approach (T. T. Nguyen et al., 2019; Read, Bifet, Holmes, & Pfahringer, 2012) to dynamically train models on newly available data from the ongoing/future disasters (NOAA National Centers for Environmental Information (NCEI) U.S., 2018). Such an incremental learning approach is likely to increase the accuracy of the model as it utilizes data from an ongoing disaster.

To achieve a better classification accuracy, more complex classification methods such as probabilistic neural networks (Ahmadlou & Adeli, 2010), dynamic neural networks (Rafiei & Adeli, 2017a), and hierarchy-based models (Cerri, Basgalupp, Barros, & de Carvalho, 2019; Wehrmann, Cerri, & Barros, 2018) can be considered. A probabilistic neural network is a fast, efficient, and flexible model to add/remove new training data and hence may be more suitable for real-time disruption prediction for an unseen disaster. Since textual data have a large feature space, a dynamic neural network might be useful in finding an optimal number of features to achieve better performance. Moreover, hierarchy-based models might be more suitable when there exists more hierarchy in the disruption types, especially considering disruptions from multiple disasters (hurricane, wildfire, snowstorm etc.). A hierarchy-based model can have classes for disaster type, disruption type, and disruption status. A hierarchical relationship can be created from disaster type to disruption type to disruption status (e.g., if a post is not disaster related it has no disruption type and disruption status).

8 CONCLUSIONS

This paper presents an approach to identify infrastructure disruptions and a dynamic disruption mapping framework using social media data. While previous research focused mainly on identifying hurricane or damage related social media posts, we consider five types (power/electricity, communication, drinking water, and wastewater) of infrastructure disruptions, their co-occurrence, and their status (whether a post is reporting an actual disruption, disruption in general, or not affected by a disruption). The result shows that our multi-label classification approach (logistic regression adopted in a label powerset approach) performs better than a baseline method (based on keyword search and sentiment analysis). Moreover, we present a method, to visualize disruptions in a dynamic map. Identifying disruption types and disruption locations is vital for disaster recovery, response and relief operations. The developed approach of identifying the co-occurrence of multiple disruptions may help coordinate among infrastructure service providers and disaster management organizations.

ACKNOWLEDGMENT

The authors are grateful to the U.S. National Science Foundation for the grants CMMI-1832578, CMMI-1832693, and CMMI-1917019 to support the research presented in this paper. However, the authors are solely responsible for the findings presented in this paper.

REFERENCES

Ahmadlou, M., & Adeli, H. (2010). Enhanced probabilistic neural network with local decision circles: A robust classifier. *Integrated Computer-Aided Engineering*, *17*(3), 197–210.

Alavi, A. H., & Buttlar, W. G. (2019). An overview of smartphone technology for citizen-centered, real-time and scalable civil infrastructure monitoring. *Future Generation Computer Systems*, *93*, 651–672.

Alinizzi, M., Chen, S., Labi, S., & Kandil, A. (2018). A Methodology to Account for One-Way Infrastructure Interdependency in Preservation Activity Scheduling. *Computer-Aided Civil and Infrastructure Engineering*, *33*(11), 905–925.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Binkhonain, M., & Zhao, L. (2019). A review of machine learning algorithms for identification and classification of non-functional requirements. *Expert Systems with Applications*.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc."

Birregah, B., Top, T., Perez, C., Châtelet, E., Matta, N., Lemercier, M., & Snoussi, H. (2012). Multi-layer crisis mapping: a social media-based approach. In *2012 IEEE 21st International Workshop on Enabling*

Technologies: Infrastructure for Collaborative Enterprises (pp. 379–384).

Bryson, K.-M. N., Millar, H., Joseph, A., & Mobolurin, A. (2002). Using formal MS/OR modeling to support disaster recovery planning. *European Journal of Operational Research*, *141*(3), 679–688.

Buldyrev, S. V, Parshani, R., Paul, G., Stanley, H. E., & Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature*, *464*(7291), 1025.

Cerri, R., Basgalupp, M. P., Barros, R. C., & de Carvalho, A. C. (2019). Inducing Hierarchical Multi-label Classification rules with Genetic Algorithms. *Applied Soft Computing*, *77*, 584–604.

Chang, P. C., Flatau, A., & Liu, S. C. (2003). Health monitoring of civil infrastructure. *Structural Health Monitoring*, *2*(3), 257–267.

Cheng, Z., Caverlee, J., & Lee, K. (2010). You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 759–768).

Cresci, S., Cimino, A., Dell'Orletta, F., & Tesconi, M. (2015). Crisis mapping during natural disasters via text analysis of social media messages. In *International Conference on Web Information Systems Engineering* (pp. 250–258).

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.

Du, E., Cai, X., Sun, Z., & Minsker, B. (2017). Exploring the role of social media and individual behaviors in flood evacuation processes: An agent-based modeling approach. *Water Resources Research*, *53*(11), 9164–9180.

Dyskin, A. V, Basarir, H., Doherty, J., Elchalakani, M., Joldes, G. R., Karrech, A., … others. (2018). Computational monitoring in real time: review of methods and applications. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, *4*(3), 235–271.

Fan, C., & Mostafavi, A. (2019). A graph-based method for social sensing of infrastructure disruptions in disasters. *Computer-Aided Civil and Infrastructure Engineering*.

Fan, C., Mostafavi, A., Gupta, A., & Zhang, C. (2018). A system analytics framework for detecting infrastructure-related topics in disasters using social sensing. In *Workshop of the European Group for Intelligent Computing in Engineering* (pp. 74–91).

Fang, Y.-P., & Sansavini, G. (2019). Optimum post-disruption restoration under uncertainty for enhancing critical infrastructure resilience. *Reliability Engineering & System Safety*, *185*, 1–11.

FCC. (2017). Communications Status Report for Areas Impacted by Hurricane Irma.

Fry, J., & Binner, J. M. (2015). Elementary modelling and behavioural analysis for emergency evacuations using social media. *European Journal of Operational Research*, *249*(3), 1014–1023. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84930690948&partnerID=tZOtx3y1

Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems*, *26*(3), 10–14.

Guan, X., & Chen, C. (2014). Using social media data to understand and assess disasters. *Natural Hazards*, *74*(2), 837–850.

Hasan, S., & Foliente, G. (2015). Modeling infrastructure system interdependencies and socioeconomic impacts of failure in extreme events: emerging R&D challenges. *Natural Hazards*, *78*(3), 2143–2168.

Homeland Security. (2019). Critical Infrastructure Sectors. Retrieved January 8, 2019, from https://www.dhs.gov/cisa/critical-infrastructure-sectors

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. Retrieved from http://www.mdpi.com/2220-9964/4/3/1549/htm

Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

Jenelius, E., & Mattsson, L.-G. (2012). Road network vulnerability analysis of area-covering disruptions: A grid-based approach with case study. *Transportation Research Part A: Policy and Practice*, *46*(5), 746–760.

Jongman, B., Wagemaker, J., Romero, B., & de Perez, E. (2015). Early flood detection for rapid humanitarian response: harnessing near real-time satellite and Twitter signals. *ISPRS International Journal of Geo-Information*, *4*(4), 2246–2266.

Kadri, F., Birregah, B., & Châtelet, E. (2014). The impact of natural disasters on critical infrastructures: A domino effect-based study. *Journal of Homeland Security and Emergency Management*, *11*(2), 217–241.

Keim, M. E., & Noji, E. (2010). Emergent use of social media : A new age of opportunity for disaster resilience. *American Journal of Disaster Medicine*, *6*(1), 47–54.

Khan, A., Baharudin, B., Lee, L. H., & Khan, K. (2010). A review of machine learning algorithms for text-documents classification. *Journal of Advances in Information Technology*, *1*(1), 4–20.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, *160*, 3–24.

Kryvasheyeu Y, Chen H, Moro E, Van Hentenryck P, C. M. (2015). Performance of Social Network Sensors During Hurricane Sandy. *PLoS One 10.2 (2015): E0117288.*, *10*(2). Retrieved from http://arxiv.org/abs/1402.2482

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Hentenryck, P. Van, Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science Advances 2.3 (2016): E1500779.*

Lambert, J. H., & Patterson, C. E. (2002). Prioritization of schedule dependencies in hurricane recovery of transportation agency. *Journal of Infrastructure Systems*, *8*(3), 103–111.

Li, C., & Sun, A. (2014). Fine-grained location extraction from tweets with temporal awareness. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 43–52).

Li, Z., Wang, C., Emrich, C. T., & Guo, D. (2018). A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. *Cartography and Geographic Information Science*, *45*(2), 97–110.

Lu, L., Wang, X., Ouyang, Y., Roningen, J., Myers, N., & Calfas, G. (2018). Vulnerability of interdependent urban infrastructure networks: Equilibrium after failure propagation and cascading impacts. *Computer-Aided Civil and Infrastructure Engineering*, *33*(4), 300–315.

Martani, C., Jin, Y., Soga, K., & Scholtes, S. (2016). Design with uncertainty: the role of future options for infrastructure integration. *Computer-Aided Civil and Infrastructure Engineering*, *31*(10), 733–748.

Martín, Y., Li, Z., & Cutter, S. L. (2017). Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew. *PLoS ONE*, *12*(7), 1–22. https://doi.org/10.1371/journal.pone.0181701

Middleton, S. E., Middleton, L., & Modafferi, S. (2013). Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*, *29*(2), 9–17.

Mitsova, D., Escaleras, M., Sapat, A., Esnard, A.-M., & Lamadrid, A. J. (2019). The Effects of Infrastructure Service Disruptions and Socio-Economic Vulnerability on Hurricane Recovery. *Sustainability*, *11*(2), 516.

Mouzannar, H., Rizk, Y., & Awad, M. (2018). Damage Identification in Social Media Posts using Multimodal Deep Learning. In *ISCRAM*.

NERC. (2018). Hurricane Irma Event Analysis Report. *Www.Nerc.Com*.

Nguyen, D. T., Ofli, F., Imran, M., & Mitra, P. (2017). Damage assessment from social media imagery data during disasters. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (pp. 569–576).

Nguyen, T. T., Dang, M. T., Luong, A. V., Liew, A. W.-C., Liang, T., & McCall, J. (2019). Multi-Label Classification via Incremental Clustering on Evolving Data Stream. *Pattern Recognition*.

NOAA. (2019). National Hurricane Center Data Archive. Retrieved from https://www.nhc.noaa.gov/data/

NOAA National Centers for Environmental Information (NCEI) U.S. (2018). Billion-Dollar Weather and Climate Disasters. Retrieved January 15, 2018, from https://www.ncdc.noaa.gov/billions/

Ouyang, M., & Fang, Y. (2017). A mathematical framework to optimize critical infrastructure resilience against intentional attacks. *Computer-Aided Civil and Infrastructure Engineering*, *32*(11), 909–929.

Oyeniyi, D. (2017). How Hurricane Harvey Changed Social Media Disaster Relief. *TexasMonthly*. Retrieved from https://www.texasmonthly.com/the-daily-post/how-social-media-managers-responded-to-hurricane-harvey/

Pant, R., Thacker, S., Hall, J. W., Alderson, D., & Barr, S. (2018). Critical infrastructure impact assessment due to flood exposure. *Journal of Flood Risk Management*, *11*(1), 22–33.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Pietrucha-Urbanik, K., & Tchórzewska-Cieślak, B. (2018). Approaches to failure risk analysis of the water distribution network with regard to the safety of consumers. *Water*, *10*(11), 1679.

Rafiei, M. H., & Adeli, H. (2017a). A new neural dynamic classification algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(12), 3074–3083.

Rafiei, M. H., & Adeli, H. (2017b). A novel machine learning-based algorithm to detect damage in high-rise building structures. *The Structural Design of Tall and Special Buildings*, *26*(18), e1400.

Rafiei, M. H., & Adeli, H. (2018a). A novel unsupervised deep learning model for global and local health condition assessment of structures. *Engineering Structures*, *156*, 598–607.

Rafiei, M. H., & Adeli, H. (2018b). Novel machine-learning model for estimating construction costs considering economic variables and indexes. *Journal of Construction Engineering and Management*, *144*(12), 4018106.

Rafiei, M. H., Khushefati, W. H., Demirboga, R., & Adeli, H. (2017). Supervised Deep Restricted Boltzmann Machine for Estimation of Concrete. *ACI Materials Journal*, *114*(2).

Ramos, J., & others. (2003). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (Vol. 242, pp. 133–142).

Read, J., Bifet, A., Holmes, G., & Pfahringer, B. (2012). Scalable and efficient multi-label classification for evolving data streams. *Machine Learning*, *88*(1–2), 243–272.

Rinaldi, S. M., Peerenboom, J. P., & Kelly, T. K. (2001). Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems Magazine*, *21*(6), 11–25.

Rosenzweig, C., & Solecki, W. (2014). Hurricane Sandy and adaptation pathways in New York: Lessons from a first-responder city. *Global Environmental Change*, *28*, 395–408.

Roy, K. C., Cebrian, M., & Hasan, S. (2019). Quantifying human mobility resilience to extreme events using geo-located social media data. *EPJ Data Science*, *8*(1), 18.

Sadri, A. M., Hasan, S., & Ukkusuri, S. V. (2019). Joint Inference of User Community and Interest Patterns in Social Interaction Networks. *Social Network Analysis and Mining*, *9*, 11. Retrieved from http://arxiv.org/abs/1704.01706

Sörensen, S., Webster, J. D., & Roggman, L. A. (2002). Adult attachment and preparing to provide care for older relatives. *Attachment & Human Development*, *4*(1), 84–106.

Sorower, M. S. (2010). A literature survey on algorithms for multi-label learning. *Oregon State University, Corvallis*, *18*, 1–25.

Sriram, L. M. K., Ulak, M. B., Ozguven, E. E., & Arghandeh, R. (2019). Multi-Network Vulnerability Causal Model for Infrastructure Co-Resilience. *IEEE Access*, *7*, 35344–35358.

Sumalee, A., & Kurauchi, F. (2006). Network capacity reliability analysis considering traffic regulation after a major disaster. *Networks and Spatial Economics*, *6*(3–4), 205–219.

Tang, Z., Zhang, L., Xu, F., & Vo, H. (2015). Examining the role of social media in California's drought risk management in 2014. *Natural Hazards*, *79*(1), 171–193. Retrieved from http://link.springer.com/10.1007/s11069-015-1835-2

The Effect of Hurricane Irma on Water Supply. (2017). *KRAUSZ.COM*.

Tsoumakas, G., & Vlahavas, I. (2007). Random k-labelsets: An ensemble method for multilabel classification. In *European conference on machine learning* (pp. 406–417).

Ulak, M. B., Kocatepe, A., Sriram, L. M. K., Ozguven, E. E., & Arghandeh, R. (2018). Assessment of the hurricane-induced power outages from a demographic, socioeconomic, and transportation perspective. *Natural Hazards*, *92*(3), 1489–1508.

Unsafe Drinking Water After Hurricane Irma. (2019). *Wayde King Water Filtration*.

Wang, Q., & Taylor, J. E. (2014). Quantifying human mobility perturbation and resilience in hurricane sandy. *PLoS ONE*, *9*(11), 1–5.

Wehrmann, J., Cerri, R., & Barros, R. (2018). Hierarchical multi-label classification networks. In *International Conference on Machine Learning* (pp. 5225–5234).

Yuan, F., & Liu, R. (2018). Feasibility study of using crowdsourcing to identify critical affected areas for rapid damage assessment: Hurricane Matthew case study. *International Journal of Disaster Risk Reduction*, *28*, 758–767.

Yuan, F., & Liu, R. (2019). Identifying Damage-Related Social Media Data during Hurricane Matthew: A Machine Learning Approach. *Computing in Civil Engineering 2019*.