REACT: The Riskmap Evaluation and Coordination Terminal

by

Abraham Quintero

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2019

Author . . . .
Department of Electrical Engineering and Computer Science
August 20, 2019

Certified by . . . .
Miho Mazereeuw
Associate Professor
Thesis Supervisor

Accepted by . . . .
Katrina LaCurts
Chair, Master of Engineering Thesis Committee

REACT: The Riskmap Evaluation and Coordination Terminal

by

Abraham Quintero

Submitted to the Department of Electrical Engineering and Computer Science
on August 20, 2019, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Computer Science and Engineering

Abstract

The United Nations Office for Disaster Risk Reduction (UNDRR) states that economic losses due to natural disasters have risen 151 percent in the past 20 years. Of these disasters, floods are the most common. The Sendai Framework for Disaster Risk Reduction was created by the UNDRR in order to chart goals for future risk mitigation; among its seven global targets is increasing the availability of disaster risk information and assessment systems. Disaster information systems use state of the art techniques such as remote sensing in order to mitigate damages from natural and man made hazards.

It is common in developed countries utilize networks of advanced sensors and ahead of time mapping in order to facilitate emergency responses; however, such systems are not available in developing countries due to cost limitations. The widespread proliferation of smart phones and social media use in developing countries means that citizens can be used as sensors by reporting disaster information online. The Riskmap system was developed by the Urban Risk Lab at MIT in order to gather citizen report streams. Such citizen disaster reports have two issues: a large influx of reports can cause information overload in emergency operations centers, which makes it difficult to summarize the situation. Machine learning has previously been used in order to analyze and simplify information for human consumption. This work seeks to use novel machine learning techniques to fully utilize crowd-sourced social media reports gathered using the Riskmap system.

Thesis Supervisor: Miho Mazereeuw
Title: Associate Professor

Acknowledgments

I would first like to thank Aditya and Miho- not one page of this thesis could have been written without your help and your guidance. Thank you so much for your patience and your expertise.

This work would have also been impossible without the support of the MIT TATA Center and the TATA trusts. Rob, Diane, and Chintan- your forward looking enthusiasm is infectious.

I would also like to thank my colleagues at the MIT Urban Risk Lab that created the data collection systems used within. They are: Tomas Holderness, Matthew Berryman, Mayank Ojha, Manaasa Priyaa Dharmapuri Sridhar, Christina Geros, Pritta Andrani Widyanarko, Dika Fadmastuti, Nashin Mahtani, Emir Hartato. The full list of libre software that facilitated this thesis is far too long to name, but I feel especially grateful to the collaborators that have created sklearn, torch, tensorflow, IPython, and numpy.

I would also like to thank my housemates, who listened with little complaint to my stumbling blocks and always offered valuable suggestions. Especially Liana, Eric and Damien, who often stayed up to help edit my writing or provide advice on system design.

To my family- it was your phone calls and encouragement that made the hard days a little easier.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Disasters triggered by natural hazards are a constant threat to societies all over the planet. Among those disasters, flooding is the most common calamity in the world, accounting for half of all disaster deaths [1, 2]. Although flooding impacts both developed and developing countries, developing nations face much higher mortality rates as a result of flooding since they lack resources to adequately mitigate hazards [3, 4]. Deltaic megacities in developing countries are particularly at risk because of unregulated urbanization, rising population, and climate change [1]. Moreover, there is little data that is available before, during, and after a disaster to help stakeholders respond to hazards [5].

Various stakeholders, such as humanitarian Non-governmental Organizations (NGOs), government emergency responders and affected citizens, have different but overlapping interests with regards to disaster management. Government and NGOs work together to provide relief and mitigate damages from flooding [6], while citizens look for relevant information and try to reduce their risk by avoiding heavily affected areas [7]. Information is at the core of disaster management; however, data scarcity makes it hard for emergency personnel to optimize their use of resources, while citizens have an abundance of information about their surroundings but must be careful not to trust incorrect or outdated information about broader areas [8].

Disaster information systems can connect affected communities with Emergency Operations Centers (EOCs), thereby bridging the information gap between responders and citizens. Many such disaster information systems have been developed, but they often suffer from a lack of institutional buy-in [9]. The lack of engagement can be partly attributed to the difficulty of adding data gathering responsibilities to emergency personnel that have little time during crises. One solution to this problem is asking citizens to submit information to allow for peer to peer and citizen to govenment information sharing [5]; however, crowdsourcing brings its own issue: in times of crisis EOCs can suffer from information overload when they are presented with too much data [10].

The REACT System is built on top of the Riskmap Platform for crowdsourcing disaster information[1]. REACT uses novel machine learning research to reduce information overload from crowd sourced data in EOCs, thereby decreasing data analysis time during disasters. It first extracts key features from each of the parts of a report (text, picture, metadata) using domain specific techniques and then uses an ensemble learner to classify the citizen report.

First we establish the motivation for using citizens as sensors and provide an overview of the Riskmap System which allows for gathering of citizen reports. We then show the need for analyzing this noisy data using machine learning. A review of different machine learning techniques that have been used in crisis information systems, including those that also utilize social media is presented. Finally we describe REACT, a novel ensemble learning model that can accurately predict large urban flood events from noisy crowdsourced data.

[1] REACT is libre software and can be found here

Chapter 2

Background

Global climate change is 'expected to increase the frequency and intensity of floods' [4]. Developing urban areas that are undergoing unchecked development and rising population regularly experience flooding [1]. Nowhere is this more apparent than in South and Southeast Asia, where the severity of floods has been increasing over the past several decades [11].

Of the world's 33 mega-cities with population over 10 million, more than 60 percent are located in developing Asian Countries [12]. These cities face a looming crisis as flood risk increases, but there are also unique opportunities for risk mitigation. Megacities are characterized by high population density, which leads to an increase in economic damages and loss of life during flood events, but it also means that there are large numbers of citizens that have disaster information they would like to share with others [1].

2.1 History of Disaster Informatics

One of the best known and earliest work in disaster informatics was John Snow's use of maps to find the source of the 1857 Cholera outbreak in London[13]. This example is taught to all students of epidemiology and illustrates the need not only for up to date information, but for systems that ease the analysis of this information. In John Snow's case, the map was the technology that allowed him to visualize the spread of the disease and effectively take action that ended the outbreak; however, the computer revolution has drastically changed the way that scientists and responders analyze disaster information. John Snow used mapping technology to track the spread of disease, but now researchers are using artificial intelligence to predict cholera outbreaks before they happen [14].

In more recent times, the need for Information Technology (IT) in disaster management has been clear since the mid 1980s when computers became user friendly enough to be used during disasters [15]. Now many Emergency Operations Centers (EOCs) use Geographic Information Systems (GIS), inventory control systems, and online messaging systems among other technology in order to organize spatial data and analyze disaster information. For example, Mozambique used an integrated disaster management system to provide early warning during the 2007 Zambezi floods, while Guatemala's inventory management helped to curb government bribes [9]. While technology has helped some EOCs to better respond to disaster events, researchers have often stated that 'there are many reasons to remain skeptical about the idea that technology will provide a panacea for emergency management problems' [16, 10, 17]. A number of potential negative effects associated with disaster management technology have been identified: primarily the potential of information overload and the dissemination of incorrect and outdated information [8, 18].

2.1.1 Social Media and Disasters

The history of online communities is firmly linked to disasters. In the early days of the internet the Internet Relay Chat (IRC) became one of the first truly global online communication systems after its adoption was 'prompted by the First Gulf War' [19]. Although radio and television broadcasts were halted by the Iraqi army shortly after the invasion, IRC communication continued for days afterward. IRC allowed users to communicate about conditions on the ground, including the Gulf War oil spill that grew to be the largest oil spill in history [20].

In more recent years, social media networks on mobile devices have taken over the role that IRC played in 1991. Social media allows users to connect to friends and family as well as celebrities and organization. The hashtag is a key part of many social media networks, including instagram, twitter, and facebook. The history of the hashtag is also invariably linked to disaster communication. It was during the San Diego bush fires of 2007 that the hashtag was first widely used on twitter [19].

Quarantelli, a pioneer of disaster sociology, emphasized that 'management of hazards is fundamentally social in nature and not something that can be achieved strictly through technological upgrading' [10] yet social media brings human behavior into a machine readable format that can provide information that was previously hidden.

Previously, work has been done in passively listening to social media streams in order to better understand how disasters unfold and how humans use social media as a communication tool during disaster events. Many of these studies use hand labeled tweets in order to classify what kind of information people talk about [21].Further work has evolved to using artificial intelligence methods to automatically label new tweets using supervised learning. For example, Patrick Meier's Haiti Crisis Map initially used volunteers to classify large number of tweets, but his more recent projects focus on the use of AI for tackling big data problems [5].

2.1.2 Big data collection: Crowdsourcing vs. Passive Listening

During disasters, survivors are the first responders until help arrives. This means that citizens are key eyewitnesses and have the most up to date disaster information. As Patrick Meier points out in *Digital Humanitarians*, 'since humanitarian organizations don't ask eyewitnesses on social media to report information', they must passively wait and 'rely on witnesses sharing relevant information by chance' [5]. Listening to twitter data streams and hoping that someone posts relevant disaster information is not always a winning strategy. Past solutions have paid workers to collect information and enter it into disaster information systems. The government of Mozambique used this technique during flooding in 2007; however, this method was found to be expensive and unscaleable [9]. An analysis of the system found that 'data processing

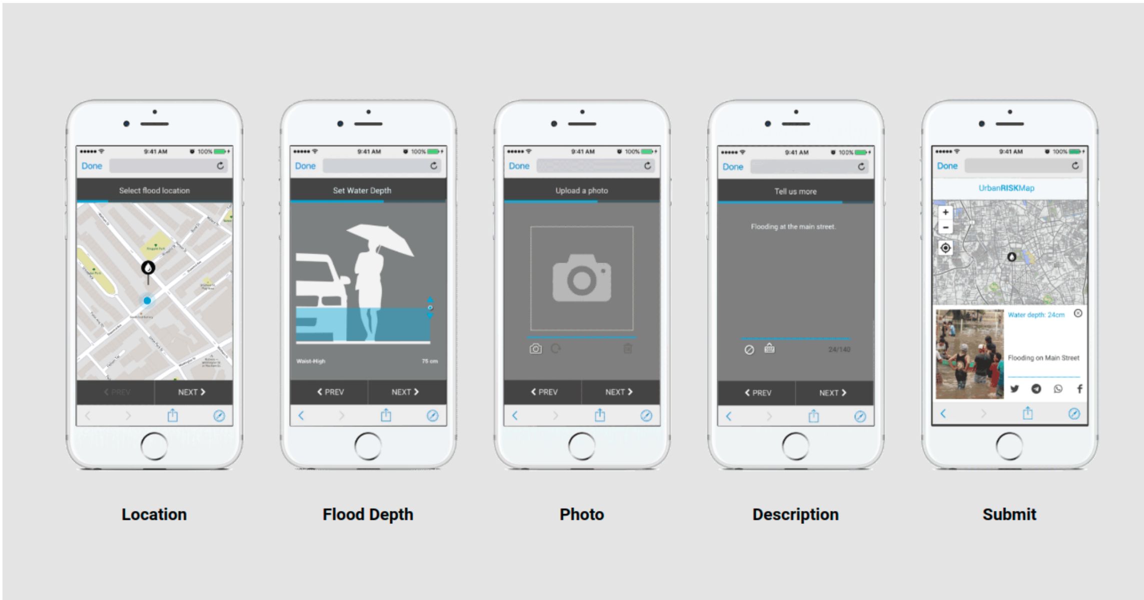

Figure 2-1: Submitting a flood report card. Graphic by MIT-URL

and consolidation [were] difficult' and that 'the few data entry clerks struggled to keep up' [9].

2.2 The Riskmap System

The Riskmap System was created by the Urban Risk Lab (MIT-URL) in 2016 to collect citizen reports on disaster events. It alleviates the load on emergency managers by centralizing reports from many social media sources. Riskmap not only makes it easy for reports to come into the response center, but it also allows emergency managers to indicate which areas of a city are most affected at any one time. The data gathered during an event is persistent and available under the Creative Commons license, which allows researchers to track the flood over time and pinpoint areas that are particularly vulnerable to flooding, thus fulfilling the need for open data [22, 23].

The system has been in place in Jakarta (as Petabencana.id) and Chennai (as Riskmap.in[1]) since 2016, and has seen hundreds of thousands of views during flood events [24, 25, 26].

[1]For more information go to riskmap.org

2.2.1 Social Media Outreach

Riskmap consists of many different social media bots that are actively filtering social media streams and looking for citizens that might be reporting flooding events. The system then reaches out to those users and asks them to submit a flood report that consists of a GPS location, the estimated flood height at that location, a picture, and a textual description. The user interface for submitting these reports is shown in Figure 2-1. These reports are displayed on a public map to help other citizens navigate the disaster zone. Furthermore, EOC personnel are able to access the Risk Evaluation Matrix (REM), a special dashboard that allows them to give even more information to citizens.

Figure 2-2: Uber drivers in Jakarta were informed via a screenshot of the public Petabencana map. Graphic by Petabencana

2.2.2 Need for open data

Disasters often affect a wide variety of services that do not have established Standard Operating Procedures (SOPs) to deal with hazardous events. In order for these services to decrease the impact of the disaster they must have open data that is available without delay or bureaucracy. One example of the Riskmap serving in this role occurred in Jakarta in 2017: Uber, a ride hailing service, used our public map to help drivers navigate the flooded city. A screen shot of the Uber drivers app is shown in Figure 2-2. This would not have been possible if the disaster data was siloed.

2.3 Conquering Information Overload

When researchers have tried using social media to track real-time disasters, they often suffer from information overload. For example, Meier states that the Haiti Crisis map was 'constantly overwhelmed with the vast amount of information that needed to be monitored and processed' and that the team never 'managed to catch up' with the backlog of social media activity [5].

It is not enough to create an advanced system for consuming citizen reports, it is also necessary to ensure that this system does not consume resources that are already scarce during a disaster event, for example the time of emergency workers [9]. Furthermore, it is also important to reduce the resources needed to create insights because if analyzing data is too difficult, then decision makers will make decisions without having fully analyzed the data [3].

Using computers to automatically make sense of disaster data has long been a goal in disaster informatics, but only recently have machine learning techniques become advanced enough to be implemented in production emergency systems [5]. Image recognition algorithms can provide summaries of objects and scenes found in user submitted photos [27, 28]. Natural language processing can estimate the probability that a textual document is overall negative or positive and thereby give EOCs a shorthand way to summarize thousands of reports in short amounts of time [27, 29]. Finally, ensemble learning methods can learn relationships between disparate datasets and synthesize a single result [30].

In this work we will experiment with different machine learning techniques for image recognition, finally showing that transfer learning at the output layer can turn off the shelf multi—label classification algorithms into classifiers for flood image classification. For textual analysis, we will show the performance of two commonly used techniques to classify report texts into ‘heavy flooding’ or ‘no heavy flooding’ classes. Flood height will first be assessed as a raw numerical feature. Finally, the output of these disparate techniques will be combined by using a small deep neural network to classify a report into one of two classes: ‘heavy flooding’ or ‘no heavy flooding’

In order to increase interpretability, the most important labels from the feature extraction machine learning algorithms are used to explain the decisions of the system.

Chapter 3

Previous Work for Machine Learning in Crisis Informatics

As discussed in Section 2.1.1, researchers have enviously eyed the panacea of information present in social media streams since the inception of the medium. Social media datasets are often very large, often hundreds of thousands or millions of data points can be collected during a single event. For example more than 1.5 million disaster related tweets were created during the 2011 Tohoku earthquake [31] and over 20 million during hurricane Sandy [5]. While all of these tweets concern the natural disaster and citizen's reaction to the event, only a few have actionable intelligence that can help Emergency Operations Systems to better respond to an event [5]— filtering out these actionable tweets and thus better understanding the ground truth is at the core of much research.

3.1 Passive Collection after Event

Because of the large number of data points and because of the difficulty of using humans to label all of the data set, researchers have turned to keyword searches to whittle down the volume of data [31]. Researchers then either hand label a small subset of the data or use keywords to aggregate data points and examine statistics on those aggregates.

Many studies of social media disaster data focused around post-event analysis using manual labeling. In [32] Starbird and Palen explore the uptake of the Tweek the Tweet (TtT) microsyntax, where citizens are encouraged to use a specific syntax in order to request or offer help during a crisis event. The adoption of the syntax by citizens on the ground was very low. Only 39 out of a total of 2,911 tweets in the TtT dataset originated within the affected areas; however, the authors saw many more tweets that were translated into the TtT syntax by other users. The authors also point out that social media enable new forms of volunteerism where humans can be used to manually 'translate' user tweets.

Other studies have analyzed social media behavior by using key word searches only and creating aggregate statistics. For example, Vieweg et al. state that they 'reduced the data sets to those user streams that included more than three tweets containing the search terms' in order to 'make samples manageable' [7]. The team then looked at easily machine readable features such as the inclusion of geotags.

3.2 Artificial Intelligence

While it is possible to gather some intelligence by using key words and digital volunteers to label data, artificial intelligence has two main advantages. Machine learning algorithms can handle much larger volumes of data than can be consumed by even a large team of people, and they can do it in a much shorter amount of time. Social media disaster data is mainly composed of images or text, therefore researchers have focused on either image recognition or Natural Language Processing (NLP) to classify data. More recently, there have been some inroads to using ensemble methods that can utilize images and text at the same time.

3.2.1 Textual Analysis

Nagy et al. used SentiWordNet, a lexical embedding for sentiment analysis, in order to classify tweets as mainly containing positive or negative feelings [29]. A dataset of 3698 tweets created during the 2010 San Bruno, CA fires was used in this study.

The authors did manually label the tweets in order to find the performance of their method and found a precision score of 90% on a held out set. After validation, the authors can use this process to classify tweets with little human interaction, allowing them to process much larger datasets.

In 2013, Chowdhury et al. used feature extraction to classify tweets into pre-incident, during-incident, and post-incident categories [33]. They showed the importance of treating crisis messages as distinct from other text based machine classification tasks. The precision score of this method was at least 96% [33].

Caragea et al. and Imran et al. compared the performance of Support Vector Machines (SVMs) and Convolutional Neural Networks (CNNs) at classifying tweets into two classes: informative, and not informative [34, 35]. The teams found that CNNs performed much better at the classification task because they were better able to learn the relationships between tokens and did not require much feature engineering in order to have good performance.

In addition to methods for classifying disaster messages, there has recently been work on creating textual datasets to help researchers conduct machine learning research. The Qatar Computing Research Institute has created the CrisisNLP datasets that contain thousands of tweets that have been human labeled [27].

3.2.2 Image Recognition

In the past decade there have been very promising results to using Convolutional Neural Networks for image recognition tasks, however training these large neural nets traditionally requires very large datasets. In 2014, the DeCAF team showed how to conduct transfer learning on deep neural nets in order to reduce the number of training images required without sacrificing performance [28]. Following this result, Nguyen et al. used transfer learning with an online component to classify social media disaster images into 3 classes: severe, mild, and little damage [36].

Another technique that researchers have used is the visual bag of words approach. In this method, each image is associated with a collection of labels for objects that are found in that scene. The collection of labels is then treated in the same way as text using a bag of words embedding scheme. Both humans and automatic means can be used to label images [37].

3.2.3 Ensemble Learning Models

The prospect of using many learning models and combine them has driven many research goals. Bagging, the practice of using many simple learners on different subsets of the training data has shown to be particularly useful [38]. Boosting is a related technique where an emsemble of simple classfiers are trained in succession, with data points misclassified by the nth classifer are weighed as more important for classifier $n + 1$. In a disaster data context, Furlanello et al. used boosting to train decision trees and geospatially predict the risk of tickborne illnesses [39].

Multimodal data

Ensemble models are not only attractive because they can increase precision on a single type of data, but also because they can enable classifiers to process many types of data. For example, Mouzannar et al. used two different CNNs to process image and text data. The team then used a variety of simple classifiers as the ensemble learners to produce a final result, classification into one of 6 classes [30]. In this manner, they were able to achieve higher than 90% accuracy with a dataset of 35,000 data points.

Researchers have also used neural networks as the ensemble learners. These models enable much more flexibility because the learner is able to encode a much more complex function and can therefore learn more information about the latent features between learners [40].

3.3 Challenges

Small Datasets

Although a limited number of larger datasets have recently become available, there has historically been a scarcity of training and validation data available for Deep learning models that are trained on small datasets tend to overfit on the training data and do not generalize well to the validation dataset [41, 27].

In many early studies only hundreds of data points were considered— combined with the small size of those data points (for example, twitter microblogs of 140 or 280 characters) and effectively using deep learning becomes very difficult. For example [29] only uses 3,698 tweets in order to train their classifier.

Cold Start

At the beginning of a disaster, there is not much data available in order to train a classifier. Some researchers have tried to remedy this problem by augmenting their datasets by adding data from other disasters [41]. This data augmentation brings its own issues as it is important to ensure that the added data comes from the same distribution as the organically generated data.

Task Subjectivity

Task subjectivity is an incredibly common issue [36, 3]. While most humans can agree on whether an object is or is not an apple, this task does not translate to defining if a picture indicates a severe event or a minor one. In other words, people's perception of risk varies widely from region to region and from citizen to citizen [3]. Furthermore, the features of the built environment change from location to location, making it difficult to execute cross location transfer learning.

Focus on technology rather than whole system design

A series of UN case studies on six disaster information systems found that while engineering and system design were essential, it was the hidden wiring of support networks that allows for technology to succeed.

There have been some notable projects that attempt to provide complete systems that can be used for different disasters. Most notably are the Sahana and the Artificial Intelligence for Digital Response (AIDR) projects; however, both of these systems have faced trouble getting on the ground attention. The Sahana foundation suspended its on the ground disaster response team that helped to mobilize volunteers to respond to disasters [42].

> An important message emerges from the case studies: an effective disaster information management system requires a good technological platform, but also much more. Software programs for storing, sharing, and manipulating data for disasters are being developed or patched together at a steady pace, often in the aftermath of disasters. The real difficulty lies in anchoring these technological approaches in an appropriate institutional context where they are supported by relevant and effective operating procedures, agreed terminology and data labeling, and a shared awareness of the benefits of proper handling of disaster information. Clearly, a disaster information management system must be supported by accepted rules, procedures, and relationships that encourage, facilitate, and guide the production, sharing, and analysis and use of data in response to disaster. In these case studies, the institutional dimension—the hidden wiring—determined the effectiveness of the systems. [9]

Chapter 4

Methodology and Results

As described in 2.2, the Riskmap system allows citizens to easily submit disaster reports. From 2016 to date, the system has allowed the Urban Risk Lab at MIT (MIT-URL) to gather 2229 reports before, during, and after flood emergencies in Jakarta. Similarly the Riskmap platform has also gathered 356 reports in Chennai. Figure 4-1 demonstrates that these data points include traffic reports, indications that an area is unsafe, and advice for other citizens in the area. Images attached to reports include a wide variety of scenes, from daylight-highways with cars and motorcycles to night time deserted alleys. The report data points include not only textual and image data but also estimated flood heights.The disparate datasets contained in the Riskmap System motivate the creation of REACT in order to process and summarize results in real time. REACT is trained on historical reports that are labeled 'heavy flooding' or 'no heavy flooding' and can then provide the probability that an unseen report is indicative of either.

4.1 System Design

It is paramount for disaster systems to be highly available and scalable during events. The REACT system has been designed such that it is modular and ready to scale. A simple but expandable configuration file interface serves as a singleton for sharing global environment variables such as database connections and logging capabilities.

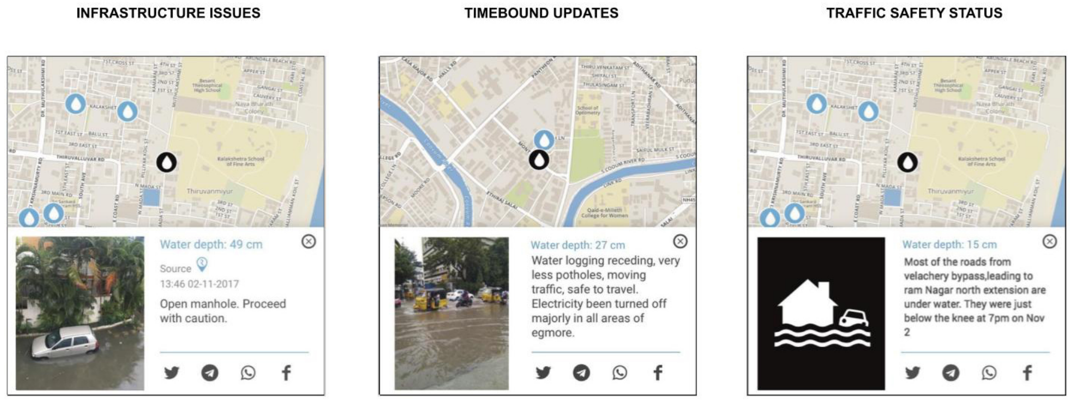

Figure 4-1: Representative reports of heavy flooding in Chennai during Nov. 2017 Monsoon. Graphic by MIT-URL

The DataLoader interface allows for flexibility in the data source that is used, while the Labeler construct allows one to use different services to create embeddings of that data. Finally, any class that follows the Learner specification is able to train and validate on those embeddings.

4.1.1 Configuration

A single configuration file allows for easy customization of the underlying training data. Figure 4-2 shows the configuration abstraction. Two default implementations of this interface include all of the data from Chennai and Jakarta in 2017. Other configurations allow the opportunity to choose only those reports that include an image, or those that come from a specific social network.

4.1.2 Data Loaders

Data Loaders are constructed by passing a configuration that implements the GenericConfig interface. They are then responsible for implementing methods that allow for retrieving the three kinds of data that the REACT system uses for prediction: text, image, and flood depth estimation. While the RiskmapLoader is the only Loader that is currently implemented, it was important to separate data loaders in order to

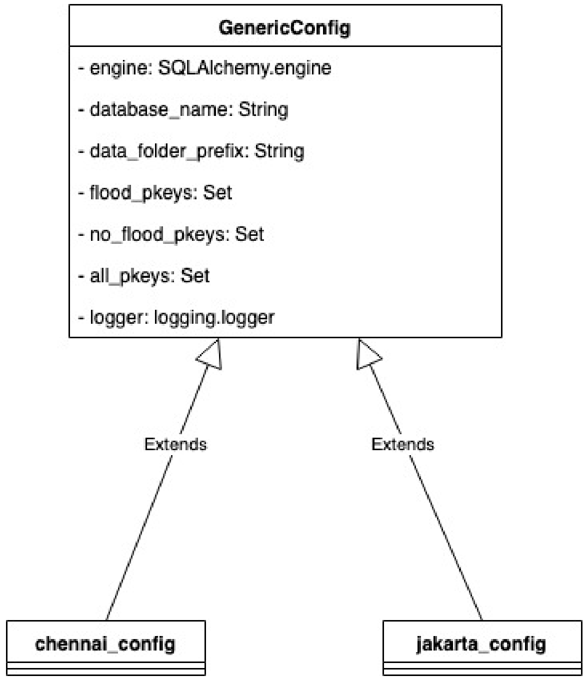

Figure 4-2: Configuration Abstraction

make REACT ready for future change.

4.1.3 Labelers

Data embeddings, or labels, turn human recognizable data into feature vectors that can be used for machine learning. There are many choices for how to label the disaster data in REACT, as such any Labeler that implements the GenericLabeler interface can be used seamlessly in the system as shown in Figure 4-4. The AwsLabeler uses AWS Rekognition to create feature vectors, while the BowLabeler uses a bag of words approach to encode textual information. An IdentityLabeler is provided in order to facilitate the passing of raw features.

4.1.4 Learners

Different learning methods are represented by different implementations of the GenericLearner interface as shown in Figure 4-4. The perceptron algorithm and the Support Vector Machine (SVM) method are implemented in REACT as simple linear classifiers. The perceptron algorithm is an iterative method that tries to find a separator

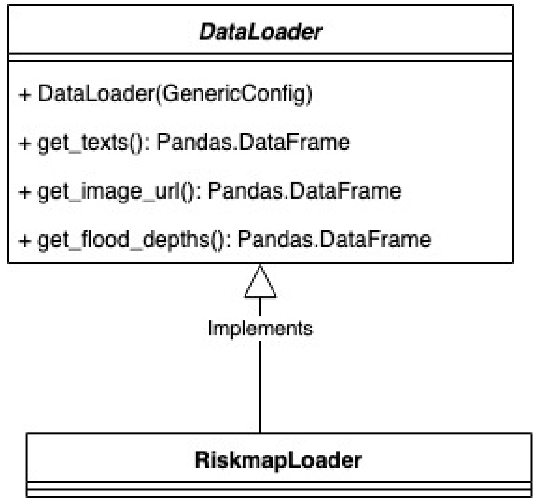

Figure 4-3: Adaptable Data Loaders

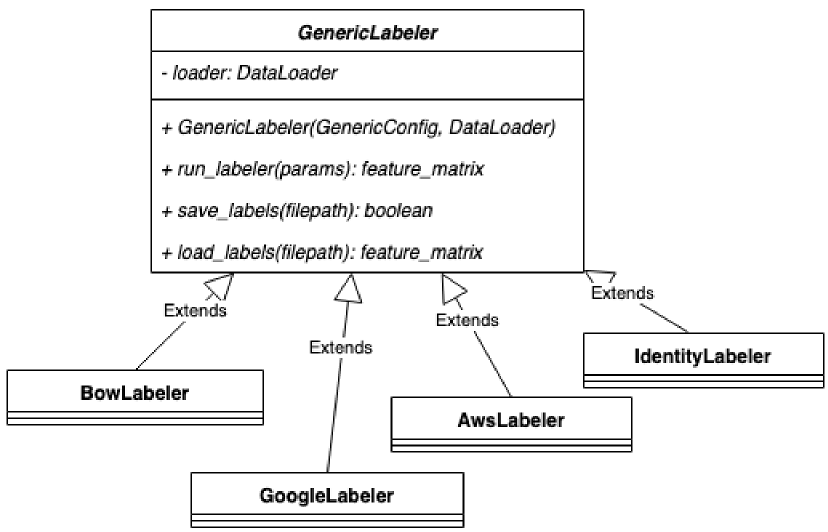

Figure 4-4: Many different labelers can be used to encode data in REACT

through simple linear transformations, meanwhile SVMs try to find the max margin separator by using a squared loss objective [43]. An IdentityLearner is also provided to act as a pass through in case raw features are desired.

4.2 Ground Truth

We explored two different methods in order to decide whether a particular report was indicative of heavy flooding or not. We first classified a report as heavy flooding if it was submitted during a time range when we knew there to be heavy flooding

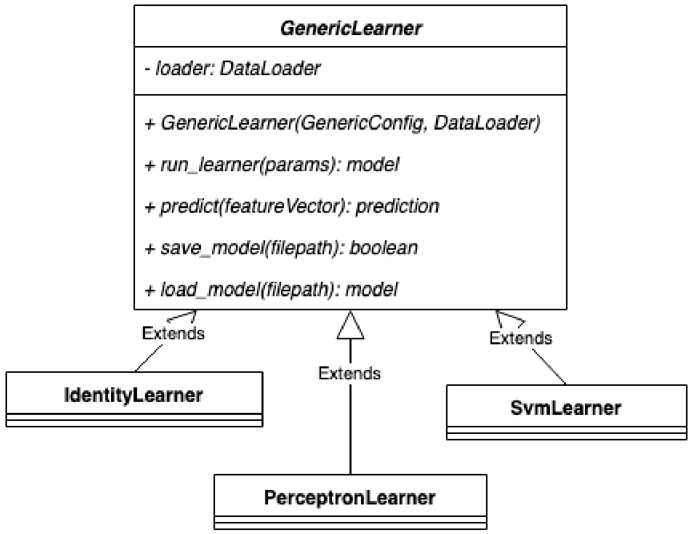

Figure 4-5: The Learner abstraction makes it easy to switch from one method to another

in the corresponding city. These shall be referred to as time-classified. Although we expected this classification to be very coarse, we hoped that it would be specific enough to create an accurate predictor. In order to test this theory we also hand labeled reports by inspecting the images that accompanied them; these will be referred to as hand-classified.

4.3 Text

As shown in Section 2.2, the Riskmap system allows citizens to provide a textual description to emergency managers. In Indonesia, most of the reports are provided in Bahasa, the local language; however, in Chennai all reports were submitted in English even though the system also supports Tamil. In both Chennai and Jakarta these reports are quite brief, with the longest reports having 140 characters. In this manner, they are quite similar to tweets which were initially 140 characters but this limit was doubled in 2017. Helpfully, this means that much of the work described in Section 3.2.1 applies to the Riskmap text corpus.

Table 4.1 contains a sample of ten reports that are indicative of those found in the Riskmap textual descriptions.

pkey	text
169	Waterlogging near cathedral road flyover
171	1st street Engineers avenue
173	Not that much water safe only
174	50cm water stagnant on the road
176	Water level rising slowly
177	Water logging
178	Model school road is completely flooded, with water almost knee deep
179	Heavy rain in West mambalam flood
180	Water on roads. Stay safe
182	4cm rainfall.. still continuing.. hope for safe .. dont come outside in night time
181	Luz signal flooded knee deep water

Table 4.1: A representative selection of report texts

4.3.1 Preprocessing

We first remove test reports, which are reports submitted in order to ensure that the system is working. The following POSTGRESQL query was executed to remove reports that were only used to test the system:

```
SELECT    pkey,
          text
FROM      riskmap.all_reports
WHERE     text IS NOT NULL
AND       Length (text) > 0
AND       text NOT similar TO '%%(T|t)(E|e)(S|s)(T|t)%%'
ORDER BY  created_at;
```

We then use python to remove punctuation and split along whitespace, thus splitting the source text into individual words without any spaces.

```
def prepare_text(report_text):
    '''
    returns a list of strings where each string is a different word
    '''
    import string
    exclude = set(string.punctuation)
```

```
s = "".join(ch for ch in inp if ch not in exclude )
return s.lower().split()
```

4.3.2 Sentiment analysis

Since each report in the Chennai dataset includes a textual description in English, we could use off the shelf sentiment analysis to gauge how negatively citizens are feeling. It might be the case that a highly negative sentiment corresponds to heavy flooding and that a positive sentiment corresponds to lighter or no flooding. We can investigate the relation between a negative sentiment and heavy flooding by using conditional probability. We set the threshhold for negative sentiment at .5 and then use the AWS Rekognition API in order to classify texts into heavy flooding when negative sentiment is greater than .5 and into light or no flooding otherwise.

We use bayes' rule in order to analyze the true positive rate— the probability that a report represents heavy flooding given a negative sentiment:

$$P(HeavyFlooding|negative) =$$
$$\frac{P(negative|HeavyFlooding) * P(HeavyFlooding)}{P(negative)} = .65$$

The false positive rate, which is the probability that there is no heavy flooding given a negative sentiment is given by:

$$P(NoFlooding|negative) = \frac{P(negative|NoFlooding) * P(NoFlooding)}{P(negative)} = .34$$

These probabilities show that while there is some relation between a negative sentiment and flooding, it is not a very strong signal. Furthermore, there are no off the shelf sentiment analysis tools for the Indonesian language, so a model based on AWS Rekogniton or Google Cloud Natural Language API would not translate to the Jakarta dataset.

4.3.3 Bag Of Words

While sentiment analysis might not be a strong enough signal of heavy flooding/ no heavy flooding, our experiment shows that the textual data contains important information. In order to train a machine learning algorithm on textual data one must first create an embedding that maps natural language into feature vectors. There are many ways of creating embeddings as discussed in Section 3.2.1, but many of them require large datasets or do not support Indonesian. For example, word2vec is a popular embedding model that produces floating point vectors and has achieved remarkable performance; however, the size of its training vocabulary is 962,000 unique words [44]. It is possible to download a pre-trained word2vec model and use it to encode new texts, but such a pre-trained model doesn't exist for Indonesian. We could train it using a different dataset of Indonesian texts, but there is no guarantee that our domain specific words would have a good embedding after having trained with a different corpus.

The bag of words encoding is particularly attractive to the Riskmap dataset because it language agnostic and can therefore work on both the Chennai and Jakarta datasets. The bag of words approach to classifying texts consists of first creating a vocabulary that maps from a token t to a unique index i. Each report text is then encoded into a feature vector by setting the ith element to 1 if the token t exists in the report text [45].

The bag of words model correctly classifies 67 percent of reports in the Chennai corpus under 5 fold cross validation. Examining the data, we see that there are many instances of reports such as 'no flooding here' which are being misclassified because the embedding is not able to understand relationships between adjacent words.

4.3.4 Bigrams

Bigrams are an embedding that allows the separator to learn relationships between adjacent words. The vocabulary is created by using pairs of adjacent words, such that ‘no flooding here’ would turn into 2 tokens: ‘no flooding’ and ‘flooding here’. Embedding the text data from the Riskmap system into a bigram vector creates a very large vector (over 3 times the size of the unitary approach), and only improves accuracy by .02 on average, as such we chose not to use a bigram embedding.

4.3.5 Results

As mentioned in section 4.2 two different classifications of the data were used:‘time-classified’ and ‘hand-classified’. The two classifications gave very similar results therefore only the results on ‘time-classified’ data are shown in Table 4.2

Dataset	Method	λ = mean 5 fold cross validation acc.
Jakarta 2017	BOW	0.707
Jakarta 2017	Bigrams	0.711
Chennai 2017	BOW	0.725
Chennai 2017	Bigrams	0.730

Table 4.2: Accuracy of Bag of Words and Bigram embeddings

4.4 Images

Visual cues are integral disaster mitigation because they allow Emergency Operations Centers (EOCs) to quickly assess the situation in different parts of the city. The Riskmap system gathered 2159 images in Jakarta and 143 in Chennai during 2017. While not every report has an image, every image contains a multitude of information that can help EOCs better understand the situation in different parts of the city. Because reports are submitted at different times and from different places the images accompanying them vary widely, from bright daylight scenes of puddles and water stagnation to barely illuminated nighttime pictures of heavily inundated alleyways. Figure 4-6 contains a representative sample of images labeled ‘heavy flooding’ from the Jakarta 2017 dataset. Figure 4-7 shows a sample of images that are labeled 'no heavy flooding'.

Figure 4-6: Examples of Jakarta 2017 reports gathered by the Riskmap System labeled 'heavy flooding'

(a) Dark Alley

(b) Daylight street

(c) Heavy flooding and heavy traffic

(d) Car submerged in water

Figure 4-7: Examples of Jakarta 2017 reports gathered by the Riskmap System labeled ‘no heavy flooding’

(a) Small puddle

(b) Drain in need of cleaning

(c) Stagnant water on the road

(d) Trash by the roadside

Dataset	Method	Best validation acc.
Jakarta 2017	Full net	0.60
Jakarta 2017	Feature extractor	0.61
Chennai 2017	Full net	0.62
Chennai 2017	Feature extractor	0.62

4.4.1 Transfer Learning

Image classification as described in Section 3.2.2 traditionally uses a deep convolutional neural network and requires millions of images. Since both the Chennai and Jakarta datasets are quite small (<5000 images), it would not be feasible to create and train our own CNN. Transfer learning uses a pretrained image classifier trained on a different dataset in order to assign labels that the original network was not trained on. We used the Resnet18 architecture pretrained on ImageNet in order to experiment with transfer learning.

4.4.2 Using Machine Learning as a Service

Results from popular machine learning as a service providers have a 'flooding' label whereas publicly available pretrained networks do not. Furthermore, as Google and Amazon Web Services (AWS) improve their own models, a classifier based on these results will also improve without the need to invest large compute resources.

AWS Image Rekognition

The boto3 python library is used to create a detect_labels API call to AWS Rekognition, which then responds with a list of (label, confidence_score) tuples. For example, AWS provides 1,319 unique labels for the Chennai 2017 dataset and 2,436 for the Jakarta 2017 images. For example, the heavy flooding report image with the unique identifier 272 shown in Figure 4-8 is labeled with upwards of 99% confidence as containing 'Nature', 'Human', and 'Flood'[1].

The report image with id 1 from the Jakarta 2017 dataset was not taken during

[1]the 'Instances' parameter is not included here for readability, but it contains bounding boxes for certain objects. It is not used in the training of the classifier.


```
[{'Confidence': 99.967,
'Name': 'Nature',
'Parents': []},
{'Confidence': 99.811,
'Name': 'Human',
'Parents': []},
{'Confidence': 99.735,
'Name': 'Flood',
'Parents': [{'Name': 'Nature'}]},
{'Confidence': 41.318,
'Name': 'Rural',
'Parents': [{'Name': 'Countryside'},
{'Name': 'Outdoors'},
{'Name': 'Nature'}]},
{'Confidence': 41.318,
 'Name': 'Outdoors',
 'Parents': []}]
```


Figure 4-8: Report image 272 and its top 5 AWS provided labels from Chennai 2017 Dataset, Riskmap India

heavy flooding. It depicts a drain that needs to be cleared ahead of the monsoon season. Labeling the image shows that the labels are quite different and that AWS correctly identifies the image as a ditch.


```
[{'Confidence': 99.723,
  'Name': 'Ditch',
  'Parents': []},
 {'Confidence': 91.929,
  'Name': 'Path',
  'Parents': []},
 {'Confidence': 91.877,
  'Name': 'Soil',
  'Parents': []},
 {'Confidence': 91.503,
  'Name': 'Vegetation',
  'Parents': [{'Name': 'Plant'}]},
 {'Confidence': 91.503,
  'Name': 'Plant',
  'Parents': []}]
```


Figure 4-9: Report image 1 and its top 5 AWS provided labels from Jakarta 2017 Dataset, Petabencana

Google Cloud Vision AI

Figure Figure 4-10 shows report image 272 as labeled by the Google Cloud Vision AI API[2].

[2]The 'mid' and 'topicality' properties have been redacted for brevity


```
label_annotations {
  description: "Water resources"
  score: 0.913
}
label_annotations {
  description: "Waterway"
  score: 0.902
}
label_annotations {
  description: "Fun"
  score: 0.801
}
label_annotations {
  description: "River"
  score: 0.784
}
label_annotations {
  description: "Flood"
  score: 0.781
}
```


Figure 4-10: Report image 272 and its top 5 Google Cloud Vision AI provided labels from Chennai 2017 Dataset, Riskmap India

Whereas the top 5 labels from AWS are very distinct between Jakarta report 1 and Chennai report 272, the label annotations are quite similar for the response from the Google API as shown in Figure 4-11

4.4.3 Visual Bag of Words

The bag of visual words as described in section 3.2.2 is used to create image embeddings out of the labels obtained from the AwsLabeler or GoogleLabeler. Here a particular feature vector has a confidence score at position i if the ith label in the vocabulary (created out of all unique labels) is present in the image [46].

Results

When the dataset size is limited and it is difficult to gather new data, it is possible to estimate future true error rates by using cross validation. Cross validation consists of splitting the dataset into k many groups. The learning model is then trained on k-1 groups and tested against the held out group. The average error over all k groups is then taken as an estimate of the true error rate of the model.

With a balanced hold out set comprising of ten percent of the data points, the


```
label_annotations {
  description: "Water resources"
  score: 0.913
  topicality: 0.913
}
label_annotations {
  description: "Waterway"
  score: 0.902
  topicality: 0.902
}
label_annotations {
  description: "Fun"
  score: 0.801
  topicality: 0.801
}
label_annotations {
  description: "River"
  score: 0.784
  topicality: 0.784
}
label_annotations {
  description: "Flood"
  score: 0.781
  topicality: 0.781
}
```


Figure 4-11: Report image 1 and its top 5 Google Cloud Vision AI provided labels from Jakarta 2017 Dataset, Petabencana

Support Vector Machine (SVM) linear classifier achieves a score upwards of $\dot{7}0$ as shown in Table 4.3. Labels from AWS Rekognition were used because AWS returned many more labels than Google did for the same image. More labels means that there are more dimensions for the classifier to differentiate between heavy flooding and no heavy flooding images.

Table 4.3: Image Classification Accuracy

Dataset	Method	Validation Method	Validation Acc.
Jakarta 2017	Perceptron	10% held out validation set	0.70
Jakarta 2017	SVM	10% held out validation set	0.71
Chennai 2017	Perceptron	mean 5 fold cross validation	0.83
Chennai 2017	SVM	mean 5 fold cross validation	0.84

4.5 Flood Height

4.5.1 Raw

The flood height parameter was added as a raw feature. Since it was a scalar, there were few degrees of freedom for a linear separator to split the space. Attempting to create a decision tree on the flood height did not create a predictor that could generalize well, neither on a held out validation set or during cross validation with random shuffles; however, the inclusion of the raw feature into the ensemble net increased performance as shown in 4.4.

4.6 Ensemble with Neural Net

In [40], Jordan and Jacobs showed that neural networks can be effectively used to vote between different classifiers that are effective only in their specific domain of the overall space. They present a case for using several weak classifiers that have middling performance in order to create a neural network that has better performance than any of the experts it uses [40]. We first train the expert SvmLearners using the BowLabeler for text and either the AwsLabeler or GoogLabeler for images. We then use the IdentityLearner in order to pass through flood height as a raw feature. The results of these learners is synthesized into a single feature vector by taking each applicable piece of data, creating its embedding, and finding the signed distance from the learned separator to the embedding. If the type of data for this separator doesn't exist, then that feature is set to zero. In this way, each datapoint that is embedded into a feature vector that has the same length as the number of Learners that were trained. The ensemble learners that had the highest accuracy were those that used all available learners as shown in 4.4. Furthermore, Table 4.4 also indicates that, the ensemble net is able to achieve scores that are higher than any one expert is able to achieve on their own.

dataset	Learners	Random 10 fold stratified cross validation
Jakarta 2017	Google,Aws,Bow,Flood	0.74
Chennai 2017	Google,Aws,Bow,Flood	0.80

Table 4.4: The learners that make the best experts and ensemble net accuracy on a randomized 10 fold cross validation split with a 10% test set

Chapter 5

Future Work and Conclusion

5.1 Future Work

As the time goes on and the Riskmap System registers more flood events, more training data will be collected. With more training data, transfer learning will become more and more accurate and the cold start problem will be less of a burden. It is also possible to use neural network embeddings for textual report data. We did not experiment with word2vec embeddings because of the difficulty of handling multilanguage datasets; however, it would be possible to create a word2vec auto embedder for Indonesian by using publicly available texts such as Wikipedia. The short coming of using Indonesian Wikipedia is that there are only 500 thousand articles compared to 5.9 million in English Wikipedia. Furthermore, it might be possible to use generative adversarial learning in order to create new example reports that fit into the distribution of user submitted reports. This would increase the size of the dataset without having to wait for disaster events to occur. One of the challenges related to generating new reports would be the multi-language problem and validating that generated report texts in other languages fit the distribution of user submitted reports in that language.

The current ensemble learner uses a very small dense neural network to reduce over fitting on small datasets. As the Riskmap System collects more data, it is likely that a larger network can be used without overfitting, thereby increasing testing accuracy.

Location information was not used because we aimed to create a per report heuristic for the whole city rather than concentrate on spatially disparate areas. In order to use spatial data it would be necessary to compensate for the sampling bias present in the dataset: richer neighborhoods have much higher rates of ownership of smart phones, which means citizens in those areas are more likely to report flooding than poorer areas— even if flooding is occurring at the same levels in both neighborhoods.

5.2 Conclusion

As global climate change continues to increase the frequency and severity of disaster events worldwide, it will become more and more important for Emergency Operations Centers (EOCs) to use all the tools at their disposal to make sense of the large amounts of data available. The REACT System has been built and trained so that it can be used as a detection tool to help EOCs find the reports that are most likely to indicate 'heavy flooding'. The REACT System sets itself apart from other disaster management systems because it is modular and adaptable to new data streams. Furthermore, the use of machine learning as a service technologies allows REACT to be less impacted by the cold start problem at the beginning of disasters and also means that the system will improve as those services improve.

Bibliography

[1] Faith Ka Shun Chan, Gordon Mitchell, Olalekan Adekola, and Adrian McDonald. Flood Risk in Asia's Urban Mega-deltas: Drivers, Impacts and Response. *Environment and Urbanization ASIA*, 3(1):41–61, March 2012.

[2] C. A Ohl. Flooding and human health. *BMJ*, 321(7270):1167–1168, November 2000.

[3] E. L. Quarantelli. Urban Vulnerability to Disasters in Developing Countries: Managing Risks. In Alcira Kreimer, Margaret Arnold, and Anne Carlin, editors, *Building Safer Cities: The Future of Disaster Risk*, pages 211–231. Disaster Risk Management Series., 2003.

[4] Mike Ahern, R. Sari Kovats, Paul Wilkinson, Roger Few, and Franziska Matthies. Global Health Impacts of Floods: Epidemiologic Evidence. *Epidemiologic Reviews*, 27(1):36–46, July 2005.

[5] Patrick Meier. *Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response*. CRC Press, Inc., Boca Raton, FL, USA, 2015.

[6] F. K. S. Chan, C. Joon Chuah, A. D. Ziegler, M. Dąbrowski, and O. Varis. Towards resilient flood risk management for Asian coastal cities: Lessons learned from Hong Kong and Singapore. *Journal of Cleaner Production*, 187:576–589, June 2018.

[7] Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. Microblogging during two natural hazards events: What twitter may contribute to situational awareness. page 10, 2010.

[8] E. L. Quarantelli. Problematical aspects of the information/ communication revolution for disaster planning and research: Ten non-technical issues and questions. *Disaster Prevention and Management; Bradford*, 6(2):94–106, 1997.

[9] Samia Amin and Markus P. Goldstein, editors. *Data against Natural Disasters: Establishing Effective Systems for Relief, Recovery, and Reconstruction*. World Bank, Washington DC, 2008.

[10] Kathleen J. Tierney, Michael K. Lindell, and Ronald W. Perry. *Facing the Unexpected: Disaster Preparedness and Response in the United States*. Joseph Henry Press, November 2001.

[11] Jacqueline Torti. Floods in Southeast Asia: A health priority. *Journal of Global Health*, 2(2), December 2012.

[12] United Nations Department of Economic and Social Affairs. *The World's Cities in 2016*. Statistical Papers - United Nations (Ser. A), Population and Vital Statistics Report. UN, September 2016.

[13] Simon Rogers. John Snow's data journalism: The cholera map that changed the world. *The Guardian*, March 2013.

[14] Kira Radinsky and Eric Horvitz. Mining the web to predict future events. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining - WSDM '13*, page 255, Rome, Italy, 2013. ACM Press.

[15] of Colorado Boulder University and of Colorado Boulder University. *Terminal Disasters: Computer Applications in Emergency Management*. Number monograph #39 in Program on Environment and Behavior. Institute of Behavioral Science, University of Colorado, Boulder?, 1986.

[16] S. Tzemos and R. A. Burnett. Use of GIS in the Federal Emergency Management Information System (FEMIS). Technical Report PNL-SA-26086; CONF-9505242-1, Pacific Northwest Lab., Richland, WA (United States), May 1995.

[17] Marcia Perry. Natural disaster management planning: A study of logistics managers responding to the tsunami. *International Journal of Physical Distribution & Logistics Management*, 37(5):409–433, June 2007.

[18] Felix Flentge, Stefan G Weber, Alexander Behring, and Thomas Ziegert. Designing Context-Aware HCI for Collaborative Emergency Management. *1987*, page 4.

[19] Eduardo Salazar. Hashtags 2.0 - An Annotated History of the Hashtag and a Window to its Future. *Revista ICONO14 Revista científica de Comunicación y Tecnologías emergentes*, 15(2):16–54, July 2017.

[20] Timeline: 20 years of major oil spills. https://www.abc.net.au/news/2010-05-03/timeline-20-years-of-major-oil-spills/419898, May 2010.

[21] Firoj Alam, Ferda Ofli, Muhammad Imran, and Michael Aupetit. A Twitter Tale of Three Hurricanes: Harvey, Irma, and Maria. *arXiv:1805.05144 [cs]*, May 2018.

[22] Philippines and PDC collaborate on supercharged disaster risk reduction programs - Philippines. https://reliefweb.int/report/philippines/philippines-and-pdc-collaborate-supercharged-disaster-risk-reduction-programs.

[23] antaranews.com. BNPB's PetaBencana.id bags UN Public Service Award. https://en.antaranews.com/news/125852/bnpbs-petabencanaid-bags-un-public-service-award.

[24] Beth Simone Noveck. Opinion | Elections won't save our democracy. But 'crowdlaw' could. *Washington Post*, 2018-10-02T09:43-500.

[25] Shruti Suresh | TNN | Updated: Oct 31, 2018, and 20:39 Ist. Chennai gets rain ready with portal for real-time mapping of flooded areas | Chennai News - Times of India. https://timesofindia.indiatimes.com/city/chennai/city-gets-rain-ready-with-portal-for-real-time-mapping-of-flooded-areas/articleshow/66452176.cms.

[26] Dharmapuri Sridhar and Maanasa Priyaa. *Real-Time Flood Mapping for Disaster Management Decision Support in Chennai.* Thesis, Massachusetts Institute of Technology, 2017.

[27] Dat Tien Nguyen, Kamela Ali Al Mannai, Shafiq Joty, Hassan Sajjad, Muhammad Imran, and Prasenjit Mitra. Rapid Classification of Crisis-Related Data on Social Networks using Convolutional Neural Networks. page 10.

[28] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. *arXiv:1310.1531 [cs]*, October 2013.

[29] Ahmed Nagy and Jeannie A. Stamberger. Crowd sentiment detection during disasters and crises. In *ISCRAM*, 2012.

[30] Hussein Mouzannar, Yara Rizk, and Mariette Awad. Damage Identification in Social Media Posts using Multimodal Deep Learning. page 16, 2018.

[31] Son Doan, Bao-Khanh Ho Vo, and Nigel Collier. An analysis of Twitter messages in the 2011 Tohoku Earthquake. *arXiv:1109.1618 [physics]*, 91:58–66, 2012.

[32] Kate Starbird and Leysia Palen. Voluntweeters:" Self-organizing by digital volunteers in times of crisis. In *Proc. of CHI (2011*, pages 1071–1080.

[33] Soudip Roy Chowdhury, Muhammad Abdullah Imran, Muhammad Rizwan Asghar, Sihem Amer-Yahia, and Carmen Castillo. Tweet4act: Using incident-specific profiles for classifying crisis-related messages. In *ISCRAM*, 2013.

[34] Cornelia Caragea, Adrian Silvescu, and Andrea H. Tapia. Identifying informative messages in disaster events using Convolutional Neural Networks. In *ICIS 2016*, 2016.

[35] Muhammad Imran, Shady Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. Practical extraction of disaster-relevant information from social media. In *Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion*, pages 1021–1024, Rio de Janeiro, Brazil, 2013. ACM Press.

[36] Dat T. Nguyen, Ferda Ofli, Muhammad Imran, and Prasenjit Mitra. 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, ASONAM '17, pages 569–576, New York, NY, USA, 2017. ACM.

[37] H. S. Jomaa, Y. Rizk, and M. Awad. Semantic and Visual Cues for Humanitarian Computing of Natural Disaster Damage Images. In *2016 12th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, pages 404–411, November 2016.

[38] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, August 1996.

[39] Cesare Furlanello and Stefano Merle. Boosting of Tree-Based Classifiers for Predictive Risk Modeling in GIS. In *Multiple Classifier Systems*, Lecture Notes in Computer Science, pages 220–229. Springer Berlin Heidelberg, 2000.

[40] Michael I. Jordan and Robert A. Jacobs. Hierarchical Mixtures of Experts and the EM Algorithm. *Neural Computation*, 6(2):181–214, March 1994.

[41] Luis Perez and Jason Wang. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv:1712.04621 [cs]*, December 2017.

[42] Sahana AGM 2018 - President's Report.

[43] Christopher Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer-Verlag, New York, 2006.

[44] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed Representations of Words and Phrases and their Compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc., 2013.

[45] Diksha Khurana, Aditya Koli, Kiran Khatter, and Sukhdev Singh. Natural Language Processing: State of The Art, Current Trends and Challenges. *arXiv:1708.05148 [cs]*, August 2017.

[46] Jun Yang, Yu-Gang Jiang, Alexander G. Hauptmann, and Chong-Wah Ngo. Evaluating bag-of-visual-words representations in scene classification. In *Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval - MIR '07*, page 197, Augsburg, Bavaria, Germany, 2007. ACM Press.