PNNL-29173

Pacific Northwest
NATIONAL LABORATORY

Multi-Formalism Modeling for Disaster Resilience, Forecasting and Response

September 2019

AM Coleman
JD Tagestad
MJ Henry
JP Almquist
JJ Harrison
JD Hendry
IR Herrera
KB Larson
AT Sligar
WC Vilwock

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-29173

Multi-Formalism Modeling for Disaster Resilience, Forecasting and Response

September 2019

AM Coleman
JD Tagestad
MJ Henry
JP Almquist
JJ Harrison
JD Hendry
IR Herrera
KB Larson
AT Sligar
WC Vilwock

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Executive Summary

Disaster response preparedness, risk mitigation, and effective post-event response are critical for national security, public health, and economic integrity. For areas impacted by extreme events, timely and accurate impact assessments of lifeline infrastructures are critical. Traditional disaster management and situational awareness platforms are currently not able to address the massive influx of available information from numerous sources that have the capability to enhance situational awareness. Next-generation analytic assessment approaches are researched here, where the use of dynamically sourced, heterogeneous data with near real-time availability are integrated into a continually-updated analytics framework.

The goal of this research is to develop a foundational multi-formalism modeling platform that can integrate heterogeneous data sources for use in improving situational awareness after a disaster event. The specific objectives to achieve this goal are to 1) develop an architecture to enable structured and unstructured multi-sensor data fusion using machine/deep-learning techniques; 2) clean, classify, geolocate, and temporally separate novel data sources for use in spatially-enabled data fusion; and 3) derive confidence-based observation data with machine-learning based prediction to provide a spatiotemporally consistent assessment of event-driven damage.

The research presented here outlines and implements an analytic workflow to bring well-known data together to develop enhanced situational awareness. These data include optical and radar satellite imaging, oblique-perspective disaster reconnaissance images, high-frequency *in situ* measurements, and numerous social media images/data sources. A use case for methods development is based on the flooding events of Hurricane Florence, largely impacting North Carolina and South Carolina in September 2018. Data fusion between authoritative geographic data (e.g., satellite imaging) and non-authoritative crowdsourced data (e.g., social media) is an underlying theme throughout this research and methods are tested bring these data to a common geographic form through data cleaning, relevancy classification, geolocation, feature extraction, feature matching, and data transformations.

To enable the efficient use and testing of various models and methodologies, this project developed a dynamic and adaptable cloud-based pipeline architecture for ingesting, storing, and analyzing incoming data, executing a range of machine/deep-learning models, and visualizing the information. The pipeline development was informed by previous PNNL investments (StreamKit from the Analysis in Motion Initiative) and built upon leading-edge technologies.

Machine-/deep-learning models were selected and tested to clean and classify social media images for relevancy and presence of flood waters. Numerous relevancy models were trained and tested and benchmark measures on a validation dataset reveal good performance, primarily with convolutional neural network (CNN) models. To validate/improve the geographic location of both social media and oblique aerial images, we tested a multi-step workflow built on a deep-learning models, Xception and Faiss, to identify similar features from source and reference scenes. Matching reference scenes to a set of test scenes with the closest similarity index had limited success in improving the precision of test scene location and additional methods have been identified for testing.

The automated extraction of supporting information for disaster response from oblique aerial images or ground-level social media images is accomplished through semantic image segmentation and transformation. The non-authoritative data are used to validate the authoritative data collections and processing and optionally, nowcast flood models. In addition, common form geographic data are combined with other static landscape features to develop spatially-explicit training vectors used in deep-generative statistical inference models. Complex, non-linear relationships in static and dynamic data are developed and ultimately derive a spatially continuous and probabilistic assessment of flood damage with the event domain. Preliminary results of this model are provided.

Acknowledgments

This work was funded by a fiscal year 2019 Laboratory Directed Research and Development (LDRD) Energy Mission Seed grant issued from the Energy and Environment Directorate at the Pacific Northwest National Laboratory. Special thanks to Dr. George Muntean, Chief Engineer in the Energy Environment Directorate, and the anonymous project peer reviewers who saw and supported the vision in this work.

Contents

Figures

Tables

1.0 Introduction

Natural disasters are categorized into four general groupings that include geophysical, meteorological, hydrological, and climatological (Below, Wirtz and Guha-Sapir 2009, Bouyerbou, Bechkoum and Lepage 2019). Further, the distinction of natural "disaster" only applies if minimum economic/insured loss, fatalities, injuries, or damages are realized, thus an intersection of natural catastrophe with human establishment (Benfield 2018). There are increasing trends of natural disasters and associated economic loss, social burden, and vulnerability. Globally in 2017, economic loss is reported as 93% higher than the 2000-2016 average, and a 35-year (1980-2014) trend analysis of loss shows hydrological, meteorological/climatological, and geophysical events an increase of 300%, 200%, 50% respectively (Benfield 2018, Hoeppe 2016). The amplification of flood frequencies is also noted for coastal areas due to sea-level rise independent of changing climatological factors (Buchanan, Oppenheimer and Kopp 2017).

Effective natural disaster resilience (i.e., preparedness, risk mitigation, damage reduction, adaptation), response and recovery are critical for national security, public health, environmental health, and economic integrity (Hoque et al. 2017, Albright and Crow 2015, USACE 2015, Haworth 2017). The evidenced increase in the frequency and magnitude of storm and flood events is causing more severe and sustained impacts across various critical infrastructure that serve as lifeline functions (Murray and Ebi 2012). Damage to critical infrastructure caused directly by wind and floodwater, and/or indirectly by wind-blown debris, downed vegetation, flood debris, and physical access barriers, can impact power generation, transmission and distribution lines, substations, refineries, oil and gas pipelines, pump stations, wastewater treatment, dikes, bridges, and transportation networks. These intense events are changing how disaster risk and resilience are perceived and how response and hazard mitigation are defined and implemented (Albright and Crow 2015). Frequent damage assessments, both pre-, peri- and post-event, for a disaster impact area are vital for effective resource planning, risk evaluation, resource planning, and recovery and restoration efforts. The disaster and infrastructure response communities have been criticized for not producing timely and accurate disaster impact assessments, which is most effective and often mandated within 24-48 hours after event initiation, daily/sub-daily throughout the event duration, and at regular but less-frequent intervals in the restoration phase (Hodgson et al. 2014). For areas impacted by natural disaster, timely, accurate and comprehensive impact assessments are critical and can impact the cadence and reliability of situational awareness and subsequently, effective response action.

1.1 Data Availability

Historically, the availability of data necessary for damage assessment and speed of analysis were relatively constrained, yet these data have been dominantly produced by trusted and authoritative sources. Increasingly, the world is being observed through complex distributed networks of formal (e.g., spaceborne and airborne imagers, stream gages) and informal sensors (e.g., personal cell phones, geotagged social media). These sensors relay a variety of data in structured and unstructured forms at unprecedented volumes, velocity, variety, and veracity (Figure 1; https://www.ibmbigdatahub.com/infographic/four-vs-big-data). In general, these data can be categorized as authoritative or non-authoritative where authoritative data are well-known, well-structured, documented and typically collected under a formalized sampling plan from a well-known source (e.g. satellite-based remotely-sensed imagery, elevation models, census data, soils, in situ sensors, etc.). Increasingly, the data from these various types of sensors are either purposely or inherently geospatial and can be linked to various geographies (Miller and Goodchild 2015, Chen et al. 2016, Goodchild 2016a, Goodchild 2016b, Li 2016, Liu et al. 2016, Shu 2016, Lewis and Park 2017). There is a growing realization about the immense amount of heterogeneous data that can potentially provide rich observations of the impacts from natural disasters, but cannot be readily used or easily integrated into automated analytical frameworks (Lee and Kang 2015).

In general, these data can be categorized as “authoritative” or “non-authoritative” where authoritative data are planned, well-known, well-structured, documented and typically collected under a formalized sampling plan from a well-known source (e.g. satellite-based remotely-sensed imagery, elevation models, census data, soils, in situ sensors, etc.). Non-authoritative data (e.g., text and image-based social media, news feeds, traffic cams, earth cams, aircraft-based oblique photos, smartphone sensing) are typically unstructured, irregular in space and time, crowdsourced or volunteered, and consequently are not commonly used in geospatial applications because these data aren’t readily usable (Schnebele et al. 2014, Cervone et al. 2016). It is estimated, however, that 80% of newly collected data fall into the realm of non-authoritative (Andriole 2015).

Presently, U.S. federal agencies spend as much as 70% of their budget on data preparation, including geospatial, suggesting the need for automated data processing pipelines to reduce the time from collection to decision making. As a major component of authoritative data, the increasing availability, frequency, and quality of remotely-sensed imagery from earth-observing sensor platforms (e.g., satellite, airborne, Unmanned Aerial Systems or UAS) continues to progress towards high spatiotemporal damage detection (e.g., physical change in ground, structure, and vegetation; flood detection; debris detection) across large geographic domains. Despite these advances, there are still inherent limitations with timing and

frequency, surface or atmospheric obstructions, and accuracy of algorithms that can rapidly translate spectral or radar signals to an actionable result.

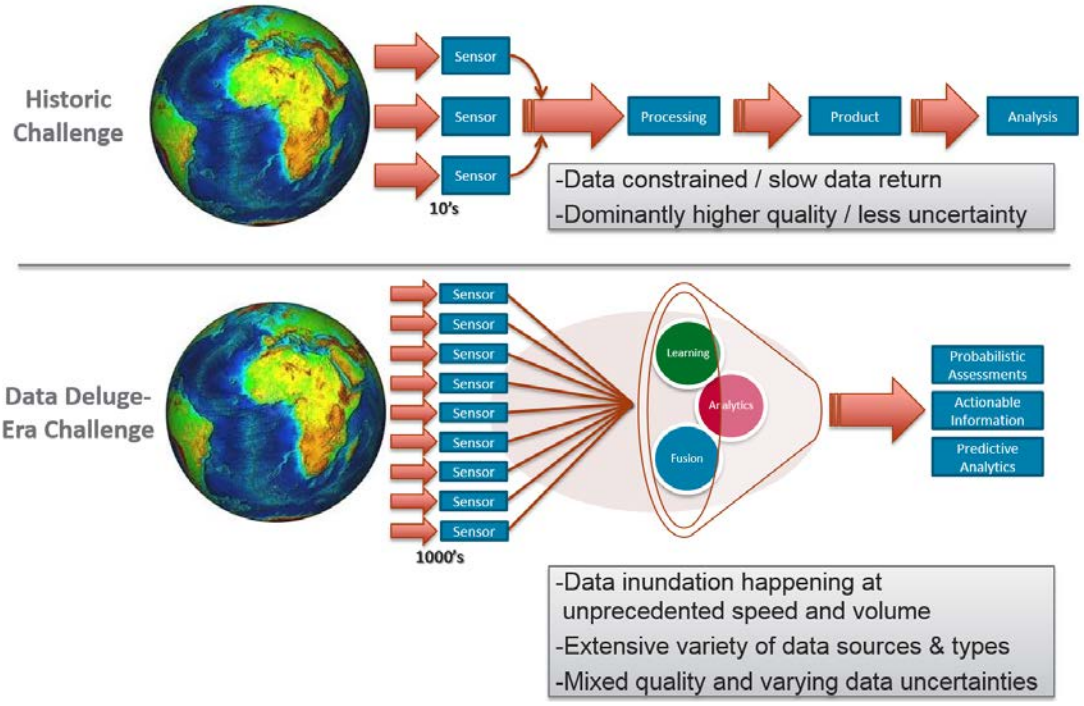

Figure 1. Conceptual illustration of historic and current data challenges with Earth Observation data collection and availability.

A growing body of research is evolving with new approaches for how to use the deluge of heterogenous data to help drive a more informed disaster management strategy. Three high-level phases of a disaster event need to be considered in this endeavor: 1) predictive risk assessments, hazard mitigation, and preparedness actions are established prior to a disaster event; 2) regularly updated information are produced for situational awareness to provide effective emergency response and resource allocation; and 3) post-event damage assessments are completed in a timely and accurate manner to inform response and recovery planning and operations (Comfort, Ko and Zagorecki 2005, Smit and Wandel 2006). Real-time analytics using heterogeneous data feeds are applicable in all three phases and will drive future initiatives in data fusion and automated analytics (Chen et al. 2016).

The human decision space can be aided by providing a continuum of automated predictive analytics and decision points resulting from a stream of temporal information. This decision space can utilize different time-steps depending on the data availability and the application requirement. However, traditional disaster management approaches are not capable of handling the massive speed and variety of available information. Next-generation analytic assessment approaches that focus on use of dynamically sourced,

multimodal data with near real-time availability, can be binned into three categories: 1) observation-based, which use various sensor data to capture damage states in pre-, peri- and post-event states; 2) inference-based, where numerical models can predict what cannot be observed; and 3) fusion-based, where both observation and inference modeling are implemented to provide system state updates. The accuracy and quantification of uncertainty in these analytical approaches are critical to effective response actions.

The practice of data fusion is well-known within the remote-sensing/image analytics domain, but it is relatively limited within the broader geospatial domain. Data fusion should not be confused with "data mashups", where multiple sources of data are brought together in a dashboard application primarily for visualization purposes. Research on geospatial data fusion, where multimodal data such as remotely observed imagery, social media and news feeds, in-situ measurements, traffic cameras, human mobility data, topographic data, land cover, are modeled to bring an analytic outcome greater than an individual dataset, is a current research challenge (Goodchild 2016b). The recent and successive advancements in neural learning and computer vision methods, including deep convolutional neural networks, deep belief networks, deep Boltzmann Machine, deep residual learning, and back propagation neural networks, have been proven effective in many multi-sensor data fusion applications (Ngiam et al. 2011, Deng and Yu 2014, NITRD 2016). The development and application of these methods in the geospatial sciences represents a new and active research field known as "Geospatial Artificial Intelligence" (GeoAI), from which disaster management can directly benefit (Mao et al. 2018). The notion of multi-formalism modelling plays a key role here, where different modelling methods (e.g., machine learning, computer vision, image processing, geospatial analysis, numerical modeling) are brought together under a common generalized framework (Gribaudo and Iacono 2014) to assess current damage conditions as a function of the data modality.

Current methods of practice for situational awareness within disaster management are rapidly becoming obsolete for dealing with the ever-increasing volume, velocity, variety, and veracity geospatial and pseudo-geospatial data. This is a vulnerability that compromises and even paralyzes our ability to improve the speed at which we can assess highly dynamic disaster events and make decisions considering the use of available information.

1.2 A New Generation of Disaster Management

The new paradigm of dynamically sourced, heterogenous geospatial "big" data is spawning a new generation of application areas, including in disaster management. The notion of volunteered geographic information (VGI) or "citizen sensors" (non-authoritative data) is an established part of geospatial big

data and has been used during natural disasters (Goodchild and Glennon 2010, Yin et al. 2012, Crawford and Finn 2015, Schnebele et al. 2015, Cervone et al. 2016, Kryvasheyeu et al. 2016, Haworth 2017). There are other invaluable data sources that come from authoritative sources, but the data are not readily usable in a geospatial context. Examples imagery from the Civil Air Patrol (CAP) and Customs and Border Protection (CBP) in the United States who consistently provide thousands of oblique- and nadir-perspective images quickly following disaster events (Sava, Clemente-Harding and Cervone 2017). Non-authoritative sources, including videos posted online taken from UAS, quickly provide unique perspectives in a disaster event though can often be unusable due to copyright issues (Lewis and Park 2017).

There is potential for spatial and temporal gaps to be filled by the use of non-authoritative data. The use of social media feeds in disaster management has been a large area of research with specific challenges defined around quality, biasing, security, and ethics, and how these can be used effectively within existing governance structures of top-down disaster management controls (Crawford and Finn 2015, Haworth 2017, Schnebele et al. 2015). Haworth (2017) further notes the potential use of VGI not only for disaster response, but also hazard mitigation and resilience. Three key findings about the use of VGI include: 1) scale determines the volumes and types of data contributed and the experience of the contributors; 2) uncertainty, trust, and unintended consequences in the contributed data are key concerns; and 3) the digital divide inherently selects a dominant demographic that is able/willing to participate in VGI and thus has the potential to bias data (Haworth 2017). Thus, there is a key need to enhance research ethics and good practice with respect to appropriate data use, rights of individuals, privacy concerns, and rights of the public (Liu et al. 2016).

Innovative approaches are being used with data not traditionally used in geospatial analysis. For example, Sava et al. (2017) developed a semi-automated supervised classification approach to determine flood/non-flood areas of images taken by handheld cameras from aircraft (e.g. Civil Air Patrol). This kind of analysis can be used alongside authoritative data sources and traditional methods to provide a source of validation and uncertainty quantification. Conversely, non-authoritative data sources can be used to inform authoritative data collections. Cervone et al. (2016) used disaster-focused Twitter feeds to help define where high-resolution commercial remote-sensing should be tasked and collected. Schnebele et al. (2015) explored including real-time mobility data to execute dynamic evacuation plans preceding an event, and using oblique aerial imagery, social media, ground observations, and numerical inundation models to aid in post-event transportation infrastructure damage assessments.

An innovative example of a geospatial big data disaster management system is described by Huang, Cervone and Zhang (2017). A generalized cloud-based automated disaster analysis system was built to

source multiple streams of social media and detect developments of a disaster event in real-time, map these events, and fuse these data with remotely-sensed data. This development begins to address several geospatial big data challenges with respect to using a rapidly scalable cloud computing platforms such as Hadoop, MapReduce, Apache Mahout, and Apache Lucene. The system consists of multiple integrated components including automated crawlers that gather a variety of data, storage systems that index retrieved data, data integration and interoperability systems, machine learning to support data mining and big data analytics, and web-based user interfaces to support data searches, analysis, and visualization (Huang et al. 2017). Initial construction of an event database is required to define and train the system for the types of events that are of interest. In addition, there are required configurations for data retrievals, the detection of relevant social media hashtags, and real-time event tracking (Huang et al. 2017). Historical event reconstruction is also possible and the authors emphasize the utility of fusing social media and remotely-sensed data in this context. This effort provides a significant step forward in the next-generation of disaster management.

1.3 Research Objectives

The work presented here documents efforts under an FY19 Pacific Northwest National Laboratory (PNNL) Lab Directed Research and Development (LDRD) project. The objectives of this research are to develop and demonstrate a foundational multi-formalism modeling platform that will:

1. Classify, geolocate, and temporally bin novel data sources for use in remotely-sensed overhead imagery validations of flood events and for spatially-enabled machine-learning based data fusion;
2. Implement a pipeline architecture that enables testing of multiple machine-learning data fusion and feature learning methodologies with various multi-sensor (authoritative and non-authoritative) datasets;
3. Derive confidence-based observation data with machine-learning based spatiotemporal prediction to fill spatial and temporal gaps; and
4. Provide observational system states to a physics-based numerical flood forecast model (PNNL's RIFT model) to enable high-value risk forecasting, reduced-time to restoration, and improved resilient capacity.

While damage from disasters exists in many forms, the focus of this initial work is on flooding that occurred primarily in North Carolina and South Carolina during Hurricane Florence in September 2018. This set of objectives knowingly provided an aggressive research agenda for a limited time and budget

allocation but provides an early exploration and tool development in this research space and establishes a roadmap for future efforts.

1.4 Research Plan

The research plan follows the objectives and is defined as follows:

1. Novel (non-authoritative) source data will be triaged, structured, inventoried, labeled (for machine learning training), temporally binned, and when necessary, geolocated using geolocation estimation method (Johns, Rounds and Henry 2017a). The classification of these data entails the use of convolutional neural network image training and classification that is domain specific to disaster environments. The work described here would constitute the initial assembly of a machine learning training library for natural disasters/damage assessments which to the knowledge of the team, does not exist. This library will eventually be made publicly available to further root the work effort.
2. Implement and test the statistical design for validation between remotely-sensed data and novel observation data for one event case.
3. Implementation and evaluation of the machine-learning based data fusion methods using a matrix of machine and deep learning models that are pre-trained, parameter tuned, or trained from scratch. The input training and validation data sources include varying remotely-sensed imagery, traditional geospatial data such as digital elevation models and derivatives such as landform, land use/land cover, and hydrography, and the novel sources of data to inform largely on ground-based observation conditions and will generate confidence-based flood observation data.
4. Implementation of a cloud-based data streaming architecture with prototype methodologies (i.e., novel data image classification; data fusion) implemented for testing.
5. A “gray box” model implementation using best performing data fusion results as observational flood states/boundary conditions for the PNNL RIFT flood forecast model. The hypothesis is the combined machine-learning and physics-based modeling will significantly improve flood forecasting results. Performance metrics will be reported with combinations of parameters used and will be fully tested over an event time-series.
6. If the “gray box” modeling proves to be of value, this will be integrated into the cloud-based streaming architecture and thus frames a prototype multi-formalism model for improvement of situational awareness and disaster management.

A high-level overview of the multi-formalism modeling (MFM) methodology and outputs is presented in Figure 2 and is described in greater detail in the following sections. The general notion is to source and retrieve authoritative for a disaster event and process each data source by its own specific workflow to determine the presence or absence of flooding. These damage determinations either exist in common geographic form (i.e., authoritative data; latitude, longitude, attribution) or must proceed with data-dependent transformations to achieve common geographic form (i.e., non-authoritative data). The non-authoritative data in common geographic form are used to validate the source authoritative data processing (remotely-sensed flood detection). Optionally, a feedback mechanism can be triggered by a minimum accuracy score to exercise a range of parameters within the remotely-sensed flood detections to test improvement in classification accuracy. On a different pathway, a majority subset of the non-authoritative data in common geographic form are combined with other relevant information (e.g., remotely-sensed flood detections, elevation, slope, land use/land cover, horizontal and vertical distance to existing waterbodies) into training datasets for a machine learning process to determine a spatially continuous and probabilistic assessment of flood damage. A portion of the non-authoritative data is set aside to perform an accuracy assessment on the damage prediction.

The remaining sections of this report detail specific research developments conducted under this LDRD project. Section 2.0 describes the use of satellite and airborne remote-sensing for rapid disaster response and the role that non-authoritative data plays in doing validations, particularly with regards to flooding. This section also documents the sources of data used in this exploratory research. Section 3.0 is focused on the numerous machine learning and deep learning models implemented to 1) determine the relevancy of non-authoritative data; 2) determine or validate the geographic location of non-authoritative social media images and oblique aerial imagery; 3) perform feature matching between non-authoritative data and reference data to enable data transforms into geographic space; and 4) run pixel-level semantic image segmentation to convert a non-authoritative image into a few relevant groupings. Section 4.0 discusses the early development of the statistical inference model, taking the native and transformed geographic data and using this information for training/testing/validation patterns to generate a probabilistic surface of flooding. Finally, Section 5.0 describes the cloud-based streaming data pipeline, architecture, API, and user-interface developed to use and process the large amount of data in for numerous models.

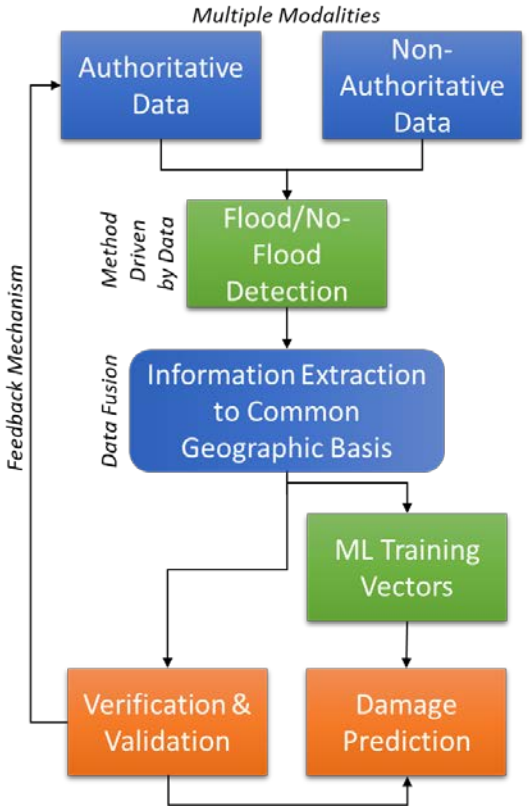

Figure 2. A conceptual workflow for multimodal data fusion to use authoritative and non-authoritative data for validation and probabilistic prediction of flood damage (ML=machine learning).

2.0 Remote-Sensing Only Rapid Response Workflow

A functional and event-exercised system developed for rapid response damage assessments should span event monitoring and action, image acquisition and retrieval, image pre-processing, algorithm deployment, geoanalytics, and analytics delivery (Coleman et al. 2017). The workflow of this system, termed "Rapid Analytics for Disaster Response" (RADR), is graphically presented in Figure 3. The motivation behind RADR is to provide actionable analytics to utilities, emergency operations centers (EOCs), emergency response organizations (EROs), and others who benefit from damage assessments throughout a disaster event. RADR is built using Python scripting and automation and focuses on the use of various remotely-sensed imagery assets from multiple sensor sources. A key goal is to best utilize available high to medium (<15m) spatial resolution sensors that collect data within an event domain, such that daily or sub-daily assessments can be achieved.

Once an event notification is received, a set of activation actions are performed to prepare for the event. The first is to define an area of interest (AOI) that is used to establish satellite imaging search domains for pre-event imagery within the past 1-2 months, and to reduce data volumes and processing time. Multiple web resources and situational awareness reports are used to understand event progression, reported damage

areas and critical areas of concern, and help develop a common operating picture. Once new imagery is retrieved, the data is unpacked and processed using algorithms specific to the sensor and the disaster event type. The system is currently designed to perform flood detection, rubble/debris detection, vegetation change detection, and pre-and-post event change detection. An overview of these damage detection algorithms is presented in Table 1 and examples of flood detection are presented in Figure 4 and Figure 5. Additional geoanalytics can be performed with assessed pixel-level damage/no-damage to identify potential impacts to specific features such as transportation networks, energy infrastructure, shipping ports and terminals, water and waste water management facilities, property, businesses, and homes. The resulting image analytics and geoanalytics are packaged and disseminated through a variety of formats and channels. RADR provides the base component for satellite image damage assessments that is supplemented with novel, non-authoritative data. For the multi-formalism modeling (MFM) effort discussed here, the flood flood detections from RADR are used. In the future, additional damage types can be integrated.

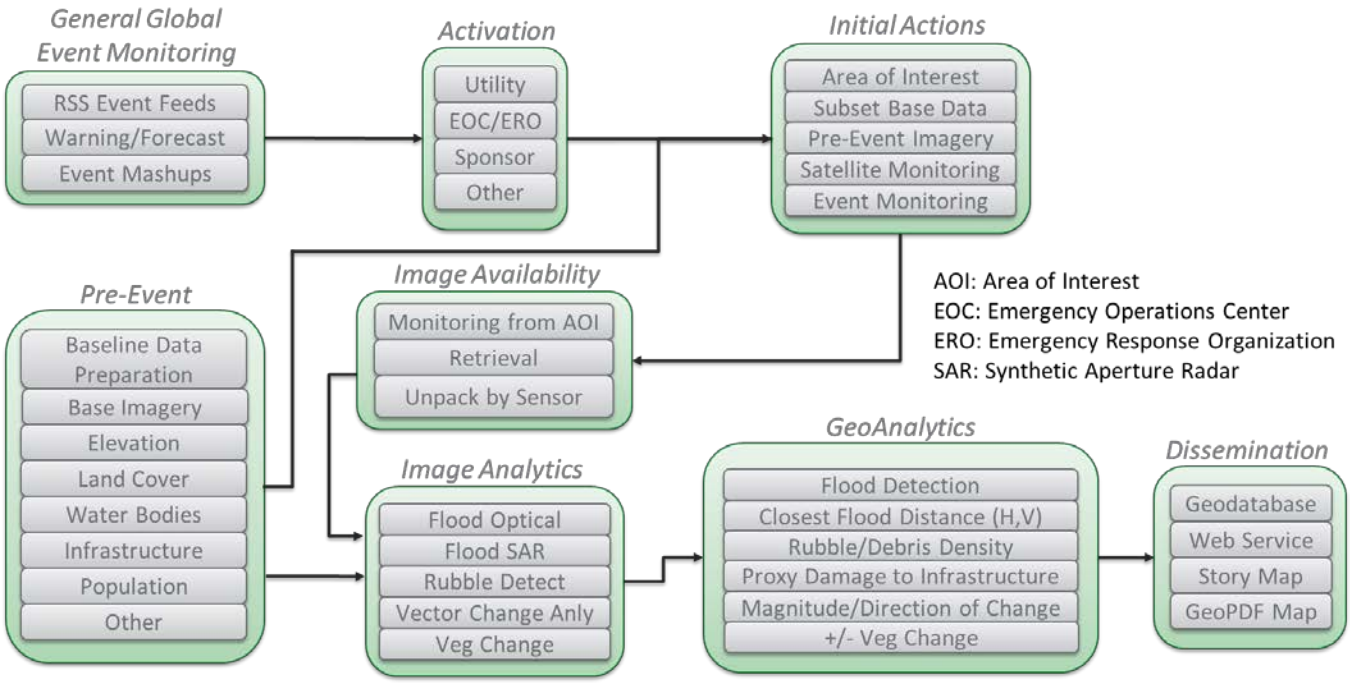

Figure 3. The "Rapid Analytics for Disaster Response" (RADR) workflow for disaster event monitoring, activation, satellite remote-sensing monitoring and data retrieval, image analytics, geoanalytics, and data dissemination (Coleman et al. 2017).

Table 1. Overview of RADR damage detection algorithms.

Algorithm	Description	Event Type	Image Type
Flood Detection (optical)	Detect flooding in single image using a modified Normalized Difference Water Index (NDWI) and fuzzy logic classifier	Inundation occurring with heavy or prolonged rains, hurricane	Multispectral imagery with a NIR band

Flood Detection (SAR)	Detect flooding from either single or pre-post event pair Synthetic Aperture Radar SAR (SAR) scene(s) using an auto-thresholding approach for single scene or change detection for pre-post pair.	Inundation occurring with heavy or prolonged rains, hurricane	SAR; C-band commonly used.
Rubble Detection	Detect rubble in a single image and use the density of detections to derive a proxy classification of the level of damage.	Structure damage due to high winds from hurricanes, tornadoes, or severe storms	High-resolution natural color or multispectral
Vector Change Analysis	Calculate magnitude and direction of change between “before” and “after” images, then apply rules to distinguish change caused by event from other types of change not associated with damage.	Destruction/loss of vegetation, and exterior damage to structures that changes shape/appearance	Mid- to high-resolution multispectral + land cover data
Vegetation Change Detection	Positive or negative change in Normalized Difference Vegetation Index (NDVI) in “before” and “after” images.	High wind events affecting vegetation	Low- to mid-resolution multispectral

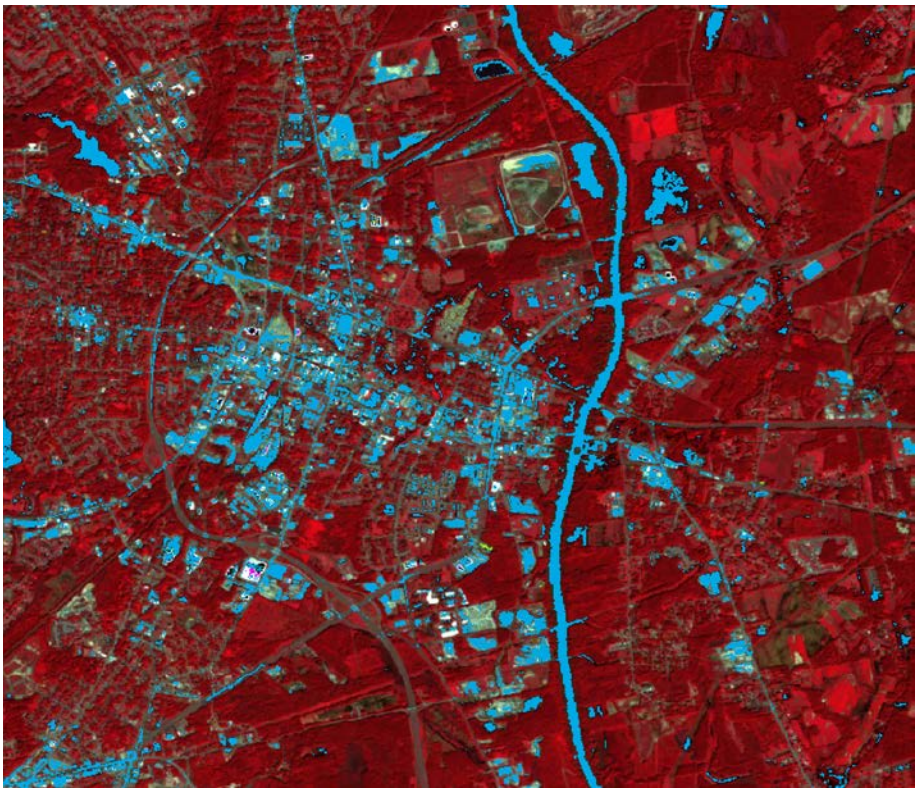

Figure 4. Example flood detection (blue) from multispectral satellite image over Fayetteville, North Carolina during Hurricane Florence, (September 17, 2018).

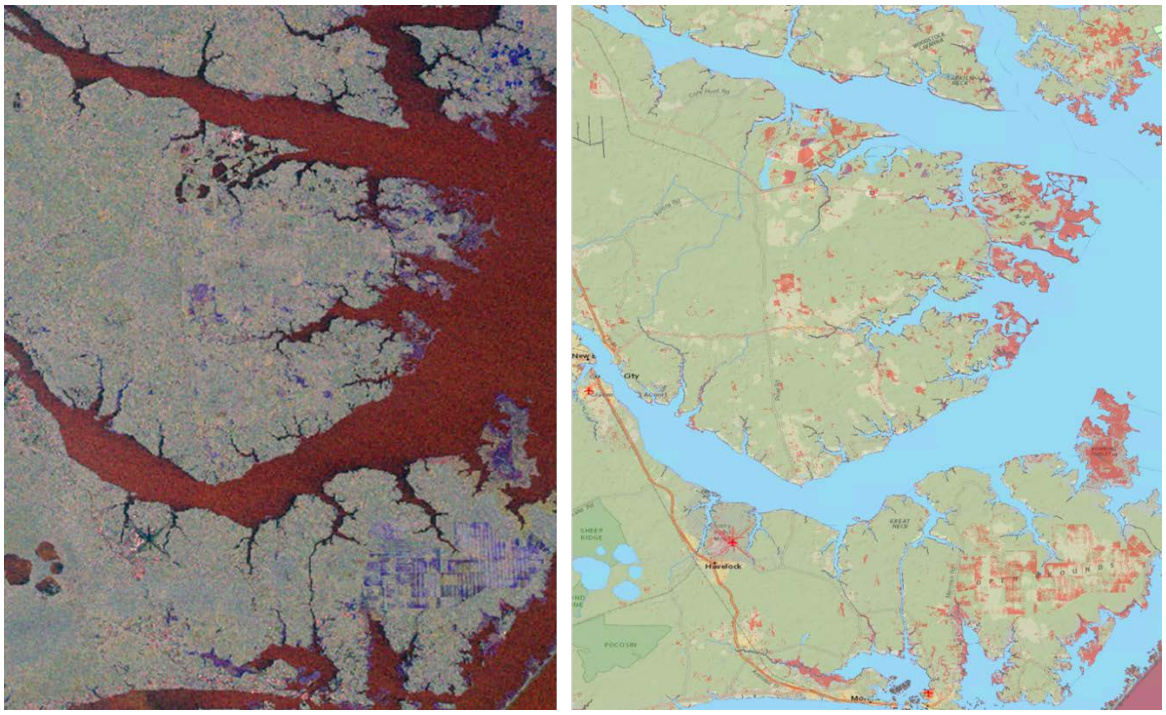

Figure 5. SAR flood detection using pre- and post-event imagery near New Bern, NC, September 15, 2018 at the onset of Hurricane Florence. The colorized SAR polarities (left) and the resulting classified flood detects are shown in red (right).

2.1 Multimodal Data Sources

2.1.1 Remote Sensing Data

Two primary types of remotely sensed imagery are used in hurricane-based flood assessments: synthetic aperture radar (SAR) (e.g., Sentinel-1, RADARSAT-2, COSMO-SkyMed, PALSAR-2, TerraSAR-X, UAVSAR) and multispectral optical (e.g., WorldView 1-4; SPOT 5-6, PlanetScope, RapidEye, Pléiades). SAR is a side-looking radar technology that uses the sensor motion and different wavelength and signal polarities to capture high-resolution image of surface physical characteristics (1-30 m ground sample distance [GSD]). A key advantage of SAR is that because it is an active sensor it can collect data day or night and effectively view through cloud cover or other obscuring atmospheric conditions. Certain SAR sensors such as UAVSAR utilize longer radar wavelengths (1-2 GHz, L-band) that provide more signal penetration through vegetation cover to detect ground-level flooding.

Multispectral sensors are multichannel detectors that measure reflectance across several wavelength ranges in the electromagnetic spectrum, including visible (e.g., blue, green, yellow, red) and non-visible wavelengths (e.g., NIR, shortwave infrared). Multispectral data provide for analytical versatility and is one of the most widely used imagery types; however, images captured during cloud-free daylight conditions are required for assessing damage.

Because each sensor follows a different orbit cycle, data availability for a given location varies daily and may not coincide with optimal times to detect damage. Additionally, each sensor and sensor pass will collect a different geographic extent, creating a patchwork of imagery over the disaster domain. Thus, it is critically important to use multiple sensors and types of sensors to improve situational awareness.

Algorithms for detecting water vary by sensor and organization, some of which are proprietary or not published. A basic description of the algorithms deployed in RADR are reported here, but further descriptions are found in Coleman et al. (2017).

2.1.1.1 SAR Flood Detection

Each SAR sensor is unique with regards to acquisition model, polarity, radar frequency, and resolution, making it challenging to design generalizable damage detection algorithms for SAR data. While sensor-specific algorithms are often required, all attempt to utilize the difference in the dielectric signal response of water versus other area land surfaces, and locations of existing waterbodies, to identify flooding. There are many analytical approaches to estimate a signal amplitude threshold that best separates water and non-water features. Much attention has been given to methods that automate selection of a threshold, and some provide an option for a human in the loop to override or fine-tune the auto-selected threshold. In the context of damage assessment, both these capabilities are considered important and have been integrated into SAR flood detection algorithms in RADR. A challenge with auto-thresholding approaches is the ability to accurately respond to different compositions of water/non-water in a given scene and maintain acceptable levels of commission or omission error. Current SAR flood detection methods in RADR have been designed with this consideration in mind because they need to be robust given the unpredictable nature of when and where flooding may occur.

Additional challenges with SAR flood detection that continue to be investigated include reducing false-positive detections from water-like features, improving flood detection in radar-obscured areas, and refining pre-processing steps to improve signal characteristics.

2.1.1.2 Multispectral Flood Detection

The Normalized Difference Water Index (NDWI) is a common method used to spectrally detect water. Deriving a classification of water using NDWI requires a secondary step of automatically or manually determining a threshold for index values that represent open water. This is a challenging process as flooding, even in the same region, can vary spectrally depending on factors affecting the color or texture of water. Other objects such as shadows, built-up surfaces, rooftop shingles, asphalt, and bright surfaces can also cause misclassifications with this method. Multispectral flood detection in RADR is performed

using a modified NDWI algorithm that makes use of a priori information about existing waterbodies, elevation, and slope and fuzzy logic techniques to identify water and non-water features. The algorithm is generally robust and capable of being used with a wide variety of sensors and scene conditions. As previously noted, visual obstructions to the ground are an inherent limitation with optical imagery and may affect the efficacy of multispectral flood detection.

2.1.2 Crowdsourced Data

Several sources of non-authoritative ground observation data are used for the purpose of validating remotely-sensed flood observations and/or multimodal data fusion to predict flood occurrence. Key sources of these data include (1) images and video from social media (e.g., Twitter, Instagram, Facebook, Flickr), news sources, traffic cams, etc.; (2) WAZE user-generated road closure data; (3) state-level department of transportation road closure notices; 4) search and rescue data; 5) Civil Air Patrol (CAP) oblique- and nadir-perspective aerial imagery; and 6) USGS high water mark surveys.

2.1.2.1 Social Media

An effort by the GISCorps (https://www.giscorps.org/) and National Alliance for Public Safety GIS Foundation (NAPSG) Foundation (https://www.napsgfoundation.org/) have begun to provide geotagged crowdsourced social media images and videos and corresponding image captions during major disaster events. See for example, the “2018 Hurricanes Crowdsourced Photos” application (https://arcg.is/1jLm4y) (Figure 6). These data provide a variety of information from wind damaged homes, fallen vegetation, downed powerlines, flooding, etc. The images and captions are used as training data for computer vision and semantic analyses designed to classify certain types of damage.

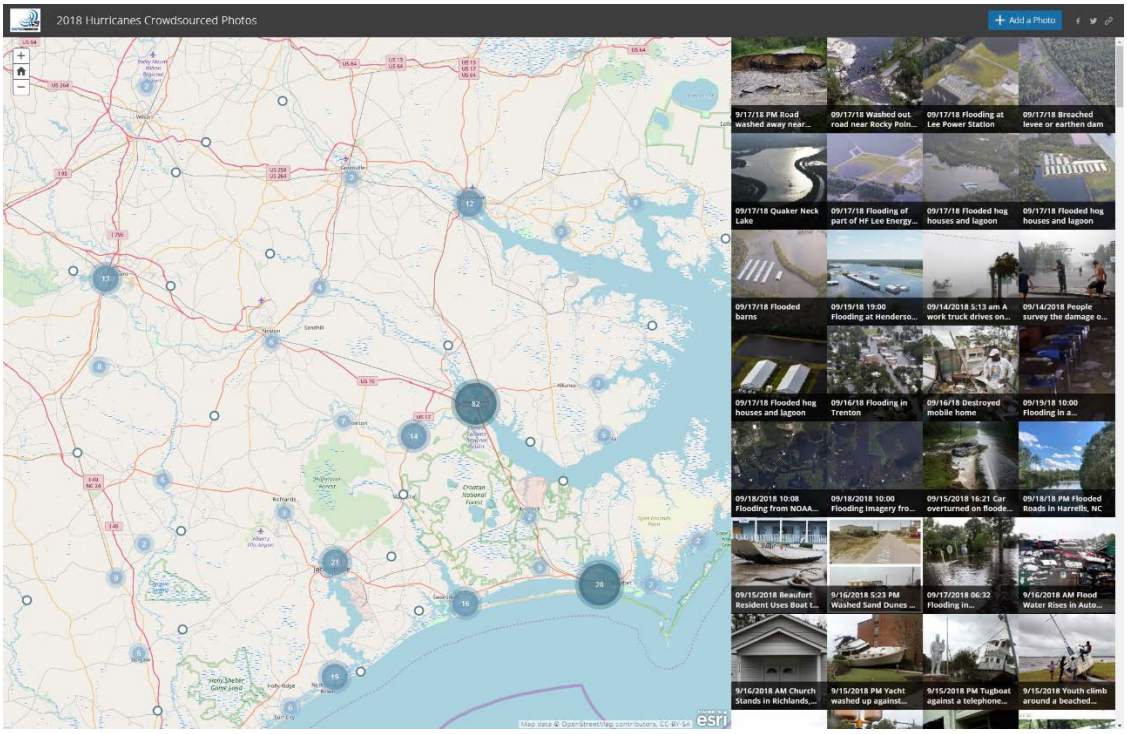

Figure 6. Example view of "2018 Hurricanes Crowdsourced Photos" application which contains geolocated time-stamped images and videos available from various social media platforms and news outlets.

2.1.2.2 WAZE, State Department of Transportation Traffic, and Real-Time Traffic Data

WAZE is a social traffic and navigation application that allows users to post traffic alert information. The WAZE Crisis Team and ESRI Disaster Response program have partnered to make these data publicly available during certain natural disasters. As an additional information resource, official state-level department of transportation road closures are often available too. For example, during Hurricane Florence (September 2018) the states of South Carolina, North Carolina, and Virginia published near real-time GIS data with road closures as both point barriers and line vectors of roads with time stamped closures. As with other non-authoritative datasets used in this research, the data are segmented by day and constrained to the remotely-sensed image analysis extents for each day. Keyword filters are also be applied to discern road closure data related to construction or other non-flood related events. In cases where a road closure can't be positively tied to flooding, the record is ignored for purposes of ground-truthing flood presence. Similarly, where roads are open, it is assumed flooding isn't a problem. In the future, tapping into a real-time traffic web service such as Google Traffic, Microsoft Bing Traffic, HERE Traffic, WAZE Data Feeds, and others, can provide up-to-date road closure information. However, programmatic access to these services require license fees.

2.1.2.3 Search and Rescue Data

A new mobile application developed by the International Association of Fire Chiefs (IAFC) and NAPSG was used by first responders and search and rescue teams to collect on-the-ground information during Hurricane Florence. The application uses the mobile device GPS to geolocate and timestamp the observation and optionally include an image with the device camera, though the data used in this research was strictly text-based. The data collection consists of several natural language entry fields, many of which were similar enough to one another that relevant information had to be gleaned from multiple fields. As with other crowdsourced data, this information is used for validation of remotely-sensed images over a common time-period and as observation data for the multimodal data fusion and flood prediction.

2.1.2.4 Civil Air Patrol

The Civil Air Patrol is typically deployed as soon as possible after a disaster event to get rapid aerial reconnaissance for areas of key concern. The flight passenger is equipped with a GPS-enabled handheld camera and after the flight mission, all imagery is uploaded to an Amazon S3 storage container and made available through a map-based web browser application (https://communities.geoplatform.gov/disasters/civil-air-patrol-cap-browser/). These data are extremely valuable for their context and wide view extents (as opposed to a ground-based social media post). The volume of imagery collected by the CAP program is also extensive. For example, ~11,000 CAP images were collected for Hurricane Florence. Significant challenges exist, however, to using CAP imagery in an analytical construct as described here. The current process requires a time-consuming (15-20 minutes per image) manual process to work through each image to identify the ground location and extent because the geographic location provided with the image at the camera body, not what is captured in the image. This is often time-consuming because the metadata often does not include sufficient information about the picture geometry including the azimuth, look angle, or use of telephoto lens. An example result of this process is shown in Figure 7. Photogrammetric solutions are tested to automate or partially automate with human-in-the-loop and are further discussed in Section 3.3.

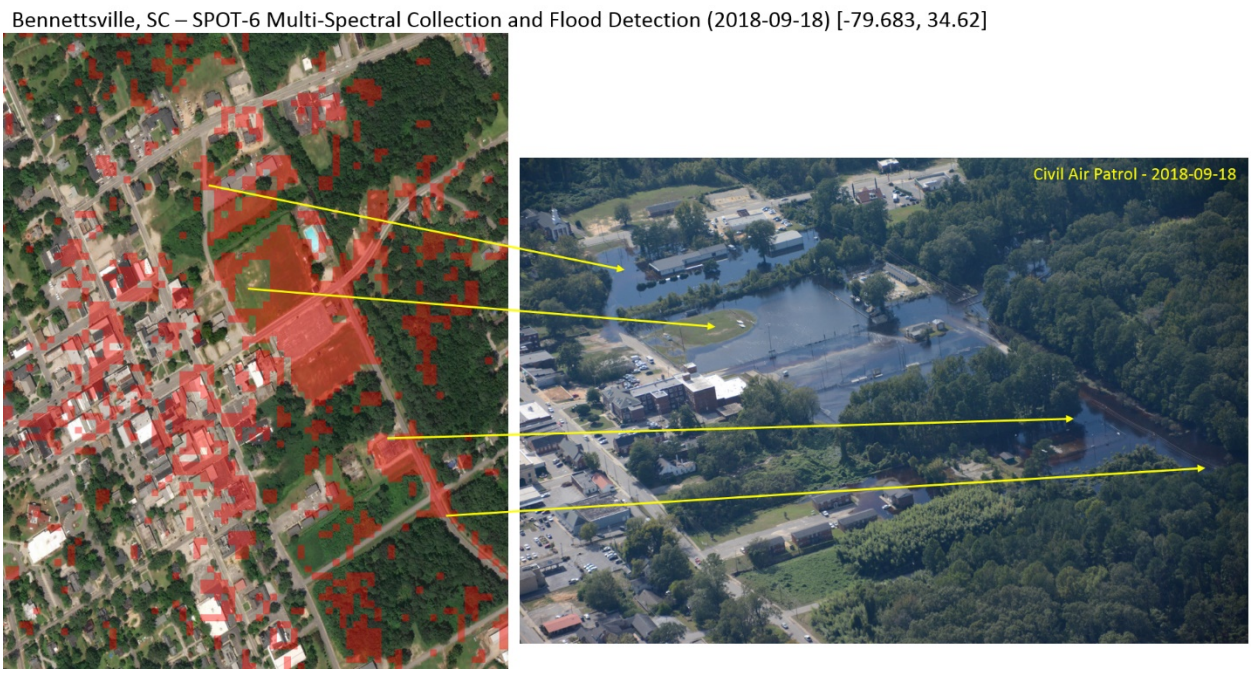

Figure 7. SPOT-6 multispectral flood detection (red polygons) related to an example oblique perspective Civil Air Patrol (CAP) photo.

2.2 Validation of Remotely-Sensed Flood Detections using Non-Authoritative Sources

A preliminary statistical approach is implemented for the use of non-authoritative observation data to validate remotely-sensed flood detections. Because remotely-sensed imagery do not have an integrated validation process, the use of independent ground observations from non-traditional and largely citizen sensor collections are proposed to serve as ground-truth for this purpose. A workflow of this process is presented in Figure 8.

Authoritative data (remotely sensed imagery) is processed to detect flooding and non-traditional ground-truth data (see Section 2.1.2) are used to validate the result. Flood detections and ground-truth data for a 6-day period spanning 2018-09-14 to 2018-09-19 are used for this assessment. A total of five sources of flood detection results are evaluated, including a NASA SAR-based process from the Advanced Rapid Imaging and Analysis (ARIA) Project for Natural Hazard (https://aria.jpl.nasa.gov/), the European Union's Copernicus Emergency Mapping Services Sentinel-1 SAR-based analysis (https://emergency.copernicus.eu/mapping), MDA Corporation's flood mapping from its RADARSAT-2 sensor and provided to the USGS Hazard Data Dissemination System (HDDS; https://hddsexplorer.usgs.gov), and PNNL's RADR system SAR (Sentinel-1) and multispectral flood detection algorithms using the SPOT-5 sensor. The data collected or generated and considered in this assessment is presented in Table 2.

Flood detection polygons and image extent boundaries were grouped by day and source/sensor type to best match ground-observations from the same day. For example, if a remotely-sensed image was collected in the morning and a non-traditional ground-truth data point was collected in the evening, they were coupled for validation. No sub-daily assessments are made, although doing so in the future may be valuable because flooding can be dynamic due to rapidly fluctuating flows.

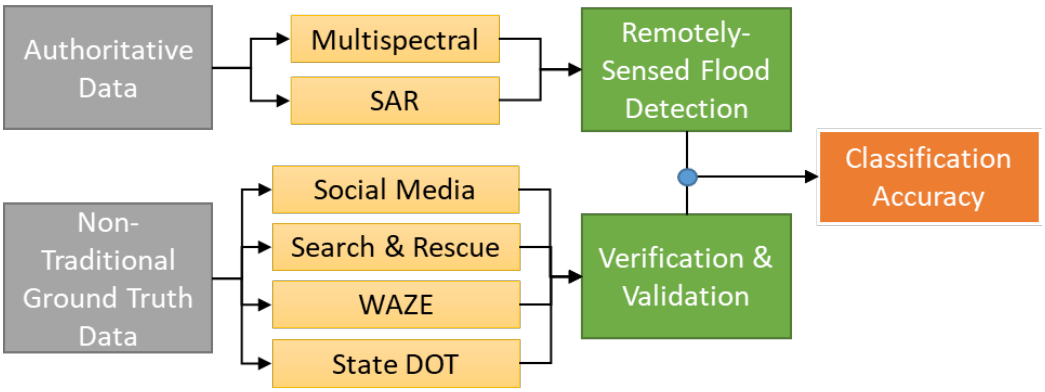

Figure 8. Workflow for validation of remotely-sensed flood detections using non-authoritative sources.

Table 2. Remotely-sensed flood assessments by source (NASA, CEMS, MDA, RADR), general sensor type (SAR, MS) and date.

Flood Assessment Source	2018-09-14	2018-09-15	2018-09-16	2018-09-17	2018-09-18	2018-09-19
NASA-SAR	X	X				
CEMS-SAR		X				X
MDA-SAR	X	X	X	X		X
RADR-SAR		X	X			X
RADR-MS					X	

Daily grouped ground-truth data and remotely-sensed flood detections were used to create a bivariate classification of flooding. To help account for positional uncertainty of flooding in ground-truth observations (i.e., images, videos, road closure points, etc. are likely not taken from within the flood water, but from a perspective adjacent to the flood water) (see Figure 9), a point-to-grid method was used (where the 'point' is the ground-based observation and the 'grid' is the remotely-sensed flood detect) to assess presence of water. Because the orientation of the crowdsourced data is not known, a 360-degree 100-m radius flood search area is used. Two iterations of point-to-grid analyses were performed. In the first iteration, ground-truth points were validated to remotely-sensed flood areas as the data were

released (i.e., all water detections; only flood detections). In the second iteration, point-to-grid analysis incorporated remotely-sensed flood detections, and existing waterbodies and wetlands instead of flood-only detections This process is repeated for each source of remotely-sensed flood observation data for each day the data exists and are formatted into contingency tables and basic descriptive findings are reported on these results. Additionally, a simple analysis was performed to assess the 'TRUE/FALSE' rate of the ground-truth data and flood detection source/algorithm.

The outcome of the validation can be implemented as a feedback mechanism to the processing of remotely-sensed flood detection to potentially adjust process input parameters and test for validation results as shown in Figure 3. This capability has not been implemented in the RADR system. Additionally, a statistical causation of covariates from the crowdsourced results could provide some explanation of errors in remotely-sensed flood detection (i.e., ground observation indicated flooding but no flood detections were present in the remotely-sensed data). Example covariates could include horizontal and vertical distance to existing waterbody, slope, land use/land cover, and canopy cover.

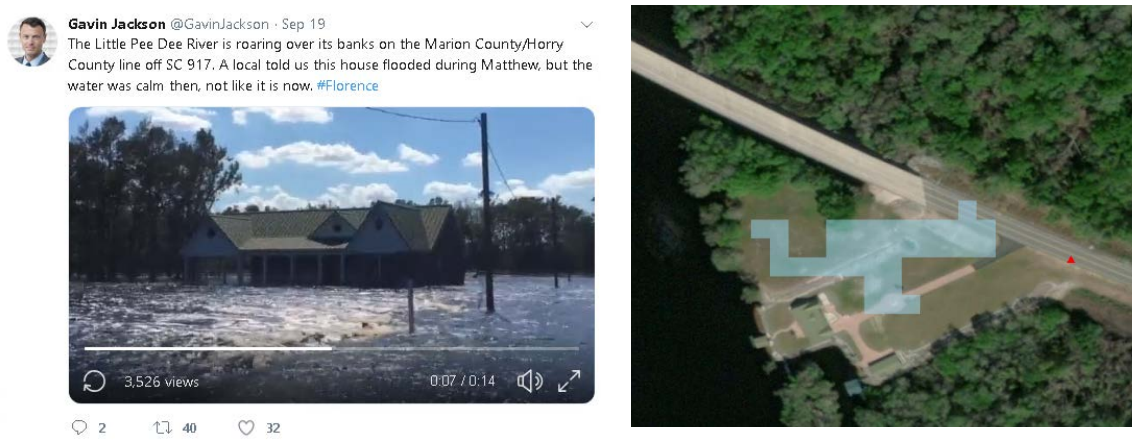

Figure 9. An example relationship of a crowdsourced video of flooding (left) and the recorded position (right; red triangle) and the SAR-based flood detection (right; blue polygon) on September 19, 2018. The distance from the video recording position to the edge of the flood polygon is 23m.

2.3 Preliminary Validation of Remotely-Sensed Flood Detections using Non-Authoritative Sources

The extents of non-traditional ground-truth data and remotely-sensed imagery for Hurricane Florence are presented in Figure 10. Approximately 8,000 ground truth samples intersected the Florence AOI, but these were reduced to 360 after removing samples unrelated to the event (Figure 11).

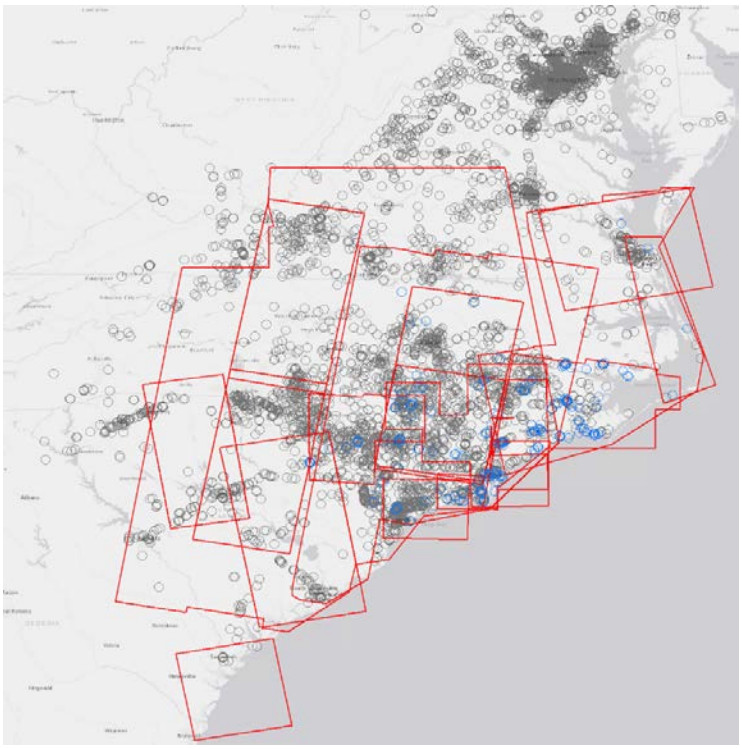

Figure 10. Map of the U.S. eastern seaboard from Savannah, GA to Atlantic City, NJ. Non-traditional ground-truth data collections assembled from September 14-19, 2018 are represented where the red outlines represent various remote-sensing image extents, grey circles are ground observations, and blue circles are filtered ground observations that indicate flooding.

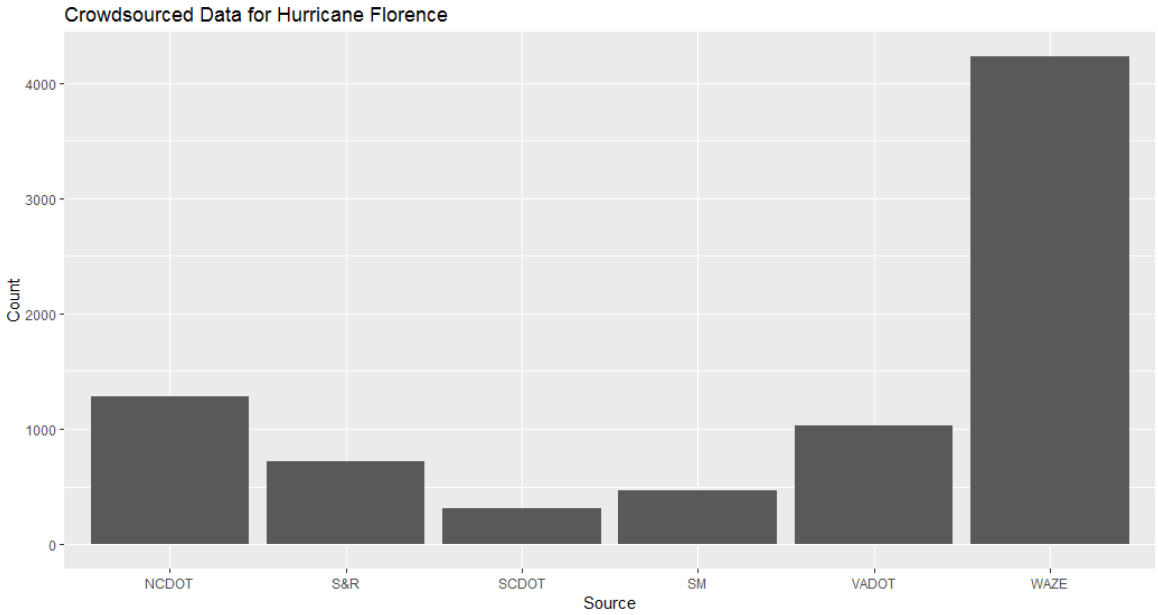

Figure 11. Crowdsourced ground observation data collected in and around the geographic domain of the Hurricane Florence storm domain. NCDOT=North Carolina Department of Transportation, S&R=Search and Rescue, SCDOT= South Carolina Department of Transportation, SM=various social media and news, VADOT=Virginia Department of Transportation, and WAZE=WAZE mobile traffic application. Total starting sample population, *n*=8,042 and final filtered population *n*=360.

The contingency table for the first iteration point-to-grid analysis is presented in Table 3. The best accuracy was observed with RADR-SAR (54.0%), followed by NASA-SAR (48.2%) and RADR-MS (45.8%) (Table 3). Conversely, the CEMS-SAR and MDA-SAR flood classifications exhibited poor agreement with ground-truth observations. While the overall accuracy of these flood detections appears to be moderate to poor, it is important to note that the more severely flooded portions of the scene are more accurately depicted than areas of less severe flooding. Additional investigation is needed to understand this issue more clearly and determine factors affecting error.

The results of the second iteration point-to-grid analysis in which remotely-sensed flood detection, the normal-condition water mask, and wetlands are considered, is presented in Table 4. Agreement between ground-truth and remotely-sensed flood detections improved in every category except MDA-SAR. The best agreement was observed with NASA-SAR (69.3%), followed by RADR-MS (66.7%) and RADR-SAR (58.9%). The CEMS-SAR agreement improved significantly compared to the first iteration results but is still low (27.3%).

Table 3. Contingency tables for point-to-grid validation results of crowdsource to remotely-sensed flood observation, analyzed daily and summed for the event period (September 14-19, 2018). The results presented here are the first-level point-to-grid analysis where remotely-sensed flood observation data is used as generated and released.

	NASA-SAR	***CEMS-SAR***	***MDA-SAR***	***RADR-SAR***	***RADR-MS***	**TOTAL**
TRUE	66	3	9	67	11	**156**
FALSE	71	30	33	57	13	**204**
TOTAL	**137**	**33**	**42**	**124**	**24**	**360**
% TRUE	**48.2%**	**9.1%**	**21.4%**	**54.0%**	**45.8%**	

Table 4. Contingency tables for second-level point-to-grid validation results of crowdsource to remotely-sensed flood observation, normal-condition water mask, and wetlands are analyzed daily and summed for the event period (September 14-19, 2018).

	NASA-SAR	***CEMS-SAR***	***MDA-SAR***	***RADR-SAR***	***RADR-MS***	**TOTAL**
TRUE	95	9	9	73	16	**156**
FALSE	42	24	33	51	8	**204**
TOTAL	**137**	**33**	**42**	**124**	**24**	**360**
% TRUE	**69.3%**	**27.3%**	**21.4%**	**58.9%**	**66.7%**	

2.3.1 Ground-Truth Data Performance by Source

Ground-truth data were investigated more closely to assess whether agreement rates with remotely-sensed flood detections varied among different sources. Figure 12 illustrates these rates by ground-truth data source. The best agreement is observed for search-and-rescue data (69.2%), followed by social media (49.1%), and WAZE (41.7%). One possible explanation for the lower agreement rate with WAZE data is that the location of observation may be greater than 100 m from actual flooding, our criteria for point-to-grid correspondence, because road closure barriers are often placed away from the actual hazard to divert traffic. Thus, a greater radius of association may be necessary to effectively utilize WAZE data in MFM modeling framework. It is believed that Search & Rescue data had the highest agreement because personnel are intently operating in flood zones when they collect observations.

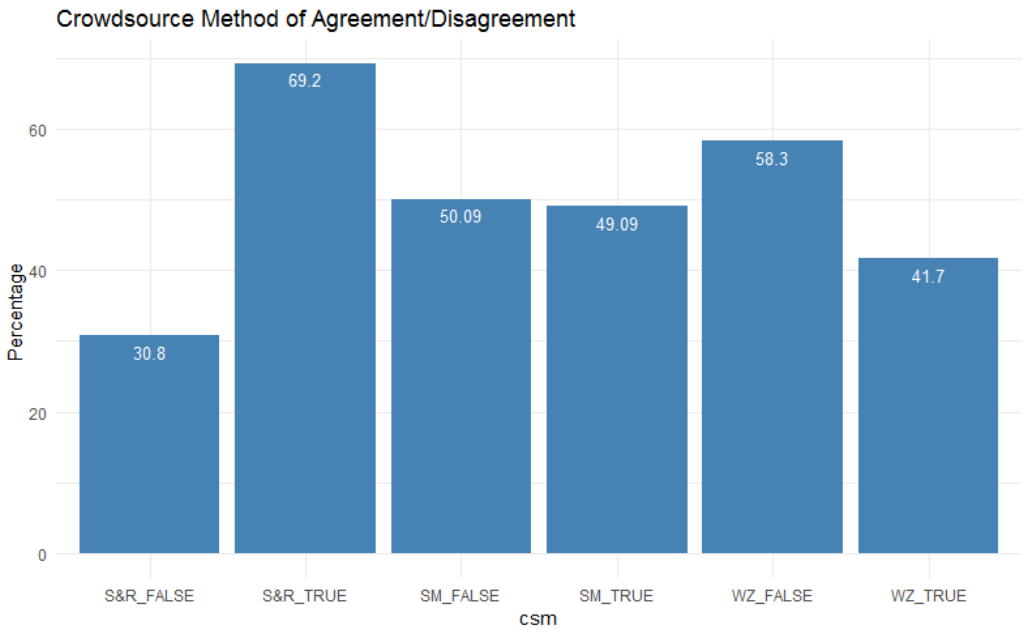

Figure 12. The agreement (TRUE) or disagreement (FALSE) of specific crowdsourced data in relation to validation of remotely-sensed flood observed data is presented. S&R = Search and Rescue, SM=various social media and news, WZ=WAZE mobile traffic data and state-level department of transportation closure data.

As part of an image similarity and geospatial validation process in this study (Section 3.2), it was determined that a large number of the social media images did not have accurate geographic information (often one to several blocks away from the actual location recorded in the image). Some of these data were thought to have been verified by a volunteer social media team (GISCorps/NAPSG), but it is unclear to what extent and rigor. The locations that are embedded in an image are based on location data collected by the user's device, which is subject to low-accuracy GPS signal, signal bounce in urban areas, or location triangulation from surrounding cell towers. This discovery impacts the current validation results for the social media data and the overall results presented in the contingency tables, thus this analysis will

need to be re-run with correct image locations. Importantly, the step of performing location validation is crucial to perform prior to running validations on non-authoritative data. These considerations highlight some of the challenges of using non-authoritative data for damage assessments, particularly how uncertainties about the spatial accuracy of observations can affect subsequent analysis. Developing methods to better validate or derive the geographic position of non-authoritative data sources represents a critical area for future research.

3.0 Deep Learning for Multimodal Data Fusion

Non-authoritative image/video data is a vibrant data resource, but in its native form is difficult to use in traditional geospatial analytics without supplemental data from GPS and inertial measurement units (IMU). In comparison to the "Validation of Remotely-Sensed Flood Detections using Non-Authoritative Sources" (Section 2.2) where a single point observation is used to validate the presence of flooding, this work aims to derive a rich sampling of point observations from multi-perspective images to detect flooding and transform these observations into a common geographic form. An overview of the workflow for this objective is presented in Figure 10 and is described in further detail below.

Because the non-authoritative source images aren't in a geospatial format, information must be extracted from the images in a manner that they can be tied to other georeferenced data. A sequenced deep learning process using image relevancy, image similarity, feature detection and matching, and semantic image segmentation, derives required information and returns it in geographic space. The non-authoritative data sources here rely on ground-level social media images and oblique aerial images primarily from the Civil Air Patrol (CAP), but Customs and Border Protection (CBP) handheld images and UAS (drone) still images or videos posted online without GPS/IMU data are also viable data resources.

3.1 Image Enrichments, Object Labels, and Relevancy

There is a necessary step to filter large volumes of non-authoritative data that doesn't contribute to the objective. For example, many of the social media images collected in the Hurricane Florence use case were not useful for assessing flood or other damage detection (e.g., selfies, birthday parties, food, etc.). Therefore, classes of images determined to be non-informative to situational awareness can be binned and removed from further use in the workflow (see Figure 13, "Bin Non-Contributing Images"). Deep neural networks can sift through the thousands of images streaming in from various sources and derive probabilistic image labels that determine the objects and features within the image (e.g., house, water, lake, flood, reflection, family, building) (see for example, Figure 14). A potential issue however, is that

training datasets and subsequent trained models may not be focused toward detection of natural disaster situations. To support this application, a disaster-based ontology, Geo-MD (Figure 12), was considered as a schema for social media image labeling and classification (Bouyerbou et al. 2019). The implementation of this ontology was kept in reserve pending baseline tests with pre-trained machine-learning models.

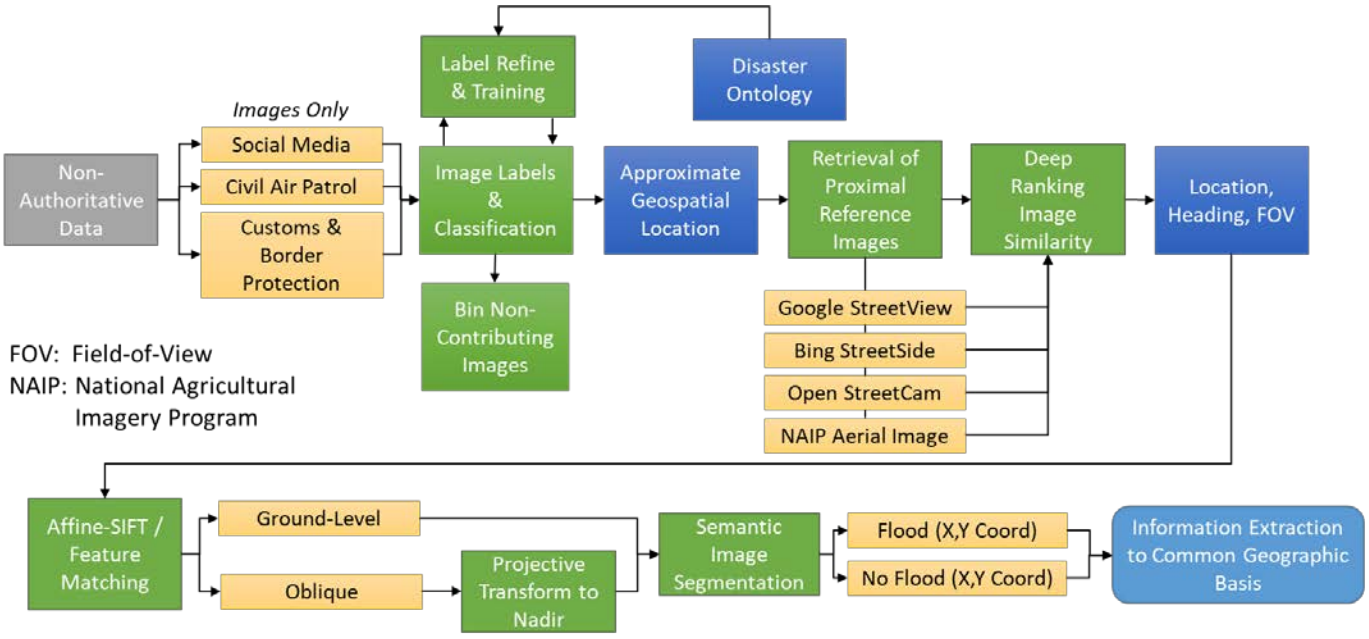

Figure 13. Process overview for 'Deep Learning Approaches for Multimodal Data Fusion in Disaster Events'.

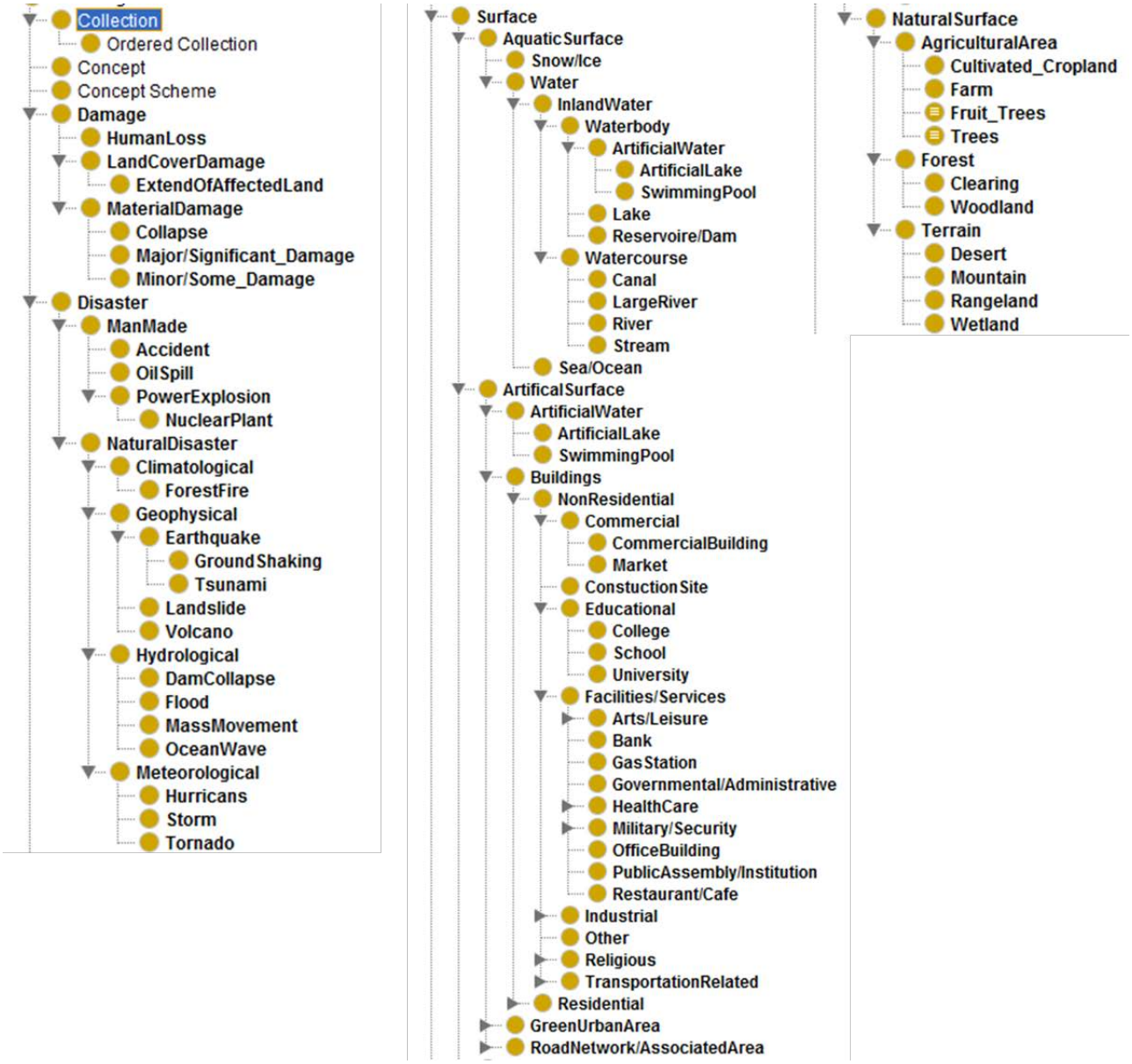

Figure 14. Geo-MD, a disaster-based ontology, was considered for use in manual image labeling to enable sorting and classification of disaster-relevant images (Bouyerbou et al. 2019).

3.1.1 Identifying duplicates

It is common for social media users to share content that they did not collect, both on the original social media platform and across social medial platforms, resulting in numerous duplicate images that need to be remove unnecessary. Rule-based screening for image size, time stamp, and reported location was evaluated but is ineffective because of cropping, platform-specific time stamping, and deletion of EXIF tags that contain information on camera type, location, etc. Successful screening of images for common features has been implemented using the perceptual hashing (pHash) method. The pHash method is often

used for digital forensics and can be applied to a wide variety of visual media. The technique creates a fingerprint based on high-dimensional image features such that the pHash similarity distance between identical photos will be zero. Cropped photos with text overlaid would have pHash values very close to the original image and the distance between altered photos and original would be very small. This approach is enabled via open-source libraries (https://www.phash.org/; https://pypi.org/project/ImageHash/) and the results are programmatically assessed to build a table of hash values that are used to reduce the input image set to a set of unique images.

3.1.2 Object Prediction in Images

One step toward understanding the content in social media images is using object prediction to generate a list of probabilistic labels that relate to objects in the image. The manual sifting of images is not feasible so an automated approach was enabled to provide image labels that could be used to remove non-relevant images. Thus, several pre-trained commercial models including Google VisionAI, Amazon Rekognition, and Clarifai were evaluated for the possibility of repurposing them for labeling information relevant to natural disaster situations. Based on a qualitative assessment of the pre-trained models we opted to use Amazon's Rekognition to rapidly label objects for all of our social media images. Rekognition is based on highly scalable, deep learning technology and can be easily implemented on any set of images stored in an Amazon S3 cloud storage container. Rekognition returns a series of probabilistic labels for each image which we used for determining the relevance of an image to our objective (see for example Figure 15).

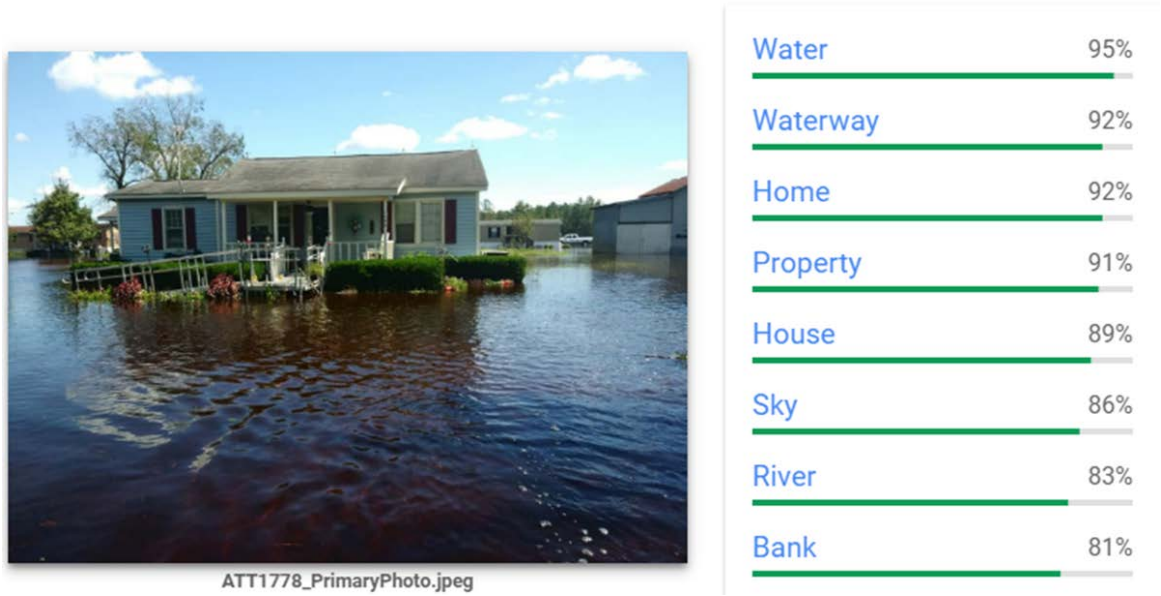

Figure 15. Example of deep-learning based probabilistic image labeling.

3.1.3 Relevance Voting for Model Development

Our social media images came from two sources; NAPSG and Flikr and were quite different in terms of image content. The NAPSG dataset was curated by volunteers to contain only images related to disasters and as such, we expected that most of the images therein were relevant. The Flickr dataset came from a massive download of all images posted in the region during the timeframe of the disaster event, many of which we suspected were not relevant. Within the UI we created the ability to vote on whether an image was relevant for assessing flood extent; either flooding visible or not visible (figure 16). Our rubric for determining relevancy was based on the clarity and content of images; an image was voted relevant if it showed visible ground (or flooded area) with enough scene context such that a human could likely identify location. Nighttime images, blurry images, and indoor scenes were voted non-relevant. A total of 3,656 images out of ~11,000 were reviewed and voted on. Table 5 summarizes the proportion of relevant and non-relevant images in our data sets.

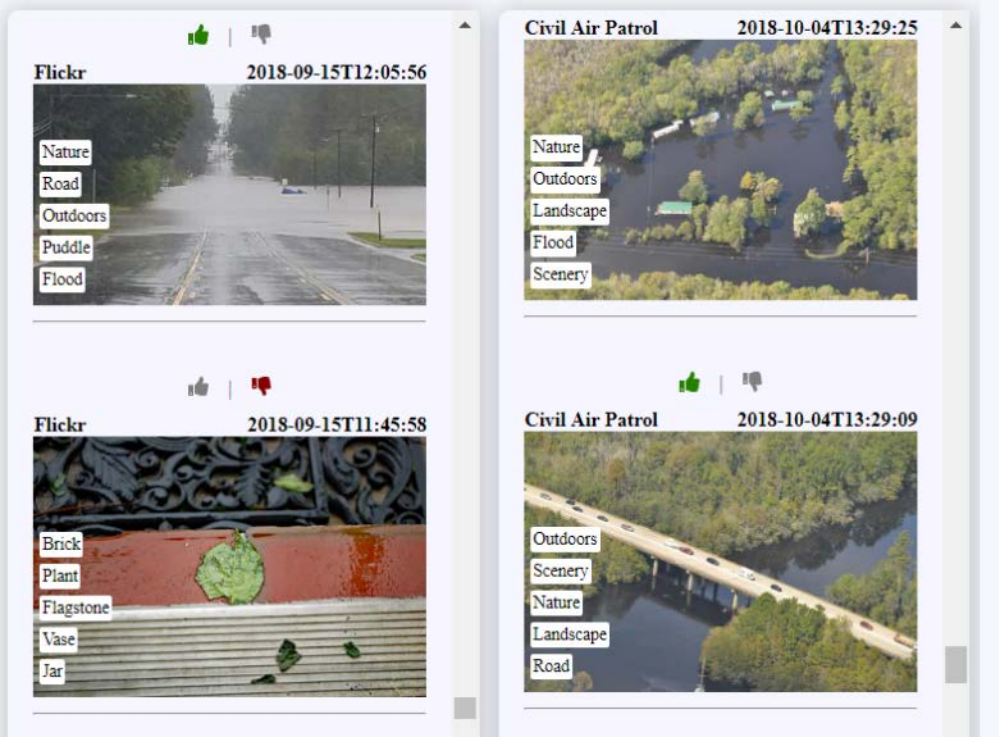

Figure 16. Example view of the custom web-based interface used to provide qualitative feedback on the relevancy of images and their Rekognition labels. The “Green Thumbs Up” icon indicates the image is relevant to the objective and the “Red Thumbs Down” indicates the image is not relevant.

Table 5. Relevancy metrics for the Flickr-collected data and the NAPSG-collected data based on manual quantitative feedback.

	Relevant	Not Relevant
Flickr	27%	73%
NAPSG	90%	10%

3.1.4 Relevancy Classification Models

We trained and tested several machine learning models to automatically predict if an image is relevant or not, including Random Forest Classifier and three CNN models (AlexNet, ResNet18, and VGG11). Using the Rekognition labels and relevancy class from the labeled imagery, we split the data 80/20% for training/testing datasets. The trained models were run on two validation datasets: GISCorps/NAPSG data collected from a separate flooding event (Hurricane Harvey, Houston, Texas region, August 2017) and a set of random Flickr images. The Hurricane Harvey data contained 523 images and the random Flickr dataset contained 998 images. Example image results from the CNN models are shown in Figure 18 - Figure 20 and output metrics are provided in Table 6.

In a stream of social media images being posted during a disaster event, there likely will be many more irrelevant images than relevant even if the posts are limited to the geographic region of impact. For the application envisioned here, there is an obvious need to screen incoming images as quickly and accurately as possible to determine which should be retained for assessing potential damage; therefore, we measured processing time and accuracy of each model (Table 6). All three CNN models exhibited faster processing times (>1 image/second) and moderately better overall validation accuracy with Flick images (>0.78) than Random Forest (0.56 images/second and 0.65 overall accuracy), although Random Forest demonstrated better overall validation accuracy with NAPSG data (0.904) than all three CNN models (AlexNet=0.78, ResNet18=0.868, VGG11=0.860)(Figure 17).

Recall was considered more closely than other accuracy metrics because it best describes the ability of a model to find all relevant cases within a dataset, which is especially important in our use case because datasets may contain proportionally few images of interest. AlexNet demonstrated the best recall (0.736) of all four models with Flickr data, whereas Random Forest exhibited the best recall (0.930) of all models with NAPSG data. These results suggest there may be dataset-specific tradeoffs among models in terms of both speed and accuracy. For example, Random Forest may be a better model for determining relevancy of NAPSG images, but it requires more processing time. This could potentially be improved, however, with fine tuning of the model and more compute power. Conversely, AlexNet was the better model in terms of both speed and accuracy for determining relevancy of Flickr images.

An additional classification step was performed to further classify the relevant images into the presence/absence of flooding. Using all the relevant Flickr and NAPSG images (~1,796) we manually determined which images had visible flooding (961) and no flooding (835). Using these data, we trained multiple CNNs to exploit image features to classify flood presence/absence. The training accuracy was encouraging ranging from 83% to 87%. It is expected that similar training can be performed for other damage types in the future.

Table 6. Output metrics from relevancy model tests against two independent validation datasets: NAPSG images from Hurricane Harvey (*n*=523) and a random image set from Flickr (*n*=998).

Model	**Dataset**	**Testing – Accuracy**	**Validation – Accuracy**	**Validation - Precision**	**Validation - Recall**	**Performance**
Random Forest	Flikr	0.903	0.646	0.433	0.675	~0.56 images/sec
	NAPSG	0.903	0.904	0.969	0.930	
AlexNet	Flikr	0.904	0.780	0.590	0.736	~6.15 images/sec
	NAPSG		0.868	0.982	0.879	
ResNet18	Flikr	0.885	0.790	0.664	0.593	~2.95 images/sec
	NAPSG		0.868	0.978	0.881	
VGG11	Flikr	0.912	0.810	0.741	0.586	~1 images/sec
	NAPSG		0.860	0.978	0.879	

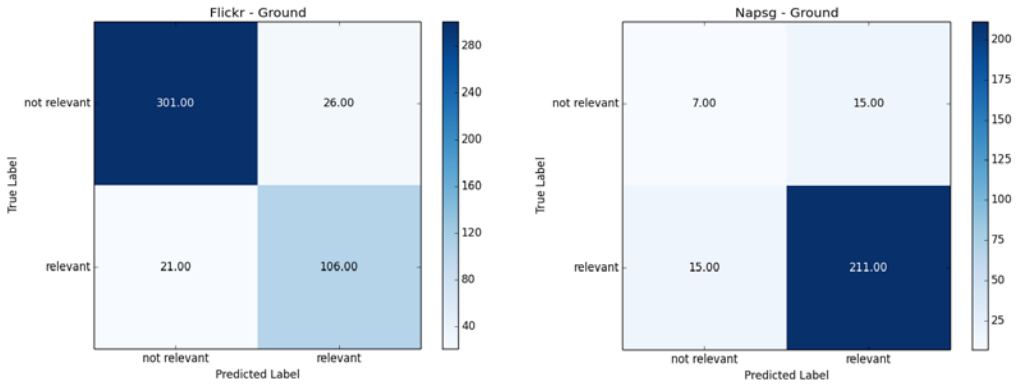

Figure 17. Confusion matrix showing the results of the Flickr and NAPSG image relevancy classifier using a Random Forest binary classification model.

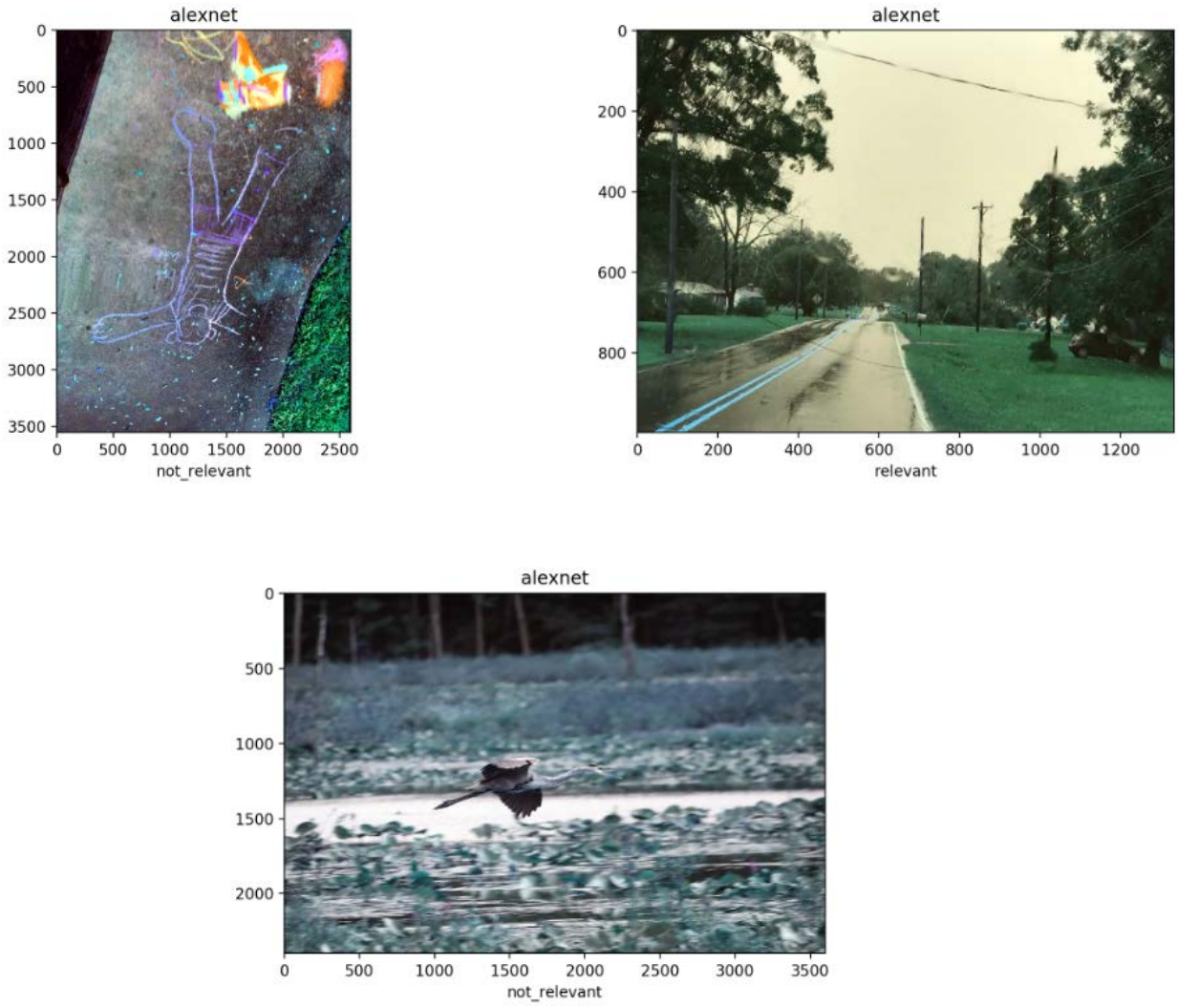

Figure 18. Classified images from the AlexNet model indicating “relevant” or “not_relevant” as defined on the x-axis.

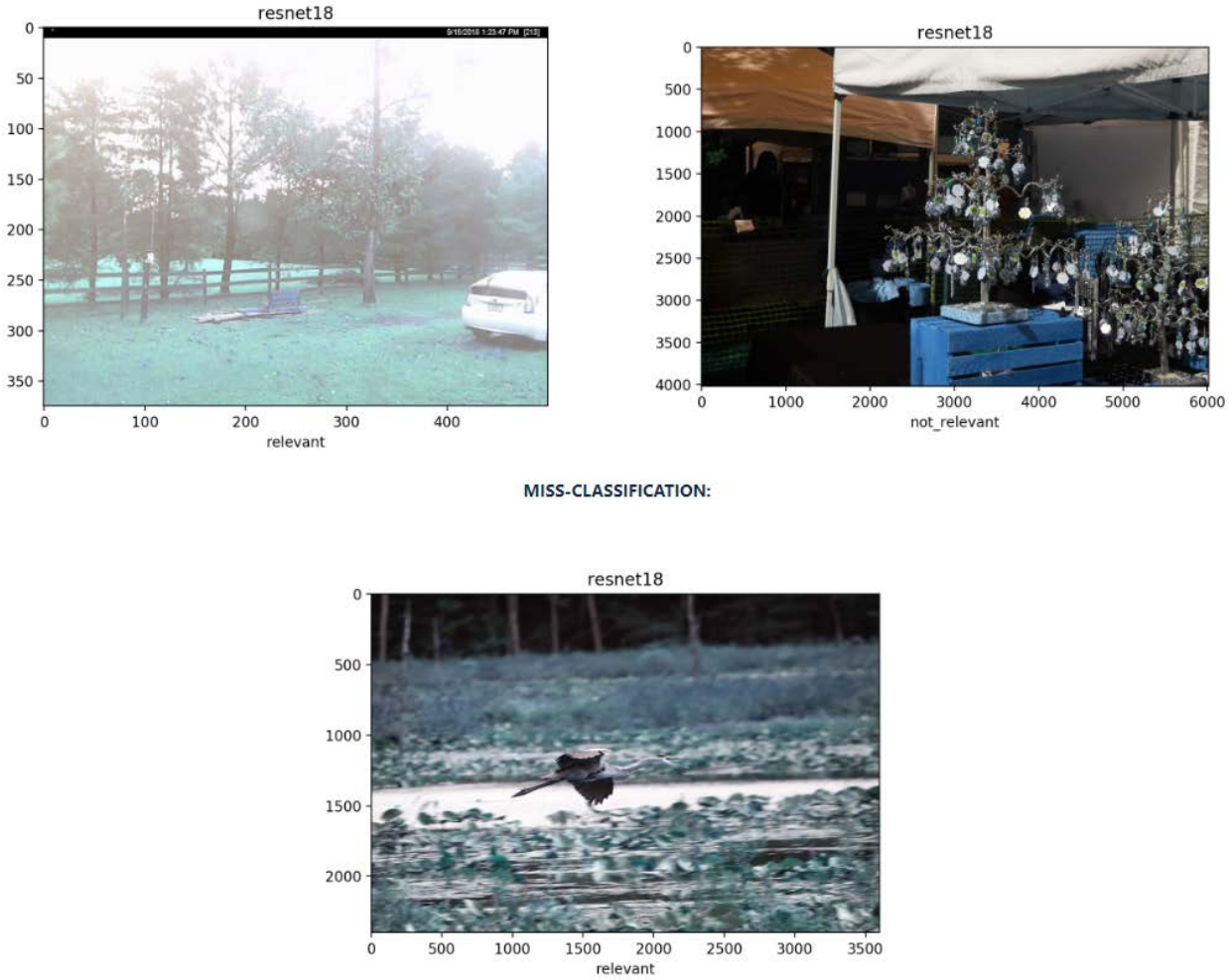

Figure 19. Classified images from the ResNet18 model indicating “relevant” or “not_relevant” as defined on the x-axis.

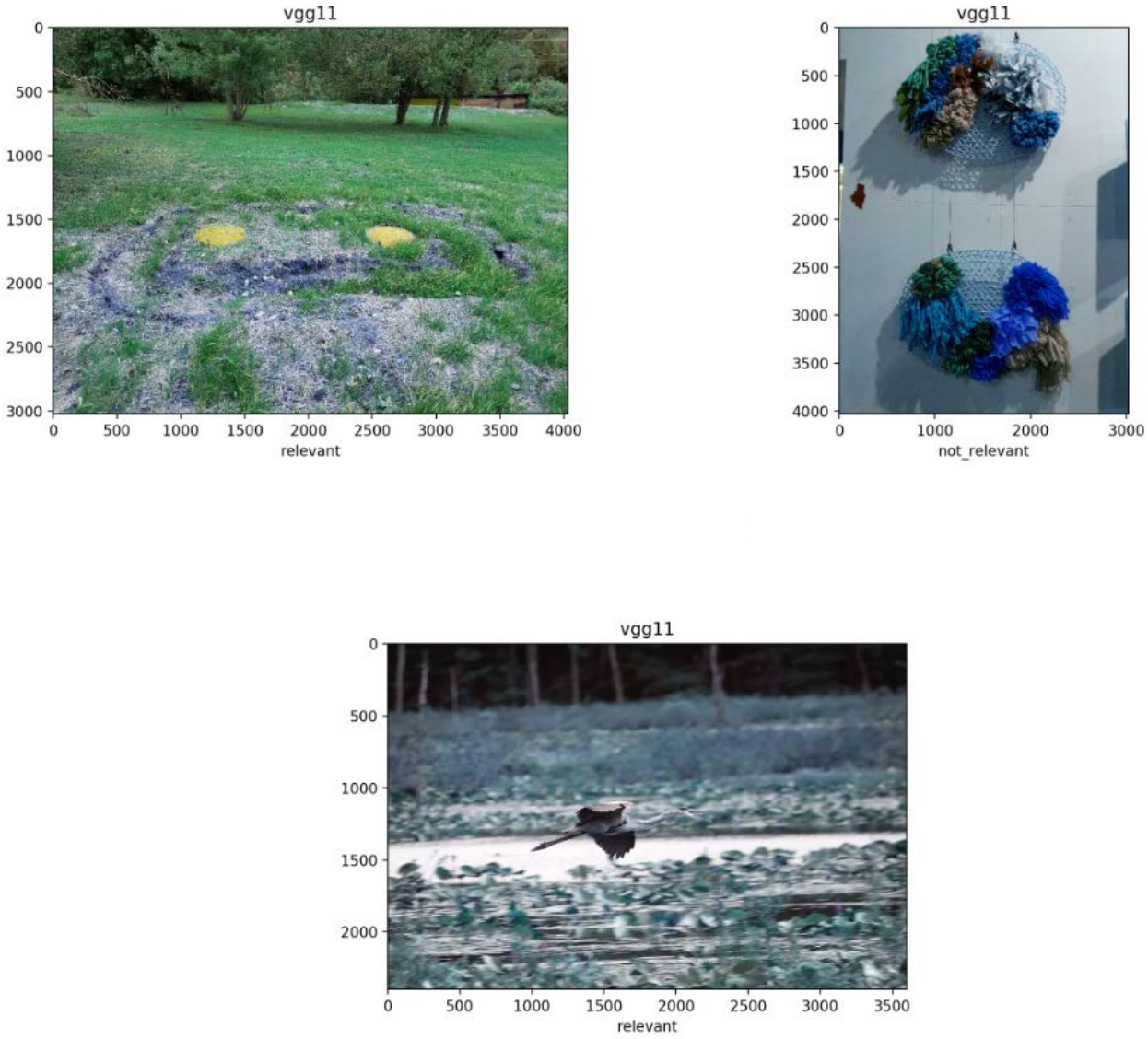

Figure 20. Classified images from the VGG11 model indicating “relevant” or “not_relevant” as defined on the x-axis.

3.2 Geographic Location and Validation of Non-Authoritative Ground-Level Images

The method of geographic validation and refinement of images relies upon the use of the best known geographic coordinates of an image (source image), a reference image dataset, and a deep-learning based image similarity process to find the closest image match between the source and the reference.

Many of the non-authoritative images contain an embedded set of geographic coordinates in the image metadata (i.e., EXIF geotag) that indicate with varying levels of accuracy where the image was taken. In cases where this information is not available, image geolocation estimation methods (Johns, Rounds and

Henry 2017b, Suresh, Chodosh and Abello 2018, Vo, Jacobs and Hays 2017) may be used to provide an approximate location. Flickr data obtained in this study used the methods of Johns, Rounds and Henry (2017a) to collect and geographically store the images. The NAPSG collected data had a geographic coordinate stored with the image data. We implemented a refined approach to use the best-known location for ground-level, social media images and refine the geographic coordinates. The approach is based on an image similarity metric between the social media image and a set of reference images with known location. This process is important for three primary reasons. First as noted in Section 2.3.1, correct image locations impact validation of remotely-sensed flood observations and validation of flood modeling. Second, location accuracy affects the ability to convert potentially valuable information in the image to geographic space for data fusion and modeling. Third, by using a set of spatially accurate source images with various look angles and perspectives, additional metadata is retrieved from the reference imagery such as the heading and field-of-view, so it understood where the user was pointing the camera; these additional metadata are important for developing the transformation process between arbitrary pixel space, 3D depth data, and geographic space through feature matching.

There are numerous image similarity techniques available. Here, we test the Xception deep-learning, depth-wise, separable convolutions model to extract image features, and the Faiss model to perform similarity searches that compares disaster event images to a corpus of reference image sources (Chollet 2017; Johnson, Douze and Jégou 2019). The methods are illustrated in the workflow in Figure 13.

Reference images with known locations were obtained by harvesting Google StreetView images. The harvesting process entails following streets in the proximity of the comparison image, and at each StreetView path increment, capturing all StreetView images in increments to capture a 360° view at two different field-of-view settings (i.e., zoom levels) (Figure 21). The objective is to provide a range of images and perspectives that are used to match what an individual captured on a smart-device camera. The entire set of Flickr and NAPSG source images, and StreetView reference images are brought into the Amazon AWS data pipeline where the image similarity analysis is performed.

Initial results of this process helped to reveal geographic location errors prevalent in the social media images. Some examples of these errors are presented in Figure 22 and Figure 23. In one case, a location in Bolivia, NC is presented as an example in Figure 24, where a seemingly uninhabited house in 2013 and 2015 had been transformed into a business making the source and reference images significantly different. Situations like this would dictate a second-tier comparison using other available street-level imagery (e.g., Microsoft StreetSide, OpenStreetCam) to help achieve a match. Though as shown in this example, even the latest imagery may be erroneous and could only be resolved by a human-in-the-loop process. There are cases too where social media images are captured far enough away from a street that

there is no street-level imagery available (e.g., images taken within a park or stadium behind a school) which would also necessitate a human-in-the-loop process.

Figure 21. Source image (top-center) and a series of 8 Google StreetView images that capture a 360° perspective of the area at a given coordinate. This process is repeated for each coordinate at two field-of-view (zoom-level) measures, where the current field-of-view captures more area in each image.

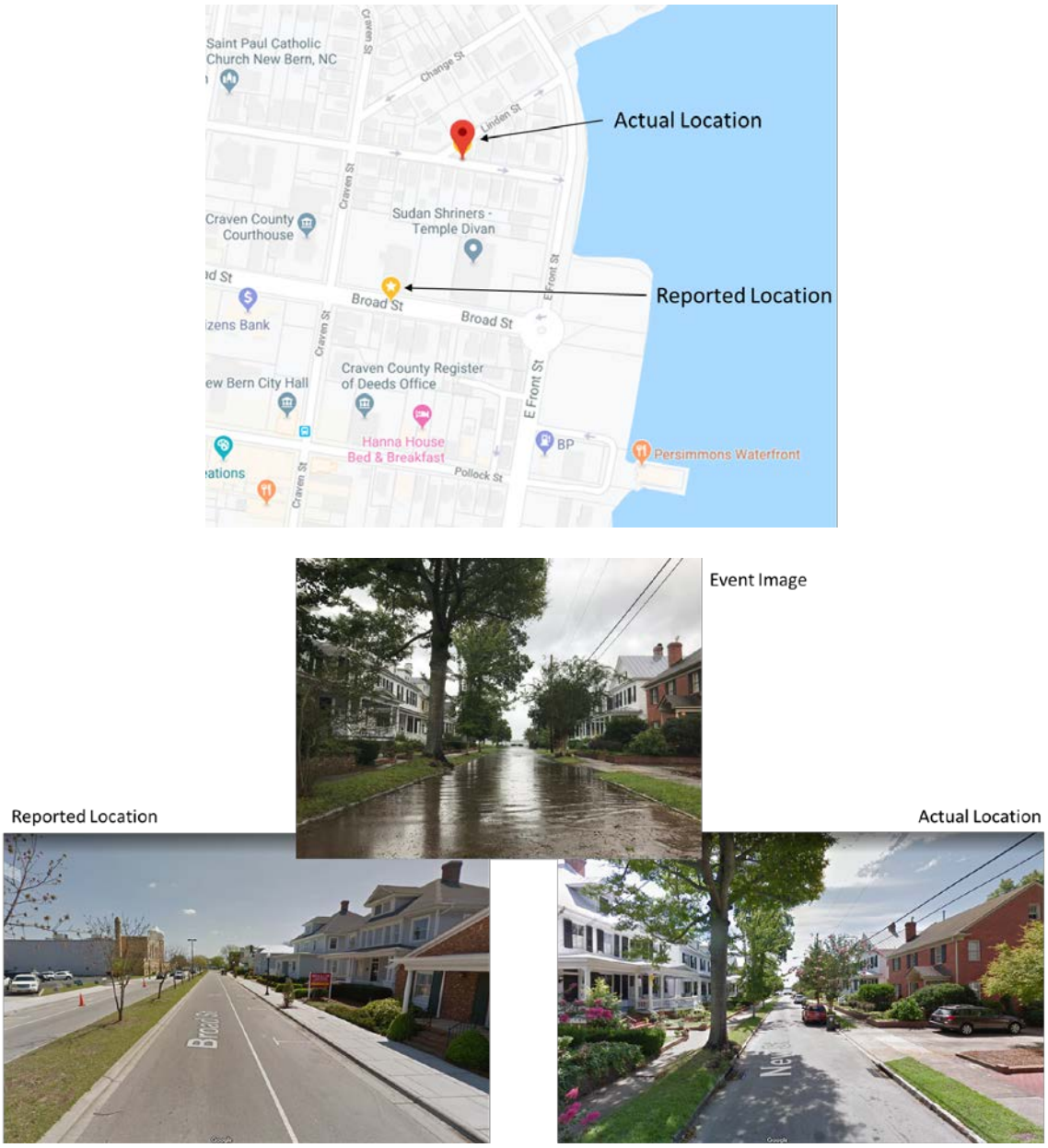

Figure 22. Example of low-accuracy location coordinates for a social media image which turned out to be positioned a block to the north.

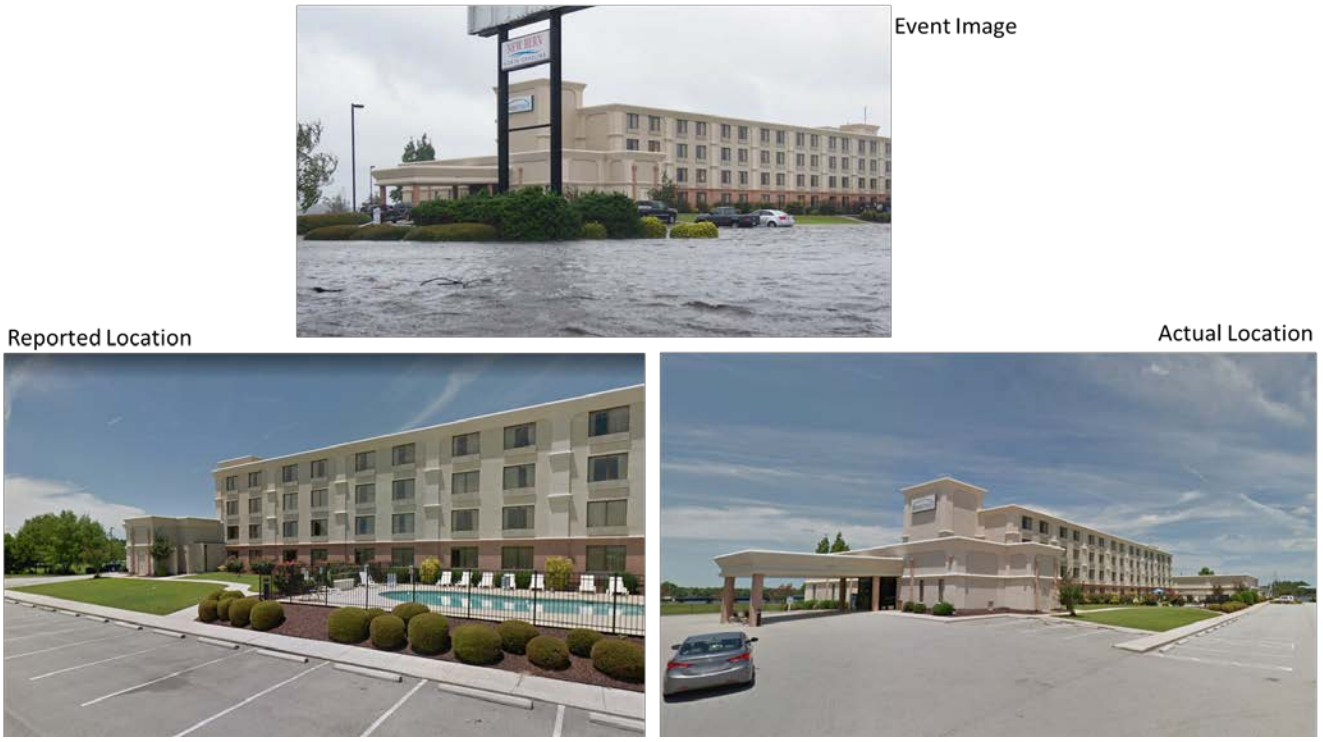

Figure 23. In the social media image example, the location was nearly correct at the hotel parking lot, but the perspective of the building was taken at the front of the building.

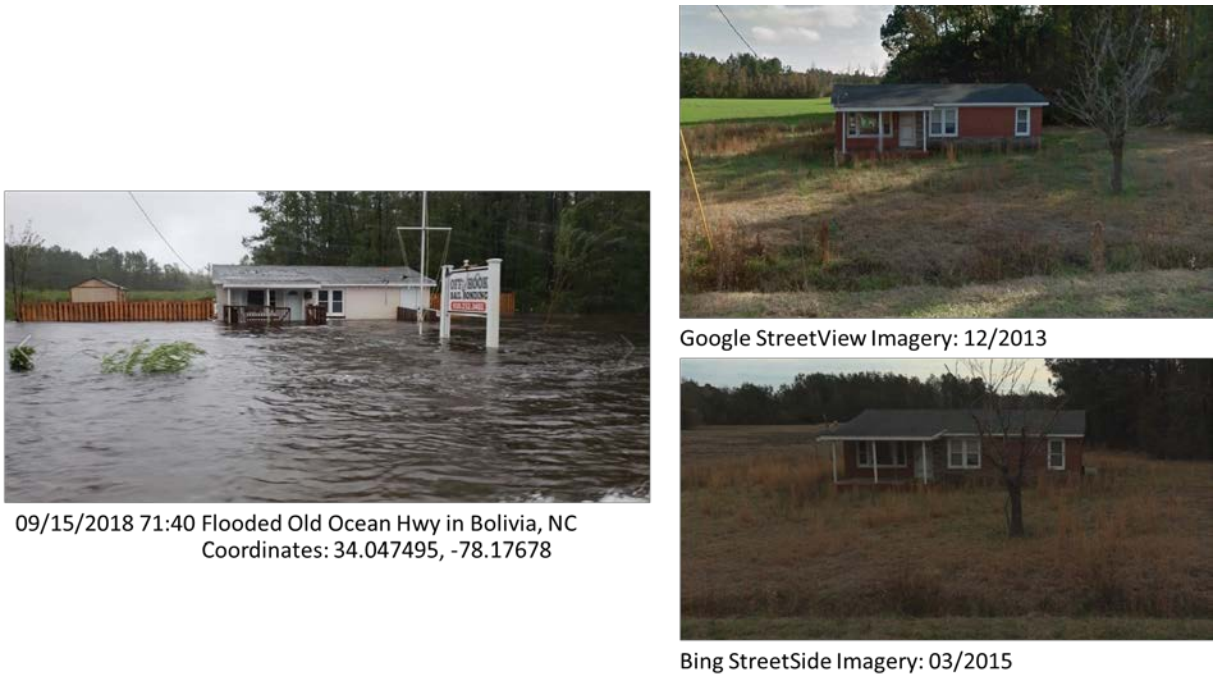

Figure 24. Even with a proper coordinate, there are cases where the feature of interest has undergone significant changes since the reference imagery was collected. In these cases, additional reference imagery such as Bing StreetSide or OpenStreetCam imagery can be used with the hope of a more recent image capture. Otherwise the image must be binned for human-in-the-loop review.

3.3 Geographic Location and Validation of Aerial Oblique Images

After a disaster event, the Civil Air Patrol and Customs and Border Protection (CBP) in the United States are commonly tasked with flight missions to capture images of damage areas or general areas of concern (see for example, the flight segment on Figure 25). The flights are generally carried out by a pilot and a photographer in a small fixed-wing aircraft. The photographer uses a high-quality camera with an integrated GPS. For each oblique photograph captured (Figure 26) a series of metadata are collected and stored with the imagery, including the date and time, focal length, GPS coordinate of the camera body, and in rare cases, the heading. While these data provide a valuable information resource with the many thousands of images collected after an event, they are not readily usable for automated image analysis. For example, the location of the plane when the photograph was taken is known, the location and extent of the photograph are unknown and may be many kilometers from the aircraft location. The goal of this part of the research is to improve the geolocation of individual oblique aerial images.

We have observed that due to the oblique nature of their acquisition, the geolocation of CAP images is often many kilometers from the ground target. Review of ancillary data about altitude, heading, angle of incidence, focal length, etc., has revealed that these data are often missing or inaccurate; thus, methods to improve photo geolocation based solely on geometry of the acquisition are not feasible. Furthermore, without reliable ancillary information it cannot be determined which side of the aircraft the photo was taken, approximately how far from the aircraft location the ground target may be, or how large the ground extent of the CAP photo may be. These challenges require new methods to identify ground target locations that do not require a priori information about the photo geometry.

The objectives in this task are 1) to test off-the-shelf algorithms for assessing the similarity of CAP image with orthorectified reference planar aerial imagery (e.g., National Agricultural Imagery Program [NAIP]); and 2) to determine if the true location of CAP image can be approximated and improved using spatial information of most similar planar images. As with the ground-level imagery, proper location identification leads to the next step in the process workflow, feature matching, which ultimately allows for a transformation of the oblique data into geographically referenced data that can fused with other sources.

Figure 25. Civil Air Patrol flight segment after Hurricane Florence. The black dots provide locations where a photograph was taken and through a recreation of the time-stamps, the general flight line (red) is reconstructed.

Figure 26. Example oblique style aerial image captured by the Civil Air Patrol after Hurricane Florence.

Two off-the-shelf algorithms were tested for the purpose of comparing CAP photos to high-resolution planar imagery (NAIP) to determine potential ground target locations based on image similarity. We applied a deep convolutional neural net, dubbed Xception (Chollet 2017), and an efficient similarity search algorithm developed by Facebook Research called Faiss (Johnson, Douze and Jégou 2019). Xception was employed to create high-dimensional feature representations of CAP and NAIP photos, which were subsequently indexed using Faiss to enable queries of the most similar NAIP images for a

given CAP photo. This process was implemented using the Amazon cloud resources implemented via the developed MFM Data Pipeline and Architecture. A visual workflow is presented in Figure 27.

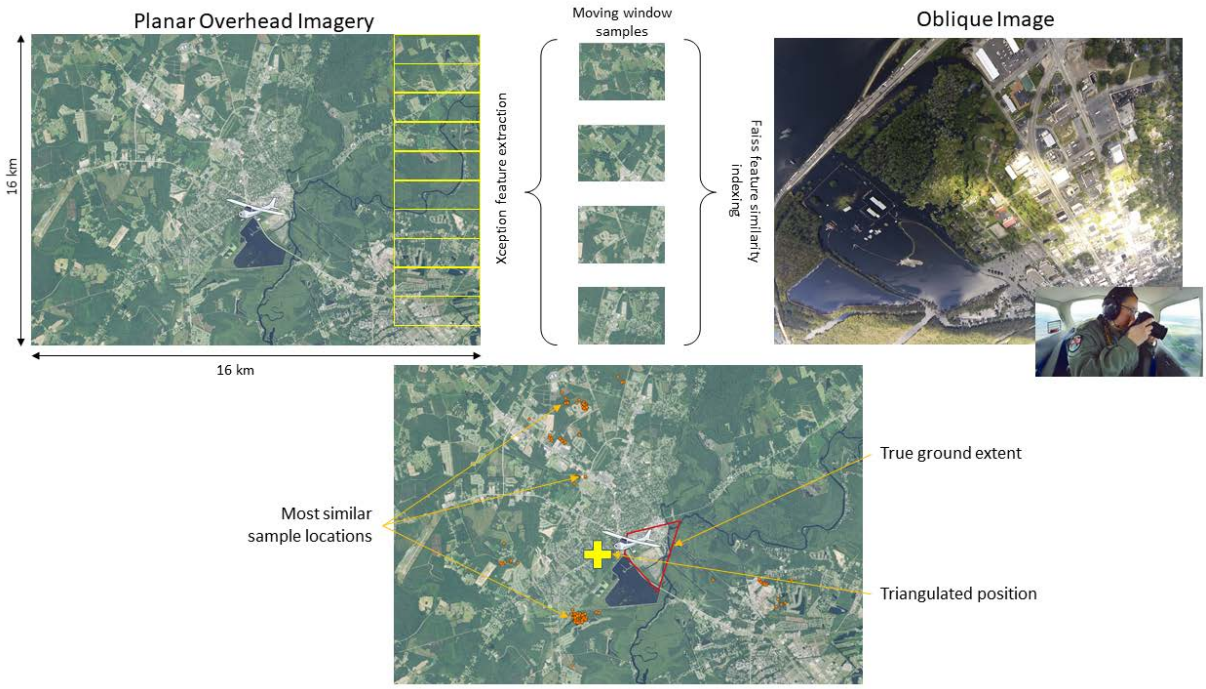

Figure 27. A visual workflow of the Civil Air Patrol image matching methodology relying on varying size image chips, the Xception and Faiss models.

NAIP imagery was selected for initial testing because it is freely available, recent (updated every three years), and is of high enough resolution (typically ≤1 m horizontal resolution) that it should provide a reasonable assessment of the approach. Because the ground extent and location of the CAP photo is unknown, a large number of NAIP images of varying sizes and landscape positions relative to the aircraft location are required to encompass the full range of potential ground locations. A Python algorithm was written to perform this task and included the following steps: 1) download enough NAIP tiles from a NAIP Web Map Service (WMS) to create a mosaic image covering ~8-km radius around a given CAP photo location; 2) chip the NAIP mosaic into smaller, overlapping images using a moving window process; and 3) save the NAIP chips to an Amazon S3 storage container. This algorithm was run multiple times to generate NAIP chips of multiple sizes ranging from 300 x 300 to 1500 x 1500 pixels square. A default step size of half the chip size was used for all iterations. Results from testing different chip sizes against return number of similar images are presented in Figure 28.

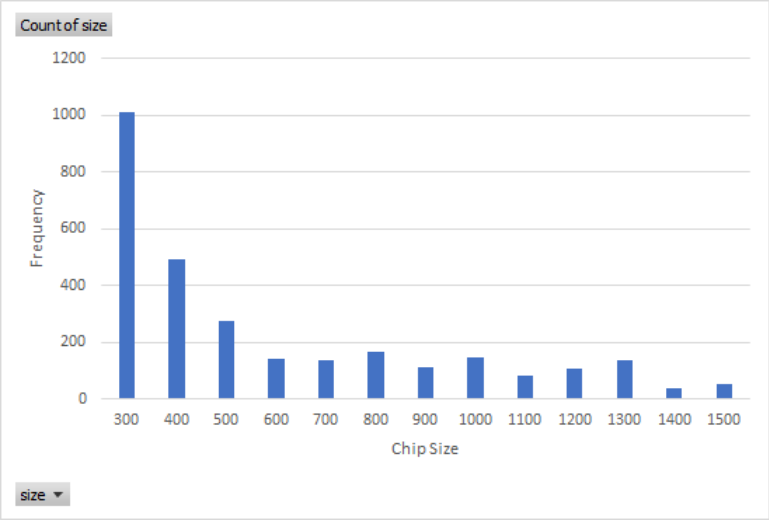

Figure 28. Tests were run to understand the ideal image chip size against similarity measures. Smaller chip sizes were more effective in this regard.

All CAP photos and NAIP chips were subsequently analyzed and indexed as described above so that similarity searches could be performed. All NAIP chips were considered in the indexing process due to limitations with the existing implementation of Faiss. This meant that any NAIP chip generated for the greater set of CAP test locations could be returned in a similarity search despite not being co-located within the 8-km radius of the target CAP photo. To limit the results of a similarity search within this area an algorithm implemented in Python was run to calculate the Euclidean distance of a NAIP chip from the target CAP photo which could be used to filter results based on maximum allowed distance.

A simple Python algorithm was written to triangulate the centroid of NAIP chip locations returned by similarity searches. The centroid and associated NAIP chip locations were then visually compared to the CAP photo and true ground location to assess the efficacy of the overall approach. Some examples of the results are presented in Figure 29 to Figure 31. The geographic location and validation of oblique aerial images is an area that will require additional approaches to be implemented and tested. One of the challenges in running image similarity is the CAP images showing flooding are significantly different than the reference imagery that doesn't have flooding, thus the use of a mixed set algorithms, including geometric matching, will likely help to improve the matching.

Figure 29. Example CAP image similarity matching. Pink square = extent of NAIP image chipping; Green triangle point = CAP photo/aircraft location; Green polygon = approximate true ground extent; Orange points = "similar" NAIP chips; Yellow cross = triangulated centroid.

Figure 30. Example CAP image similarity matching. Pink square = extent of NAIP image chipping; Green triangle point = CAP photo/aircraft location; Green polygon = approximate true ground extent; Orange points = "similar" NAIP chips; Yellow cross = triangulated centroid.

Figure 31. Example CAP image similarity matching. Pink square = extent of NAIP image chipping; Green triangle point = CAP photo/aircraft location; Green polygon = approximate true ground extent; Orange points = "similar" NAIP chips; Yellow cross = triangulated centroid.

A manual assessment of the resulting NAIP image chips showed that the landcover patterns within them were often similar to those in the CAP image, but the specific locations were not associated. In general, this technique lacked skill in identifying image chips that were coincident with the actual ground location covered by the oblique images, and the triangulated location of the images tested did not improve the understanding of the CAP image location. There are a few potential improvements that could be enabled in future efforts. One key concern is that the version of the Xception model that we employed was trained on the ImageNet dataset. ImageNet is dominated by images taken at ground-level; thus, the features therein may not be easily associated with those derived from aerial images. Other potential modifications to the approach might include: apply standard, random transformations to the NAIP imagery before feature extraction (rotation, flipping, etc.); employ more complex models to address train-test mismatch, including generative (VAE, GAN etc.) or techniques from few-shot learning; implement a geometric matching approach for detecting image similarity; use pre-trained baseline model from Functional Map of the World (FMoW); and test a weighted *k*-nearest neighbor approach for approximating center location based on the similarity of target images to reference image.

3.4 Feature Matching

A method of matching features in a non-authoritative disaster event image (both ground-level and aerial oblique) with corresponding features in a reference image was explored. The method applies an affine scale-invariant feature transform (Affine-SIFT) and maximally stable extremal regions (MSER) approach to establish pixel-based georeferenced coordinates in the non-authoritative image (Geniviva, Faulring and Salvaggio 2014, Bansal et al. 2011). The basic notion with this method is similar to an image-to-map or image-to-image georectification process where recognizable features on a georeferenced image are located

on the same features in a non-georeferenced image, after which a polynomial transformation is applied to generate a final georeferenced image (Figure 32).

In the case here, both ground-level and oblique image perspectives are used. Because of the large number of images to consider, the process of common feature detection and matching needs to be automated. Overlapping feature detects occur and classic photogrammetric stereoscopic principles are applied to generate a 3D model from 2D images because the reference set of images being used can provide many different look angles. For example, the Google StreetView positions the observer perspective +/- 20m from the reported position of the ground-level image, creating sufficient overlap. A small set of oblique and ground-level images were tested for automated feature matching. Results are considered marginal in that there are a high number of feature matches, but few are correct. Examples of both ground-view and oblique perspective feature matchings are provided in Figure 33 and Figure 34, respectively. This is an area that requires additional research where a geometric matching approach is expected to provide a more positive result (Rocco, Arandjelovic and Sivic 2017).

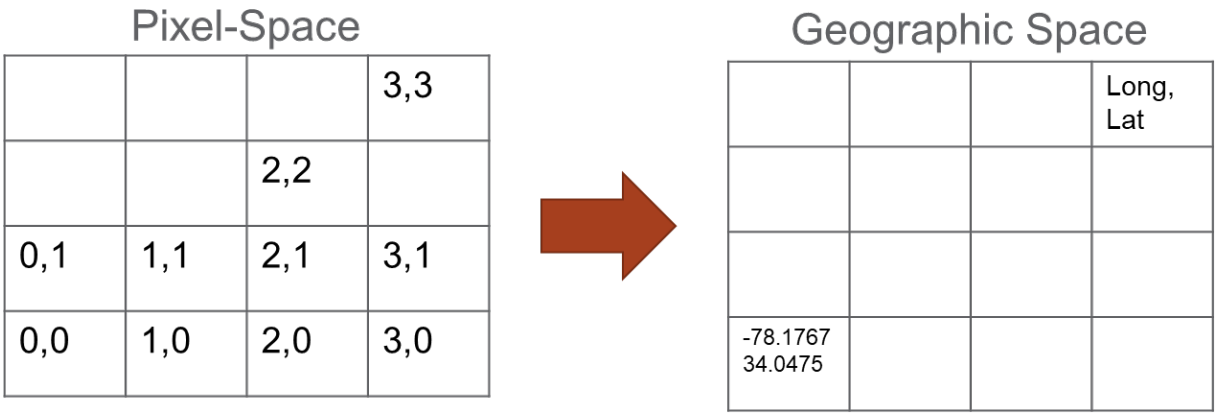

Figure 32. The notion behind the feature matching between a source image and reference image is to enable a transformation between the image arbitrary pixel space to geographic space where each pixel is assigned a coordinate.

Figure 33. Ground-view perspective and image feature matching between the Google StreetView reference image (top) and the social media image (bottom) taken during Hurricane Florence.

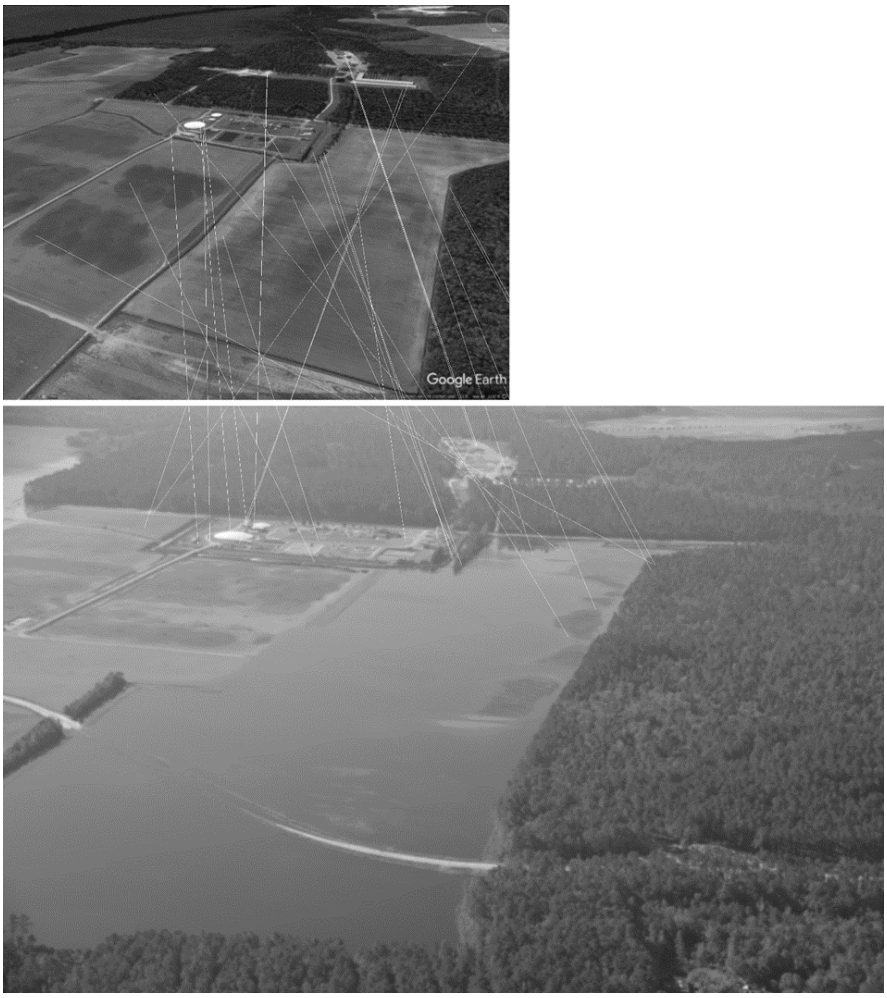

Figure 34. Oblique perspective and image feature matching between the Google Earth reference image (top) and the Civil Air Patrol image (bottom) taken during Hurricane Florence.

3.4.1 Distance Mapping

As described in Section 3.2, Google StreetView data are used as a reference source for determining or validating proper location and heading for ground-level social media images. For the feature matching capability described above, determining common features between the social media image and the reference image provides key information for a transformation of the images from arbitrary pixel space to a geographically referenced pixel space. To further aid in this transformation we add an additional dimension in the z-space. An undocumented and hidden feature of the Google StreetView API is the ability to retrieve 3D distance information for a given coordinate and heading location. The presumption in using the distance mapping data is that the image similarity process has confirmed the source to reference image mapping and there is confidence in extracting additional data for transformation. The returned data is a low-density 3D point-cloud. As part of the Google StreetView metadata retrieval, the

location and heading are known, and the distance data is calculated as linear distance from the source location with explicit consideration of both the azimuth and elevation for each pixel in the 2D array. The native spherical coordinates are converted to a cartesian coordinate system and distance from the camera to a given pixel in the array is stored. The distance data are processed and stored as JSON files. An example of these data are presented in Figure 35.

Figure 35. Google StreetView images including a panoramic view (top) and the associated distance map (bottom) that is retrieved and processed.

3.5 Semantic Image Segmentation

Semantic image segmentation is a deep-learning computer-vision method to regionalize and label pixels in an image to a set of defined classes that are often object descriptors (e.g., building, road, water, car). The purpose of segmentation is to reduce the complexity of the image to a set of delineated objects or features. The segmentation output is stored to a 2D array of attributed coordinate pairs based on the

image transformation defined in the feature matching process. The resulting image segmentation work is the last step of the data fusion process of creating validation and/or training data for the statistical inference model to predict spatial extents of flooding. We implemented and tested several pre-trained deep learning models and applied these to both ground-level and oblique aerial images. A representation of the semantic segmentation process and feature transfer to geographic space is demonstrated in Figure 36.

DeepLabv3 is a new deep convolutional neural network semantic image segmentation model that utilizes dilated (atrous) convolutions. We tested this model using the PASCAL VOC 2012 as a training dataset and used a collection of ground-level and oblique aerial images for testing. Example results are presented in Figure 37 with semantic segmentation performed on oblique aerial images. Model performance was poor, likely because the training dataset does not include aerial images, thus the perspectives and content in an aerial image are largely unknown to the model. Additionally, the class labels within PASCAL VOC 2012 were also limiting for use in our application. Using ground-level images, the model behaves quite well (Figure 38). The model is also trained against the fMoW dataset which has application to overhead remote-sensing classification and segmentation. Initial training runs performed with this dataset yield more promising results, but additional work is still required to achieve acceptable levels of accuracy in the predictions.

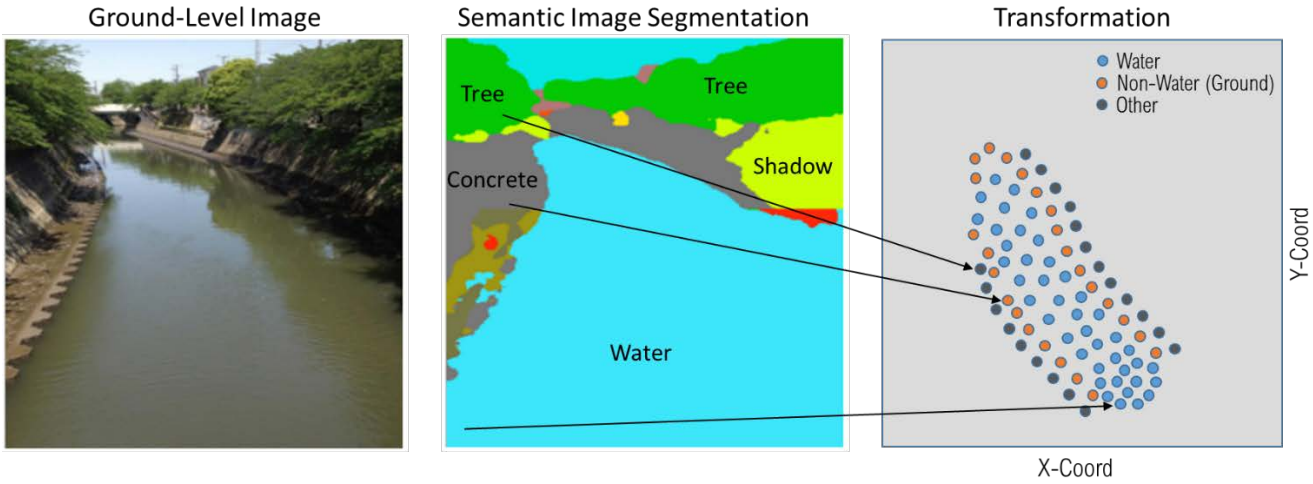

Figure 36. Representation of the semantic segmentation process and the feature transfer to geographic space.

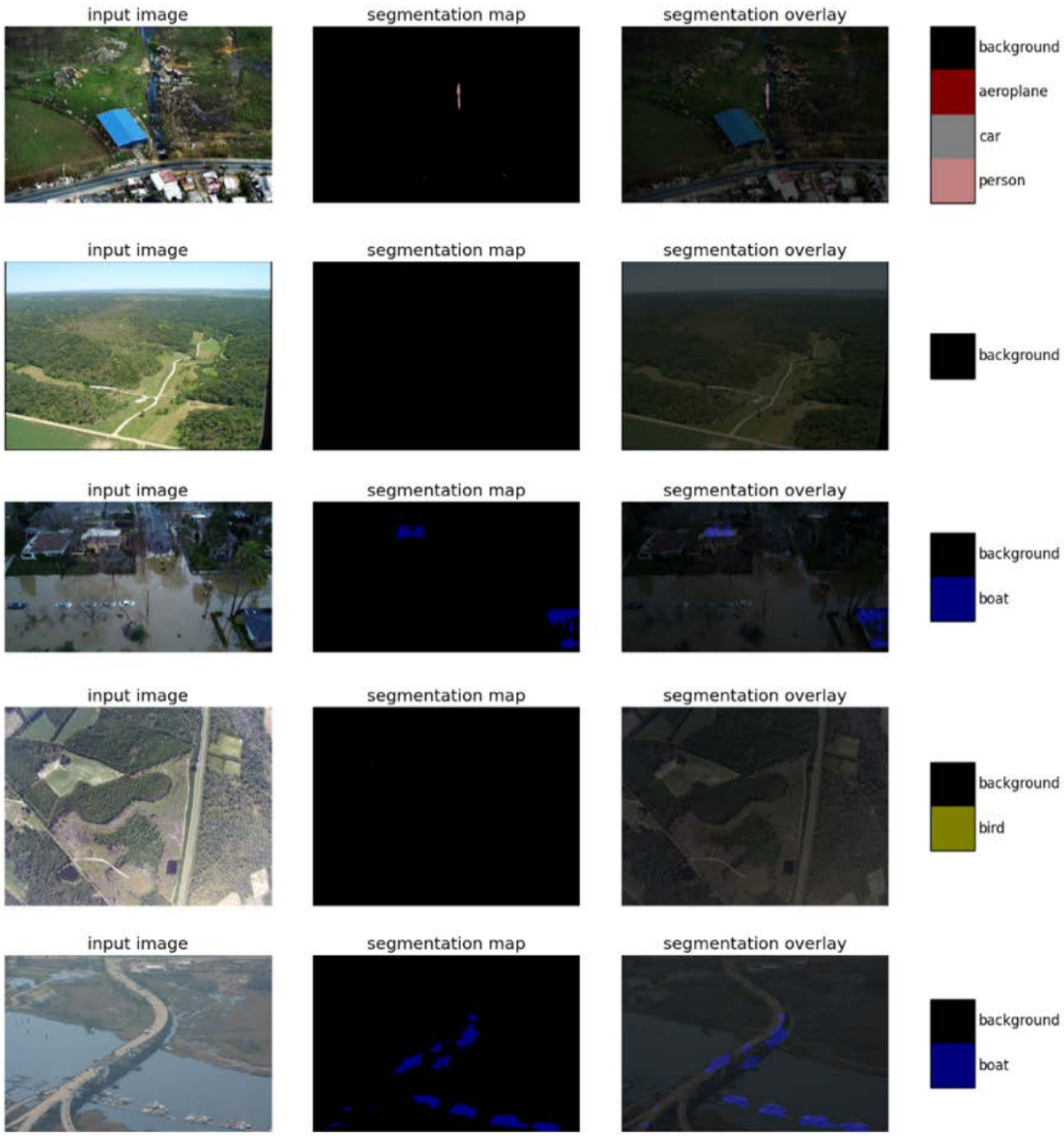

Figure 37. Example semantic segmentation results from Civil Air Patrol images run through DeepLabv3 trained against the PASCAL VOC 2012 dataset.

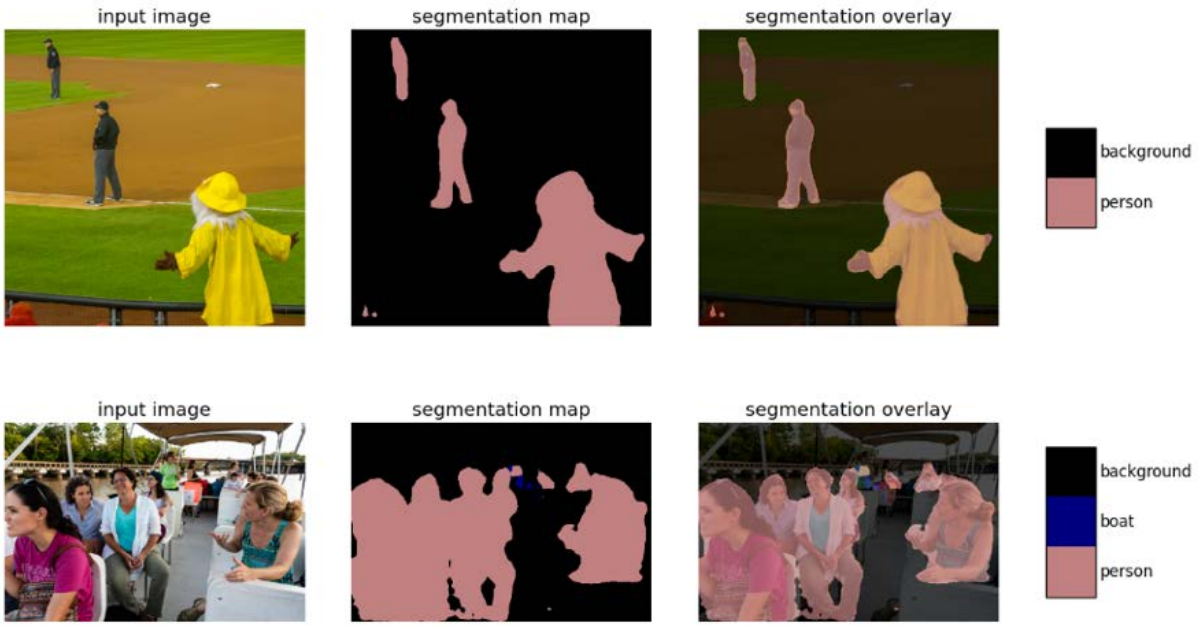

Figure 38. Example semantic segmentation results from ground-level images run through DeepLabv3 trained against the PASCAL VOC 2012 dataset.

In another test, we implemented semantic segmentation tools from the RasterVision deep-learning computer vision library, which has a more direct focus on satellite and aerial imagery. Example results from this approach for classifying building and road classes in nadir-perspective aerial images are presented in Figure 39. The model was also tested on oblique aerial images for building and road class types (Figure 40).

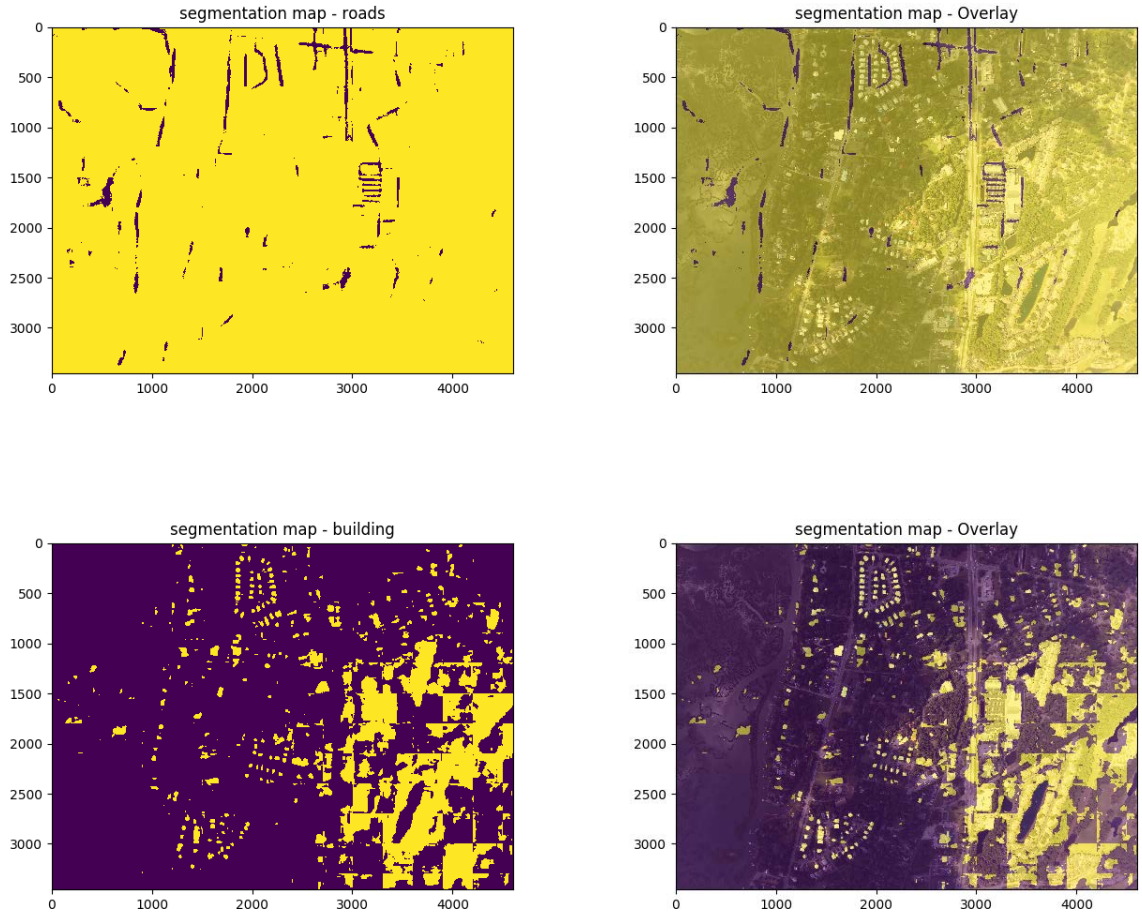

Figure 39. Example semantic segmentation results on nadir-oriented aerial imagery using the RasterVision deep-learning library.

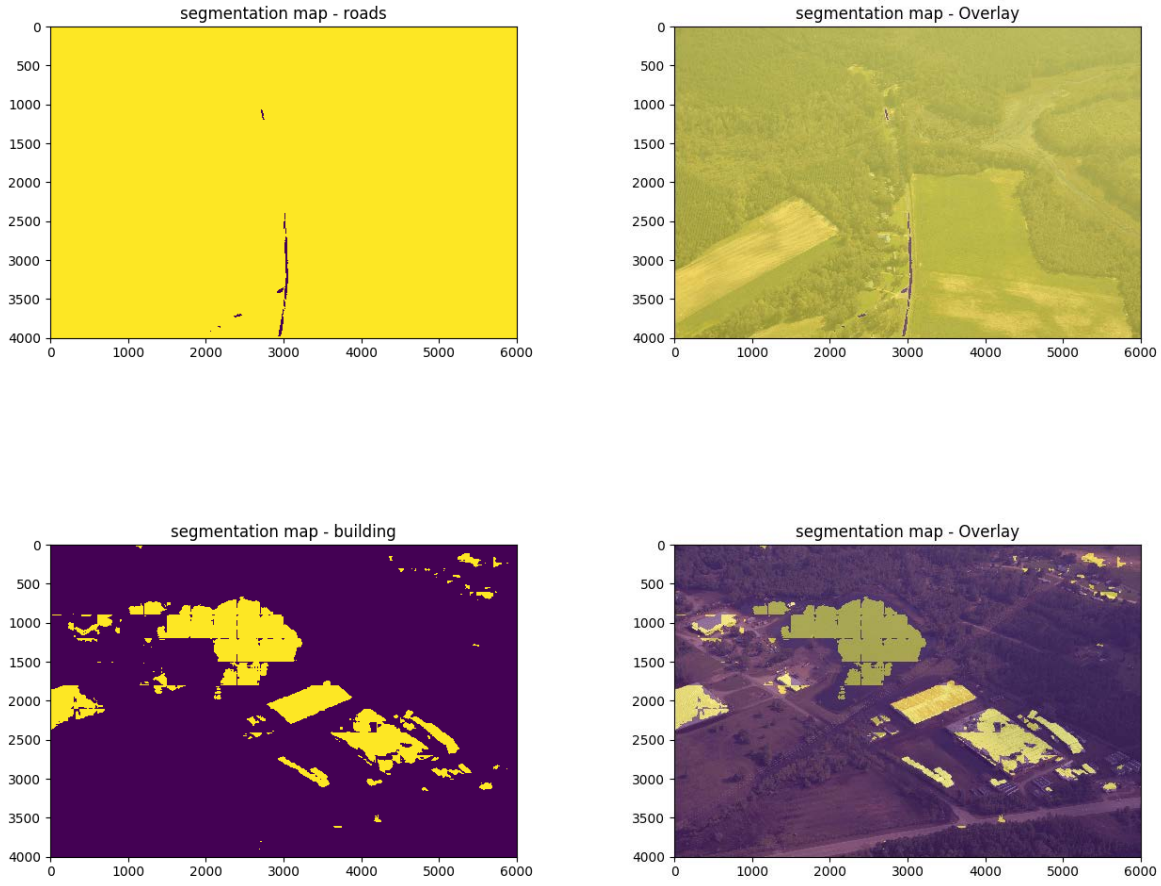

Figure 40. Example semantic segmentation results on oblique aerial imagery using the RasterVision deep-learning library.

Based on testing performed with existing models and training datasets, it was apparent that a new labeled dataset was required for training. This labeling process was implemented using Intel's open-source computer vision annotation tool (CVAT) using the following basic classes: Building, Water, Forest, Field, Road, and Other. CVAT provides a web-based interface (Figure 41 and Figure 42) for managing images in a queue and performing the delineations and associated labels. This process is time-consuming and by the project end we did not have enough images to perform a new model training. Future work in this space may be better implemented through Amazon's Mechanical Turk.

Figure 41. The image queue in CVAT for annotating imagery.

Figure 42. The image delineation and annotation tools within CVAT.

4.0 Statistical Inference Model

The capstone objective of this work is to improve situational awareness during disaster events by using novel sources of data as a means of ground-truth to enable rapid validation of remotely-sensed damage assessments and provide training data for a statistical learning inference model to achieve probabilistic predictions of flooding. The basic notion of inference modeling is to provide the model what is known in regard to where flooding has (and has not) occurred and allow it to derive complex non-linear relationships amongst predictor data to predict flooding at a given location in the event domain. Figure 43 illustrates the project workflow that moves from multiple disparate types of source data to transformation of that data into a common geographic basis, then to the concepts and workflow for the inference modeling. The work described herein describes progress and lessons learned along each of the various steps in this workflow. While the body of work represents significant progress toward the last step, implementation of a culminative statistical inference model could not be fully tested since the non-authoritative source data processing requires further work. As a surrogate for the transformed non-authoritative data, hindcast model results from PNNL's RIFT model were run and flood presence/absence samples were randomly sampled over the training domain.

In addition to the authoritative and surrogate non-authoritative data source, an additional step is to couple these data with landscape descriptors that help define the physical processes that affect flood behavior (e.g., elevation, slope, topographically-driven drainage patterns, land cover, etc.) (Figure 44). The combination of these data is assembled into spatially-explicit training vectors for testing multiple deep generative models to determine a spatially continuous and probabilistic assessment of flooding. Inherent to this process is a pixel-level measure of uncertainty or confidence in the result.

An additional objective is to be able to predict flooding in areas where information may be obscured or simply outside a data collection area. This is an important goal because it reflects the dynamic nature of data availability during an actual disaster event. For example, remotely-sensed flood detections from overhead imagery data can provide a significant amount of information about the extent of flooding, but such data may be incomplete due to cloud cover (optical-based imagery), noncoincident timing of acquisition, or partial coverage of an area of interest. The dynamic nature of flood events and relevant data collection requires inference modeling be trained on past events to better learn the complex associations among predictor data and flooding. During an actual event, the trained model should be run daily at a minimum. If successful, there may be impetus to move toward sub-daily model capability and near real-time data streaming into the system. An expectation that needs to be tested is that model training will need to be regionally specific as the non-linear processes may differ enough between different hydrometeorological regions.

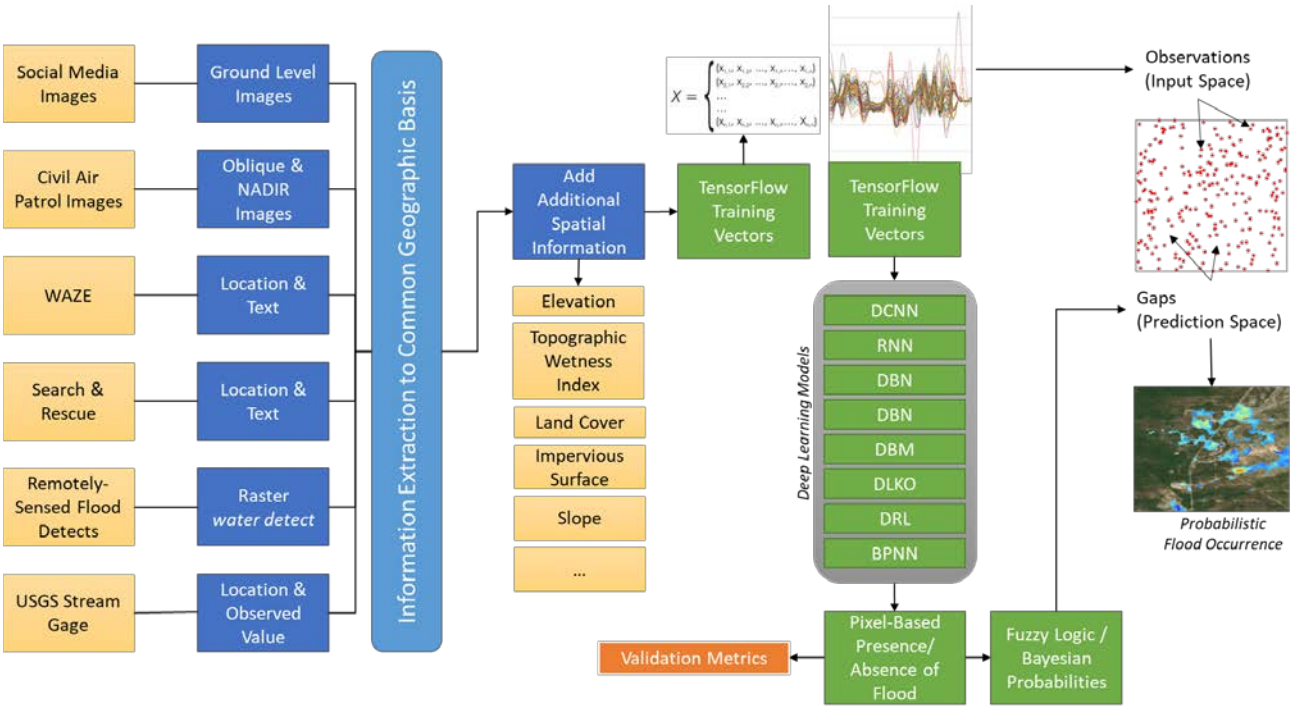

Figure 43. Workflow from source data to statistical inference modeling to achieve the result of probabilistic occurrence of flooding.

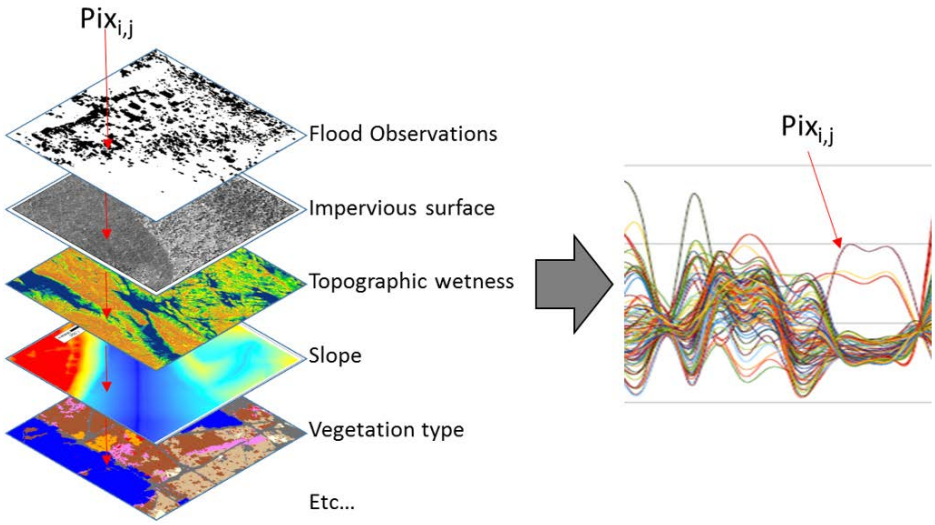

Figure 44. Development of spatially-explicit 'static' datasets to supplement non-authoritative datasets for training vectors in statistical learning.

4.1 Inference Model Setup

The geographic domain for training the inference model was established in the Pee Dee and Cape Fear river basins, which were impacted by Hurricane Florence. Additional impacted basins, Esisto-Santee and

Neuse Pamlico, are set aside for model validation (Figure 45). Some examples of the input data for these areas are presented in Figure 46 - Figure 50.

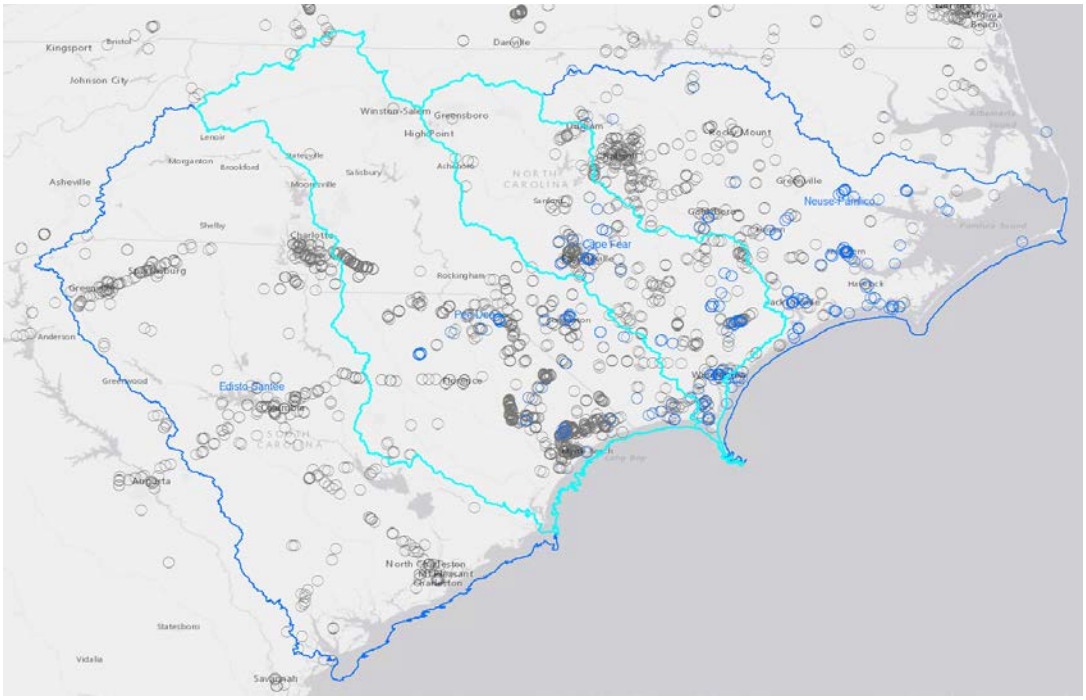

Figure 45. The Pee Dee and Cape Fear River basins were used to assemble training data and the Esisto-Santee and Neuse Pamlico basins were reserved for validation. The circles indicate ground-level observations, where the gray circles indicate no-flooding and the blue circles indicate confirmed flooding.

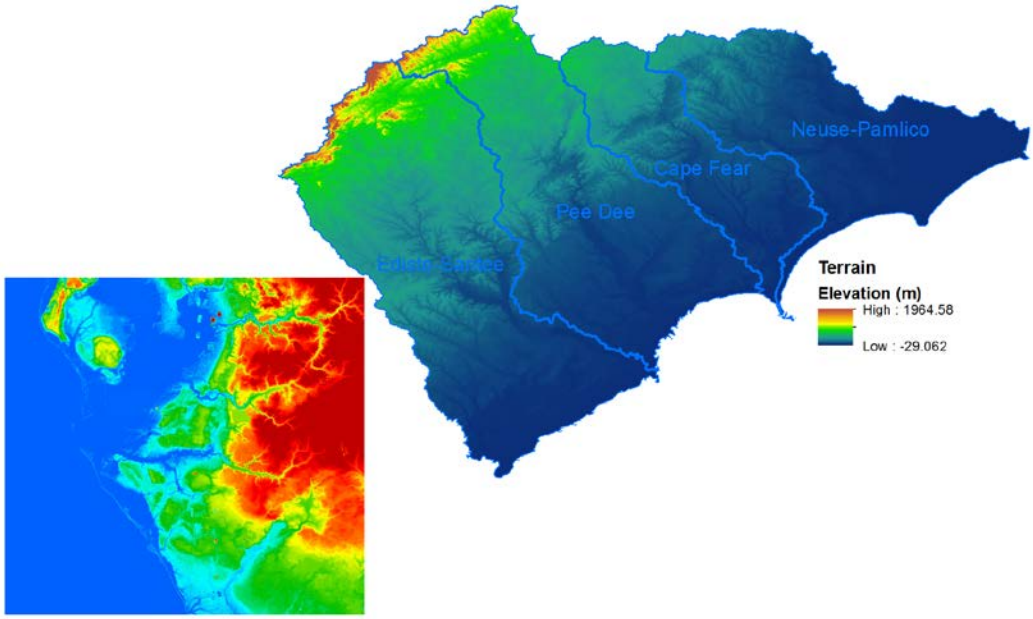

Figure 46. Digital Elevation Model (DEM) representing elevation at 10 meter spacing over the event domain.

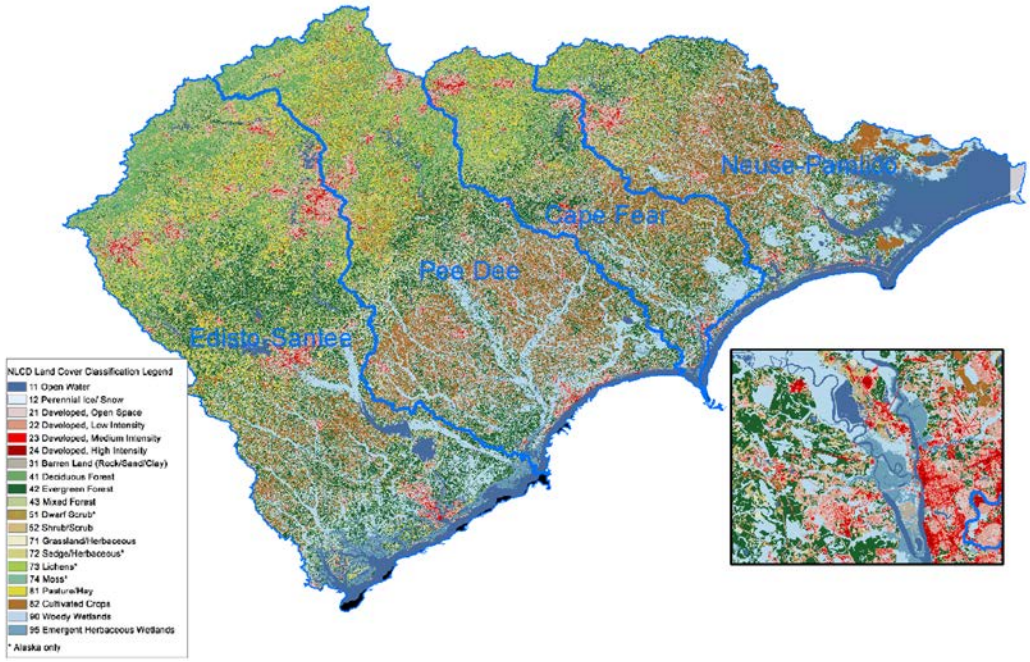

Figure 47. Land use/land cover in the event domain derived from the 2016 National Land Cover Dataset.

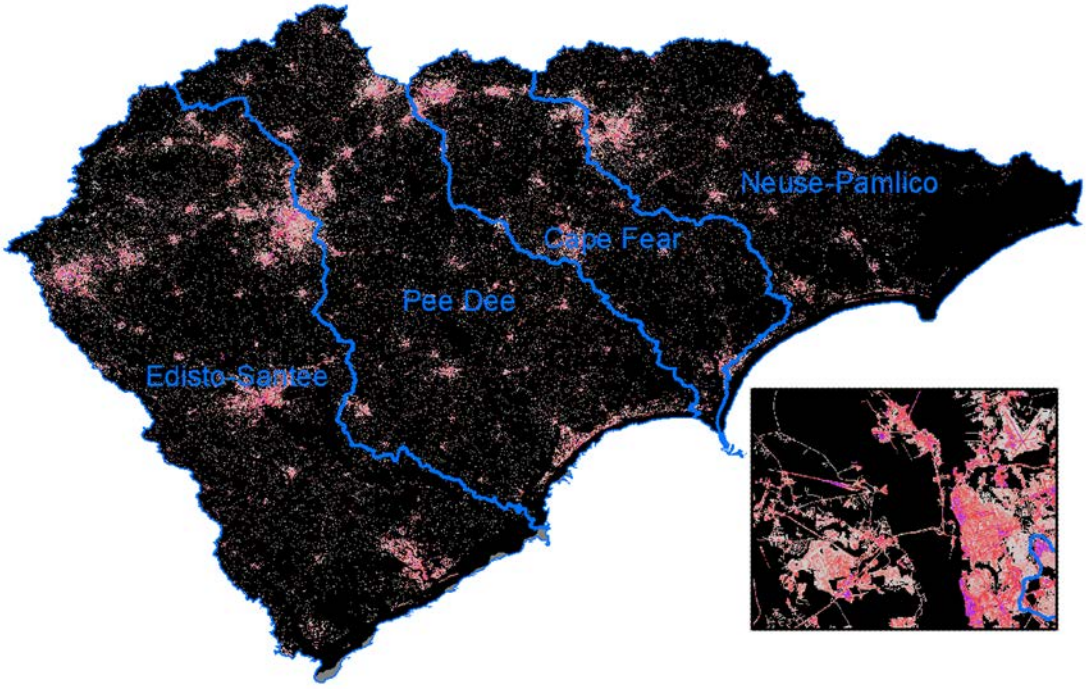

Figure 48. Impervious surfaces in the event domain derived from the 2016 National Land Cover Dataset.

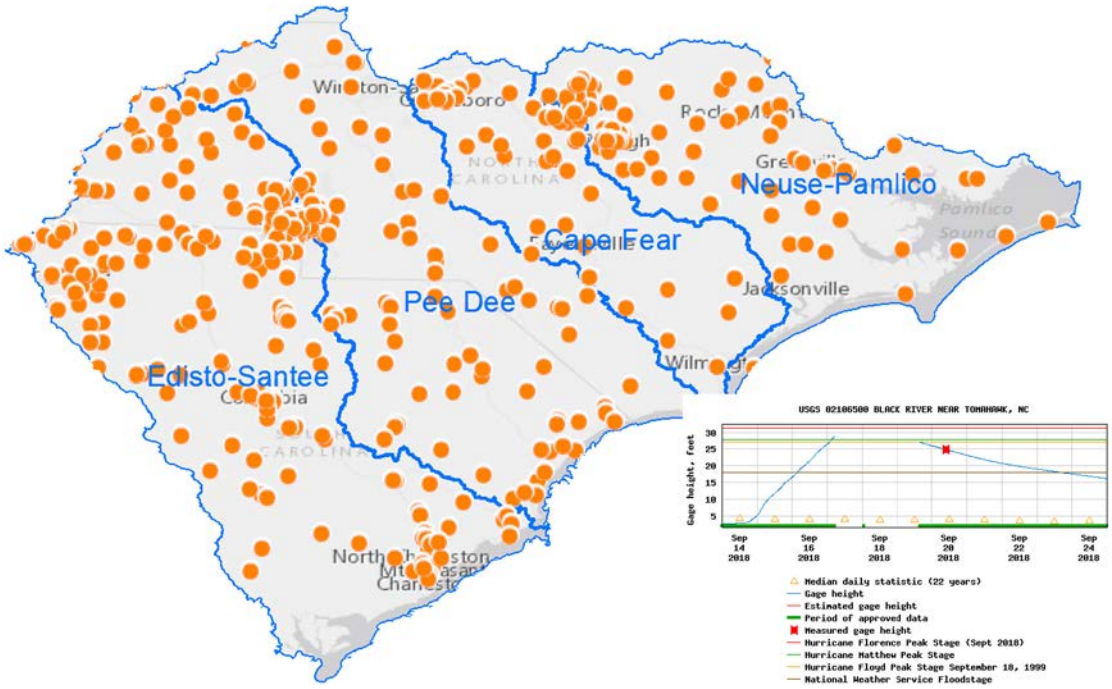

Figure 49. USGS gages within the event area providing stage (water height) measurements over time.

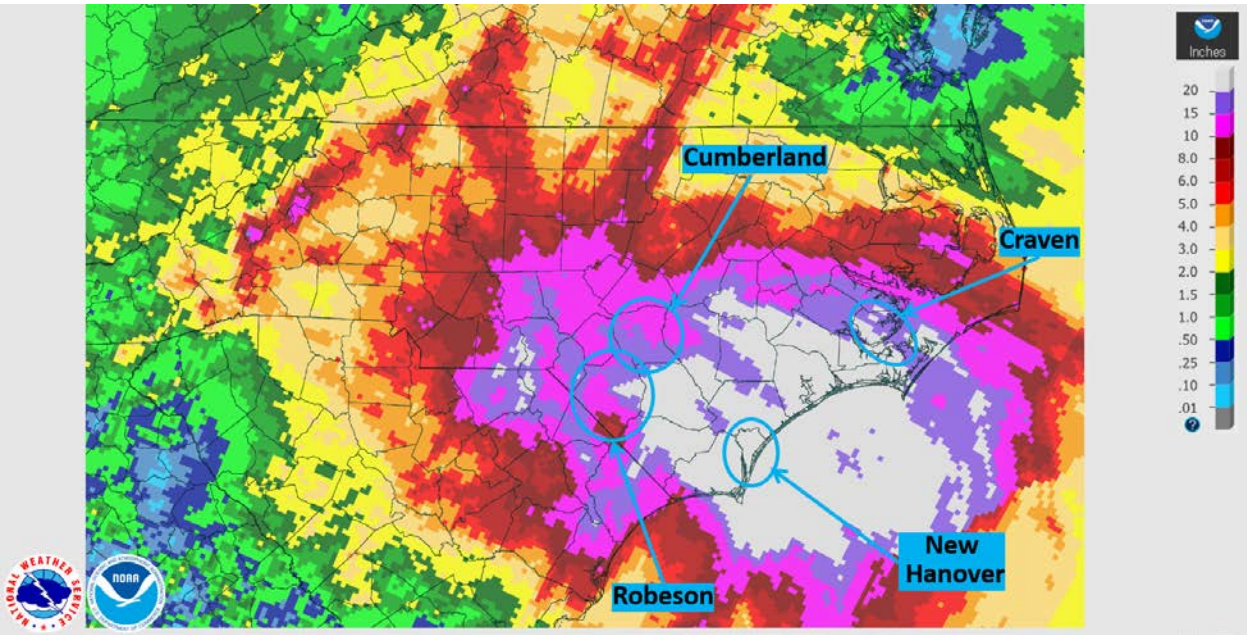

Figure 50. Quantitative Precipitation Estimates (QPE) provided by the National Weather Service provide estimated rainfall totals over time and over space for Hurricane Florence.

The assembled training vectors took the form:

UniqueID, Lat, Long, Elevation, Slope, LandCover (class#), TopoWetnessIndex, Horizontal Distance to Nrml Water, Vertical Distance to Nrml Water, Impervious_Surface (%), Precip_T-24, Precip_T-48,

Precip_T-72, Near Gage Height_T-1, Near Gage Height_T-2, Near Gage Height_T-3, Flood/No Flood (0/1)

The initial model was built using the probabilistic programming language / toolkit, Pyro (Bingham et al. 2019) and PyTorch. Toolkits such as this provide a range of options for developing complex probabilistic models and this particular toolkit is able to implement automatic enumerations over discrete latent variables. Using this toolkit, a variational autoencoder is setup based on a variational inference model (VIM). This VIM implementation was used due to its relatively straightforward setup and reliance on non-linear relationships. A complete validation result was not ready at the time of writing; however, a preview probabilistic result is presented in Figure 51 and demonstrates a functioning model framework. Further validation and testing of additional models will ensue under future research.

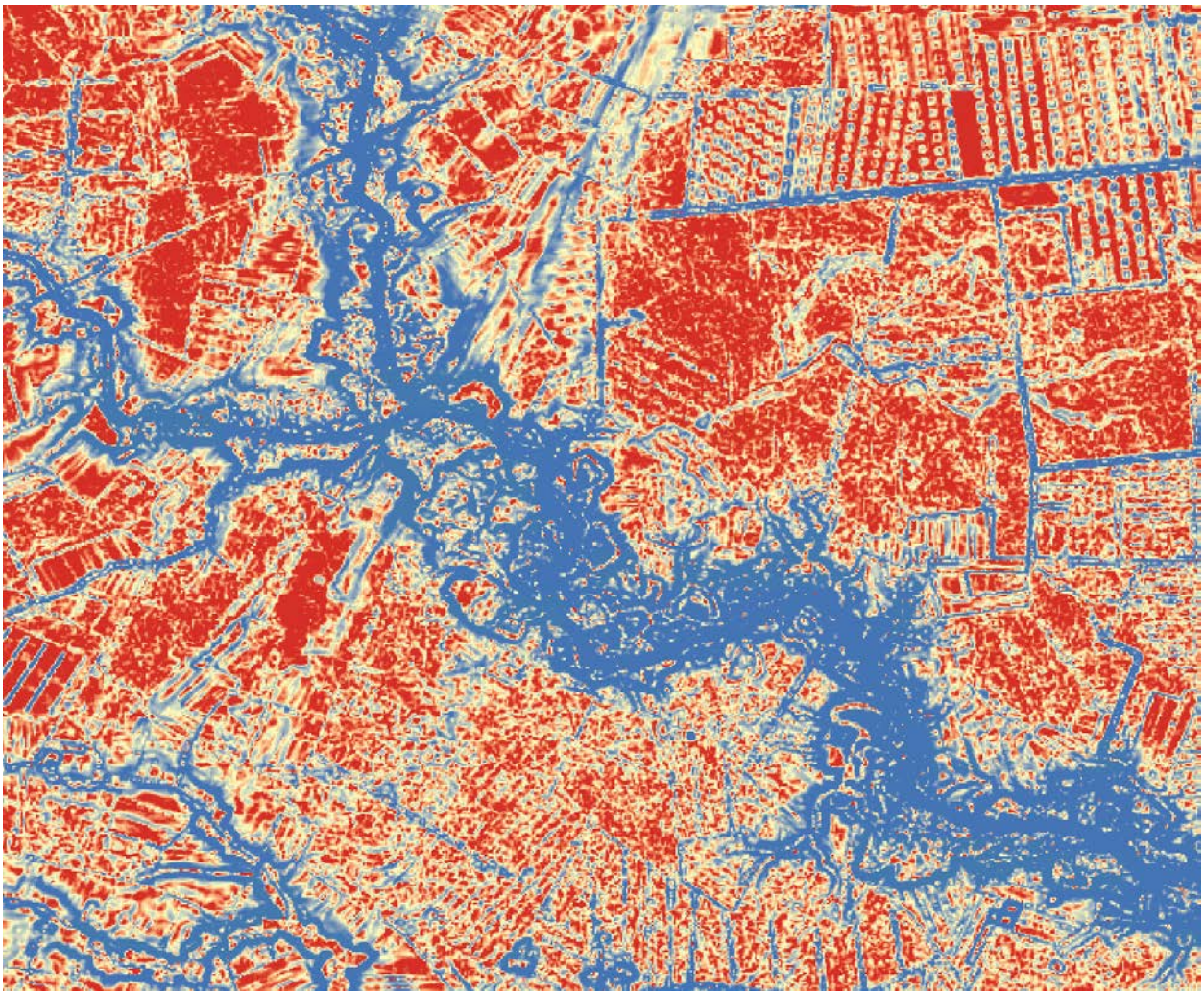

Figure 51. A preliminary probabilistic output produced using a variational inference model. The cool colors indicate a higher probability of flooding and the warmer color indicate a lower probability of flooding.

5.0 Data Pipeline, Architecture, API and User-Interface

Ingesting, analyzing, and displaying large amounts of geographical data related to flooding presents a unique challenge. Our project team developed an architecture for dealing with this challenge in a manner that could help a potential end-user find relevant data, provide feedback to the data ingest pipeline, determine emerging threats, and visualize the information geographically. We employed several mature technologies to facilitate efficient data management through the analysis lifecycle. The architecture consisted of 4 primary components: 1) a raw datastore of data sources of interest; 2) a data enrichment pipeline (orchestrated by Apache NiFi) with custom and third-party analytics; 3) enriched data storage (S3 and Elastic Search); and 4) a custom UI to display the enriched data (Figure 51).

We transferred the corpus of Hurricane Florence related data acquired from the RADR project to an AWS Simple Storage System (S3) location. A data pipeline was created using Apache NiFi software to ingest and classify the S3 data holding (Figure 52). As part of the ingestion process, images from Flickr, NAPSG, and Civil Air Patrol were labeled using AWS Rekognition. In addition to static data, the pipeline was engineered to handle real-time data ingestion. Additional third-party and in-house classification models, such as our relevancy models, are also integrated into the pipeline. Output from the pipeline was stored both in S3 and in Elasticsearch for easier searching from the UI (Figure 53).

Several cloud technologies were leveraged to build a lightweight and quickly deployable application programming interface (API). The API allows the UI to access and query the ingested data. The API also can be used to store additional user provided data. We developed the UI using React, a popular JavaScript framework, to display data in a useful manner to a potential end-user.

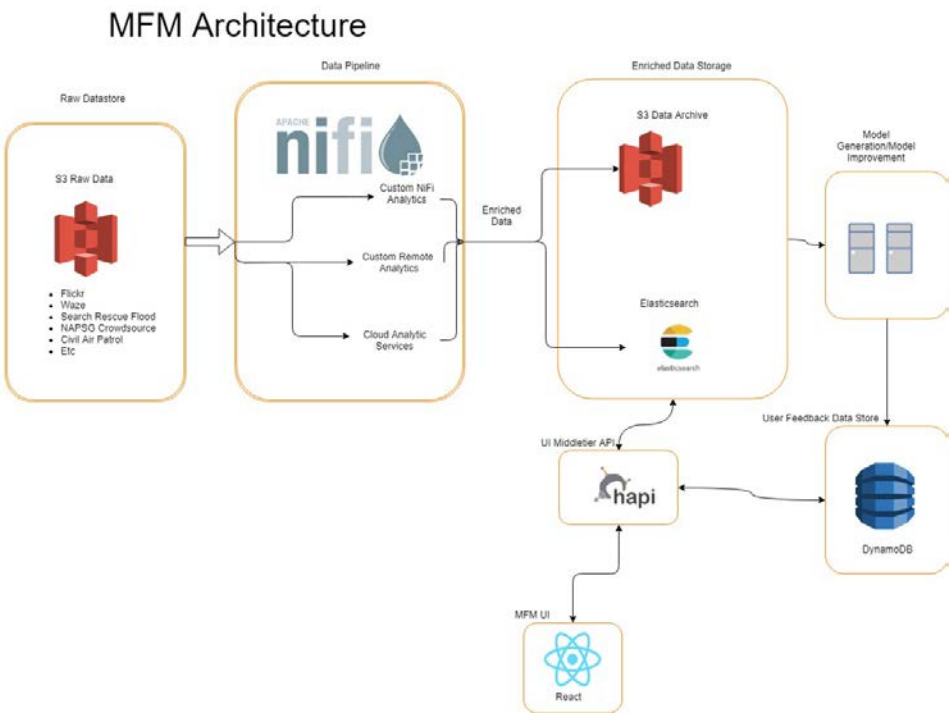

Figure 52. The Multi-Formalism Modeling (MFM) Architecture is a cloud-based data pipeline, storage, multi model implementation, and web-based user-interface.

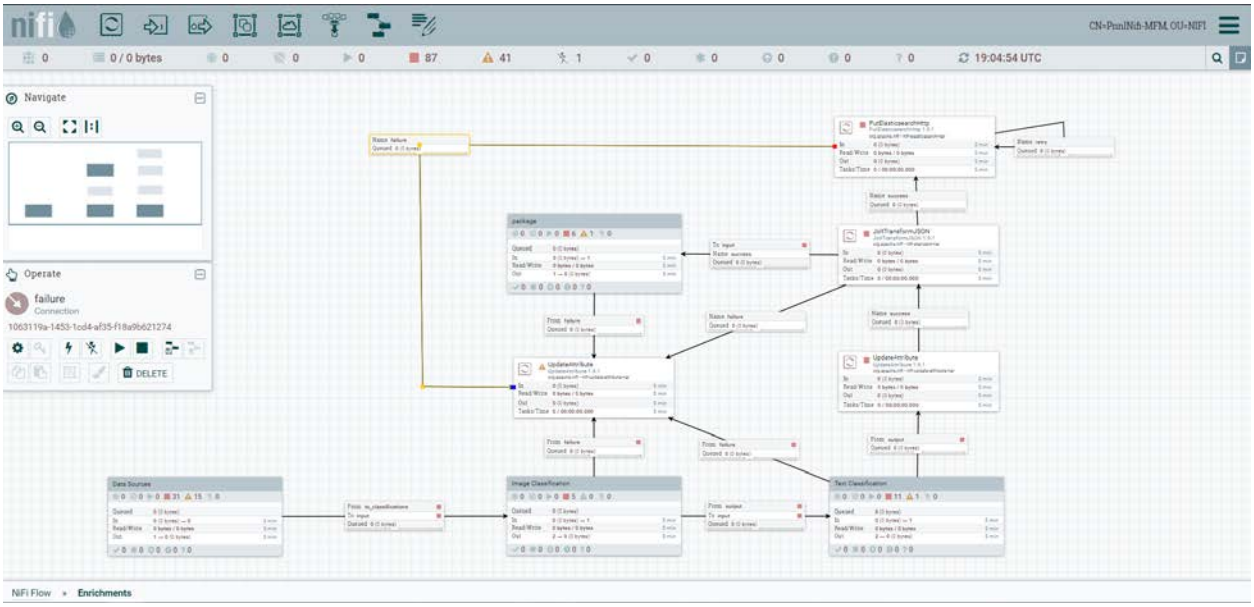

Figure 53. The Apache NiFi ETL pipeline tool ingests data from the raw data store and performs custom analytics.

The data ingestion pipeline, API, and UI assisted in the development of the relevancy model described in Section 3.1 and was a common architecture to deploy and test models after a proof-of-concept was completed. For example, the UI was used by team analysts to conduct and catalogue image relevancy, and the API was used to aggregate this data along with the image labels. In the future, it is envisioned that an

end-user could provide feedback within the UI for improving the relevancy model. The data pipeline is extensible, allowing the relevancy classifier and other potential models to be added to help sort and enrich the data.

The UI demonstrates how the capabilities being developed under this work could be delivered to potential end-users such as first responders, utilities, transportation authorities, and other disaster mitigation entities, to monitor and manage response to floods. Using Leaflet, an open-source mapping tool, a user could quickly visualize emerging threats or on-going events. Map markers and aggregation layers using the location saved during data ingestion help the user to quickly pair data points to a location on the map. Finally, displaying a map overlay of a probabilistic flood model would help the end user to pair their knowledge of the flooding event with our system.

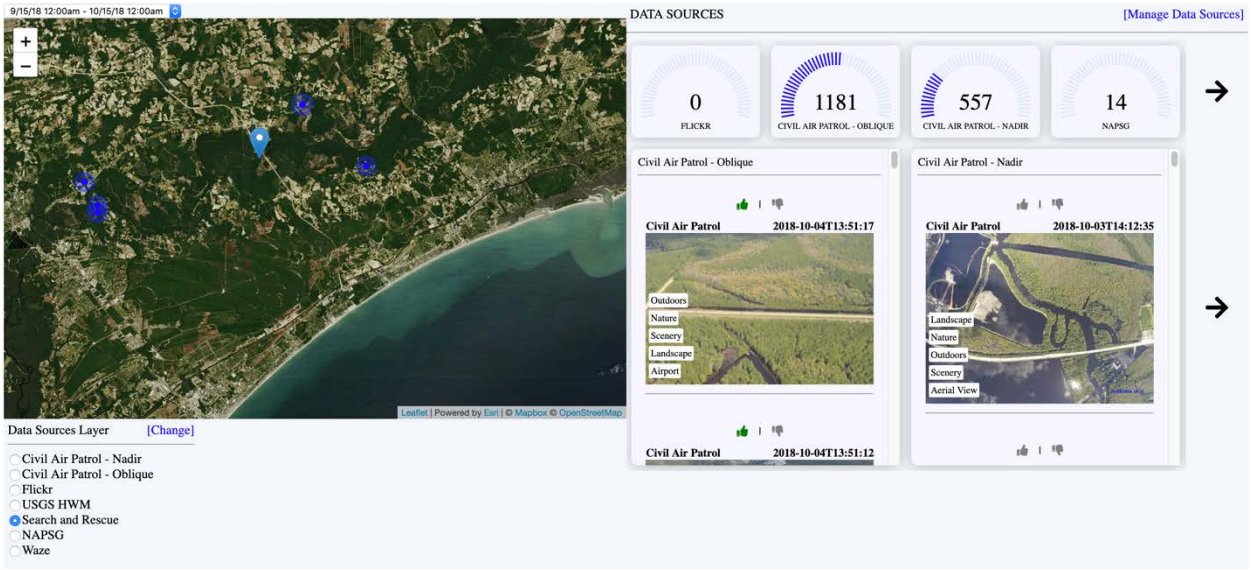

Figure 54. Example of the user-interface developed to interact with the data in multiple ways.

6.0 References

Albright, E. A. & D. A. Crow (2015) Learning processes, public and stakeholder engagement: Analyzing responses to Colorado's extreme flood events of 2013. *Urban Climate,* 14, Part 1**,** 79-93.

Andriole, S. 2015. Unstructured Data: The Other Side of Analytics. . In *Forbes*.

Bansal, M., H. S. Sawhney, H. Cheng & K. Daniilidis. 2011. Geo-localization of street views with aerial image databases. In *Proceedings of the 19th ACM international conference on Multimedia*, 1125-1128. ACM.

Below, R., A. Wirtz & D. Guha-Sapir. 2009. Disaster category classification and peril terminology for operational purposes.

Benfield, A. 2018. Weather, Climate & Catastrophe Insight: 2017 Annual Report. London: Aon Benfield. Available from: http://thoughtleadership. aonbenfield ….

Bingham, E., J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall & N. D. J. T. J. o. M. L. R. Goodman (2019) Pyro: Deep universal probabilistic programming. 20**,** 973-978.

Bouyerbou, H., K. Bechkoum & R. Lepage (2019) Geographic ontology for major disasters: Methodology and implementation. *International Journal of Disaster Risk Reduction,* 34**,** 232-242.

Buchanan, M. K., M. Oppenheimer & R. E. Kopp (2017) Amplification of flood frequencies with local sea level rise and emerging flood regimes. *Environmental Research Letters,* 12.

Cervone, G., E. Sava, Q. Y. Huang, E. Schnebele, J. Harrison & N. Waters (2016) Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. *International Journal of Remote Sensing,* 37**,** 100-124.

Chen, J., I. Dowman, S. N. A. Li, Z. L. Li, M. Madden, J. Mills, N. Paparoditis, F. Rottensteiner, M. Sester, C. Toth, J. Trinder & C. Heipke (2016) Information from imagery: ISPRS scientific vision and research agenda. *Isprs Journal of Photogrammetry and Remote Sensing,* 115**,** 3-21.

Coleman, A., S. Matzner, J. D. Tagestad, K. B. Larson, T. E. Warfel, S. A. Montgomery, A. R. Maxwell, P. J. Bruillard, A. Veeramany & E. O. Jones. 2017. Improving Situational Awareness Through Automated Damage Analysis from Overhead Imagery.

Comfort, L. K., K. Ko & A. Zagorecki (2005) Coordination in rapidly evolving disaster response systems: The role of information. *Agent-Based Simulation: From Modeling Methodologies to Real-World Applications,* 1**,** 208-219.

Crawford, K. & M. Finn (2015) The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters. *GeoJournal,* 80**,** 491-502.

Deng, L. & D. Yu (2014) Deep learning: methods and applications. *Foundations and Trends® in Signal Processing,* 7**,** 197-387.

Geniviva, A., J. Faulring & C. Salvaggio. 2014. Automatic georeferencing of imagery from high-resolution, low-altitude, low-cost aerial platforms. In *Geospatial InfoFusion and Video Analytics IV; and Motion Imagery for ISR and Situational Awareness II*, 90890D. International Society for Optics and Photonics.

Goodchild, M. F. 2016a. Fundamentals of Big Data for CyberGIS. In *CyberGIS Curriculum Workshop Position Papers*. CyberGIS Center for Advanced Digital and Spatial Studies: University of Illinois at Urbana-Champaign.

--- (2016b) GIS in the Era of Big Data. *Cybergeo : European Journal of Geography*.

Goodchild, M. F. & J. A. Glennon (2010) Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth,* 3**,** 231-241.

Gribaudo, M. & M. Iacono. 2014. An introduction to multiformalism modeling. In *Theory and Application of Multi-Formalism Modeling*, 1-16. IGI Global.

Haworth, B. T. (2017) Implications of Volunteered Geographic Information for Disaster Management and GIScience: A More Complex World of Volunteered Geography. *Annals of the American Association of Geographers*, 1-15.

Hodgson, M. E., S. E. Battersby, B. A. Davis, S. Liu & L. Sulewski. 2014. Geospatial Data Collection/Use in Disaster Response: A United States Nationwide Survey of State Agencies. 407-419. Springer Berlin Heidelberg.

Hoeppe, P. (2016) Trends in weather related disasters – Consequences for insurers and society. *Weather and Climate Extremes,* 11**,** 70-79.

Hoque, M. A.-A., S. Phinn, C. Roelfsema & I. Childs (2017) Tropical cyclone disaster management using remote sensing and spatial analysis: A review. *International Journal of Disaster Risk Reduction,* 22**,** 345-354.

Huang, Q., G. Cervone & G. Zhang (2017) A cloud-enabled automatic disaster analysis system of multi-sourced data streams: An example synthesizing social media, remote sensing and Wikipedia data. *Computers, Environment and Urban Systems,* 66**,** 23-37.

Johns, J. M., J. Rounds & M. J. Henry (2017a) Multi-modal Geolocation Estimation Using Deep Neural Networks. *arXiv preprint arXiv:1712.09458*.

Johns, J. M., J. Rounds & M. J. Henry (2017b) Multi-modal Geolocation Estimation Using Deep Neural Networks. *CoRR,* abs/1712.09458.

Johnson, J., M. Douze & H. J. I. T. o. B. D. Jégou (2019) Billion-scale similarity search with GPUs.

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

Lee, J.-G. & M. Kang (2015) Geospatial Big Data: Challenges and Opportunities. *Big Data Research,* 2**,** 74-81.

Lewis, Q. W. & E. Park (2017) Volunteered Geographic Videos in Physical Geography: Data Mining from YouTube. *Annals of the American Association of Geographers*, 1-19.

Li, D. (2016) Towards geo-spatial information science in big data era. *Cehui Xuebao/Acta Geodaetica et Cartographica Sinica,* 45**,** 379-384.

Liu, J. Z., J. Li, W. F. Li & J. Z. Wu (2016) Rethinking big data: A review on the data quality and usage issues. *Isprs Journal of Photogrammetry and Remote Sensing,* 115**,** 134-142.

Mao, H., Y. Hu, B. Kar, S. Gao & G. McKenzie (2018) GeoAI 2017 workshop report: the 1st ACM SIGSPATIAL International Workshop on GeoAI:@ AI and Deep Learning for Geographic Knowledge Discovery: Redondo Beach, CA, USA-November 7, 2016. *SIGSPATIAL Special,* 9**,** 25-25.

Miller, H. J. & M. F. Goodchild (2015) Data-driven geography. *GeoJournal,* 80**,** 449-461.

Murray, V. & K. L. Ebi (2012) IPCC Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX). *Journal of Epidemiology and Community Health,* 66**,** 759-760.

Ngiam, J., A. Khosla, M. Kim, J. Nam, H. Lee & A. Y. Ng. 2011. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, 689-696.

NITRD. 2016. Federal Big Data Research and Development Strategic Plan. ed. N. a. I. T. R. a. D. Program, 39. Washington, D.C.: National Science and Technology Council.

Rocco, I., R. Arandjelovic & J. Sivic. 2017. Convolutional neural network architecture for geometric matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6148-6157.

Sava, E., L. Clemente-Harding & G. Cervone (2017) Supervised classification of civil air patrol (CAP). *Natural Hazards,* 86**,** 535-556.

Schnebele, E., G. Cervone, S. Kumar & N. Waters (2014) Real Time Estimation of the Calgary Floods Using Limited Remote Sensing Data. *Water,* 6**,** 381-398.

Schnebele, E., C. Oxendine, G. Cervone, C. M. Ferreira & N. Waters. 2015. *Using Non-authoritative Sources During Emergencies in Urban Areas*. Springer International Publishing.

Shu, H. (2016) Big data analytics: six techniques. *Geo-spatial Information Science,* 19**,** 119-128.

Smit, B. & J. Wandel (2006) Adaptation, adaptive capacity and vulnerability. *Global Environmental Change-Human and Policy Dimensions,* 16**,** 282-292.
Suresh, S., N. Chodosh & M. Abello (2018) DeepGeo: Photo Localization with Deep Neural Network. *CoRR,* abs/1810.03077.
USACE. 2015. North Atlantic coast comprehensive study: Resilient adaptation to increasing risk. Final Report. 140.
Vo, N. N., N. Jacobs & J. Hays (2017) Revisiting IM2GPS in the Deep Learning Era. *2017 IEEE International Conference on Computer Vision (ICCV)***,** 2640-2649.
Yin, J., A. Lampert, M. Cameron, B. Robinson & R. Power (2012) Using Social Media to Enhance Emergency Situation Awareness. *Ieee Intelligent Systems,* 27**,** 52-59.

Pacific Northwest
NATIONAL LABORATORY

www.pnnl.gov

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)

U.S. DEPARTMENT OF
ENERGY