Towards Using Remote Sensing and Social Media Data for Flood Mapping

Zainab Akhtar
Qatar Computing Research Institute, HBKU
Doha, Qatar
zakhtar@hbku.edu.qa

Ferda Ofli
Qatar Computing Research Institute, HBKU
Doha, Qatar
fofli@hbku.edu.qa

Muhammad Imran
Qatar Computing Research Institute, HBKU
Doha, Qatar
mimran@hbku.edu.qa

ABSTRACT

Ghana's capital, the Greater Accra Metropolitan Area, is vulnerable to flooding. This paper proposes a fusion of satellite imagery and social media data to derive informed flood extent maps and to understand affected population reports during a flood disaster. We use a change detection technique and present an automatic thresholding approach for flood extent mapping using Sentinel-1 images. We explore four different speckle filters and compare them using the VV, VH and VV/VH polarizations to determine the best polarization(s) for delineating flood extents. The VV and VH bands together on Perona-Malik filtered images achieved the highest accuracy with an F1-score of 81.6%. Moreover, we add social media data layers representing tweet text and images posted from the region to highlight different ways these heterogeneous data sources can be used. The obtained signals complement each other and help find flooded roads and areas which are not identified by the satellite imagery analysis.

Keywords

Flood mapping, social media, satellite imagery, remote sensing

INTRODUCTION

Flooding in Ghana is an inevitable socio-natural disaster that has wreaked havoc over the last decade. In 2017 alone, 117 flood events have been recorded by Ghana's National Disaster Management Organisation which affected 4124 people whilst damaging 5700 houses.[1] In particular, the Greater Accra Metropolitan Area (GAMA)–Ghana's national capital and largest urban agglomeration–is the largest and fastest growing industrial hub in Ghana, and has been experiencing unprecedented torrential rain in the form of intensive storm events. These storm events consist of heavy rainfall over a short period of time which result in flash floods that devastate low-lying areas. With a total population of almost 4.7 million people as of 2020, the number of people vulnerable to flooding is intensified in GAMA.[2] Some of the factors attributed to the flooding problem include changes in rainfall and temperature patterns, rapid urbanization, poor physical planning, flaws in the drainage network, and coastal inundation and erosion as a result of climate change (Appeaning Addo et al. 2011; Rain et al. 2011; Karley 2009).

Satellite-based flood inundation mapping is becoming increasingly important during a flood event. Not only does it provide better situational awareness by drawing attention to the areas that require immediate action, but it also assists crisis managers with flood risk preparedness, management, response, and mitigation at the time of disaster (Tiwari et al. 2020). Whilst satellite-based flood extent maps can offer timely information right after a disaster, it fails to provide information in real-time. This is where the potential of social media comes into play, as it provides insights in real-time. With this advantage there comes the drawback of requiring verification and this weakness is actually

[1] https://www.desinventar.net/
[2] https://ghana.opendataforafrica.org/kzspqfg/population-size-ghana-by-region-2010-2020

the strength of remote sensing, where satellite imagery is usually reliable and accurate. Since remote sensing data and social media data complement each other well, as the strength of one is the weakness of the other, the fusion of both sources for rapid flood mapping can be a powerful tool for crisis responders. This is because both the wide perspective provided by the satellite view combined with the ground-level perspective from locally collected textual and visual information can help in planning a rapid response. Obtaining this comprehensive view of the flood situation is crucial because satellite images only give a bird's-eye view of a flood event, but does not reveal the impact the flood had on people's lives. In such situations, social media information can be used to support the conclusions one might be reaching from satellite images, and vice versa; thus allowing for the quick verification of one data set against another.

This paper plans to leverage satellite imagery for flood detection and Twitter data to understand human activity during a flood event. Satellite imagery, particularly synthetic-aperture radar (SAR) imagery, with the help of flood detection algorithms, is the preferred method to delineating flood extent in near real-time as it can penetrate through clouds, unlike optical sensors (Joyce et al. 2009). Likewise, many studies show the use of Twitter social network during flooding, as locals inform one another about casualties, damages, and alerts during a disaster (Velev and Zlateva 2012). Research studies show people closely located to a disaster produce more informative disaster-related tweets (Peters and De Albuquerque 2015; De Albuquerque et al. 2015; Herfort et al. 2014). When it comes to prioritizing relief efforts, it is crucial that the damage assessment maps visualize which buildings, roads, and facilities are completely damaged or partially affected, along with the number of lives at risk. Altogether, this paper aims towards using useful information from different heterogeneous data sources to build a map that would assist crisis managers identify flood extent, determine most vulnerable areas and population and in prioritizing response for communities at high risk.

We use the change detection technique on Sentinel-1 imagery taken before and after the Ghana October 10, 2020 flood event. We determine the best speckle filter for Sentinel-1 images when used with the Otsu's threshold detection technique. Moreover, we determine the best polarization(s) in delineating flood extent when using Otsu's automatic threshold detection technique. Due to the lack of ground-truth data from Ghana, we apply our proposed approach on imagery taken from Enugu Otu in Nigeria and show promising results. Moreover, on top of the remote sensing layer, we add two social media layers representing relevant tweets and images classified using deep learning models. With the help of our consolidated map, we highlight the following three types of signals useful for crisis managers (more details in the results section):

(1) Confirmatory signals: when remote sensing, tweets, and images show flooding in the same area.

(2) Complementary signals: when tweet text or images bring additional contextual information and report/show the kind of damage (flooded roads, bridges, damaged building).

(3) Novel signals: when one of the sources (either remote sensing, or tweet text, or tweet images) shows flooding or other useful information.

The rest of the paper is organized as follows. In the next section, we summarize related work followed by a section on study area and data. We present our approach in the *methodology section* and show results in the *results section*. Finally, we provide a discussion of the results and conclude the paper in the last section.

RELATED WORK

Remote Sensing for Flood Detection

Satellite-based remote sensing is a cost-effective way to delineate floods in near real-time due to the availability of timely multi-temporal images. Two different types of satellites have been used to map floods, the first one being optical followed by multispectral. Synthetic-aperture radar (SAR) is a form of radar imagery which is unaffected by weather conditions such as cloud coverage during flooded periods, thus making it suitable for detecting floods (Carreño Conde and De Mata Muñoz 2019). However, SAR data produces flood extents with limited accuracy in densely populated areas due to the increased amount of backscattering from buildings.[3] Optical data on the other hand, is able to separate water from non-water regions much easily compared to SAR data due to the distinct reflectance in water. However, getting access to zero cloud coverage optical data during a flood event is very unlikely as the flooded areas are usually covered with clouds.

Based on the summaries by Liang and Liu (2020), various SAR-based flood detection techniques have been proposed in literature, including histogram thresholding, fuzzy classification, interferometric coherence calculation, region

[3]https://un-spider.org/advisory-support/recommended-practices/recommended-practice-google-earth-engine-flood-mapping/in-detail

growing and active contour model, object-oriented classification, and change detection. Among these methods, thresholding (manual and automated) is the most commonly adopted method due to its simplicity and ability to achieve similar accuracies compared to more complex algorithms. Liang and Liu (2020) further divides thresholding methods into two categories, the first one being global thresholding which uses a single threshold for the entire image, and the second category being local thresholding which uses varying thresholds for different parts of the image. Most thresholding methods used in research rely on a single threshold to separate water from non-water, where some researchers like Martinis et al. (2009) considered only a single SAR flood image of the flood event to extract water regions, whilst others like Giustarini et al. (2012) make use of a non-flood reference image as well to improve the flood extent. Ruzza et al. (2019) and Tiwari et al. (2020), for example, computed a global threshold value using Otsu's method (Otsu 1979), which is a widely used automatic thresholding technique to differentiate water from non-water regions. According to Ruzza et al. (2019), the best results are obtained when the image histogram is characterized by a bimodal or multimodal distribution.

A few studies have investigated the fusion of both SAR and optical data where Amoah Addae (2018) focused on a floodplain in Accra. Results showed that the fusion of both optical (PlanetScope) derived inundation and radar (Sentinel-1) derived flood extent resulted in an Intersection-over-Union (IoU) of 88% as compared to 75% for the optical image alone and ~50% for the radar image alone. This shows that the combination of flood inundation from different satellite sources can be beneficial in achieving near-accurate flood extents.

Social Media to Detect Human Activity

Volunteered geographical information (VGI) is defined as the collaborative user-generated content through crowdsourcing where users voluntarily contribute geographic data via the web (Goodchild 2007). Based on the systematic literature review conducted by Horita et al. (2013), VGI has been commonly used to manage floods and fires with the predominant research area being disaster response. Social media in particular is being utilized in studies for the early detection of floods, the derivation of flood maps in real-time, and flood depth map generation.

Jongman et al. (2015) analyzed flood-related Tweets in Pakistan and the Philippines to determine its use in early detection of floods. Results showed that flood-related tweets were posted one to seven days prior to the event being reported to the humanitarian organization. For example, flooding due to an overflowing dam in Pambujan in Northern Samar, was reported to the national Philippines Red Cross Society on December 8, 2014. However, on Twitter around 200 tweets and photos were posted two days earlier (December 6, 2014) discussing the start of the flood, thus proving the potential for early flood detection.

Huang et al. (2018) proposed a novel approach on integrating post-event remote sensing optical data with real-time VGI from Twitter to develop near real-time flood maps. A kernel-based weighting algorithm is used to assign a higher weight to a VGI point if the satellite-extracted wetness, i.e., Normalized Difference Water Index (NDWI), is also higher. The proposed methodology was tested on a 2015 South Carolina Flood, where results showed an overall 72.04% match with the ground truth inundation map. Eilander et al. (2016) also studied the use of Twitter data for developing near real-time maps in the city of Jakarta, Indonesia. Tweets that mentioned the physical characteristics of a flood, such as flood depth and the location were utilized in combination with a Digital Elevation Model (DEM) to derive flood probability maps. Results showed that the peaks in the number of tweets correlated with the water level observations at the gauging station where 69% of the validation points were within 500m from the modelled flood extent.

Fohringer et al. (2015) generated flood inundation depth maps from photos posted via Twitter and Flickr within the city of Dresden, Germany during the June 2013 flood. Water depths from the photos were manually estimated, and with the help of a DEM layer and water level observations from gauges, water level estimates were interpolated throughout the area. Similarly, Li et al. (2018) also generated near real-time flood maps by using water height points derived from tweets and stream gauges along with a DEM layer. The model was applied on the 2015 South Carolina flood case study where the results provided a consistent and comparable estimation of the flood situation in near real-time.

The use of social media to assist flood mapping has not yet been investigated in Ghana, and thus GAMA is a new region to be investigated in this paper. Moreover, adding a remote sensing layer will highlight new opportunities to better understand flood extent and associated contextual information from tweets and images.

STUDY AREA AND DATA

Study Area and Flood Event

In this study, we focus on the severe flooding event that happened on October 10, 2020 in various parts of Ghana. The study area selected is the Greater Accra Metropolitan Area (GAMA), as seen in Figure 1, which has a land area of approximately 1585 km^2 (Addae and Oppelt 2019). On 10 October 2020, the GAMA region was severely affected by flash floods that lasted for over six hours. With raging floodwaters up to one metre deep, roads turned into rivers which caused vehicles to be swept away, resulted in traffic chaos, and also damaged many homes. Many locals shared their sentiments on Twitter by posting images and videos to show the devastation caused by the rains and flooding.[4] Figure 2 shows the devastating impact of the flash flood at the time of the event, where cars are submerged and trees have been uprooted. Two days after the event, Figure 3 shows that the flood waters have also inundated houses.

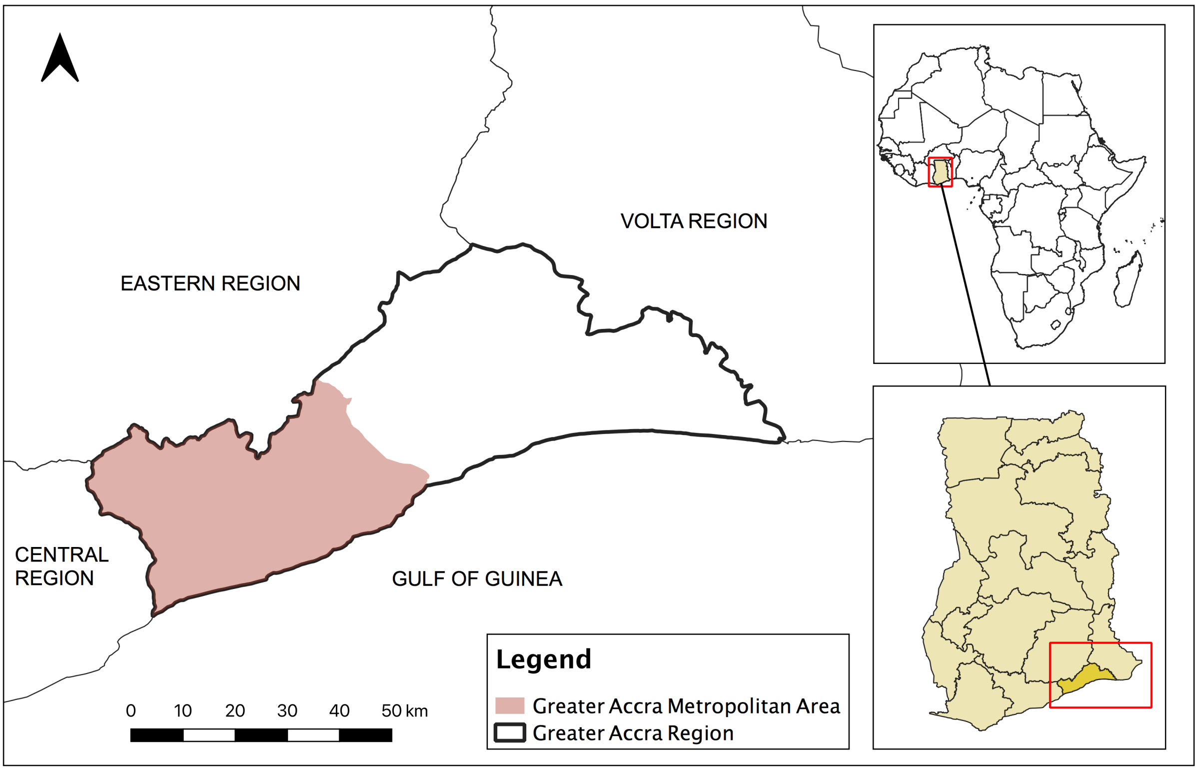

Figure 1. Map of Study Area

Figure 2. Tweet showing roads affected

Figure 3. Tweet showing a house inundated

[4] http://floodlist.com/africa/ghana-flash-floods-cause-traffic-chaos-in-accra

Dataset

Remote Sensing Data

Level-1 Ground Range Detected (GRD) Sentinel-1 SAR images from the Google Earth Engine (GEE) cloud computing platform (Gorelick et al. 2017) were used for flood detection (Table 1). The pre-flood image, i.e., the reference image, represents a day from the driest period in GAMA, between December 4 and February 10, according to Weather Spark[5] analysis of historical data from the years 1980 to 2016 for Accra. Since there was no Sentinel-1 imagery available on the day of the flood event (10 October 2020), the earliest post-flood imagery available was three days after the event (13 October 2020). Both VV (Vertical transmit and Vertical receive) and VH (Vertical transmit and Horizontal receive) polarization's were investigated to determine whether one single band or a combination of both is best at delineating flood extent.

Specifications	**Sentinel-1 (SAR Data)**	
Acquisition Date	Before Flood	12 Dec 2019
	After Flood	13 Oct 2020
Polarization Mode	VV and VH	
Observation Mode	Interferometric Wide (IW)	
Pass Direction	ASCENDING	
Spatial Resolution	10 m	

Table 1. Sentinel-1 data specifications for GAMA

Social Media Data

In order to collect flood-related tweets about GAMA flooding events, we used the Artificial Intelligence for Disaster Response (AIDR) system (Imran et al. 2014). Specifically, we collected 30,266 tweets from October 10–13, 2020 using different keywords and hashtags ($N = 16$) related to Accra floods, including *#AccraFloods, #floodsAccra, Accra flood, Accra flooding, Accra rained, Accra heavy downpour, Accra torrential flood, etc.* Moreover, we downloaded images corresponding to the collected tweets. Table 2 shows the daily distribution of tweets and images.

	Oct 10	**Oct 11**	**Oct 12**	**Oct 13**
Number of tweets	7209	9974	6666	6417
Number of images	865	934	908	906

Table 2. Twitter data collection details: daily distribution of tweets and images

METHODOLOGY

Processing Remote Sensing Data

United Nations Platform for Space-based Information for Disaster Management and Emergency Response (UN-SPIDER) recommends a practice for Flood Mapping and Damage Assessment using Sentinel-1 SAR data in Google Earth Engine.[6] We fine-tuned this methodology to include an automatic Otsu thresholding algorithm (Otsu 1979) rather than a static threshold. Moreover, a difference of the before and after imagery is not taken directly after applying the threshold, but rather it is applied as the last step. A flow chart of our detailed methodology adopted for flood mapping using Sentinel-1 imagery is shown in Figure 4. Next, we describe these methodological steps in detail.

[5] https://weatherspark.com/y/42322/Average-Weather-in-Accra-Ghana-Year-Round#Sections-Precipitation
[6] https://un-spider.org/advisory-support/recommended-practices/recommended-practice-google-earth-engine-flood-mapping/in-detail

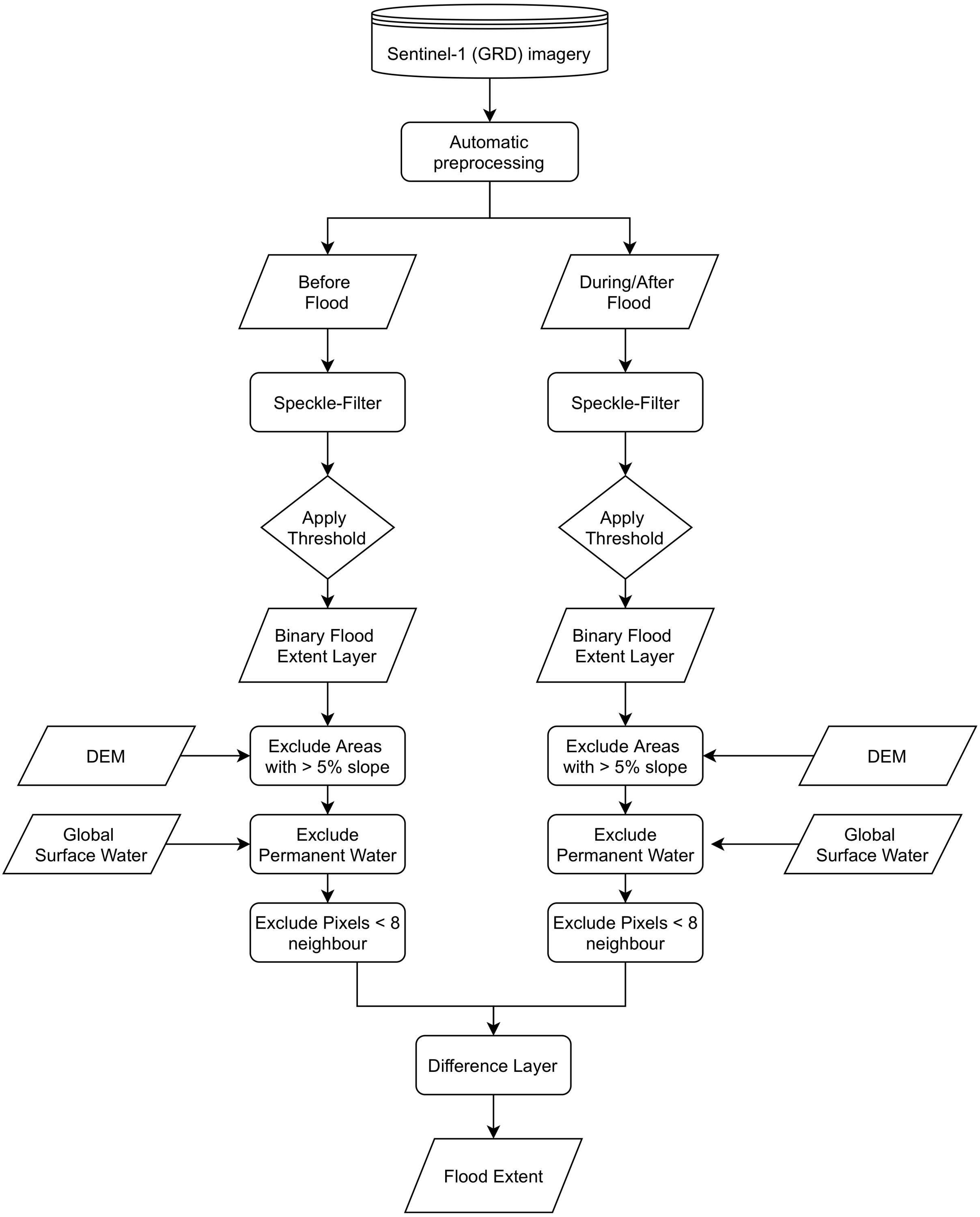

Figure 4. Flood detection methodology using Sentinel-1 (SAR) imagery

Pre-processing

Sentinel-1 Ground Range Detected (GRD) images in GEE[7] include several pre-processing steps as shown in Figure 5. The first step involved applying an orbit file correction to update the product metadata with accurate satellite position and velocity information (Filipponi 2019). Next, the GRD border noise removal step removes low-intensity noise and invalid data on scene edges (Filipponi 2019). To further eliminate noise, thermal noise removal is applied to normalize the backscatter signal within the entire Sentinel-1 scene (Filipponi 2019). Now that

[7] https://developers.google.com/earth-engine/guides/sentinel1

majority of the noise has been removed, a radiometric calibration is performed to convert image intensity values into sigma nought values (Filipponi 2019). Lastly, the image is terrain-corrected where the image is converted from slant range geometry to a map coordinate system (Filipponi 2019).

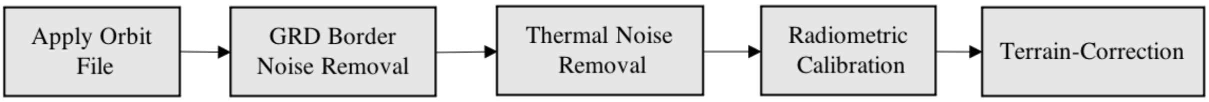

Figure 5. Pre-processing steps done in GEE

Speckle-Filter

All radar images come with speckle noise which is variation in backscatter from nonuniform cells which gives a grainy appearance to the image thus making the segmentation classification of radar imagery complicated (Mansourpour et al. 2006). Whilst many speckle filters have been proposed in literature, the Gamma MAP (Lopes et al. 1990), Perona-Malik (Perona and Malik 1990), Lee (Lee 1980; Lee 1981b) and Refined Lee (Lee 1981a) have been tested in the GEE environment to evaluate the performance of each filter in reducing speckle noise in SAR images.

Otsu Threshold and Binarize Flood Extent

The next step involved classifying the Sentinel-1 SAR images into two classes: water and non-water, to determine the water inundated areas. Otsu's automatic threshold detection method assumes that the distribution of image pixels follow a bi-modal histogram with two varying peaks thus allowing it to determine an optimum threshold for separation. The optimum threshold value is determined by minimizing the weighted sum of within-class (intra-class) variances of the foreground and background pixels (Moothedan et al. 2020). Next, we apply the thresholds identified by the Otsu algorithm on the before and after images to get a binary layer of water-inundated regions.

Refine Flood Extent

In order to further refine the flood extent layer, three post-processing steps have been applied. First NASA's Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) which has a resolution of approximately 30m (Farr et al. 2007), was used to remove flood classified pixels that have a slope greater than 5 degrees to further reduce speckle noise (a similar approach has been used by (Cian et al. 2018)). Next, we use a high-resolution Global Surface Water Dataset (Pekel et al. 2016), particularly the 'Seasonality' layer, to mask out flooded pixels that intersect with permanent/perennial water bodies that have water present for more than 10 months of the year. Lastly, flooded pixels which are connected to eight or fewer neighbours are eliminated to get rid of further noise in the flood extent, thus resulting in a final flood delineation layer.

Processing Social Media Data

Tweets geolocation and classification

In this study, we are interested in tweets which report useful information about Accra flooding on October 10, 2020 and ideally posted from within the Accra region of interest or, at least, mentions places, street names, or points of interest from Accra. Although the tweets were collected using keywords related to Accra floods, many of them might still be content-wise irrelevant or not geographically related to Accra. For this purpose, next we extract geographical information from tweets using the approach proposed in (Qazi et al. 2020). The geolocation inference technique uses various meta-data fields from a tweet to geolocate a tweet. Table 3 lists all the fields and our priority order (1=highest) to decide which meta-field, when available, is used to put a tweet on the map. Given the output of the geo-inference algorithm and our prioritization, we add tweets on the map using the approach shown in Figure 6. We discard tweets outside our region of interest (i.e., GAMA). Table 4 shows the remaining tweets geolocated in the region of interest (column "Geolocated in GAMA").

In order to filter out irrelevant tweets, we used two deep learning classifiers from CrisisDPS system (Alam et al. 2019), namely, *disaster type* (reported F1-score=0.93) and *informativeness* (reported F1-score=0.93). These classifiers help us filter out tweets which are not related to flood disaster and do not report any useful information. Table 4 shows the distribution of the number of geolocated tweets classified as related to flood disaster and are informative.

Priority	Meta-data field	Description
1	Geo-Coordinates	Exact coordinates from the user's device
2	Place	The user tags specific locations in their tweets
3	Tweet Text	Actual tweet content. We extract toponyms from it
4	User Location	Users shared location (free-form text)
5	User Profile Description	Profile description provided by users. We extract toponyms from it

Table 3. Tweets meta-data fields used for geo-inference and their prioritization

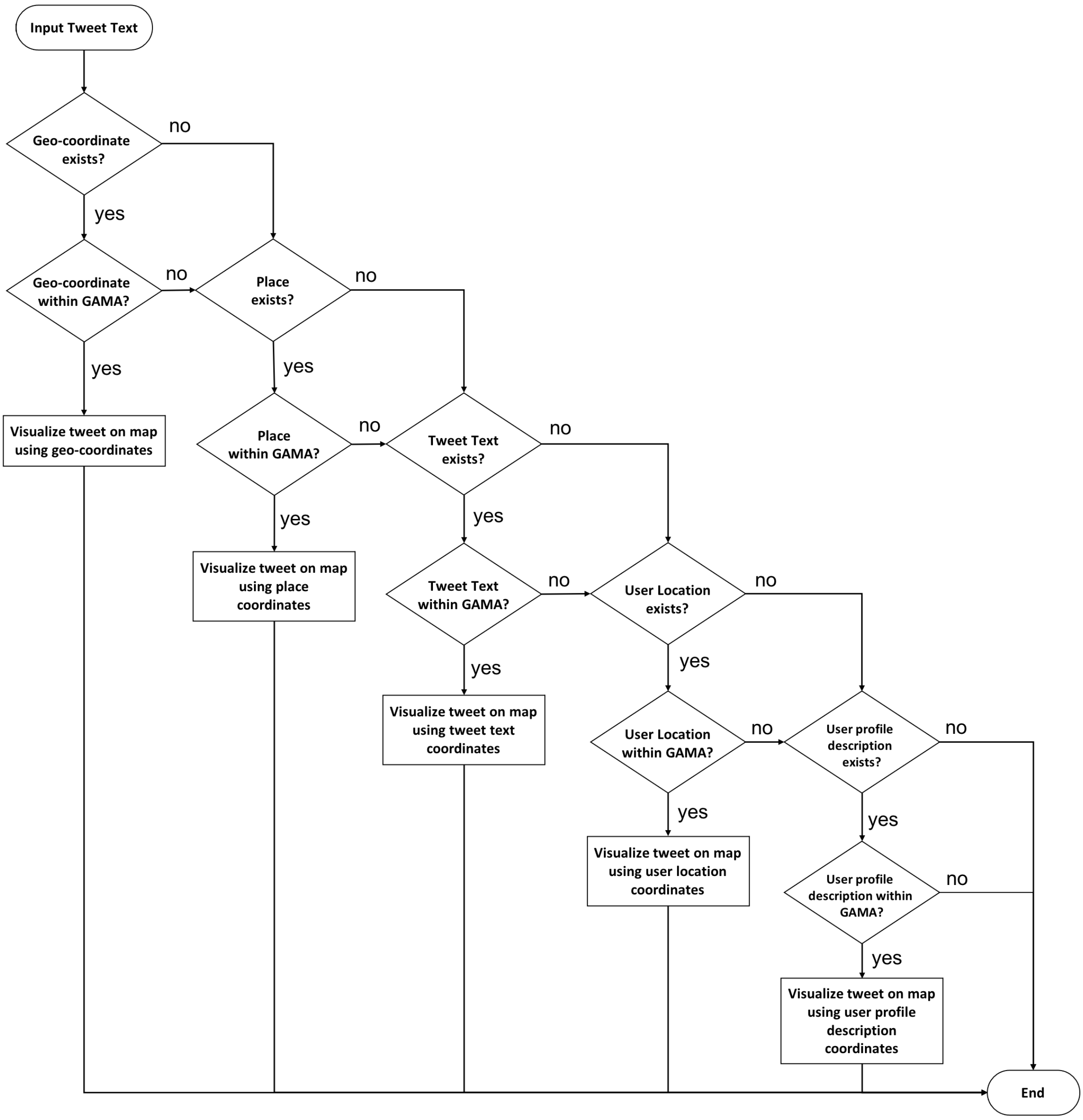

Figure 6. Approach to determine prioritization of the location fields

Tweets image processing

As stated earlier, we downloaded images from the collected tweets (N=3,613). To extract only flood related images, we used a deep learning image classification model from (Weber et al. 2020). This model is trained on one million images related to 43 crisis incidents. We ran the model on our images and selected the ones classified as "heavy rainfall", "storm surge", "mudslide", "flooded", "mudflow", and "landslide". Table 5 shows the total number of images, proportion of flood classified images, and flood-related images geolocated in the GAMA region.

Date	Total Tweets	Geolocated in GAMA	Flood-related Tweets	Informative Tweets
10 Oct 2020	7209	6335	1378	1349
11 Oct 2020	9974	7023	1754	1739
12 Oct 2020	6666	6185	377	352
13 Oct 2020	6417	5817	174	165
Total	**30266**	**25360**	**3683**	**3605**

Table 4. Tweets data processing details: tweets geolocated in GAMA, related to flood disaster, and informative

Date	Total Images	Flood-related Images	Flood-related Images in GAMA
10 Oct 2020	865	135	115
11 Oct 2020	934	132	97
12 Oct 2020	908	94	89
13 Oct 2020	906	63	59
Total	**3613**	**424**	**360**

Table 5. Tweet images statistics: total images, flood-related images, and flood-related images geolocated in GAMA

RESULTS

Remote Sensing

Accuracy Assessment

Due to the lack of flood extent ground truth in data-scarce regions like Ghana, the flood detection algorithm was validated on a test site in another country in Africa characterized with an urban morphology. This is because the GAMA region is densely populated, so it was important to ensure that case study selected is similar in nature. This morphology factor is important because the detection of flooded areas is more complicated in urban areas as compared to rural areas covered with crops and vegetation. The country selected was Nigeria, with the area of interest being Enugu Otu. The Copernicus Emergency Management Service (EMS)[8] was used to download the vector ground truth for free. EMS provides post-event information rapidly for many major disasters through their Rapid Mapping activation service, where EMS activation code EMSR314 was used for the Nigeria ground truth. A potential limitation of testing the algorithm on another country poses a question on whether the flood detection algorithm can generalize well on different land areas classified with an urban morphology. This is because the Enugu, Otu region used as the case study is smaller as compared to the GAMA region. However, since a change detection technique was used for flood detection, it is known that change detection algorithms are better at handling spatial heterogeneity of land surface, compared to flood detection models that use a single-image (Matgen et al. 2019).

Enugu Otu, Uganda Case Study: Torrential rains on September 18, 2018 triggered devastating floods in different parts of Nigeria causing the Niger and Benue rivers to burst their banks. As a result many communities and farms were inundated which affected hundreds of thousands of people.[9] The ground truth is derived from Sentinel-1 images on September 22, 2018 using a semi-automatic extraction method where the exact same image is also used to validate the flood detection algorithm. The validation was done using well-known measures, i.e., precision, recall, F1-score and Goodness of Fit (i.e., Intersection over Union). Sentinel -1 SAR classified images of August 21, 2018 for the Engu, Otu region of Nigeria and the corresponding Sentinel-1 ground truth of August 21, 2018 for Engu were used for the validation task. Looking at Table 6, the VV band was found to perform better than the VH band across all filters in determining flood extent. This is because the VV band produces a bi-modal distribution as seen in Figure 7 which allows the automatically detected threshold to separate the water regions from the non-water regions more accurately. Moreover, the combination of the VV and VH band together resulted in an increased accuracy when compared to the VV band alone across all speckle filters. In particular, the Perona-Malik filter performed the best with an F1-score of 79.2% when both the VV and VH bands are used. Thus, the combination of VV and VH bands on Perona-Malik filtered images for GAMA will be used when delineating the flood extent.

[8] `https://emergency.copernicus.eu/`
[9] `https://emergency.copernicus.eu/mapping/list-of-components/EMSR314`

	Gamma			Perona-Malik			Lee			Lee Sigma		
	VV	VH	Both	VV	VH	Both	VV	VH	Both	VV	VH	Both
Precision	94.3 %	94.6 %	94.4 %	95.6 %	95.2 %	95.7 %	91.3 %	95.4 %	91.5 %	94.2 %	95.4 %	94.7 %
Recall	56.1 %	43.0 %	60.1 %	69.7 %	47.2 %	71.1 %	69.0 %	44.6 %	71.8 %	47.8 %	44.7 %	54.3 %
F1-score	70.3 %	59.1 %	73.5 %	80.6 %	63.1 %	81.6 %	79.2 %	60.8 %	80.5 %	63.4 %	60.9 %	69.0 %
Goodness of Fit	54.2 %	42.0 %	58.0 %	67.6 %	46.1 %	68.9 %	65.5 %	43.7 %	67.3 %	46.5 %	43.8 %	52.7 %

Table 6. Comparison of statistical measures for all speckle filters

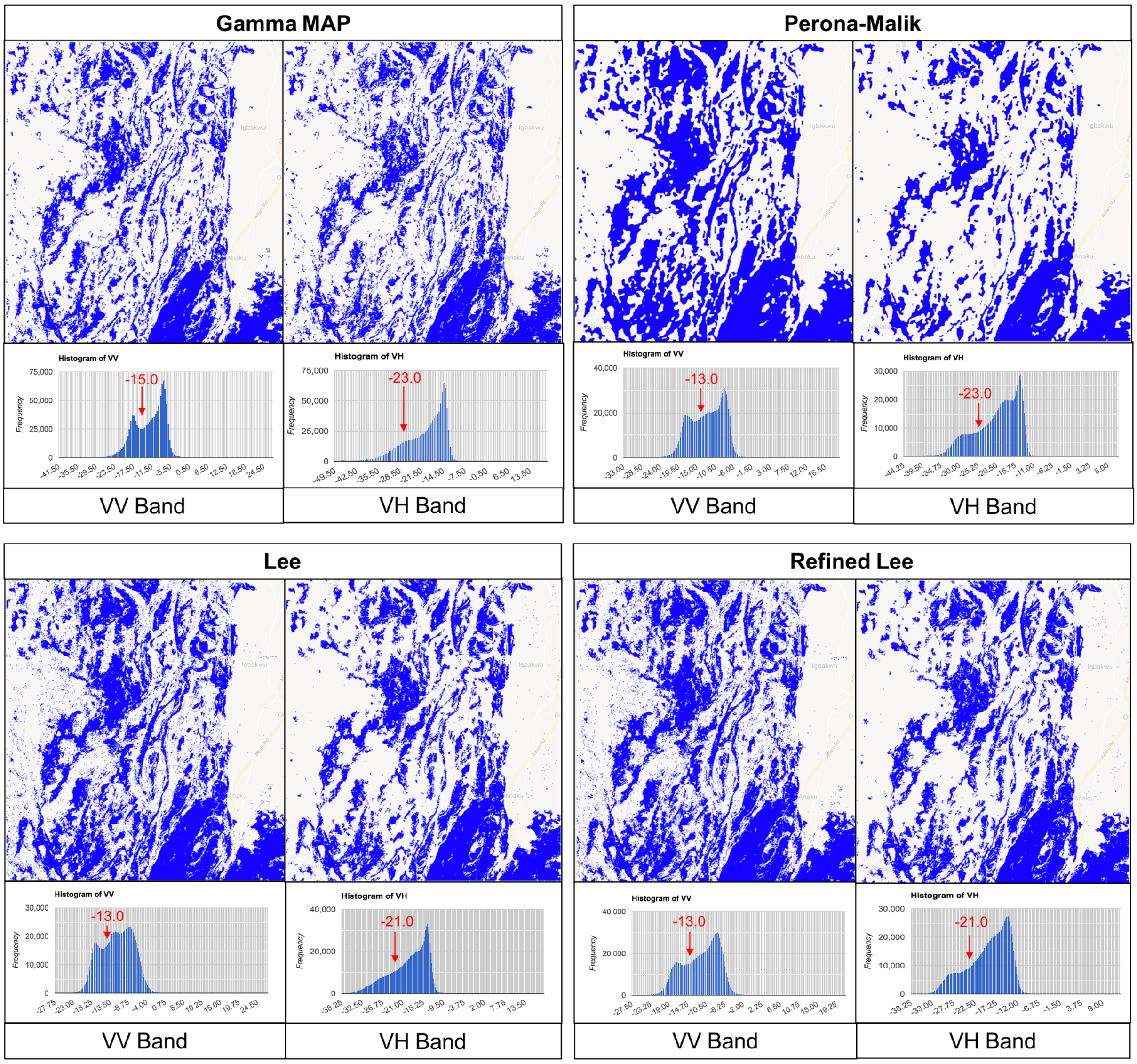

Figure 7. Comparison of speckle filters

Flood Mapping on GAMA region

We apply the flood detection algorithm on Perona-Malik filtered images from the GAMA region using both the VV and VH bands. Figure 8 shows the flood extent. The black regions show the permanent water bodies captured by applying the flood detection algorithm on the before image from a dry day. The flooded regions as a result of the 10 October 2020 flash flood are shown in blue, which shows that the flooding was indeed severe.

Furthermore, we remark that freely available Sentinel-1 SAR images in the GEE platform has a great potential for flood mapping as it avoids applying extra pre-processing steps which require high computing power and large images do not have to be downloaded. Instead, the analysis of the inundation maps can be done directly in GEE

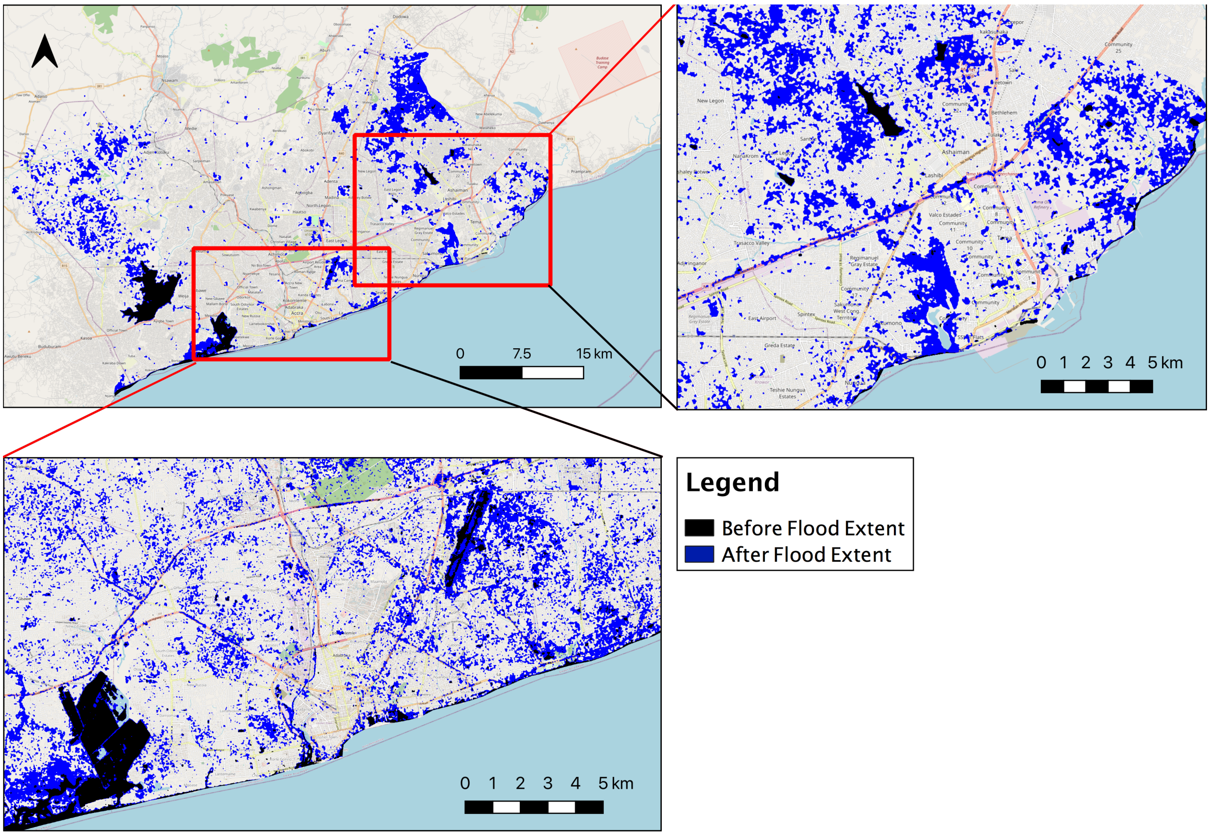

Figure 8. Overview of flood extent in GAMA

with the help of predefined functions. Moreover, the high classification accuracy using both the VV and VH bands on Perona-Malik filtered images demonstrates the potential of the Otsu's algorithm in automatically detecting thresholds for separating water and non-water pixels without requiring user input.

Social Media

All the relevant tweets and images are overlaid onto the flood extent layer to visualize their geographic distribution. The markers in green represent the tweet text and the markers in red are for tweet images. Figure 9 shows that majority of the tweets are located in the south of the GAMA which is situated towards the coastal region. In Figure 10, we show a tweet reporting flooding on the roads. Figure 11 shows the devastating effects as a result of the flash flood, where a truck is submerged in deep waters and locals are forced to use their boats to navigate their way out. This information can be used by crisis responders to understand the severity of the flood situation in different areas, where they will be able to prioritize their actions accordingly.

With these results, we observe that tweet images can provide crisis managers with a quick visual insights on the flood situation (i.e., flood depth, damages) and the severity of damage across areas which can immediately assist relief teams with their response operations. Moreover, textual content of tweets can assist crisis managers in understanding the sentiments of the affected people and the alerts which are communicated to the public by affected locals. This valuable information can be valuable for taking the right response strategy.

Combining Remote Sensing and Social Media

When both remote sensing and social media data sources are combined, three types of signals can potentially be derived to assist crisis managers with their relief efforts. These signals are listed below:

(i) Confirmatory signals: When all three data points (remote sensing, tweet text, tweet image) show flooding in the same area, they confirm the flooding situation. The close proximity of flood relevant tweets and images on or near flooded regions derived from the flood detection algorithm prove that remote sensing, tweets, and images can provide conclusive information to crisis managers about flood extent. Figure 12 shows a closeup of a region in

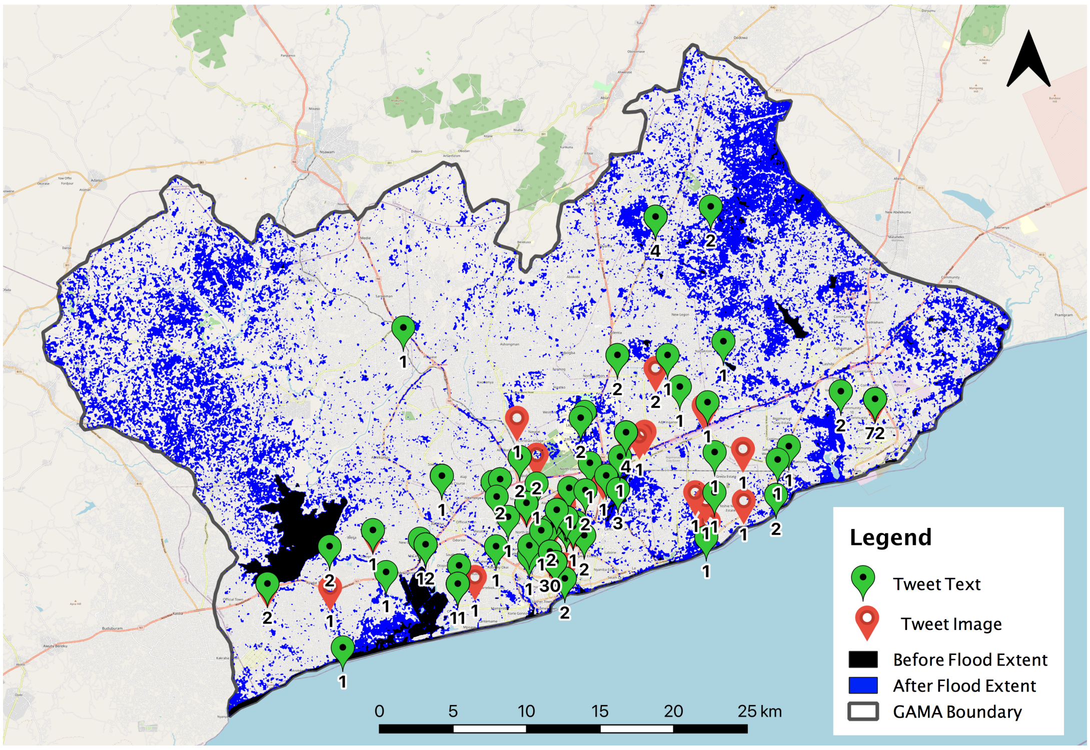

Figure 9. Overview of tweet text and image markers

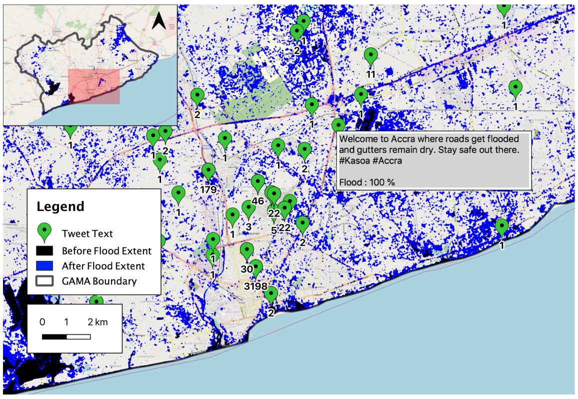

Figure 10. Tweet text

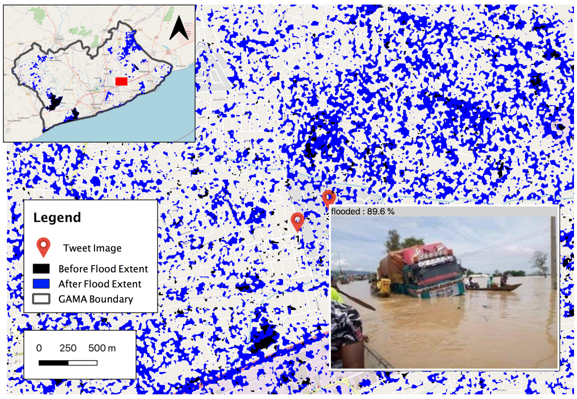

Figure 11. Tweet image

GAMA where the tweet text and images are located on flooded regions identified from the satellite imagery which indicates that the region is definitely vulnerable and requires immediate action from crisis responders.

(ii) Complementary signals: Tweet text and images can potentially provide extra contextual information to crisis managers. Such information can be on the damage incurred to buildings, roads and other infrastructure, as well as, difficulties that the affected people face due to the disaster. As shown in Figure 13, *Kaneshie*, *Dansoman*, *Mallam* areas are reported as flooded (see tweet text) and the image shows a specific community having raging floodwaters.

(iii) Novel signals: When flood extent mapping is not possible using remote sensing, tweets and images can provide novel signals about flooding. Higher number of tweets clustered in an area as seen in Figure 14 provides an indication that people are vulnerable to flooding even though it's not visually seen on the flood extent map.

DISCUSSION

Real-time satellite imagery during flood events is often not available due to poor weather conditions causing difficulties for decision-makers to get a comprehensive picture of the situation. An alternative source of data which can provide insights into the ground-level perspective during a flood, is social media posts (textual messages and images) from locals who are directly affected. However, social media data on it's own raises the limitation of not

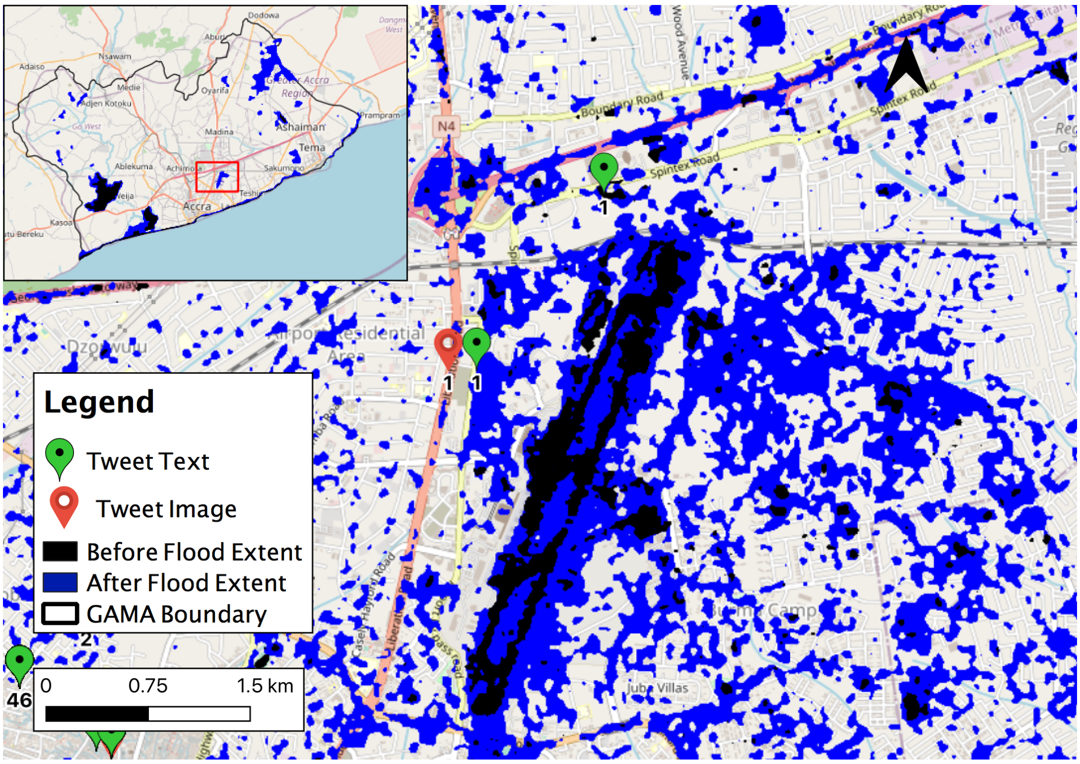

Figure 12. Tweet distribution and Flood Extent

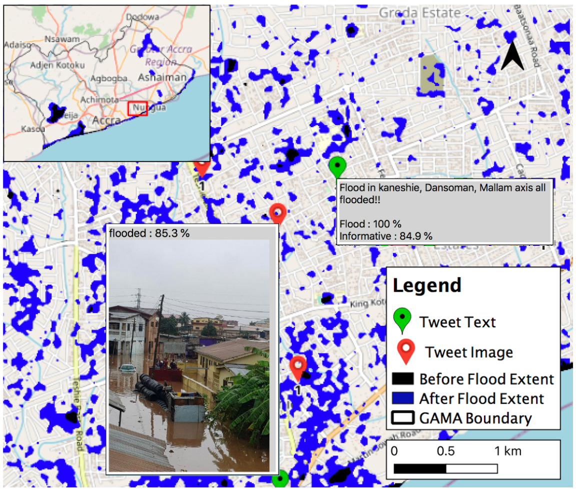

Figure 13. Tweet text and image information

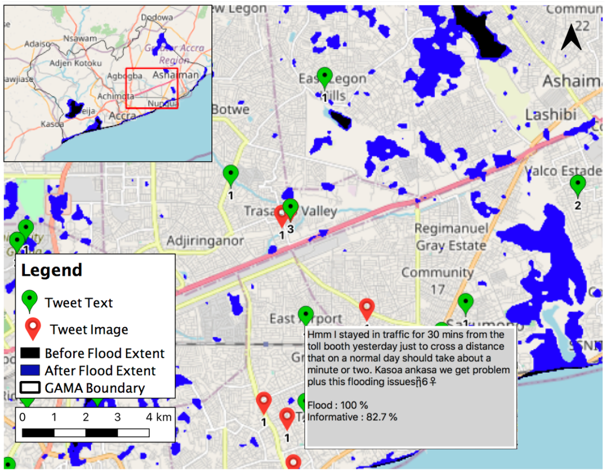

Figure 14. Novel signs of flooding from tweets

being fully represented by the general population and thus identifying floods through social media solely is not an ideal approach. To tackle this problem, social media data can be fused with near real-time flood maps derived from satellite imagery to obtain a comprehensive picture of the flood situation from both the bottom-up view represented by social media and the top-down view captured by satellites. This will in turn provide situational awareness needed for emergency response and the coordination of rescue efforts.

Flood hazard, flood vulnerability and flood risk maps have already been developed at the local level for hotspots in 10 pilot districts in Ghana, which was part of a Community Resilience through Early Warning (CREW) project, implemented by National Disaster Management Organization (NADMO) in collaboration with United Nations Development Programme (UNDP).[10] Whilst these maps are expected to improve risk mitigation where government and non-governmental organizations can identify risk factors and allocate resources, build infrastructure, and ensure

[10] https://www.gh.undp.org/content/ghana/en/home/presscenter/articles/2015/10/20/ghana-launches-disaster-risk-maps.html

early warning systems are put in place for disaster resilience, there are no maps implemented for post flooding response and damage assessment. Therefore, this paper proposes an opportunity for Ghana where near real-time maps fused with social media can provide the following advantages to key stakeholders in Ghana:

(i) The ability to identify regions severely affected by floods with greater confidence through confirmatory signals.

The spatial distribution of tweets can provide an indication that a specific area is severely affected by flooding when there is a cluster of tweets at a close proximity. However, if a flood map based on satellite data shows that the same area is also flooded, then this will turn the indication into a confirmatory signal that the area is indeed severely affected; and that attention should be given to this region to understand what relief operations are required.

(ii) Response teams (e.g., ambulatory services, rescue teams etc.) can be immediately directed to the affected areas in a timely manner based on the additional complementary signals from tweet images and/or text.

The contextual information in tweets can enhance humanitarian organizations' understanding of the problems specific to a local region—which is useful for them to launch right relief operations in a timely manner e.g., emergency vehicles directed to flooded areas for rescuing trapped people. Moreover, images posted from an area can further enhance situational awareness e.g., estimating flood water-level from submerged buildings and cars, whilst tweet text reporting specific roads that are impassable.

(iii) A comprehensive picture of the flood situation means that flooded areas that are hard hit and not easily noticeable from the flood map, are now visualized onto the flood map with the help of novel signals from tweet text and/or tweet images.

Tweets and images which are spatially distributed away from severely flooded areas detected via remote sensing can potentially provide novel signals. Moreover, flash flood events, which lasts a few hours with devastating impact, may not be visible and detectable from remote sensing techniques. During such situations, social media data can provide pertinent information in near real-time.

Whilst the proposed flood mapping approach requires no additional costs as the satellite imagery used is free and open-source, potential technical challenges may arise if there is a lack of computing infrastructure capacity needed to retrieve satellite imagery in real-time. Moreover, a lack of personnel to maintain and support the system and lack of access to uninterrupted internet service may hinder the ongoing generation of flood maps in near real-time.

Large-scale socio-natural disasters like flooding, poses a big challenge for crisis responders as they need fast access to comprehensive and accurate post-damage assessments to best allocate their limited resources. In order to conduct a quick assessment of flood damage and provide support in areas with immediate attention, one potential approach we plan to explore in the future is to intersect a land use/land cover map with the flood inundated layer extracted from the satellite imagery to understand what area of land (i.e crop, urban etc.) is affected in comparison to the land use/land cover map before the flood event. Similarly, we plan to intersect various geospatial data like population density, building density and road network with the flood map to identify vulnerable populations, affected buildings and roads which will assist crisis responders with the prioritization of their efforts.

CONCLUSIONS

This paper explored the fusion of two heterogeneous data sources, namely, satellite imagery and social media for building more informed flood extent maps. As for the remote sensing data, we applied a change detection technique to identify flooded regions. Specifically, we assessed the Otsu's thresholding approach for flood mapping on four different speckle filtered images using Gamma MAP, Perona-Malik, Lee and Refined Lee. Moreover, the different polarizations were also investigated to determine if one specific band or the combination of both bands, accurately estimates water inundated areas. In addition to the remote sensing layer, we added two more layers from social media data representing tweet text reports, which provide more contextual information about the crisis, and tweet images that show flooded roads, streets, and damaged buildings. We used deep learning models to filter out irrelevant tweets and images as well as those outside the region of interest. We remark that the fusion of these two data sources produce informative signals for crisis managers to not only understand flood extent, but also the consequences of flooding, type of incurred damages, and the severity of damages.

ACKNOWLEDGMENTS

This publication was made possible by NPRP grant #BFC03-0630-190011 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

The authors would like to gratefully acknowledge the financial support of the project Re-Energize Governance of Disaster Risk Reduction and Resilience for Sustainable Development (Re-Energize DR3) provided by the Belmont

Forum's first disaster-focused funding call DR3 CRA Joint Research, which was supported by the Ministry of Science and Technology (MOST) of Chinese Taipei in partnership with funders from Brazil (FAPESP), Japan (JST), Qatar (QNRF), the UK (UKRI), and the US (NSF).

REFERENCES

Addae, B. and Oppelt, N. (2019). "Land-use/land-cover change analysis and urban growth modelling in the Greater Accra Metropolitan Area (GAMA), Ghana". In: *Urban Science* 3.1, p. 26.

Alam, F., Ofli, F., and Imran, M. (2019). "CrisisDPS: Crisis Data Processing Services." In: *ISCRAM*.

Amoah Addae, R. (2018). "Satellite-based flood mapping for hydronamic flood moddel assessment: Accra, Ghana". MA thesis. University of Twente.

Appeaning Addo, K., Larbi, L., Amisigo, B., and Ofori-Danson, P. K. (2011). "Impacts of coastal inundation due to climate change in a cluster of urban coastal communities in Ghana, West Africa". In: *Remote Sensing* 3.9, pp. 2029–2050.

Carreño Conde, F. and De Mata Muñoz, M. (2019). "Flood Monitoring Based on the Study of Sentinel-1 SAR Images: The Ebro River Case Study". In: *Water* 11.12.

Cian, F., Marconcini, M., and Ceccato, P. (2018). "Normalized Difference Flood Index for rapid flood mapping: Taking advantage of EO big data". In: *Remote Sensing of Environment* 209, pp. 712–730.

De Albuquerque, J., Herfort, B., Brenning, A., and Zipf, A. (Feb. 2015). "A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management". In: *International Journal of Geographical Information Science* 29.

Eilander, D., Trambauer, P., Wagemaker, J., and Van Loenen, A. (2016). "Harvesting social media for generation of near real-time flood maps". In: *Procedia Engineering* 154, pp. 176–183.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., et al. (2007). "The shuttle radar topography mission". In: *Reviews of geophysics* 45.2.

Filipponi, F. (2019). "Sentinel-1 GRD preprocessing workflow". In: *Multidisciplinary Digital Publishing Institute Proceedings*. Vol. 18. 1, p. 11.

Fohringer, J., Dransch, D., Kreibich, H., and Schröter, K. (2015). "Social media as an information source for rapid flood inundation mapping". In: *Natural Hazards and Earth System Sciences* 15.12, pp. 2725–2738.

Giustarini, L., Hostache, R., Matgen, P., Schumann, G. J.-P., Bates, P. D., and Mason, D. C. (2012). "A change detection approach to flood mapping in urban areas using TerraSAR-X". In: *IEEE transactions on Geoscience and Remote Sensing* 51.4, pp. 2417–2430.

Goodchild, M. F. (2007). "Citizens as sensors: the world of volunteered geography". In: *GeoJournal* 69.4, pp. 211–221.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone". In: *Remote Sensing of Environment*.

Herfort, B., De Albuquerque, J., Schelhorn, S.-J., and Zipf, A. (Jan. 2014). "Exploring the Geographical Relations Between Social Media and Flood Phenomena to Improve Situational Awareness". In: pp. 55–71.

Horita, F., Degrossi, L., Assis, L. F., Zipf, A., and De Albuquerque, J. (June 2013). "The use of Volunteered Geographic Information and Crowdsourcing in Disaster Management: a Systematic Literature Review". In: vol. 5.

Huang, X., Wang, C., and Li, Z. (2018). "A near real-time flood-mapping approach by integrating social media and post-event satellite imagery". In: *Annals of GIS* 24.2, pp. 113–123.

Imran, M., Castillo, C., Lucas, J., Meier, P., and Vieweg, S. (2014). "AIDR: Artificial intelligence for disaster response". In: *Proceedings of the 23rd international conference on world wide web*, pp. 159–162.

Jongman, B., Wagemaker, J., Romero, B. R., and De Perez, E. C. (2015). "Early flood detection for rapid humanitarian response: harnessing near real-time satellite and Twitter signals". In: *ISPRS International Journal of Geo-Information* 4.4, pp. 2246–2266.

Joyce, K. E., Belliss, S. E., Samsonov, S. V., McNeill, S. J., and Glassey, P. J. (2009). "A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters". In: *Progress in physical geography* 33.2, pp. 183–207.

Karley, N. K. (2009). "Flooding and physical planning in urban areas in West Africa: situational analysis of Accra, Ghana". In: *Theoretical and Empirical Researches in Urban Management* 4.4 (13, pp. 25–41.

Lee, J.-S. (1980). "Digital image enhancement and noise filtering by use of local statistics". In: *IEEE transactions on pattern analysis and machine intelligence* 2, pp. 165–168.

Lee, J.-S. (1981a). "Refined filtering of image noise using local statistics". In: *Computer graphics and image processing* 15.4, pp. 380–389.

Lee, J.-S. (1981b). "Speckle analysis and smoothing of synthetic aperture radar images". In: *Computer graphics and image processing* 17.1, pp. 24–32.

Li, Z., Wang, C., Emrich, C. T., and Guo, D. (2018). "A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods". In: *Cartography and Geographic Information Science* 45.2, pp. 97–110.

Liang, J. and Liu, D. (2020). "A local thresholding approach to flood water delineation using Sentinel-1 SAR imagery". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 159, pp. 53–62.

Lopes, A., Nezry, E., Touzi, R., and Laur, H. (1990). "Maximum A Posteriori Speckle Filtering And First Order Texture Models In Sar Images". In: *10th Annual International Symposium on Geoscience and Remote Sensing*, pp. 2409–2412.

Mansourpour, M., Rajabi, M., and Blais, J. (2006). "Effects and performance of speckle noise reduction filters on active radar and SAR images". In: *Proc. ISPRS*. Vol. 36. 1, W41.

Martinis, S., Twele, A., and Voigt, S. (2009). "Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data". In: *Natural Hazards and Earth System Sciences* 9.2, pp. 303–314.

Matgen, P., Martinis, S., Wagner, W., Freeman, V., Zeil, P., and McCormick, N. (2019). "Feasibility assessment of an automated, global, satellite-based flood-monitoring product for the Copernicus Emergency Management Service". In:

Moothedan, A. J., Dhote, P. R., Thakur, P. K., Garg, V., Aggarwal, S., and Mohapatra, M. (2020). "AUTOMATIC FLOOD MAPPING USING SENTINEL-1 GRD SAR IMAGES AND GOOGLE EARTH ENGINE: A CASE STUDY OF DARBHANGAH, BIHAR". In: *National Seminar on Recent Advances in Geospatial Technology & Applications*.

Otsu, N. (1979). "A Threshold Selection Method from Gray-Level Histograms". In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1, pp. 62–66.

Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S. (2016). "High-resolution mapping of global surface water and its long-term changes". In: *Nature* 540.7633, pp. 418–422.

Perona, P. and Malik, J. (1990). "Scale-space and edge detection using anisotropic diffusion". In: *IEEE Transactions on pattern analysis and machine intelligence* 12.7, pp. 629–639.

Peters, R. and De Albuquerque, J. (Jan. 2015). "Investigating images as indicators for relevant social media messages in disaster management". In:

Qazi, U., Imran, M., and Ofli, F. (2020). "Geocov19: a dataset of hundreds of millions of multilingual covid-19 tweets with location information". In: *SIGSPATIAL Special* 12.1, pp. 6–15.

Rain, D., Engstrom, R., Ludlow, C., and Antos, S. (2011). "Accra Ghana: A city vulnerable to flooding and drought-induced migration". In: *Case study prepared for cities and climate Change: Global Report on Human Settlements* 2011, pp. 1–21.

Ruzza, G., Guerriero, L., Grelle, G., Guadagno, F. M., and Revellino, P. (2019). "Multi-method tracking of monsoon floods using Sentinel-1 imagery". In: *Water* 11.11, p. 2289.

Tiwari, V., Kumar, V., Matin, M. A., Thapa, A., Ellenburg, W. L., Gupta, N., and Thapa, S. (2020). "Flood inundation mapping-Kerala 2018; Harnessing the power of SAR, automatic threshold detection method and Google Earth Engine". In: *Plos one* 15.8, e0237324.

Velev, D. and Zlateva, P. (June 2012). "Use of Social Media in Natural Disaster Management". In:

Weber, E., Marzo, N., Papadopoulos, D. P., Biswas, A., Lapedriza, A., Ofli, F., Imran, M., and Torralba, A. (2020). "Detecting natural disasters, damage, and incidents in the wild". In: *European Conference on Computer Vision*. Springer, pp. 331–350.