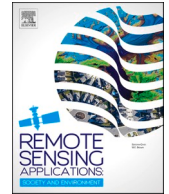




Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

Integrating remote sensing and social sensing for flood mapping

Rizwan Sadiq^{*}, Zainab Akhtar, Muhammad Imran, Ferda Ofli

Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

ARTICLE INFO

Keywords:

Flood mapping
Deep learning
Social media
Satellite imagery
Remote sensing

ABSTRACT

Flood events cause substantial damage to infrastructure and disrupt livelihoods. Timely monitoring of flood extent helps authorities identify severe impacts and plan relief operations. Remote sensing through satellite imagery is an effective method to identify flooded areas. However, critical contextual information about the severity of structural damage or urgent needs of affected population cannot be obtained from remote sensing alone. On the other hand, social sensing through microblogging sites can potentially provide useful information directly from eyewitnesses and affected people. Therefore, this paper explores the integration of remote sensing and social sensing data to derive informed flood extent maps. For this purpose, we employ state-of-the-art deep learning methods to process heterogeneous data obtained from four case-study areas, including two urban regions from Somalia and India and two coastal regions from Italy and The Bahamas. On the remote sensing side, we observe that deep learning models perform generally better than Otsu in flood water prediction. For example, for highly urban areas from Somalia and India, U-Net achieves better F1-scores (0.471 and 0.310, respectively) than Otsu (0.297 and 0.251, respectively). Similarly, for coastal areas, FCN yields a better F1-score for Italy (0.128) than Otsu (0.083) while FCN and Otsu perform on par for The Bahamas (0.102 and 0.105, respectively). Then, on the social sensing side, we add two data layers representing relevant tweet text and images posted from the case-study regions to highlight different ways these heterogeneous data sources complement each other. Our extensive analyses reveal several valuable insights. In particular, we identify three types of signals: (i) confirmatory signals from both sources, which puts greater confidence that a specific region is flooded, (ii) complementary signals that provide different contextual information including needs and requests, disaster impact or damage reports and situational information, and (iii) novel signals when both data sources do not overlap and provide unique information.

1. Introduction

Climate change is increasing the intensity and frequency of floods around the globe and crisis responders are turning towards satellite-based flood inundation mapping to get a comprehensive overview of the flooded regions (Shen et al., 2019; Mateo-Garcia et al., 2021; Liu et al., 2019). While these maps provide situational awareness about the affected areas, they fail to provide important contextual information such as affected population and their urgent needs in near real-time to assist crisis managers with flood response. This is where Online Social Networking (OSN) platforms like Twitter and Facebook can provide insights in real-time. The collection of large amounts of data and information from these online human interactions has led to the emergence of a new field called Participatory Sensing or Social Sensing where the information is used for sensing a physical phenomenon like natural disasters (Aulov

^{*} Corresponding author.

E-mail address: rsadiq@hbku.edu.qa (R. Sadiq).

and Halem, 2012). With the advantage of acquiring information in real-time, there comes the drawback of requiring verification and this weakness is actually the strength of remote sensing, where satellite imagery is usually reliable and accurate (Marty Graham, 2019; Li et al., 2017). Since remote sensing data and social sensing 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.

This paper aims to leverage satellite imagery for flood detection and social sensing data from Twitter 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 for delineating flood extent in near real-time as the signal can penetrate through clouds, unlike optical sensors (Joyce et al., 2009). While there are many common techniques relying on thresholding methods to differentiate between water and surface pixels in satellite imagery, such as Otsu (Otsu (1979) and NDWI (Gao, 1996), deep learning algorithms have emerged to be reliable approaches in detecting floods due to their generalization capability over different geographic locations and geologies (Mosavi et al., 2018). Social sensing data, on the other hand, provides important contextual information such as flood severity, affected population and their urgent needs, and other impacts.

The motivation behind this study stems from a gap in the research of flood delineation via fusing remote sensing and social sensing data. This is because majority of the researchers use thresholding techniques for flood delineation. Only a few studies have used supervised deep learning models but they perform tests only on a single flood event. Therefore, there is a need to further investigate how these models perform across different flood events with varying land characteristics. Moreover, existing studies utilize only geolocated tweets which come in very small numbers. To alleviate this limitation, we incorporate a geolocation inference pipeline to increase the number of tweets utilized in our analysis.

Fig. 1 shows an overview of our approach where Sentinel-1 imagery was used to detect flood water. We experimented with two deep learning methods, which were trained and tested on an open source, labeled satellite imagery dataset called Sen1Floods11 (Bonafilia et al., 2020). We employed Fully Convolutional Network (FCN) and U-Net due to their better performance in segmentation, as reported in recent literature (Bonafilia et al., 2020; Mateo-Garcia et al., 2021). We then compared the performance of these two deep learning models with a conventional thresholding-based flood segmentation model, Otsu. The flood extent from the model which performs the best was then displayed on a map. On top of this remote sensing layer, two social media layers representing relevant tweets and images classified using deep learning models were added to create an informative flood map with social media overlay as markers. With the amalgamation of information coming from different sources (i.e., satellite imagery, tweet text and images), we are able to highlight three different types of signals useful for crisis managers: (i) *confirmatory signals*: when remote sensing and social sensing signals provide matching flood predictions, (ii) *complementary signals*: when social sensing data provides extra context (e.g., incurred damages) to a flood prediction obtained from remote sensing, and (iii) *novel signals*: when social sensing report flooding and/or provide extra contextual information, but do not overlap with remote sensing predictions.

2. Related work

2.1. Remote sensing

Satellite-based remote sensing is an effective way to delineate floods due to the availability of multi-temporal images. Synthetic-aperture radar (SAR) is a form of satellite imagery which is unaffected by weather conditions such as cloud coverage during flooded

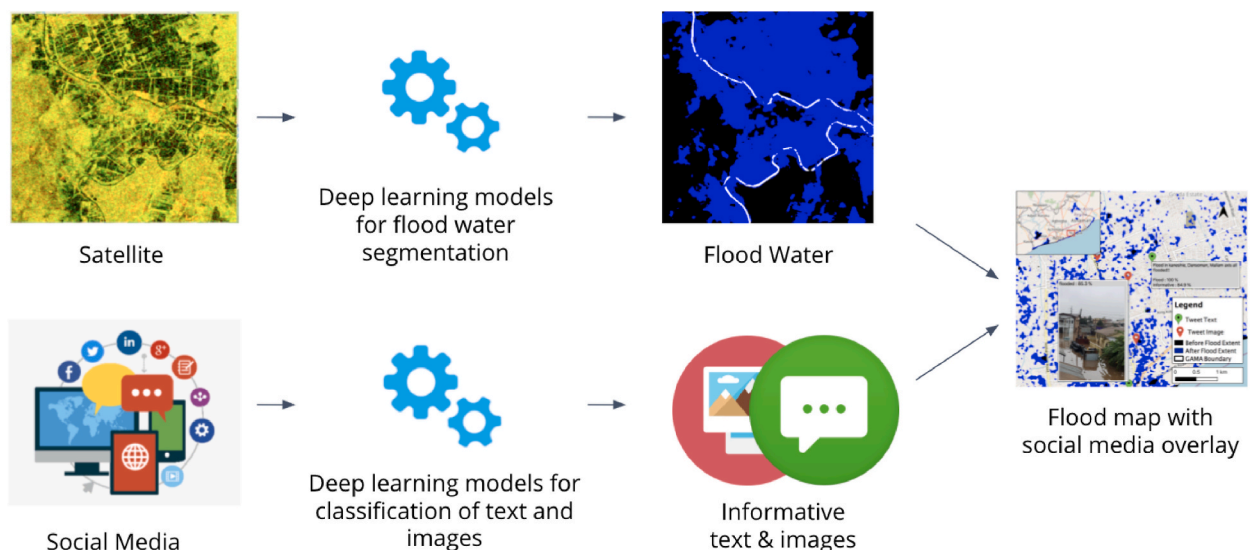


Fig. 1. Overview of the approach fusing remote sensing and social sensing to create more informative flood maps.

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 back scattering from buildings (United Nations, 2021).

The most commonly adopted method to detect floods from both satellite imagery is the thresholding technique which can rapidly create a binary classification of water and non-water regions. However, with it comes a major limitation of generalizability. For example, the detection of floods in urban areas is a challenge as images can be unimodal or can have an inconspicuous bimodal histogram distribution which makes it difficult in determining an optimal threshold (Zhan and Zhang, 2019). Moreover, Lee et al. (1990) and Matgen et al. (2011) found that the performance of the thresholding technique rapidly declined when the ratio of water is very small for images in urban areas as well as regions with varying terrain. To address these limitations, researchers have shifted their focus to deep learning methods which learn from a varied set of images, thus allowing it to generalize across different regions and ratios of water.

One of the recent studies on SAR imagery, Bonafilia et al. (2020) showed that FCN can perform better for segmentation of temporary water when compared to the Otsu thresholding method. Katiyar et al. (2021) reported their findings on flood segmentation using different combination of bands when using Sentinel-1 (SAR). Their results also supported that deep learning models (U-Net and SegNet) performed better than thresholding methods. While there are many studies which use either optical or SAR imagery, more recent research has focused on the fusion of multiple images coming from different sources. For example, Muñoz et al. (2021) analyzed multispectral Landsat imagery (optical) and dual-polarized SAR imagery to assess the performance of integrating convolutional neural networks (CNN) with data fusion framework for generating flood maps. Bai et al. (2021) further explored the capabilities of fusing Sentinel-1 (SAR) and Sentinel-2 (Optical) imagery to train a water segmentation model via a boundary aware deep network. The results showed that deep learning models were superior than thresholding models for flood water segmentation.

2.2. Fusion of remote sensing and social sensing data

While social media sources alone can provide useful insights, the fusion of social media with remote sensing in the field of disaster management is much more compelling and have been utilized in studies targeting mainly two phases of a disaster: during (monitoring) and after (damage assessment).

On the monitoring side, timely disaster detection is essential for reducing the severity of damage caused by the disaster. Using Gaussian kernel functions, a study was conducted by Huang et al. (2018), integrating post-event remote sensing optical data with real-time data from Twitter to produce near real-time flood likelihood maps. Rosser et al. (2017) also developed flood probability maps by integrating terrain data alongside socially sensed images and satellite imagery into a Bayesian statistical model. Schnebele et al. (2014) generated flood estimation time series maps over a six day period for long flood events, by fusing traditional SAR imagery, Digital Elevation Model, Twitter messages, geolocated photographs and online news with a geostatistical interpolation technique called kernel density estimation. Schnebele and Cervone (2013) also used the same technique to create flood hazard maps by fusing social sensed data (photos, videos and news), remote sensing data, DEM, meteorological data and river gauge data. Another study by Sun et al. (2016) found that the fusion of social sensing and remote sensing were in line with each other as 95% of Flickr images were relative to the spatial distribution of the satellite-derived flood extent. Interestingly, Jongman et al. (2015) conducted three different analyses where one of them focused on event understanding to enhance response efforts. In particular, a qualitative situational analysis is conducted on the social sensing data where a few insights are highlighted such as public discussions about evacuation, requests made by the local for emergency aid and evidence of flooding from images. While these insights are very limited, there is potential for an extensive analysis of all signals that come from fusing these two heterogeneous data sources which could help humanitarian organizations with their response efforts, thus serving as a motivation for our study.

On the damage assessment side, humanitarian organizations need to prioritize the immediate needs for emergency which can be related to humans, infrastructure or the environment. Majority of the research in the domain of flood damage assessment (Pregolato et al., 2017) focuses on transportation or road impacts. Cervone et al. (2017) utilized a supervised machine learning model to identify water in Optical images. Geolocated tweets ($n = 130$) and photographs ($n = 80$) were then interpolated using a kernel smoothing application to create a damage assessment surface. An additional road network was overlaid to identify impassable roads. Results showed that, when remote sensing data is limited, social media can provide useful information about damage at the street or block level. Similarly, Oxendine et al. (2014) also identified road damages where the findings were also consistent. Instead, a different interpolation method (Kriging) was used to fuse crowdsourced photos, social media videos and remote sensing data. To further enhance this research, Schnebele et al. (2015) utilized a similar approach for road damage assessment but instead used a neural network classifier to segment floods from satellite imagery. One challenge identified in this study was the limited number of photos ($n = 25$) and videos ($n = 15$) used in the analysis due to lack of geolocated data. Therefore, to remedy the data scarcity issue, there is a dire need for a geolocation inference algorithm that can geotag tweets using various metadata fields.

Our contributions. To alleviate the limitations of the aforementioned studies, this paper extends our previous work (Akhtar et al., 2021) with the following contributions in the field of remote sensing, social sensing, and the fusion of both data modalities:

- Two state-of-the-art deep learning models (FCN and U-Net) have been used to segment floods in SAR imagery and shown to outperform the Otsu thresholding baseline for flood detection purposes.
- The generalization ability of the deep learning models was tested on four carefully selected case study events, two from urbanized areas and two from coastal regions.
- An improved geolocation inference algorithm presented in (Akhtar et al., 2021) was employed to increase the number of tweets across the four events as the number of geotagged tweets is often almost negligible.

- All flood relevant tweets that belonged to an humanitarian category were included in our analysis to allow for an extensive qualitative analysis of tweet information.
- We derived flood extent maps and social media data separately and then overlaid them to perform an extensive qualitative analysis, where three different types of information signals (confirmatory, complementary and novel) have been identified useful for crisis responders.
- We propose further categorization of complementary signals into three groups, i.e., urgent needs or offers, impact assessment and situational awareness, which are in line with the information needs of humanitarian organizations.

3. Study areas and data

Machine learning models, especially deep neural networks, can perform well on the dataset they are trained with. However, they do not always generalize well to unseen data or other contexts when applied *in the wild*. For this purpose, in addition to performing standard training and evaluation of our models on a public satellite imagery dataset (i.e., Sen1Floods11), four flood events across the world were selected to test the models' generalization strength.

3.1. Selected flood events

To achieve geographical diversity in our study areas, severe flooding events from four continents (i.e., Africa, Asia, North America, and Europe) with different topographic characteristics were selected. Specifically, two of the events were from regions characterized with a high-density urban morphology whereas the other two events were from regions with a coastal morphology. Fig. 2 shows the areas of the selected flood events. Below we provide details of the events and their corresponding remote sensing and social sensing data.

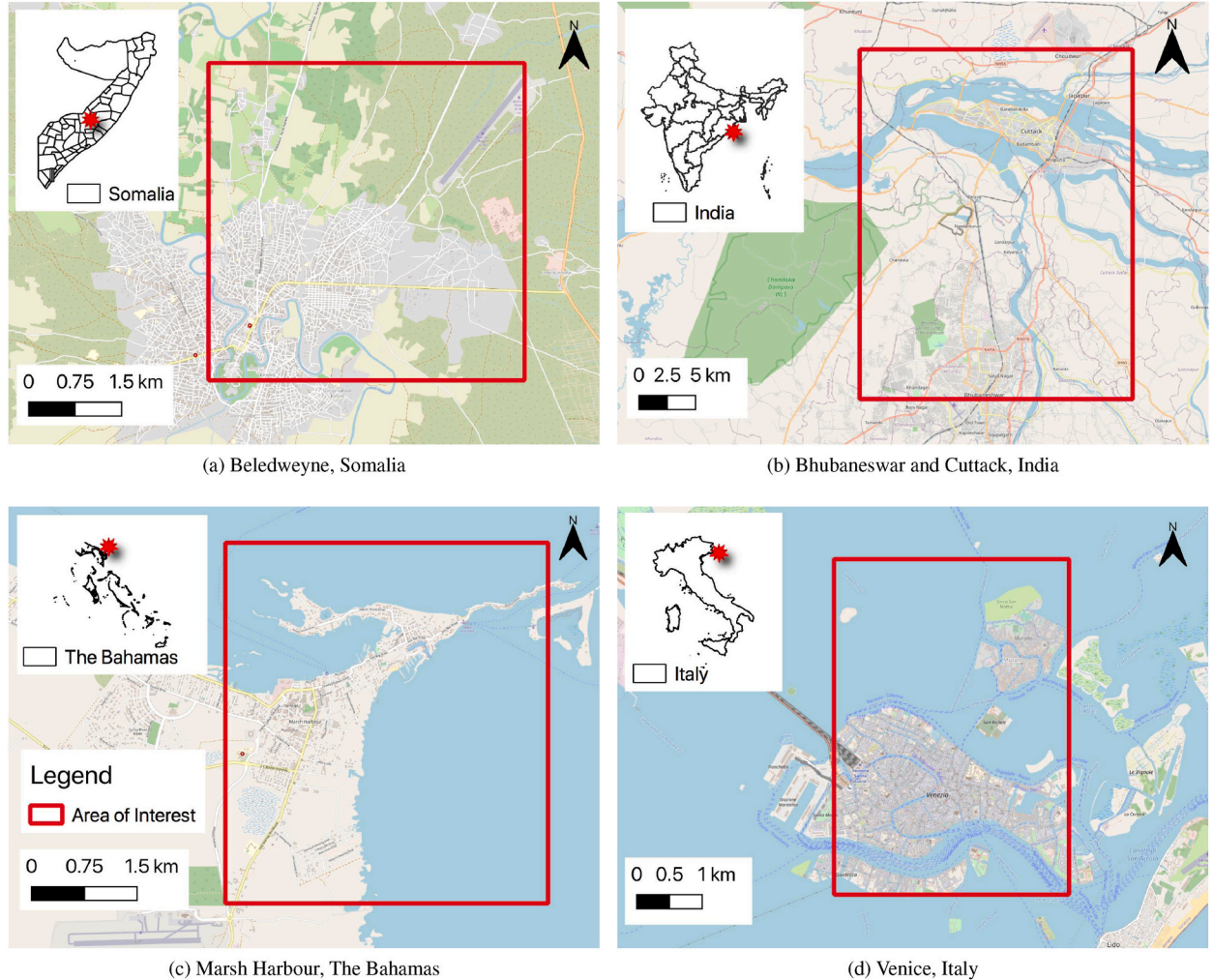


Fig. 2. Area of interest for all four events.

- (1) *Beledweyne, Somalia*: On 24 October 2019, heavy rains along Somalia's two major rivers, Shabelle and Juba, affected thousands of people in Beledweyne town (ReliefWeb, 2019). According to the Somalian government, ten people went missing or drowned and more than 2,000 ha of agriculture was destroyed. Fig. 2(a) shows the area of interest for Beledweyne, Somalia.
- (2) *Bhubaneswar and Cuttack, India*: Tropical Cyclone Fani, which formed on 26 April 2019, is one of the strongest cyclones (Category 4) to hit the poorest coastal states of India, Odisha (Wikipedia, 2019a). The cyclone devastated many coastal districts including the capital city of Bhubaneswar and its neighbouring city Cuttack (Fig. 2(b), which together are often referred to as the Twin-Cities of Odisha.
- (3) *Marsh Harbour, The Bahamas*: Tropical Cyclone Dorian, a Category 5 hurricane caused flooding and mass destruction on the northwest islands of The Bahamas from 1 September 2019 (Wikipedia, 2019b). High winds of up to 240 km/h in magnitude and a storm surge of over 5 m resulted in extensive damage of around \$3.4 billion, which is equal to one-quarter of The Bahamas' GDP. Fig. 2(c) shows the area of interest for The Bahamas.
- (4) *Venice, Italy*: On 12 November 2019, severe weather conditions caused a combination of high spring tides and a storm surge in Italy, where Venice in particular was declared a state of emergency (FloodList, 2021). The Venice city was engulfed by 1.87m (6 ft) high water levels, which cut power to homes and caused widespread damage to boats and buildings. Fig. 2(d) shows the area of interest for Venice, Italy.

For all four events, we used level-1 Ground Range Detected (GRD) Sentinel-1 SAR images from the Google Earth Engine (GEE) (Gorelick et al., 2017). These images were acquired by the Interferometric Wide (IW) mode which use a dual polarization (VV and VH) to transmit and receive electromagnetic waves in a region of interest. VV represents vertical transmit and vertical receive, whereas VH represents vertical transmit and horizontal receive. Table 1 provides data specifications in terms of acquisition dates, polarization modes, spatial resolution, etc. The corresponding ground truth flood extent maps for the events were obtained from Copernicus' Rapid Mapping Activations (Copernicus, 2021) and UNOSAT's Rapid Mapping Service (UNOSAT, 2021). For Somalia, the ground truth was created and sourced by UNOSAT on 1 September 2019. For India, The Bahamas, and Italy, the ground truth was conducted by Copernicus on 2 May 2019, 30 October 2019, and 14 November 2019 for the three events, respectively.

Social sensing data for all four events were collected from Twitter in the form of tweet messages and images. Several event-specific hashtags and/or keywords were used to collect data using the publicly available AIDR system (Imran et al., 2014). The AIDR system also allows one to specify what language of tweets should be collected, where only English tweets were specified for Somalia, India, and Bahamas. Whereas for Italy, both English and Italian languages were selected. Specifically for the Somalia event four hashtags and two keywords were used: #floods, #Somalia, #Beledweyne, #somaliafloods, river, inundated. For the India event, only three hashtags were used: #Fani, #FaniCyclone, #CycloneFani. The Bahamas event consisted of 13 keywords: HurricaneDorian, Dorian, alerts_dorian, dorianalert, puertorico, DorianMissing, DorianDeaths, HurricaneDorianMissing, HurricaneDorianDeaths, Dorian Missing, Hurricane Dorian Missing, Dorian Found, DorianFound. Lastly, for the Italy event, two hashtags and four keywords were used: #Venice, Venice, St Mark's Square, #Venezia, Piazza San Marco, St Mark's Basilica. Table 2 presents social media data details in terms of the total number of tweets and images collected across all events.

3.2. Remote sensing data for model training

For training and evaluation of remote sensing models, we used a publicly available dataset called Sen1Floods11 (Bonafilia et al., 2020). The dataset provides raw Sentinel-1 (IW-mode and GRD product) and Sentinel-2 MSI-level imagery for 11 flood events across the globe. The dataset comprises a total of 4,831 images of dimension 512x512 pixels, spanning over 120,046 km² area which were sampled at 10m spatial resolution with 2 bands (VV and VH) for Sentinel-1 images and 13 bands for Sentinel-2.

The images in the dataset are labeled using three approaches. First set of annotations was obtained for 4,370 areas of interest (AoIs) by using the histogram thresholding method (Otsu) on Sentinel-1 images, hereinafter called "Weak S1". Second set of annotations was obtained for the same AoIs by labeling the Sentinel-2 images using the Modified Normalized Difference Water Index (MNDWI), hereinafter called "Weak S2". The final set of annotations contains expert-labeled ground truth for 446 AoIs, which are different than the previous set of AoIs, hereinafter called "hand-labeled". The hand-labeled set contains segmentation masks of "all water", "land" and "no data", where "no data" represents clouds or areas where the analysts could not confidently identify water or land. Each image on the hand-labeled set is also accompanied by a binary permanent water mask obtained from the European Commission's Joint Research Centre (JRC) (Pekel et al., 2016).

Like many other remote sensing datasets, Sen1Floods11 also suffers from the class imbalance issue between flooded and non-flooded regions. The dataset provides train, validation, and test splits (60:20:20), which we use for model training. In addition to the defined test split, the dataset contains a separate test event, called Bolivia flood event, which is only used to evaluate the generalization of the models.

Table 1
Sentinel-1 data specification for the selected four events.

Specification	Somalia	India	Bahamas	Italy
Acquisition date	30 Oct 2019	4 May 2019	2 Sep 2019	15 Nov 2019
Polarization mode	VV and VH	VV and VH	VV and VH	VV and VH
Observation mode	IW	IW	IW	IW
Pass direction	Descending	Descending	Ascending	Descending
Spatial resolution	10 m	10 m	10 m	10 m

Table 2

Distribution of tweet text and images collected for all four events.

	Somalia	India	Bahamas	Italy
Number of tweets	144,220	558,130	5,942,478	325,679
Number of images	20,328	29,788	279,819	16,008

4. Methodology

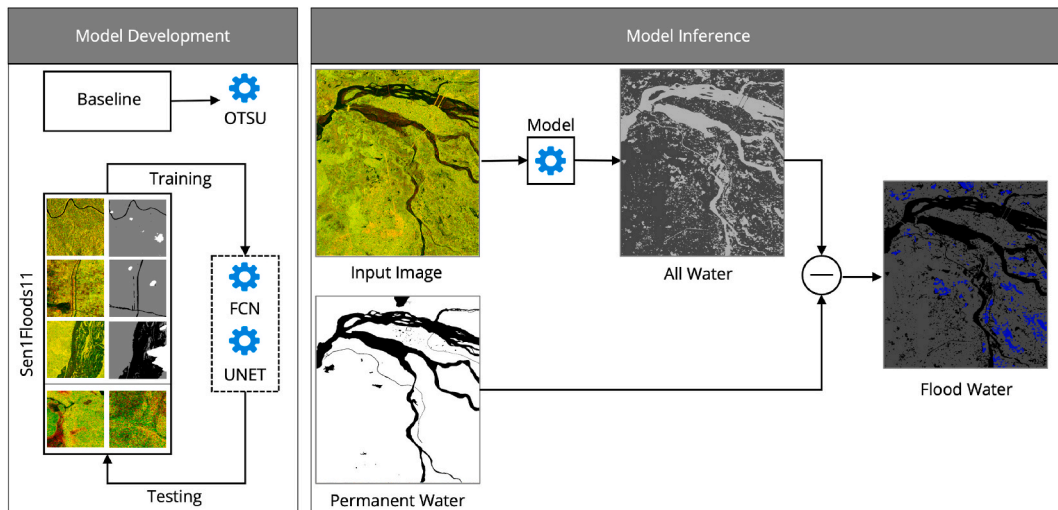
Fig. 1 presents an overview of the proposed methodology. Specifically, we first performed flood extent mapping on Sentinel-1 images using deep neural networks. Social sensing data was then processed using deep learning models. Pertinent information from both sources was fused onto a map to provide enriched information for humanitarian organizations.

4.1. Processing remote sensing data

Flood extent mapping is a segmentation task, where the objective is to assign a label (*water* or *land*) to each pixel in a given satellite image. The segments assigned with the *water* label represent *all types of water*, i.e., permanent and flood water. The permanent water represents lakes, ponds, rivers, and coastal water. Obtaining flood water segments requires subtracting permanent water segments from all water segments. Fig. 3 presents an overview of our remote sensing pipeline for processing satellite imagery. The pipeline mainly consists of two steps. The first step corresponds to the training of remote sensing models for predicting water and land segments from Sentinel-1 images. Next, the trained models are applied on four case study events to obtain their *water* segments (representing *all water* shown in light gray color) and *land* segments (shown in dark gray color). Finally, we use permanent water masks (shown in black color) to obtain flood water (shown in blue color).

Three types of labeled sets provided in Sen1Floods11 (i.e., *weak S1*, *weak S2*, and *hand-labeled*), as discussed in Section 3.2, were used for training deep learning models. Specifically, we used a Fully Convolutional Network (FCN) with ResNet50 backbone (Long et al., 2015) and a standard U-Net with ResNet34 backbone (Ronneberger et al., 2015). Each network was trained separately on three labeled sets producing a total of six models. The trained models were always tested on the hand-labeled set. As a baseline, we used the Otsu thresholding approach (Otsu, 1979), as reported by Bonafilia et al. (2020).

To avoid the problem of distributional differences between Red, Green, Blue (RGB) images and satellite imagery, the models were trained from scratch. We used a batch size of 16 with AdamW optimizer (Loshchilov and Hutter, 2017) and an initial learning rate of $5e^{-4}$ along with the cosine annealing learning rate scheduler. For data augmentation, images were randomly cropped to the patch size of 256x256 pixels. Furthermore, random horizontal and vertical flipping were also applied. Pixel values were scaled between 0 and 1 and normalized via mean–standard deviation normalization (using the mean and standard deviation computed over the training set of hand-labeled data). In order to compensate for imbalance between land and water classes, we used weighted cross entropy loss function with weights 1 and 8 for land and water, respectively. We trained models on *hand-labeled*, *Weak S1* and *Weak S2* datasets while the number of epochs was selected based on training progress (i.e., convergence and over-fitting). Models trained on weak S1 overfitted after 2 epochs, while models trained on hand-labeled and weak S2 datasets converged after 100 and 200 epochs, respectively. The PyTorch framework was used for the implementation. Model training and evaluation were performed on a GPU server equipped with two Nvidia Tesla V100 GPUs.

**Fig. 3.** Remote sensing data processing approach.

4.2. Processing social sensing data

Tweets which reports useful information for humanitarian response and impact assessment as well as those posted within the area of interest are preferred. Since only a small percentage (1–2%) of tweets contains exact geolocation information, it is necessary to perform geotagging of tweets that are without GPS-coordinates to tackle the problem of insufficient number of tweets seen in existing research. Furthermore, tweets often contain a lot of noise and irrelevant content, therefore, determining tweets' relevancy and the type of information they contain is an essential step. For these reasons, we performed rigorous processing of tweets and selected the ones which met a certain criteria.

Fig. 4 depicts our social media data processing pipeline, which mainly consisted of five steps. First step was the data collection, where event-specific keywords or hashtags were used to gather tweets related to an event. Next, tweets that were posted from areas outside the region of interest were discarded using the geotagging approach proposed in (Qazi et al., 2020). The geotagging technique analyzes information from various metadata fields (i.e., geo-coordinates, user location, place, user profile description, and tweet text) and assigns city, county, state, and country information to tweets. The tweets and images, which were geotagged inside the areas of interest, were then processed through several text and image classification models. Specifically, three deep learning text classifiers, namely *disaster type detection* ($F1 = 0.93$), *informativeness* ($F1 = 0.93$), and *humanitarian* ($F1 = 0.76$), were used from CrisisDPS (Alam et al., 2019). The first classifier predicts a disaster type of a tweet among several types. We kept the ones predicted as “Flood type”. The second classifier (i.e., “Informativeness”), determines if the tweet is informative or not, whereas the third classifier determines the type of humanitarian information (e.g., damage report, urgent needs, etc.) an informative tweet contains. This process does not limit tweets to a specific category like damages, which is often found in literature. This allowed us to qualitatively analyze all types of humanitarian tweets to derive potential information needs useful for humanitarian organizations. The images are classified using a disaster type prediction model proposed by 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”, “flooded”, and “tropical cyclone”. Finally, exact-duplicate tweets were discarded and the remaining pertinent tweets and images were used to overlay on flood water segments predicted through the satellite imagery to offer enriched information for humanitarian decision making.

5. Results and discussion

The trained models were validated against three test datasets: (i) Sen1Floods11 test set, (ii) Bolivia test set, and (iii) four case study flood events (i.e., Somalia, The Bahamas, India, and Italy). The four study areas contain *all water*, *permanent water*, and *flood water* masks, while the Sen1Floods11 dataset contains *all water* and *permanent water* masks. The *permanent water* masks for our four case study areas were obtained from Open Street Map (OpenStreetMap, 2017).

5.1. Validation of models on Sen1Floods11 and Bolivia test sets

The test set split (20%) from the Sen1Floods11 dataset is used to compute the mean Intersection-Over-Union (mIOU) metric. For image segmentation tasks, mIoU is a common evaluation metric defined as the mean value across all IOU scores computed for each class in the dataset as follows:

$$IOU = \text{TruePositive} / (\text{TruePositive} + \text{FalsePositive} + \text{FalseNegative}) \quad (1)$$

Table 3 presents the mIOU scores of the FCN and U-Net models trained on three datasets (i.e., hand-labeled, Weak S1, and Weak S2). The Otsu baseline results (as reported in (Bonafilia et al., 2020)), are shown in the last row. The results clearly show that, deep learning models outperform Otsu except for the permanent water case where Otsu achieved a higher mIOU score. Moreover, the results of the two deep learning models were comparable, however, FCN yielded better performance when trained on the weak S2 training data.

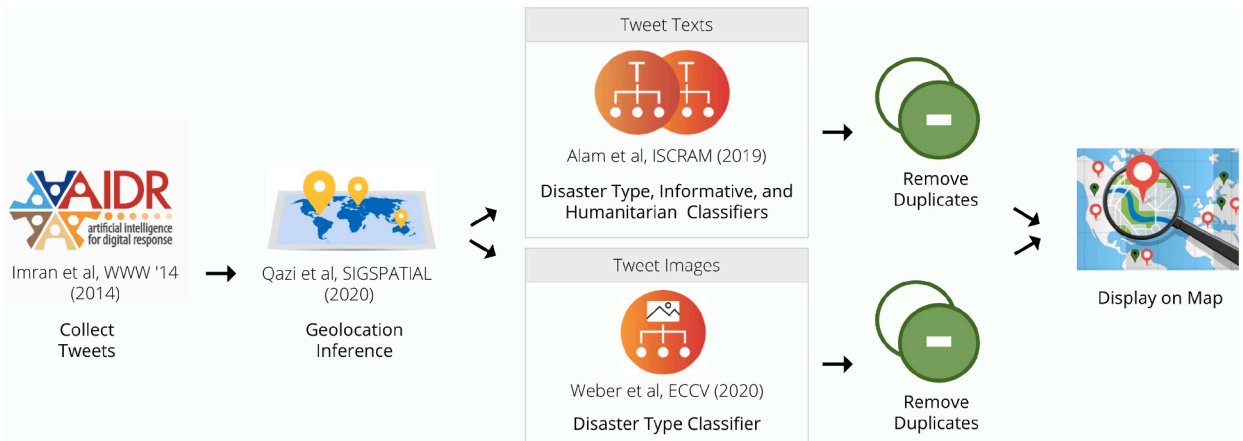


Fig. 4. Social Sensing data processing approach.

Next, we applied the models on the Bolivia test set, which is provided with the Sen1Floods11 dataset. Table 4 presents the mIOU scores for all the models and the Otsu baseline. The U-Net model trained on weak S1 outperformed others on all three water segment classes. In some cases, the FCN performance was on par with U-Net, but Otsu did not perform well in any of the cases.

5.2. Validation of models on case study events

In order to further investigate the generalization capability of the models, next we applied them on the four case study areas. These locations and events are chosen to introduce diversity in terms of terrain characteristics, continents, and the variable amount of permanent and flood water ratios. For example, the selected two coastal study areas (The Bahamas and Italy) are dominated with high permanent water ratio (Table 5). Whereas, the two urban areas (Somalia and India) have relatively high flood water ratio (Table 5).

The output of each method is a binary mask of “all water” (AW) and “land”. Flood and permanent water ground truth masks were used to separate the flood water (FW) and permanent water (PW) from the all water predictions. For false water predictions, it is impossible to determine whether the falsely detected water pixel belongs to flood water or the permanent water, so while computing the performance, we accounted false positive water predictions both in permanent water and flood water metrics. Algorithm 1 provides the overall flow of computing performance metrics for the study areas. The “ignore” pixels are masked out from predictions and the target ground truth while computing the performance metrics. Table 6 lists IOU scores of all case study events. Deep learning models performed better in most cases except for the permanent water predictions for India and flood water for The Bahamas, where higher IOU values for Otsu are observed. Due to the high imbalance in our datasets, we further evaluated the performance of models using omission and commission error rates along with models’ accuracy and F1 scores (Table 7). The extent of false negative detections is generally represented by the omission rate, whereas the extent of false positive detections is represented by the commission rate.

Algorithm 1. Performance evaluation of model inference

Algorithm 1 Performance evaluation of model inference

Input:

\hat{Y}_{AW} : All water predictions

Y_{AW} : All water ground truth mask

Y_{PW} : Permanent water ground truth mask

Y_{FW} : Flood water ground truth mask

Output:

IOU_{AW} : IOU for all water

IOU_{PW} : IOU for permanent water

IOU_{FW} : IOU for flood water

- 1: $IOU_{AW} \leftarrow \text{computeIOU}(\hat{Y}_{AW}, Y_{AW})$
- 2: $\hat{Y}_{PW} \leftarrow \hat{Y}_{AW} [Y_{FW} == 1] = \text{ignore}$ Ignore flood water from predictions
- 3: $IOU_{PW} \leftarrow \text{computeIOU}(\hat{Y}_{PW}, Y_{PW})$
- 4: $\hat{Y}_{FW} \leftarrow \hat{Y}_{AW} [Y_{PW} == 1] = \text{ignore}$ Ignore permanent water from predictions
- 5: $IOU_{FW} \leftarrow \text{computeIOU}(\hat{Y}_{FW}, Y_{FW})$

The results in Table 7 provide insights about the generalization capabilities of deep learning based segmentation models. For The Bahamas and Italy, which are both coastal areas, high accuracy for permanent, flood and all water are observed. On the contrary, low F1-scores, particularly for flood water, indicate models’ weakness, potentially due to high imbalance between flood and permanent water ratios. In general, the F1-scores of deep learning models, specifically FCN, are better in all three water categories for the coastal region of Italy while for The Bahamas Otsu is having a slight edge over deep learning models. This is due to slightly higher omission rate for deep learning models. A similar trend for the Indian event is observed, where deep learning methods (U-Net for flood water and FCN for all water) yield high F1-scores, while for the permanent water class, Otsu produce higher F1 scores. The most challenging area seems to be Somalia, where flood water covers large urban area. For Somalia, the deep learning method (U-Net) outperformed Otsu by a huge margin through achieving much higher F1-scores, accuracy, and lower omission rates.

Figs. 5 and 6 depict the water segmentation maps of each model for Somalia and India, respectively, both of which are highly populated urban areas. Fig. 5 shows that Otsu struggled significantly to differentiate between water and non-water pixels whereas the deep learning models, especially U-Net, performed exceptionally well, highlighting the generalization capabilities of the deep learning models. In Fig. 6, both Otsu and deep learning models performed well on detecting permanent water bodies of river areas. However, for flooded areas, deep learning models, especially U-Net, performed better in distinguishing flood water regions. Figs. 7 and 8 depict the water segmentation maps of each model for The Bahamas and Italy, respectively, both of which are coastal areas. Both Otsu and deep

Table 3
mIOU scores computed on Sen1Floods11 test set.

Model	Training set	Permanent Water	Flood Water	All Water
FCN	Hand Label	0.28	0.27	0.34
	Weak S1	0.35	0.26	0.33
	Weak S2	0.37	0.36	0.44
U-Net	Hand Label	0.35	0.31	0.37
	Weak S1	0.36	0.27	0.34
	Weak S2	0.35	0.35	0.42
Otsu		0.45	0.28	0.35

Table 4
mIOU scores computed on Bolivia test set.

Model	Training set	Permanent Water	Flood Water	All Water
FCN	Hand Label	0.25	0.28	0.25
	Weak S1	0.29	0.39	0.45
	Weak S2	0.24	0.27	0.32
U-Net	Hand Label	0.26	0.34	0.39
	Weak S1	0.34	0.40	0.45
	Weak S2	0.22	0.35	0.42
Otsu		0.28	0.32	0.38

Table 5
Permanent to flood water ratios in case study regions.

Region	Permanent Water	Flood Water
Somalia	8.36%	92.20%
India	62.17%	39.38%
The Bahamas	99.14%	0.86%
Italy	99.16%	0.84%

Table 6
Validation results (IOU) for case study events.

Event	Permanent Water			Flood Water			All water		
	Otsu	FCN	U-Net	Otsu	FCN	U-Net	Otsu	FCN	U-Net
Somalia	0.017	0.026	0.027	0.174	0.259	0.309	0.186	0.270	0.317
India	0.704	0.697	0.519	0.141	0.193	0.193	0.540	0.540	0.497
Bahamas	0.940	0.945	0.935	0.055	0.035	0.054	0.936	0.939	0.931
Italy	0.930	0.943	0.928	0.016	0.040	0.019	0.920	0.934	0.918

learning models yielded similar prediction performances for permanent and flood water regions. But, as shown in Fig. 7, Otsu showed a slight edge by correctly identifying the small land region on top right corner of the image while the deep learning models falsely categorized the same region as water. On the other hand, deep learning models performed better in detecting small water bodies, as seen in Fig. 8 for Italy. Otsu missed the middle canal water which was almost fully detected by the deep learning methods.

5.3. Social sensing data processing results

The approach described in Section 4.2 is used to process tweets and images collected during the four case study events. Table 8 shows the details in terms of tweets geotagged inside the area of interest and those classified as related to flood disaster, regarded as informative, and containing some humanitarian information. The last column shows the remaining unique tweets after removing the exact duplicates such as retweets.

Table 9 shows the details of the image processing approach. Specifically, images geotagged in the events AOI and those classified as heavy rainfall, flooded, or tropical cyclone are retained. In addition, all relevant tweets and images are overlaid onto the flood extent layer taken from the best performing flood water segmenting model. Maps with all layers are shown in Figs. 9–12 for India, Somalia, The Bahamas and Italy flood events, respectively.

5.4. Fusing remote sensing and social sensing data

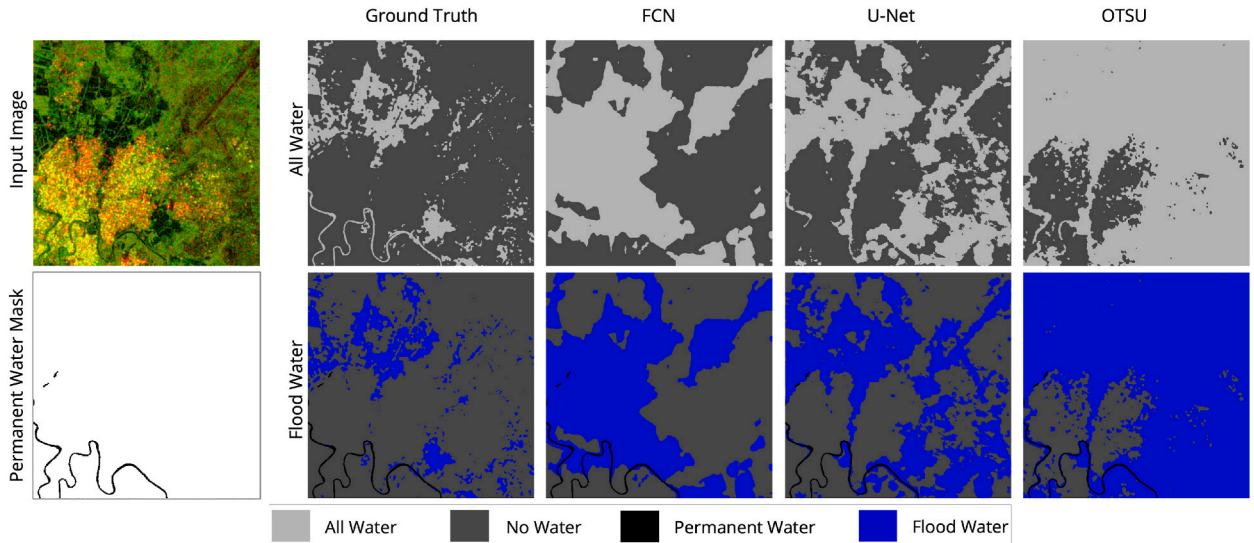
An extensive qualitative analysis of the four case study areas revealed compelling insights and ways through which both sources, when jointly used, demonstrate better understanding of the disaster situation. Unlike (Cervone et al., 2017), the proposed method leverages the improved pipeline for geotagging social media content (tweet text and images), which solves the problem of data scarcity. Moreover, the deep learning models for remote sensing provides the benefit of generalization over thresholding method like NDWI used by (Huang et al., 2018). 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:

- Confirmatory signals:** Machine learning models achieve a certain accuracy level and are prone to errors. When multiple models are predicting the same outcome, it becomes more reliable than a single model. Specifically, in our scenario, when remote sensing and social sensed data (tweet text and/or image) show flooding in the exact same area, they confirm the flooding situation in that area. Looking at Fig. 12 for Italy, the image in Post-8 overlaps with the flooded regions around the river. This image shows only flooding without any additional context. The tweet text in Post-5 for the same event reports that the streets have started to flood, and provides no other information, thus making it a confirmatory signal as it is located on a flooded region. Moreover, Fig. 11 for The Bahamas also shows a confirmatory signal where the tweet text in Post-1 reports that the community of Mudd in Abaco has been flooded, and the marker position is also located on top of the flood layer.

Table 7

Omission, Commission, Accuracy and F1 Scores for case study events.

Task	Area	Method	Omission ↓	Commission ↓	Accuracy ↑	F1-Score ↑
Permanent Water	Somalia	Otsu	0.049	0.786	0.224	0.033
		FCN	0.313	0.385	0.649	0.042
		U-Net	0.407	0.329	0.669	0.048
	India	Otsu	0.122	0.046	0.944	0.830
		FCN	0.174	0.034	0.943	0.822
		U-Net	0.050	0.154	0.861	0.681
	Bahamas	Otsu	0.011	0.108	0.957	0.969
		FCN	0.002	0.117	0.960	0.971
		U-Net	0.003	0.139	0.953	0.966
	Italy	Otsu	0.061	0.024	0.952	0.960
		FCN	0.021	0.061 1	0.959	0.967
		U-Net	0.023	0.086	0.948	0.959
Flood Water	Somalia	Otsu	0.003	0.786	0.325	0.297
		FCN	0.145	0.385	0.648	0.410
		U-Net	0.081	0.329	0.706	0.471
	India	Otsu	0.803	0.046	0.875	0.251
		FCN	0.751	0.034	0.889	0.302
		U-Net	0.559	0.154	0.802	0.310
	Bahamas	Otsu	0.616	0.108	0.882	0.105
		FCN	0.738	0.117	0.871	0.06
		U-Net	0.537	0.139	0.853	0.102
	Italy	Otsu	0.916	0.024	0.958	0.083
		FCN	0.747	0.061	0.923	0.128
		U-Net	0.785	0.086	0.948	0.085
All Water	Somalia	Otsu	0.007	0.786	0.333	0.314
		FCN	0.158	0.385	0.649	0.425
		U-Net	0.107	0.329	0.704	0.482
	India	Otsu	0.377	0.046	0.876	0.701
		FCN	0.399	0.034	0.879	0.701
		U-Net	0.250	0.154	0.822	0.664
	Bahamas	Otsu	0.016	0.108	0.954	0.967
		FCN	0.009	0.117	0.956	0.968
		U-Net	0.008	0.139	0.950	0.964
	Italy	Otsu	0.065	0.024	0.949	0.958
		FCN	0.031	0.061	0.957	0.966
		U-Net	0.033	0.086	0.946	0.957

**Fig. 5.** Remote sensing results for Somalia event showing the input Sentinel-1 image and ground truth masks as well as the output predictions from FCN, U-Net, and Otsu models.

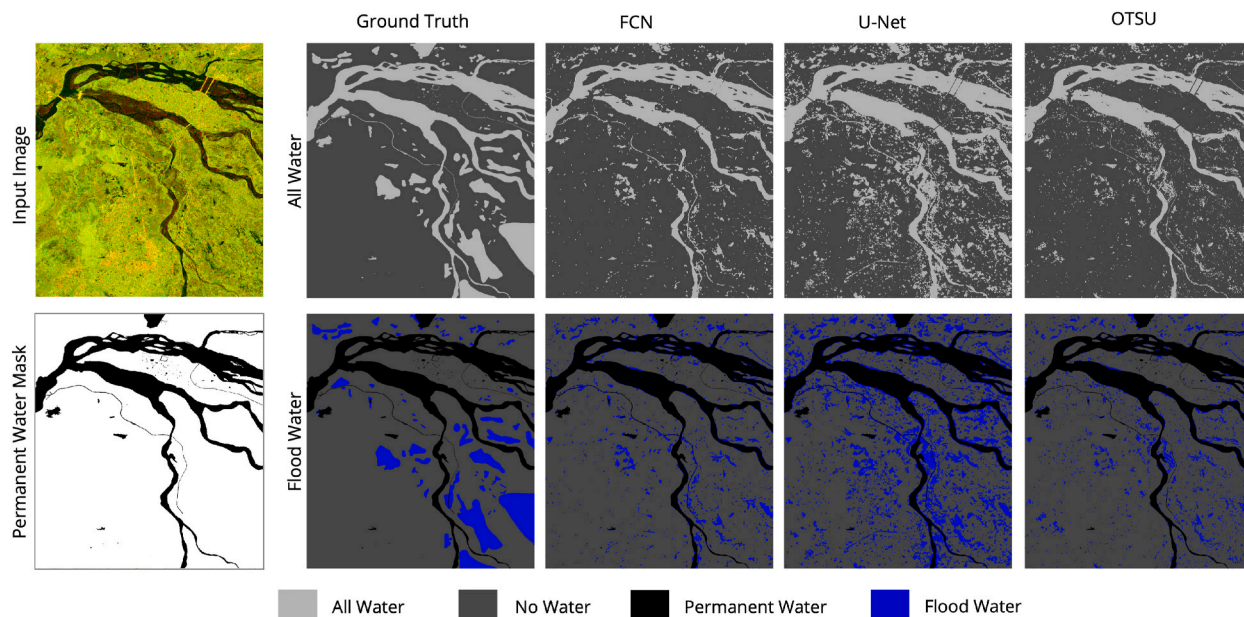


Fig. 6. Remote sensing results for India event showing the input Sentinel-1 image and ground truth masks as well as the output predictions from FCN, U-Net, and Otsu models.

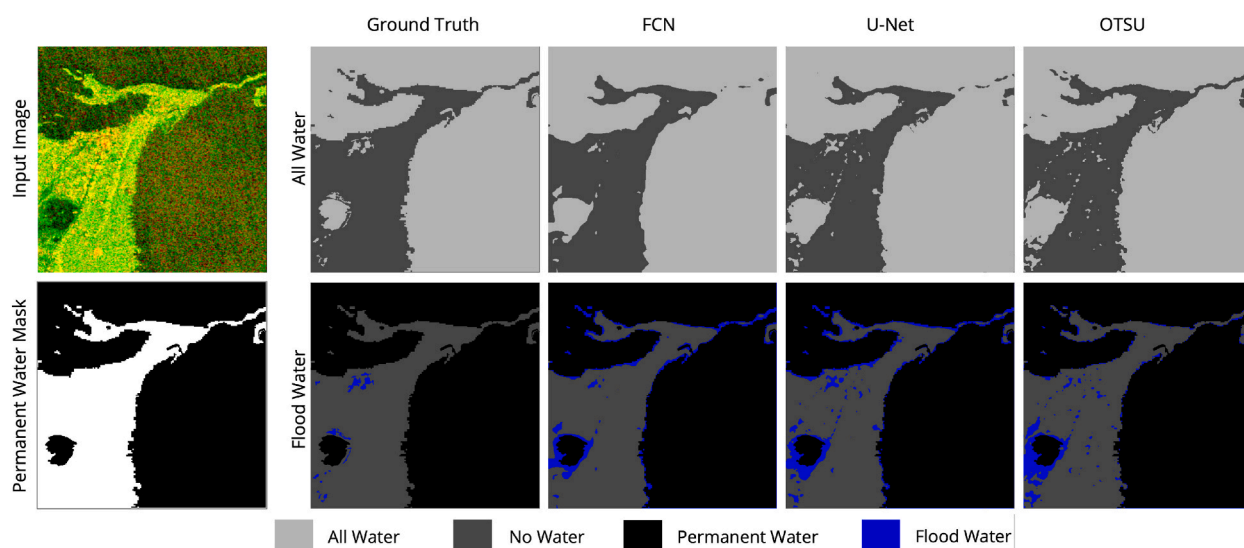


Fig. 7. Remote sensing results for The Bahamas event showing the input Sentinel-1 image and ground truth masks as well as the output predictions from FCN, U-Net, and Otsu models.

- (ii) *Complementary signals*: Tweet text and/or images which intersect or are proximate to the flood extent layer from remote sensing can potentially provide extra contextual information to crisis managers.

The analysis of contextual information carried by these social sensing data reveal several types of information. Contextual information reports were divided into three main categories, namely: (i) needs and requests, (ii) disaster impact or damage reports, and (iii) situational information.

- *Needs and requests*: Understanding people's needs and urgent requests is an important task for response agencies. The analysis found several such reports. For example, The tweet text in Post-5 of Fig. 11, which is located on the flood water states that, around 70,000 people in The Bahamas need access to food, water and medicine which all must be prioritized with high urgency. Furthermore, the tweet text in Post-1 of Fig. 12 is located on top of the flood extent area, which is requesting for specialists to help save manuscripts and sheet music that have been damaged in the flood. Moreover, it was observed that some request

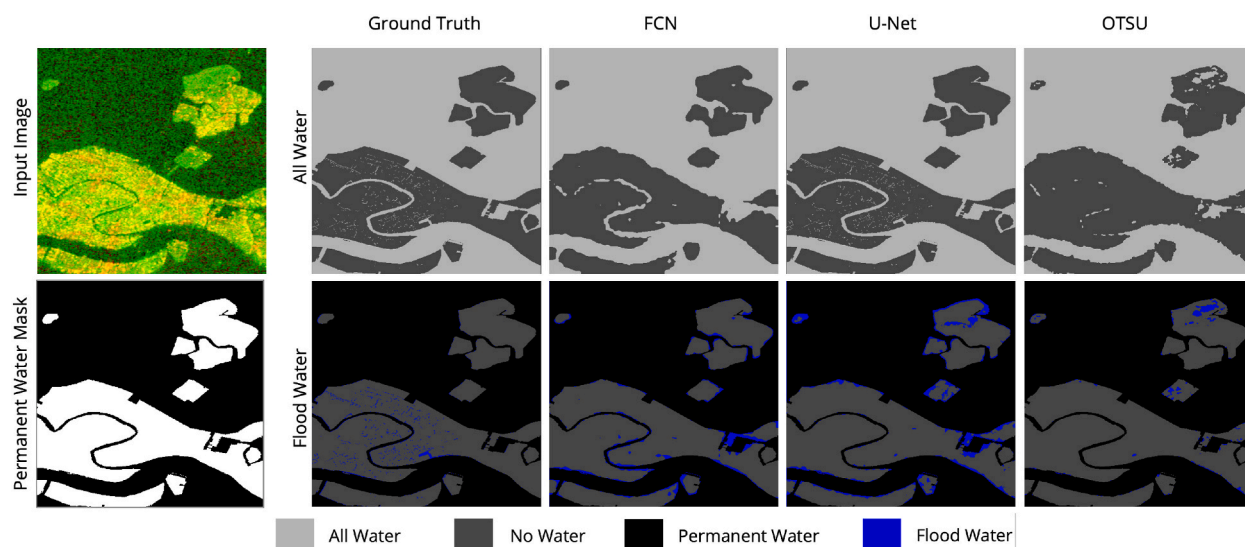


Fig. 8. Remote sensing results for Italy event showing the input Sentinel-1 image and ground truth masks as well as the output predictions from FCN, U-Net, and Otsu models.

Table 8

Processing of tweet messages for all case study events.

	Geolocated in AOI	Flood type tweets	Informative tweets	Humanitarian tweets	Unique tweets
Somalia	115	84	84	84	27
India	15,666	10,231	7,715	5,981	2,359
Bahamas	4,885	4,152	4,083	2,365	239
Italy	2,113	299	275	224	122

Table 9

Processing of tweet images for all case study events.

	Images geotagged in AOI	Classified as heavy rainfall	Classified as flood	Classified as tropical cyclone	Total relevant
Somalia	16	1	13	0	14
India	548	103	24	28	155
Bahamas	24	8	1	1	10
Italy	19	0	13	0	13

reports that have been partially or fully completed, which does not involve the locals on the ground. For example, the tweet text in Post-2 of Fig. 11 mentions a rescue request which has already been fulfilled as 19 patients were airlifted from Marsh Harbour hospital. With this information, humanitarian organization can keep track of the number of people that have already been rescued, in order to communicate and inform the public.

- Disaster impact and damages:** Several reports and images carrying vital information about different kinds of impacts are found. For example, during the flooding event in India, (Fig. 9), the tweet text in Post-1 reports damage done to a crane as it collapsed due to the strong winds. Likewise, Tweet text in Post-3, which is superimposed onto the flood extent, reported potential damage to wildlife habitat where 4,000 deers were at risk of being affected. Humanitarian organizations can use this information to send their animal rescue teams in a timely manner. Along with tweet texts, tweet images like the one in Post-8 show infrastructure damage done to a petrol station. This is useful for damage assessment teams who can inspect the site to ensure safety measures are met. Similarly, the tweet text in Post-3 from the Italy event 12 reports damage to centuries-old Jewish cemeteries as a result of toppled trees.
- Situational reports:** Social sensing data gathered during flood events carry different types of situational updates. For example, the tweet text in Post-5 of Fig. 9, which is located around the river Mahandi in India, reports cautionary information to inform the public to stay alert in the next 24 h. The tweet text in Post-6, on the other hand, reports on weather updates where the city is expected to experience heavy rainfall on May 3. Along with tweet texts, images (e.g., the image in Post-7) show the negative impact the flood has on the locals, as many people have been displaced from their homes. The image in Post-6 of Fig. 11 shows the affected individuals are lacking basic needs and necessities, and when posted on social media this can encourage others to donate as users can empathize with the affected people. Moreover, the tweet text in Post-4 of Italy, which is situated along the edge of the river, reports on the three different areas affected as a result of the flood.

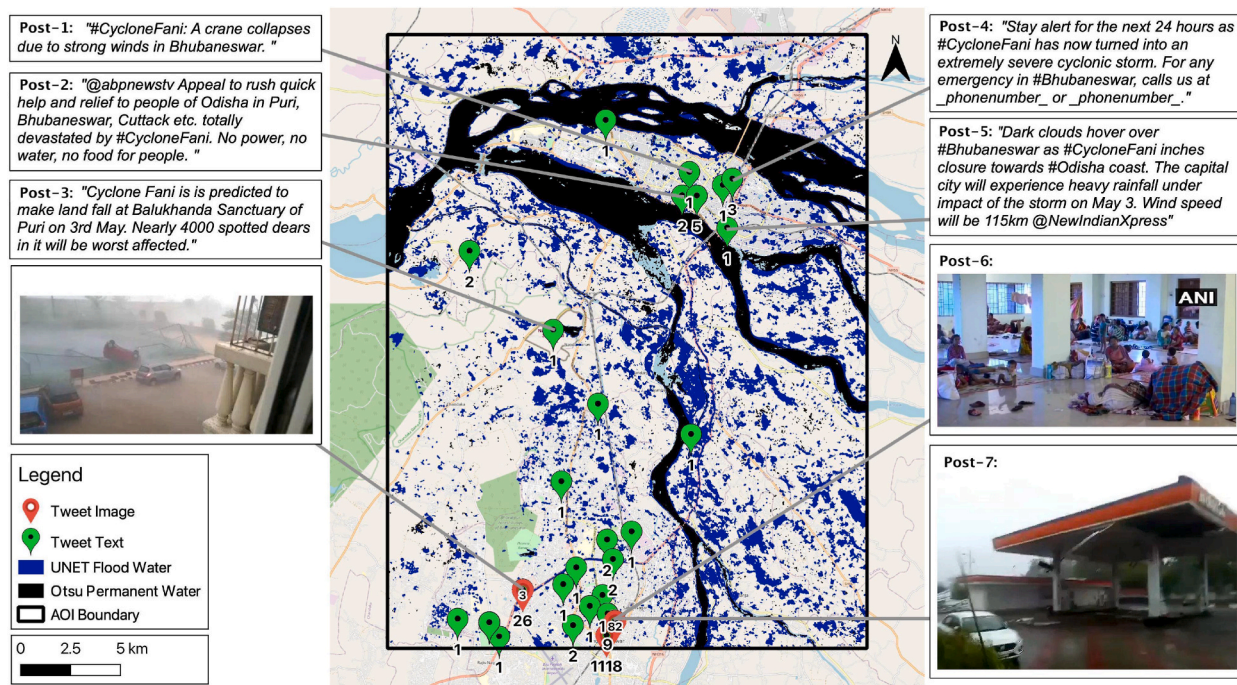


Fig. 9. Map of the India event with all layers. Numbers underneath the green markers represent tweet counts posted from the same location. Image counts are put inside the red markers. Tweet messages with phone numbers are masked with _phonenumber_. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

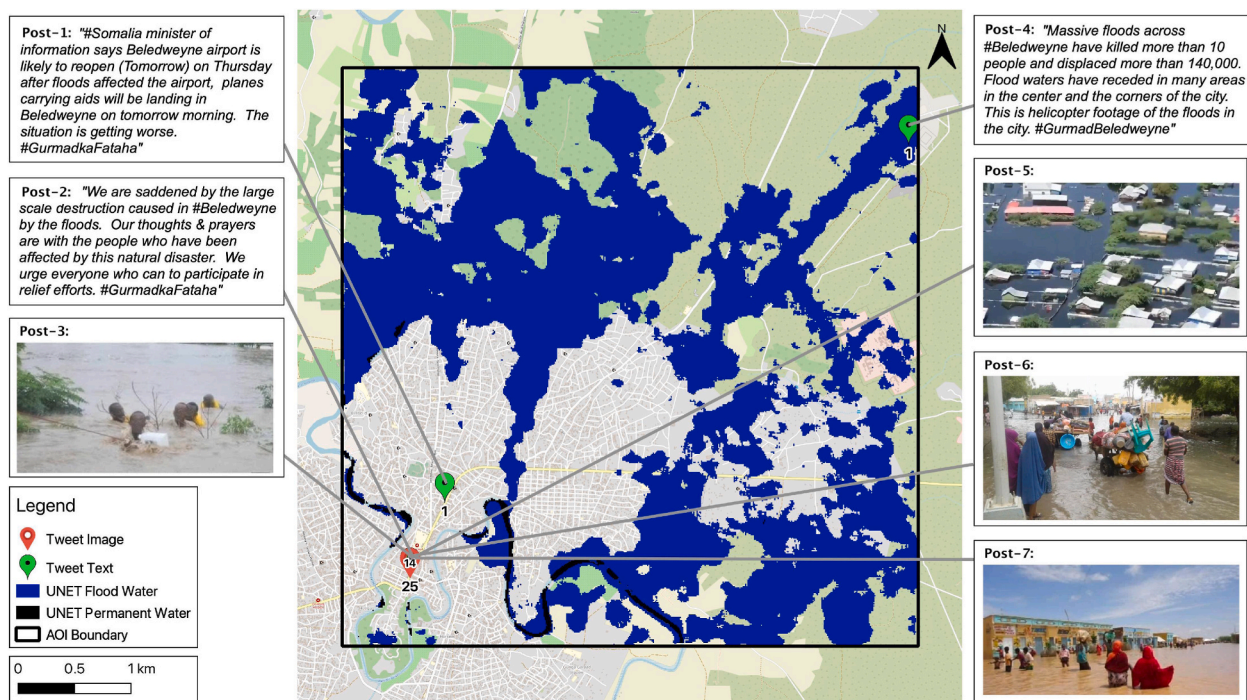


Fig. 10. Map of the Somalia, Beledweyne event with all layers. Numbers underneath the green markers represent tweet counts posted from the same location. Image counts are put inside the red markers. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

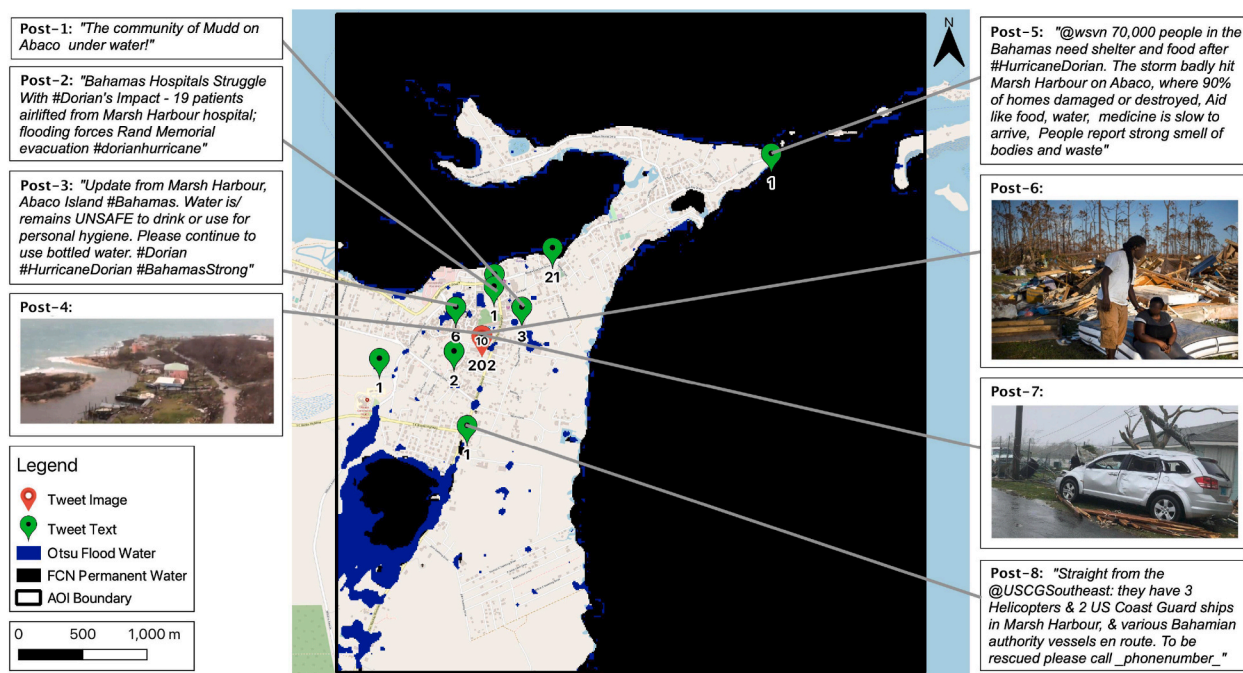


Fig. 11. Map of The Bahamas event with all layers. Numbers underneath the green markers represent tweet counts posted from the same location. Image counts are put inside the red markers. Tweet messages with phone numbers are masked with _phonenumber_. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

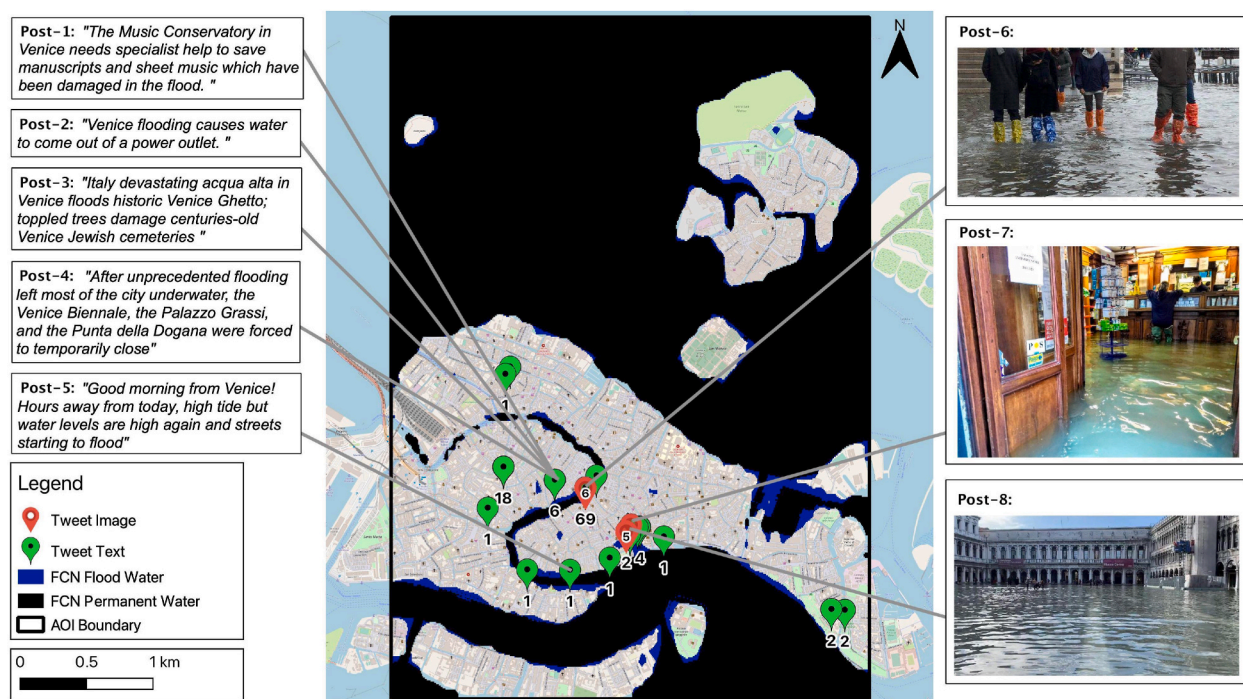


Fig. 12. Map of the Venice, Italy event with all layers. Numbers underneath the green markers represent tweet counts posted from the same location. Image counts are put inside the red markers. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

- (iii) *Novel signals*: Informative tweets and images which are not overlapping with any flood water pixels are vital to detect flooded areas which are not detected by the remote sensing method. For example in Somalia (Fig. 10), all the tweet text and images are spatially distributed away from the river. The image in Post-5 shows an aerial view of all the flooded houses in the area. The image in Post-3 shows locals making their way through the flood water which is waist level, and this high depth level can be risky for children who could potentially drown. The tweet text in Post-1 reports on public information from Somalia's Minister of Information who announced that the Beledweyne airport would reopen tomorrow after it got flooded, and that planes carrying aid would land in the city the following morning. This information has been distributed via Twitter as a form of crisis communication to ensure that the public does not panic during the onset of the disaster.

6. Conclusion

This paper explored the fusion of two heterogeneous data sources, namely, satellite imagery and social sensing data for building more informed flood extent maps for analysis. As for remote sensing, two deep learning models, FCN and U-Net were trained using existing labels and were tested on four carefully selected disaster events. The results of these deep learning models were compared to a traditional thresholding method, Otsu. Deep learning models were found to perform the best across both urban and coastal regions. In addition to the remote sensing layer, we added two social sensing layers representing tweet text and images, which provide more contextual information about the crisis such as the ability to understand the urgent needs of the affected people, the different types of impact and the overall situation. This enhanced understanding would not have been possible through satellite imagery analysis only, and thus, we remark that the fusion of these two data sources will help crisis managers in carrying out their operations more effectively.

7. Future directions and limitations

Supervised deep learning methods perform better with a large-scale and high-quality annotated dataset. However, there is a scarcity of high-quality ground truth data and standardized benchmark datasets for flood disasters which is yet to be tackled. With the existing datasets available, there lies an imbalance between different terrain types and geographical locations. These limitations open an opportunity for researchers to focus on the development of high quality, balanced datasets that can be standardized for benchmarking. Moreover, remote sensing data with low spatial resolution causes difficulty in detecting small water segments like flooded streets and vegetation even for deep learning methods. To enable accurate flood mapping in urban areas, future studies can consider enhancing the resolution of satellite imagery or even exploring state-of-the-art edge-aware deep learning models. Lastly, deep learning models can be developed for classifying social sensed content (text and images) tailored for flooding only, as the current study utilized generic classifiers for all disasters.

While remote sensing and social sensing have shown great potential for flood mapping and analysis, there are certain unavoidable limitations. These limitations typically originate from the remote sensing data sources where researchers have no control over. Some common limitations include low spatial resolution, high revisit time, or low temporal resolution. For accurate flood mapping, high spatial and temporal resolutions are desired, but majority of satellites offering free imagery have medium to low spatial resolution and relatively high revisit time. On the other hand, services offering high resolution imagery are expensive.

Funding

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). In particular, this research was funded by NPRP grant #BFC03-0630-190 011 from the Qatar National Research Fund (a member of Qatar Foundation).

Ethical statement

The authors declare that all ethical practices have been followed in relation to the development, writing, and publication of the article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Open Access funding provided by the Qatar National Library.

References

- Akhtar, Zainab, Ofli, Ferda, Imran, Muhammad, 2021. In: *Towards Using Remote Sensing and Social Media Data for Flood Mapping*. ISCRAM, pp. 536–551.
 Alam, Firoj, Ofli, Ferda, Imran, Muhammad, 2019. In: *CrisisDps: Crisis Data Processing Services*. ISCRAM.
 Aulov, Oleg, Halem, Milton, 2012. Human sensor networks for improved modeling of natural disasters. *Proc. IEEE* 100 (10), 2812–2823.

- Bai, Yanbing, Wu, Wenqi, Yang, Zhengxin, Yu, Jinze, Zhao, Bo, Liu, Xing, Yang, Hanfang, Mas, Erick, Koshimura, Shunichi, 2021. Enhancement of detecting permanent water and temporary water in flood disasters by fusing sentinel-1 and sentinel-2 imagery using deep learning algorithms: demonstration of sen1floods11 benchmark datasets. *Rem. Sens.* 13 (11), 2220.
- Bonafilia, Derrick, Tellman, Beth, Anderson, Tyler, Isenberg, Erica, June 2020. Sen1floods11: a georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In: *CVPR Workshops*.
- Conde, Francisco Carreño, Muñoz, María De Mata, 2019. Flood monitoring based on the study of sentinel-1 sar images: the ebro river case study. *Water* 11 (12).
- Cervone, Guido, Schnebele, Emily, Waters, Nigel, Moccaldi, Martina, Rosa, Sicignano, 2017. Using social media and satellite data for damage assessment in urban areas during emergencies. In: *Seeing Cities through Big Data*. Springer, pp. 443–457.
- Copernicus, 2021. Copernicus emergency management service, 2021-08-01. <https://emergency.copernicus.eu/>.
- FloodList, 2021. Reporting floods and flooding news since 2008, 2021-08-01. <https://floodlist.com/europe/italy-venice-floods-november-2019>.
- Gao, Bo-Cai, 1996. Ndw— a normalized difference water index for remote sensing of vegetation liquid water from space. *Rem. Sens. Environ.* 58 (3), 257–266.
- Gorelick, Noel, Hancher, Matt, Dixon, Mike, Simon, Ilyushchenko, Thau, David, Moore, Rebecca, 2017. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sensing of Environment*.
- Huang, Xiao, Wang, Cuizhen, Li, Zhenlong, 2018. A near real-time flood-mapping approach by integrating social media and post-event satellite imagery. *Spatial Sci.* 24 (2), 113–123.
- Muhammad, Imran, Castillo, Carlos, Lucas, Ji, Meier, Patrick, Vieweg, Sarah, 2014. Aidr: artificial intelligence for disaster response. In: *Proceedings of the 23rd International Conference on World Wide Web*, pp. 159–162.
- Jongman, Brenden, Wagemaker, Jurjen, Romero, Beatriz Revilla, De Perez, Erin Coughlan, 2015. Early flood detection for rapid humanitarian response: harnessing near real-time satellite and twitter signals. *ISPRS Int. J. Geo-Inf.* 4 (4), 2246–2266.
- Joyce, Karen E., Belliss, Stella E., Samsonov, Sergey V., McNeill, Stephen J., Glassey, Phil J., 2009. A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. *Prog. Phys. Geogr.* 33 (2), 183–207.
- Katiyar, Vaibhav, Tamkuan, Noppawan, Nagai, Masahiko, 2021. Near-real-time flood mapping using off-the-shelf models with sar imagery and deep learning. *Rem. Sens.* 13 (12), 2334.
- Lee, Sang Uk, Chung, Seok Yoon, Park, Rae Hong, 1990. A comparative performance study of several global thresholding techniques for segmentation. *Comput. Vis. Graph Image Process* 52 (2), 171–190. ISSN 0734-189X.
- Li, Jun, He, Zhi, Plaza, Javier, Li, Shutao, Chen, Jinfen, Wu, Henglin, Wang, Yandong, Liu, Yu, 2017. Social media: new perspectives to improve remote sensing for emergency response. *Proc. IEEE* 105 (10), 1900–1912.
- Liu, Bin, Li, Xiaofeng, Zheng, Gang, 2019. Coastal inundation mapping from bitemporal and dual-polarization sar imagery based on deep convolutional neural networks. *J. Geophys. Res.: Oceans* 124 (12), 9101–9113.
- Long, Jonathan, Shelhamer, Evan, Trevor Darrell, 2015. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440.
- Loshchilov, Ilya, Hutter, Frank, 2017. Decoupled Weight Decay Regularization arXiv preprint arXiv:1711.05101.
- Marty, Graham, 2019. Social media + satellite imagery = predictive power, 2021-08-01. <https://www.delltechnologies.com/en-us/perspectives/social-media-satellite-imagery-predictive-power/>.
- Mateo-García, Gonzalo, Veitch-Michaelis, Josh, Smith, Lewis, Oprea, Silviu, Guy, Schumann, Gal, Yarin, Backes, Dietmar, 03 2021. Towards global flood mapping onboard low cost satellites with machine learning. *Sci. Rep.* 11 <https://doi.org/10.1038/s41598-021-86650-z>.
- Matgen, P., Hostache, R., Schumann, G., Pfister, L., Hoffmann, L., Savenije, H.H.G., 2011. Towards an automated sar-based flood monitoring system: lessons learned from two case studies. *Phys. Chem. Earth, Parts A/B/C* 36 (7), 241–252. ISSN 1474-7065.
- Mosavi, Amir, Ozturk, Pinar, Chau, Kwok-wing, 2018. Flood prediction using machine learning models: literature review. *Water* 10 (11), 1536.
- Muñoz, David F., Paul, Muñoz, Moftakhari, Hamed, Hamid, Moradkhani, 2021. From local to regional compound flood mapping with deep learning and data fusion techniques. *Sci. Total Environ.* 782, 146927.
- OpenStreetMap, 2017. Planet dump retrieved from, 2021-08-01. <https://planet.osm.org>. <https://www.openstreetmap.org>.
- Otsu, Nobuyuki, 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* 9 (1), 62–66.
- Oxendine, Christopher E., Schnebele, Emily, Guido, Cervone, Waters, Nigel, 2014. In: *Fusing Non-authoritative Data to Improve Situational Awareness in Emergencies*. ISCRAM.
- Pekel, Jean-François, Cottam, Andrew, Gorelick, Noel, Belward, Alan S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature* 540 (7633), 418–422.
- Pregnotato, Maria, Ford, Alistair, Wilkinson, Sean M., Dawson, Richard J., 2017. The impact of flooding on road transport: a depth-disruption function. *Transport. Res. Transport Environ.* 55, 67–81.
- Qazi, Umair, Imran, Muhammad, Ofli, Ferda, 2020. Geocov19: a dataset of hundreds of millions of multilingual covid-19 tweets with location information. *SIGSPATIAL Special* 12 (1), 6–15.
- ReliefWeb, 2019. Somalia floods update (31 october 2019), 2021-08-01. <https://reliefweb.int/report/somalia/somalia-floods-update-31-october-2019>.
- Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas, 2015. U-net: convolutional networks for biomedical image segmentation. In: *MICCAI*. Springer, pp. 234–241.
- Rosser, Julian F., Leibovici, D.G., Jackson, M.J., 2017. Rapid flood inundation mapping using social media, remote sensing and topographic data. *Nat. Hazards* 87 (1), 103–120.
- Schnebele, Emily, Guido, Cervone, 2013. Improving remote sensing flood assessment using volunteered geographical data. *Nat. Hazards Earth Syst. Sci.* 13 (3), 669–677.
- Schnebele, Emily, Guido, Cervone, Kumar, Shamanth, Waters, Nigel, 2014. Real time estimation of the calgary floods using limited remote sensing data. *Water* 6 (2), 381–398.
- Schnebele, Emily, Oxendine, Christopher, Guido, Cervone, Ferreira, Celso M., Waters, Nigel, 2015. Using non-authoritative sources during emergencies in urban areas. In: *Computational Approaches for Urban Environments*. Springer, pp. 337–361.
- Shen, Xinyi, Anagnostou, Emmanouil, Allen, George, Brakenridge, Robert, Albert, Kettner, 02 2019. Near-real-time non-obstructed flood inundation mapping using synthetic aperture radar. *Rem. Sens. Environ.* 221, 302–315. <https://doi.org/10.1016/j.rse.2018.11.008>.
- Sun, Donglian, Li, Sanmei, Zheng, Wei, Croitoru, Arie, Anthony, Stefanidis, Goldberg, Mitchell, 2016. Mapping floods due to hurricane sandy using npp viirs and atms data and geotagged flickr imagery. *Int. J. Digit. Earth* 9 (5), 427–441.
- UnitedNations, 2021. In detail: recommended practice: flood mapping and damage assessment using sentinel-1 sar data in google earth engine, 2021-08-01. <https://un-spider.org/advisory-support/recommended-practices/recommended-practice-google-earth-engine-flood-mapping/in-detail>.
- UNOSAT, 2021. Unosat rapid mapping service, 2021-08-01. <https://www.unitar.org/maps/unosat-rapid-mapping-service>.
- Weber, Ethan, Nuria Marzo, Papadopoulos, Dim P., Biswas, Aritro, Lapedriza, Agata, Ofli, Ferda, Imran, Muhammad, Torralba, Antonio, 2020. Detecting natural disasters, damage, and incidents in the wild. In: *European Conference on Computer Vision*. Springer, pp. 331–350.
- Wikipedia, 2019a. Cyclone fani, 2021-08-01. https://en.wikipedia.org/wiki/Cyclone_Fani.
- Wikipedia, 2019b. Hurricane dorian, 2021-08-01. https://en.wikipedia.org/wiki/Hurricane_Dorian.
- Zhan, Yantong, Zhang, Guoying, 03 2019. An improved otsu algorithm using histogram accumulation moment for ore segmentation. *Symmetry* 11, 431.