Received: 12 June 2024 | Accepted: 23 July 2024

DOI: 10.1002/aaai.12196

SPECIAL TOPIC ARTICLE

Fusing remote and social sensing data for flood impact mapping

Zainab Akhtar | **Umair Qazi** | **Aya El-Sakka** | **Rizwan Sadiq** | **Ferda Ofli** | **Muhammad Imran**

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

Correspondence
Muhammad Imran, Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar.
Email: mimran@hbku.edu.qa

Funding information
Qatar National Research Fund, Grant/Award Number: BFC03-0630-190011

Abstract

The absence of comprehensive situational awareness information poses a significant challenge for humanitarian organizations during their response efforts. We present Flood Insights, an end-to-end system, that ingests data from multiple nontraditional data sources such as remote sensing, social sensing, and geospatial data. We employ state-of-the-art natural language processing and computer vision models to identify flood exposure, ground-level damage and flood reports, and most importantly, urgent needs of affected people. We deploy and test the system during a recent real-world catastrophe, the 2022 Pakistan floods, to surface critical situational and damage information at the district level. We validated the system's effectiveness through various statistical analyses using official ground-truth data, showcasing its strong performance and explanatory power of integrating multiple data sources. Moreover, the system was commended by the United Nations Development Programme stationed in Pakistan, as well as local authorities, for pinpointing hard-hit districts and enhancing disaster response.

INTRODUCTION

Flooding is a prevalent natural disaster that can inflict significant harm on communities by causing infrastructure damage, loss of life, disruption of essential services, and environmental damage (Milly et al. 2002). The main obstacle that impedes disaster response efforts is the absence of timely situational awareness information on damage impacts, urgent needs of the population and flood extent (Vieweg 2012).

Traditional methods of obtaining situational information include conducting damage and needs assessment surveys on the ground, which are resource-intensive, time-consuming, and difficult to obtain in hard-to-reach areas (Macabuag et al. 2022). The unavailability of critical information delays decision-making and thus hinders relief efforts, which can result in further loss of life, property damage, and economic disruption.

To tackle the above-mentioned issue, we leverage timely information from multiple nontraditional data sources and artificial intelligence to develop an end-to-end system, called Flood Insights. The system employs remote sensing, social sensing, and geospatial data to overcome the limitations of existing rapid assessment methods and tools whilst answering six vital questions for enhanced situational awareness. These questions have been derived based on our consultations with humanitarian organizations and an analysis of their official reports outlining their information needs. This includes Disaster Emergency Needs Assessment Reports (Government of Somalia

2020; Malawi Government, The United Nations, The World Bank, GFDRR 2019), Flood Response Plans (FAO Representation in the Sudan 2020; Shelter Cluster 2017), Impact Assessment Reports (Ministry of Agriculture and Irrigation; Ministry of Livestock, Fisheries & Rural Development, FAO, WFP 2015; NDC Solomon Islands, City of Honiara National Capital 2018), and Flood Assessment Reports (Centre for Peace and Democracy 2018; Department of Civil Protection, UN-Agencies, NGOs 2019; Kenya National Disaster Operations Centre, UNDP 2013). The questions guiding our system include:

1. Which regions (province or district level) are flooded?
2. What is the extent of flooding in urban and rural regions?
3. How many people are exposed to flooding?
4. Which age groups and genders are at higher risk?
5. What type of damage is caused by floods?
6. What are the urgent needs of the affected people?

To answer these questions, the Flood Insights system comprises three main data processing pipelines: (i) satellite and geographic information systems (GIS), (ii) social media text, and (iii) social media images. These pipelines incorporate various types of computational components responsible for data fetching, processing, analysis, and visualization. Specifically, the satellite pipeline can download both Synthetic Aperture Radar (SAR) and Optical imagery from freely available data sources (i.e., Sentinel-1 and Sentinel-2) and process it through deep learning-based models to identify flood segments. The pipeline also relies on nightlight imagery and population estimates to determine exposed populations in urban and rural regions. Social sensing pipelines collect textual and imagery posts from disaster-affected areas on Twitter (now X) and employ several state-of-the-art deep learning models, including natural language processing and image processing, to perform flood, damage, and needs assessment.

In this paper, we describe the architecture of the Flood Insights system and report our deployment experience during a recent devastating flood disaster in Pakistan. Over a 1-month period (25 August to 25 September), the system downloaded and processed 1608 satellite images (size = 890 GB) covering 98% of Pakistan and all 160 districts. Moreover, the system collected and analyzed around 9.4 million text messages and 411k images posted on Twitter by locals. The insights gained from integrating these multiple data sources are presented in the Flood Insights Dashboard[1] as seen in Figure 1. This dashboard has several components that update dynamically based on four filters that require user selection. This includes province, district, time period, and region type. The layout of the dashboard is divided into three major sections. The left-hand side is related to the satellite and geospatial analysis followed by interactive choropleth maps showing the distribution of needs and flood/damage reports in the middle. The right-hand side is dedicated to social media, where the tweet text and image analysis are detailed.

Satellite and geospatial visualizations

The first column is subdivided into two subsections, where "Overall Country Level Statistics" indicates country-level flood extent percentage and population exposure, along with an interactive flood map. A detailed breakdown of this is shown in the "Detailed Flood Extent & Exposure" where the user can use all four filters to dynamically update the "Flood Extent," "Gender Exposure," and "Age Group Exposure" charts.

Tweet text and image distributions

The middle section has two multilayer map visualizations. The top map shows the percentage of Shelter, Donation, Rescue, Volunteering, Basic Necessities, Medical, and Food requests in different regions, which are normalized by the percentage of total tweets in that region. The bottom map shows Flood and Damage reports based on images per region, normalized by total images from that region. Both these maps are dynamically updated based on the following three global filters: province, district, and time period.

Other social media components

This section displays several cards highlighting the total number of reports about the affected people, needs, damage, and flood reports. There are two timelines indicating the daily distribution of tweet text and image reports where users can focus on their desired date range through the scrollbar functionality. Moreover, there is a snapshot of the damage and flood images in the rightmost panel of the dashboard where users can view the images in detail by clicking on an image. Below these panels, there is a list of the top-five locations with the highest number of flood reports, damage reports, flood extent, and population exposure. All these components dynamically update based on province, district, or time period selection.

To validate the reliability of our system, we used official ground-truth data from the government of Pakistan to obtain a comprehensive impact score comprising human and infrastructure impacts. Similarly, we constructed impact scores for remote sensing and social sensing and compared them against the official ground truth using correlation and regression techniques. When examining

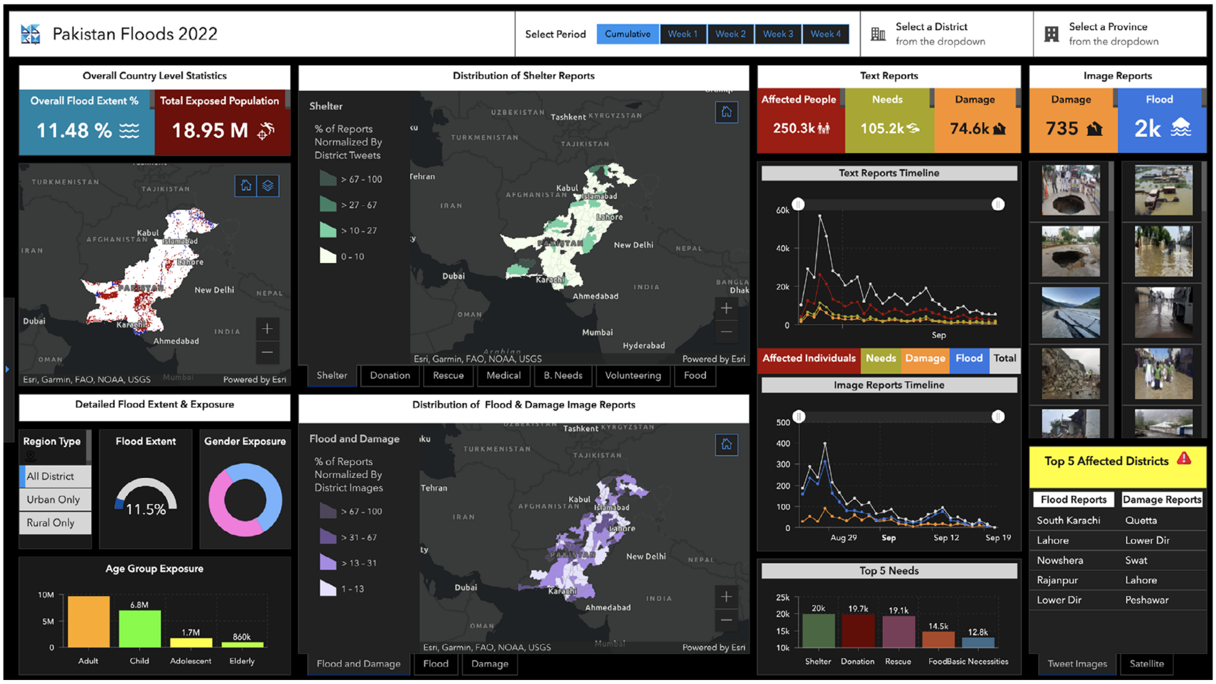

FIGURE 1 Flood Insights system's dashboard showing different interactive visualization components.

the predictive power of each data source separately, remote sensing was found to be effective for infrastructure impact, whereas social sensing was more reliable for human impact. More importantly, when combining both data sources, the predictive power for overall disaster impact was significantly stronger.

Furthermore, the Flood Insights dashboard was shared with the United Nations Development Programme (UNDP) and other local authorities in Pakistan, who found the information valuable for their disaster response strategy.

A summary of our main contributions is as follows:

1. We design and develop a novel working system, Flood Insights, that integrates nontraditional data sources and state-of-the-art AI models for mapping flood extent, damages, and urgent needs of the affected population.
2. We deploy the system during a real-world disaster and conduct qualitative assessments to reveal its high effectiveness and usefulness.
3. We reveal the individual and collective power of remote sensing and social sensing in predicting ground truth disaster impacts through comprehensive statistical analyses.
4. We release the Flood Insights code publicly on GitHub.

RELATED WORK

Humanitarian organizations typically use a combination of technology (i.e., remote sensing, crowdsourcing, GIS and mapping), field assessments, and household surveys to perform rapid flood, damage, and needs assessments. With regards to flood assessments, the European Space Agency's Copernicus Emergency Management Service (EMS) and United Nations Operational Satellite Applications Programme (UNOSAT) are focused on utilizing satellite imagery for flood location and extent mapping. UNOSAT has developed its own deep learning model for segmenting floods in Sentinel-1 imagery (Nemni et al. 2020). These organizations typically disseminate their information through static maps and reports, except for UNOSAT, which also develops FloodAI S-1 Dashboards to showcase their satellite data and GIS analysis. Whilst these existing services pinpoint flood location and extent, crisis responders lack insight into the scale of urban and rural flooding. This information is needed to customize response interventions based on the differing needs across regions.

With respect to damage assessment, several organizations (i.e., American Society of Civil Engineers, United States Federal Emergency Management Agency, International Federation of Red Cross and Red Crescent Societies) have developed tools to assess flood impacts on infrastructure, facilities, communities, food, and agriculture. These methods are typically time-consuming, particularly in hard-to-reach areas, and also lack standardization, making it difficult to compare results from different assessments. Therefore, crisis responders are unable to get a holistic overview of all damage impacts needed to prioritize their response efforts.

Likewise, with needs assessment, several guides and frameworks have been developed, that is, Joint

Assessment Mission (JAM) tool (Oman 2011), Multi-cluster/Sector Initial Rapid Assessment (MIRA) tool (Ziolkovska, Ali, and Qureshi 2017), Multisector Needs Assessment (MSNA) tool (REACH 2023), which aim to set standards for disaster experts who are assessing the needs of affected populations during a disaster. Since this requires extensive human involvement on the ground, getting access to rapid needs information during the time of a disaster becomes challenging. This results in crisis responders being unable to provide assistance to the impacted people in a timely manner.

By leveraging AI technology and integrating multiple nontraditional data sources, the Flood Insights system is the first deployable solution, to the best of our knowledge, that overcomes the limitations of existing rapid flood, damage and needs assessment tools by utilizing off-the-shelf models (Akhtar et al. 2023). The Flood Insights system offers a two-fold advantage. First, the system provides an in-depth view of flood extent in both urban and rural regions to help crisis responders prioritize their response efforts and target their resources where they are needed most. Second, the system incorporates machine learning algorithms and artificial intelligence to analyze social media text and image reports about all types of damages and needs information in real-time allowing for an up-to-date understanding of the situation on the ground. Overall, the Flood Insights system provides a valuable tool for disaster responders, allowing them to make more informed and effective decisions.

SYSTEM ARCHITECTURE

To perform flood exposure, damage, and population needs mapping, the Flood Insights system relies on four types of data including satellite imagery, geospatial data, social media text, and images. The system architecture, drawn in Figure 2, consists of three data processing pipelines. The first pipeline manages the satellite and geospatial data and includes modules for downloading, processing, and analyzing the data. The social media text pipeline (depicted in the middle) performs the processing of the textual data using various natural language processing components. Finally, the third pipeline deals with social media image processing and consists of components that use computer vision techniques. The Flood Insights system code is publicly available on GitHub.[2]

Satellite and geospatial data processing

To enable satellite imagery acquisition, we developed a module utilizing Google Earth Engine's (GEE) API. The module supports three types of inputs: (i) name selection of the area of interest (AOI), which queries Global Administrative Areas (GADM) database[3]; (ii) shapefile upload with one or more AOIs; (iii) custom-drawn AOI by the end-user. The module simplifies the AOI, divides it into multiple tiles, and employs a multithreaded approach for individual tile downloading, thus bypassing the 32MB download limit of the GEE API. Additionally, the corresponding permanent water from the European Commission's Joint Research Centre (JRC) global surface water (Pekel et al. 2016) is also downloaded for the same tile dimension. In particular, six layers are downloaded, including occurrence (intra- and interannual variations of surface water), change (changes in water occurrence), transitions (type of transition between the first and last year), seasonality (months with water in a year), recurrence (frequency with which water returns from year to year), and extent (water was detected or not).

Water segmentor

For water segmentation (for both permanent and flood water), a state-of-the-art flood segmentation model is employed based on U-Net architecture (Ronneberger, Fischer, and Brox 2015). The model is trained on stacked dual-band SAR images (VV and VH) using the Cloud to Street Microsoft Flood Dataset (Cloud to Street - Microsoft 2022), NASA Digital Elevation Model's single-band elevation data, and the six JRC permanent water layers. The output is a binary mask of water and land pixels. During inference, with a batch size of 1, the model consumes 4.6GB memory of a single Nvidia P100 GPU and can process 1.6 tiles/s, where each tile represents $512 \times 512 \times 9$ pixels.

Flood water extractor

The flood segmentation model predicts all water, including permanent and flood water. By subtracting the JRC seasonality layer, representing permanent water, from all water predictions, we obtain flood-only water regions.

Urban and rural boundary extractors

We use harmonized nighttime lights (2019) (Zhao et al. 2022) to identify urban regions based on the difference in light intensity. The nighttime light data are clipped to the user's region of interest and converted into a vector geospatial format to represent the urban boundaries. To retrieve rural boundaries in a vector format, we subtract the urban boundary from the original region of interest.

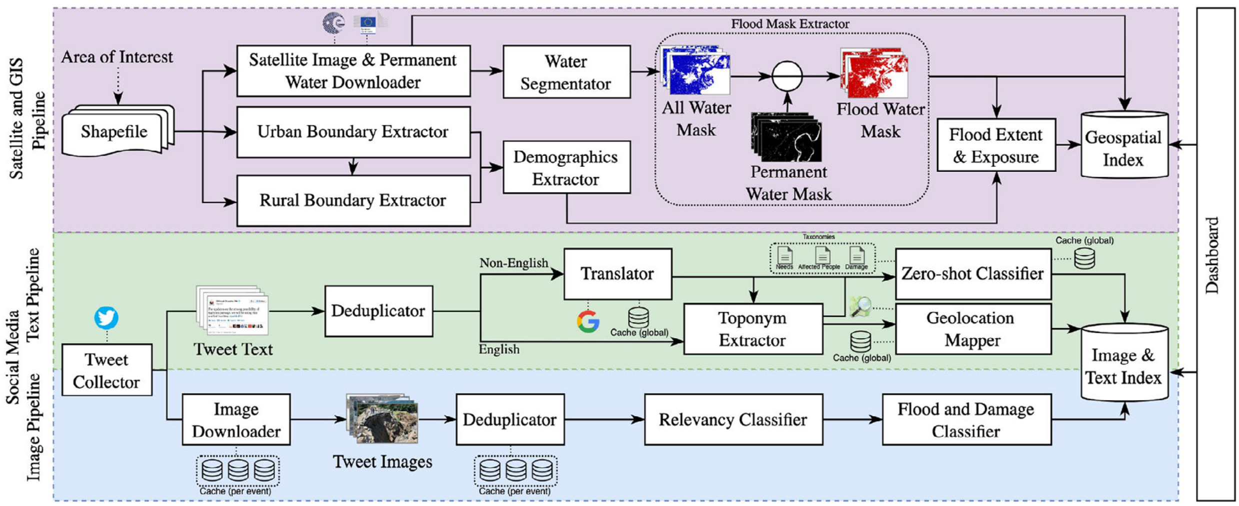

FIGURE 2 Flood Insights system architecture showing three data processing pipelines.

Demographics extractor

All three regions of interest (original, urban and rural) are passed to GEE where the WorldPop Estimated Age and Sex Structures of Residential Population per 100 × 100-m Grid Square dataset[4] is queried separately for population breakdown by gender (male and female) and age group, children (0–14), adolescents (15–19), adults (20–64), and elderly (65+).

Flood extent and exposure

Once each satellite image tile is processed, the flood extent is determined by dividing the total number of flooded pixels by the total number of tile pixels. This calculation is performed separately for the three regions of interest (original, urban and rural). This flood extent information, when multiplied with population numbers from WorldPop, reveals information about the different populations exposed to flooding.

Social media text processing

Tweet collector

The system uses the X (formerly Twitter) Streaming API to collect tweets matching user-specified keywords, hashtags, or geographical areas in JSON format.

Deduplicator

Duplicate tweet text, usually due to retweets, causes excessive computational load. Modules that rely on textual features, such as text classifiers, can be prevented from processing duplicates. The deduplication module preprocesses tweet text (removes emojis, URLs, emails, new lines, and special characters) and uses a cache (ElasticSearch) to determine duplicates. Unique tweets are indexed in the cache specific to the crisis event.

Translation

This module translates non-English tweet text into English using the Google Translate API. The Translator module feeds two modules in the text processing pipeline, namely the Toponym extractor and Zero-shot classifier, which require English data. The module uses a cache to avoid unnecessary translations of identical tweet text.

Toponym extractor

We aim to geolocate tweets at various geographical levels, including subnational, district, and city levels, to facilitate aggregations at different geographical granularities. Since the ratio of geotagged tweets (with geo-coordinates) is very low (<1%), we rely on toponym mentions in the tweet text. These toponyms represent different types of location names such as names of cities, point-of-interest, or districts. To this end, we employ a state-of-the-art BERT-based model fine-tuned on disaster tweets for extracting toponyms from tweet text (Suwaileh et al. 2022).

Geolocation mapper

The location tokens identified by the Toponym extractor module are used to find georeferencing information (i.e., complete address with latitude and longitude) from OpenStreetMap (OSM). Additionally, tagged places (either

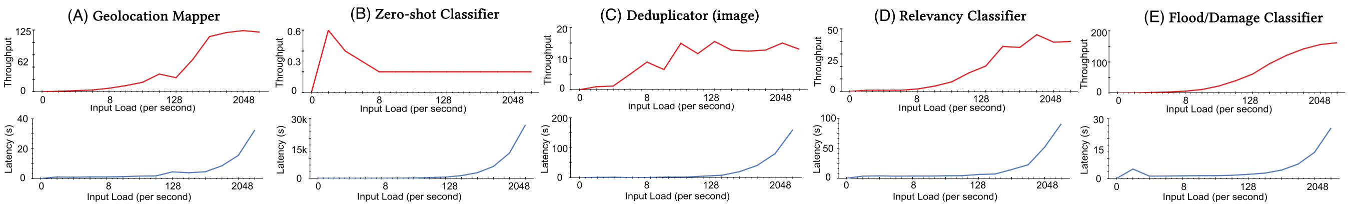

FIGURE 3 Throughput (top) and latency (bottom) of different components.

bounding box or point) are used to make a reverse geocall to OSM to retrieve georeferencing information. Tweets with multiple distinct location mentions are mapped to all mentioned locations. Figure 3A shows the performance of this module as a throughput of 40 items/s.

Zero-shot classification

To categorize tweets into flood impacts, we utilize zero-shot classification by employing a BART-large transformer model (Lewis et al. 2020). We devised a hierarchical taxonomy consisting of three top-level classes: (i) damage reports, (ii) affected people reports, and (iii) urgent needs. These classes are mapped to a total of 38 low-level classes, including five for damage reports (i.e., infrastructure, sewage, utility damage), 11 for affected people (i.e., affected, missing, or complaints, etc.), and 22 for urgent needs (i.e., food, shelter, money, etc.). All 38 classes are then expanded with various prompts to capture semantically analogous expressions. For example, "infrastructure damage" is expressed in 22 different ways (i.e., building damage, building destroyed, house damage, railway track damage). We then use the following template: "The message in this text is related to {prompt}," to classify unseen tweet texts, where "prompt" assumes a specific value associated with the class. We evaluate the model's performance (F1-score at top 1–3 predictions) using HumAID data (Alam et al. 2021). The top 1–2 predictions of the model are on-par with (Alam et al. 2021), with F1-score of 77.3% (ours) versus 78.1% (Alam et al. 2021), whereas top-three yields best F1-score of 84.1% (ours). Figure 3B shows the throughput and latency of this module (5 s/item).

Social media image processing

Image downloader

The Image Downloader extracts image URLs from tweets with attached images. It maintains a collection-based cache to store image URLs and prevent downloading duplicate images by checking for URL existence in the cache. This reduced the noise of duplicate image URLs. This module then passes the image paths and collection code to other modules for processing.

Deduplication

To reduce heavy computational load, this module removes near-or-exact duplicates by comparing the distance between the deep features of two images using a threshold of 7.1. A ResNet-50 model (He et al. 2016) pretrained on the Places dataset (Zhou et al. 2017) is used to extract image features[5], resulting in an average throughput of 13 items/s. The module processes approximately 100 items in 8 s on average, as depicted in Figure 3C.

Relevancy classification

Social media images, even during a disaster event, are usually noisy. The relevancy module is responsible for identifying and removing noisy images, such as those showing ads, cartoons, banners, and so forth. We employ a binary classification model (ResNet-50) pretrained on ImageNet (Russakovsky et al. 2015) and fine-tuned on a custom dataset (Nguyen et al. 2017). This module can process 100 images in 4 s (Figure 3D).

Flood and damage classification

This module processes a filtered stream of potentially relevant images. The images are classified into two broad categories, namely *flood* and *damage*, using the Incidents model (Weber et al. 2020). The flood category includes four labels (i.e.,'flooded,' 'heavy rainfall'), while the damage category includes seven labels (i.e., 'damaged,' 'collapsed'). As seen in Figure 3E, the module can classify 100 images in 2 s, with a throughput of 50 items/s.

SYSTEM DEPLOYMENT AND USE

The Flood Insights system was deployed during the recent Pakistan flooding in 2022. We activated the system on

August 25th, when the Pakistani government declared a country-wide emergency. The system ran for 1 month, from 25 August to 25 September 2022.

Data

Satellite and geospatial

The system was activated to download satellite imagery (Sentinel-1) for all districts in Pakistan ($N = 160$). The global urban extent (from 2019) was queried where urban boundaries were downloaded for 92 districts. The remaining 68 districts had no urban settlement and were considered to be entirely rural. The 92 districts with urban boundaries mainly belong to Punjab, Sindh, and Khyber Pakhtunkhwa with 35, 28, and 18 districts, respectively. With respect to rural boundaries, 159 districts contained rural settlements except for Central Karachi.

Social media

Around 160 event-specific keywords in both Urdu and English languages were used to collect relevant tweets. Over the 1-month period, 9.4 million tweets posted by more than one million users were collected. Around 65% of the tweets were in Urdu and 32% in English. A tweet can be (i) original, (ii) retweet, (iii) reply, or (iv) quoted. We consider replies and quoted tweets as original tweets since they offer additional content as opposed to retweets. Based on this, the overall data contain 1.15 million (12%) original (original + replies + quotes) tweets, which were considered for the Flood Insights system, whereas the remaining 8.25 million (88%) retweets were discarded. The system also downloaded images associated with the original tweets and obtained 411k images in total.

Satellite and geospatial insights

Estimated flood extent

Over the 1-month duration, around 11.48% of the entire country was estimated to be flooded. Figure 4 shows the geographic distribution of flooding across Pakistan for the entire 1-month duration, which is also visualized on the Flood Insights dashboard as an interactive map that supports a zooming functionality. This feature is intended to answer the first question: *Which regions (province or district level) are flooded?* It can be seen that Sindh and Balochistan provinces are impacted the most. This is validated through the estimated flood extent percentages as seen in Table 1, where 25.5% of the entire Sindh is under floodwaters, whereas 12.4% of Balochistan is inundated. Within both these provinces, and across all provinces, rural regions have been impacted more severely as compared to urban regions. The information from Table 1 is visualized on the Flood Insights dashboard as an interactive gauge that dynamically updates based on the region type, period, and province/district selection. This will allow crisis responders to answer the second question: *What is the extent of flooding in urban and rural regions?* Moreover, the dashboard also indicates the top-five districts with the highest estimated flood extent percentage across the entire region, which dynamically updates based on the period and province selection. This is beneficial for local disaster management authorities to prioritize hard-hit districts.

Vulnerable populations exposed to flooding

Our system shows that approximately 18.95 million people were directly exposed to the 2022 floods in Pakistan, which answers the third question: *How many people are exposed to flooding?* In terms of the gender distribution of the exposed population, 9.3 million females and 9.7 million males are at risk due to flooding. The age group distribution reveals that adults and children are highly vulnerable, where 9.6 million adults and 6.8 million children are at risk. The remaining 2.55 million constitute of elderly and adolescents. Both the gender and age group breakdown for the population exposed to flooding is visualized on the dashboard as seen in Figure 1, which answers the fourth question: *Which age groups and genders are at higher risk?* This feature is critical to the District Disaster Management Authority (DDMA) of Pakistan, where they can identify vulnerable age groups exposed within each district. For example, in Larkana district, which has the highest number of people exposed in total (1.26 million) and a flood extent of 75%, the number of children and elderly exposed together make up half of the population exposure (0.53 million), which is important information for DDMA for the prioritization of response efforts. Likewise, the proportion of male and female populations exposed is also displayed on the dashboard as a pie chart, where this information can help DDMA with prioritizing the provision of aid and resources for men/women, who may have specific needs and vulnerabilities.

Social sensing insights

Types of impact incurred through images

The system identified images showing flood and damage scenes to help answer the fifth question: *What type of*

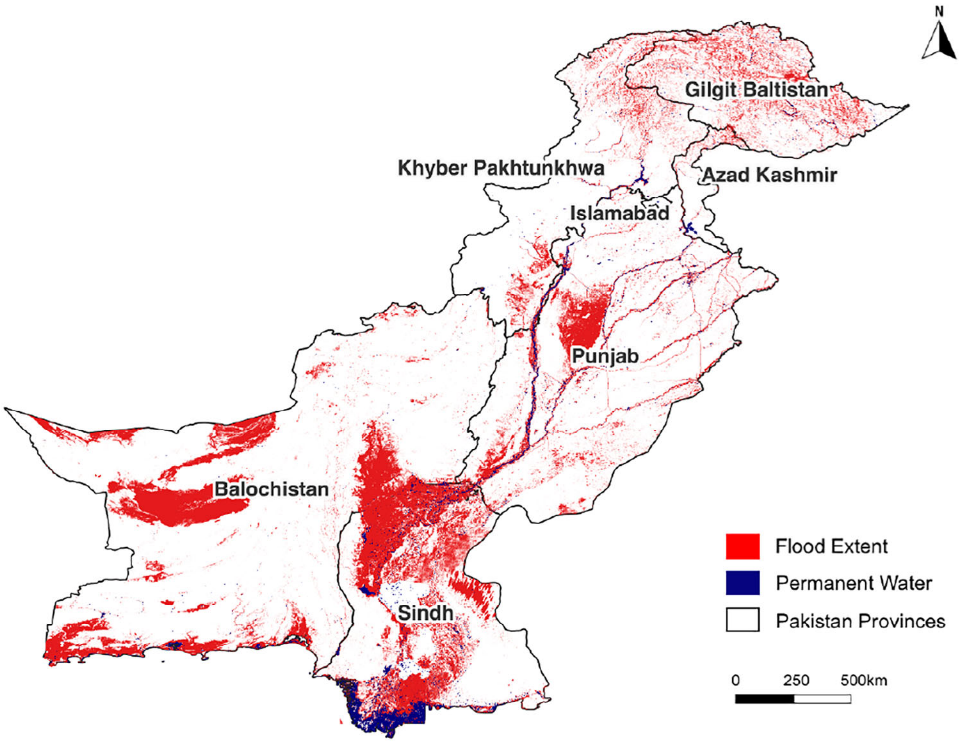

FIGURE 4 Country-level flood extent over the 1-month period.

TABLE 1 Estimated flood extent percentage per province and per region type.

	All region			Urban region			Rural Region		
	Area (km^2)	Flood extent		Area (km^2)	Flood extent		Area (km^2)	Flood extent	
Azad Kashmir	11,422	242	(2.1%)	49	0.1	(0.3%)	11,373	242	(2.1%)
Balochistan	346,724	43,104	(12.4%)	177	2.4	(1.4%)	346,547	43,102	(12.4%)
Gilgit Baltistan	69,760	5424	(7.8%)	0	0	(0%)	69,760	5424	(7.8%)
Islamabad	902	3	(0.3%)	174	0.2	(0.1%)	728	2	(0.3%)
Khyber Pakhtunkhwa	100,897	3760	(3.7%)	391	0.5	(0.1%)	100,506	3760	(3.7%)
Punjab	205,401	12,031	(5.9%)	2352	3.6	(0.2%)	203,049	12,027	(5.9%)
Sindh	141,028	36,029	(25.5%)	938	63.9	(6.8%)	140,091	35,965	(25.7%)

damage is caused by floods? In total, 2000 flood images and 735 damage images are observed across all districts. Such reports contain contextual information about the flood situation, which can help disaster responders identify different types of damages, including flooded scenes, infrastructure damage (e.g., damaged buildings, roads, bridges), agricultural damage, and transportation disruptions. A qualitative analysis of flood and damage images across districts reveals the devastation of floods in the country, as illustrated in Figure 5. The top two rows show the intensive flooding across various parts of the country, including densely populated areas and critical infrastructures such as streets, communities, railways, and bridges. The bottom two rows show images of various damages during and post-flooding, such as ruptured roads, power failures, and destroyed houses. From the 1-month duration, South Karachi district received the highest number of flood reports (194 reports), whereas Quetta topped the list for the highest damage reports (82 reports).

Top needs of the affected population

Over the 1-month duration and across all 160 districts, shelter requests were the highest with 20k reports followed by donation and rescue requests, with 19.7k and 19.1k requests, as seen in Figure 1. The top-five needs are visualized on the dashboard in the form of a bar graph, which dynamically updates based on province, district, and time period. This allows for the last question to be answered:

FIGURE 5 Images showing flood and damage impact in different scene contexts.

What are the urgent needs of the affected people? With this information, the National Disaster Management Authority (NDMA) can know which province/districts have the highest shelter reports by referring to the "Shelter" choropleth map in the center of the dashboard. This map shows the percentage of shelter reports normalized by total reports from each district, where Panjgur and Bajaur districts had all shelter reports. Another way to utilize these maps is to filter the dashboard to the top-five affected districts where all need types can be investigated as well as the daily distribution of total need reports.

Timeline distribution of reports from the ground

The dashboard visualizes the daily trends of tweet text and image reports. For instance, as the number of affected individual reports increases, the number of damage and urgent needs reports also increases, particularly during the peak of the disaster on August 27th to August 28th. With regards to images, between August 25th and September 2nd, the number of flood images is higher than damage images, but then they merge on September 3rd. This real-time information can help crisis responders identify trends and take immediate action.

EVALUATION AND PAYOFF

To validate the reliability of our system, we evaluate it using ground-truth data related to human impact (i.e., deaths and injured people reports) and infrastructure impact (i.e., houses damaged/destroyed) from Pakistan's official government assessment of the flood disaster.[6] This official ground-truth data are available for 53 districts out of 160 districts for the same 1-month duration in which our system was deployed. Of these 53 districts, six belonged to the Punjab province, 22 from Sindh, and the remaining 25 were from Khyber Pakhtunkhwa. To effectively compare our system's output with the official data, we compute three metrics that encapsulate ground-truth, remote, and social sensing data. These metrics include scores for different impact types, such as human impacts (e.g., deaths and injuries) and infrastructure impacts (e.g., damage to houses). Next, we outline how these metrics are calculated.

Ground truth metric

To develop a ground truth metric, we first compute two types of scores using the official data at the district level and then combine them to get the total score as follows:

1. **Ground truth human impact score (gH_i)**: This score aggregates the number of deaths and injuries to quantify the human toll for each district, using the equation:

$$gH_i = \frac{d_i + j_i}{|P_i|},$$

where d_i represents the total number of deaths in the ith district, j_i indicates the count of injured people in the district, P_i signifies the total population of the district under consideration.

2. **Ground truth infrastructure impact score (gI_i)**: This score aggregates the number of houses damaged and destroyed to quantify the impact on infrastructure for each district. The equation is defined as follows:

$$gI_i = \frac{Hm_i + Hd_i}{|B_i|},$$

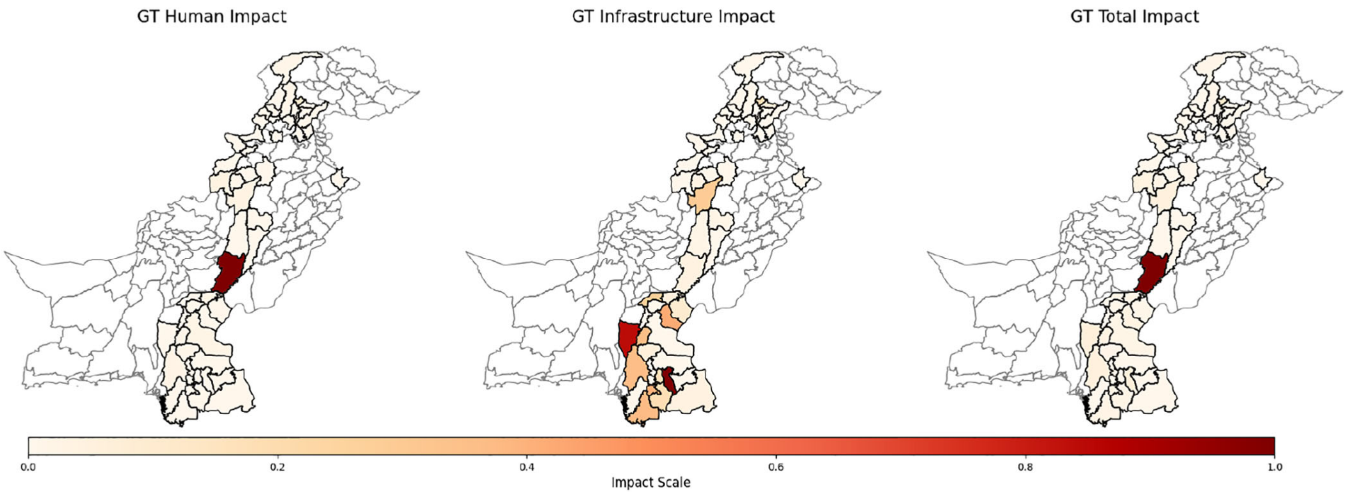

FIGURE 6 Geographic distribution of ground truth human (left), infrastructure (center), and total (right) impact scores.

where Hm_i corresponds to the number of partially damaged houses whereas Hd_i corresponds to the fully destroyed houses in the ith district, and B_i represents the total buildings within the district. We used Microsoft Building Footprints[7] to extract the total number of buildings in each district.

3. **Ground truth total impact score (g_i)**: The individual ground truth human impact score (gH_i) and ground truth infrastructure impact score (gI_i) are then summed up to obtain the ground truth total impact score (g_i) for each district using the following equation:

$$g_i = gH_i + gI_i, 1 \leq i \leq 53.$$

We then apply min–max normalization to standardize the scales of all three scores separately. This normalization technique serves to harmonize the ranges of all the ground truth impact scores to facilitate comparisons, as visually represented in Figure 6.

Having established the benchmark ground truth impact scores, we then collect data from the same 53 districts over the 1-month period (August 25th to September 25, 2022) and derive relevant metrics for two distinct domains: remote sensing and social sensing.

Remote sensing metric

We focus on three key metrics to assess flood extent:

1. **Total flood extent percentage (tF_i)**: This metric quantifies the overall flood impact by calculating the ratio of the total flooded area (in km^2) to the total area of the district (in km^2).
2. **Urban flood extent percentage (uF_i)**: This metric specifically targets urban areas, calculating the ratio of the urban flood area (in km^2) to the total urban area (in km^2).
3. **Rural flood extent percentage (rF_i)**: Similarly, this metric focuses on rural regions, determining the ratio of the rural flood area (in km^2) to the total rural area (in km^2).

The distribution of these three flood extent metrics is graphically depicted in Figure 7.

Social sensing metric

We then formulate two distinct social sensing impact scores and combine them to get the total social sensing impact score as follows:

1. **Social sensing human impact score (sH_i)**: This score is derived from textual messages that report on deaths, injuries, and other forms of human suffering using the following equation:

$$sH_i = \frac{a_i * (dt_i/pt_i)}{|P_i * C|},$$

where a_i represents the total number of affected people reports from text messages in the ith district, dt_i represents the total number of tweets posted from the district, pt_i represents the total number of tweets posted from the province of the district, P_i represents the total population of the district, and $C = 10k$ is the constant used normalize the human toll per 10k people in each district.

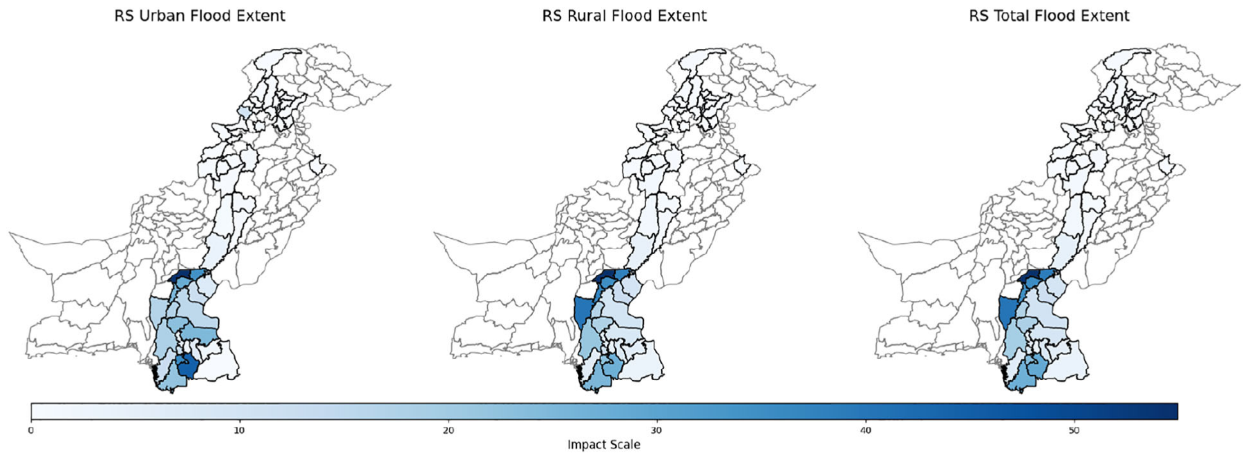

FIGURE 7 Geographic distribution of urban (left), rural (center), and total (right) flood extent percentage.

2. **Social sensing infrastructure impact score (sI_i)**: This score is based on textual messages and images that depict various types of infrastructure damage, such as roads, bridges, and buildings. The following equation is used:

$$sI_i = \frac{(Dt_i + Dm_i) * (dt_i/pt_i)}{|P_i * C|},$$

where Dt_i corresponds to the damage reports collected using tweet text in the ith district, Dm_i represents the damage reports from images in the district, dt_i represents the total number of tweets posted from the district, pt_i represents the total number of tweets posted from the province of the district, and P_i represents the total population of the district, and $C = 10k$ is the constant used to normalize the infrastructure damage per 10k people in each district.
3. **Social sensing total impact score (s_i)**: These individual social sensing human impact score (sH_i) and social sensing infrastructure impact score (sI_i) are then combined for each district to compute the social sensing total impact score (s_i) for each district based on the following equation:

$$s_i = sH_i + sI_i, 1 \le i \le 53.$$

Similar to the ground truth metric, we utilize min–max normalization to bring all three social sensing impact scores onto a standardized scale separately. This normalization technique equalizes the ranges of these diverse impact scores to facilitate comparisons, as illustrated in Figure 8.

Validation

We aim to answer two key questions through a comparative evaluation:

1. **Individual domain effectiveness**: To what extent does each individual domain (i.e., remote sensing and social sensing) offer reliable indicators of disaster impact when measured against ground truth data?
2. **Collective domain effectiveness**: Does the integration of data from these two domains yield more robust and comprehensive indicators of disaster impact compared to individual domains?

By addressing these questions, we seek to explain the potential advantages and limitations of employing nontraditional data sources for disaster impact assessment.

Individual domain effectiveness

To assess the relationships between each domain's metrics and the corresponding ground truth impact score, we used Spearman's correlation test. This method is suitable for non-normal and nonlinear disaster impact data and is less influenced by outliers.

The remote sensing domain is investigated first, where all three metrics (tF_i, uF_i, rF_i) are compared against the individual and combined ground truth metrics (gH_i, gIi, g_i). The results in Table 2 indicate that remote sensing analysis shows significant positive correlations between all flood extents (tF_i, rF_i, uF_i) and ground truth total impact score (g_i) and ground truth infrastructure impact score (gI_i), with the strongest correlations for total and rural flood extents (tF_i, rF_i). Whereas, correlations with ground truth human impact score (gH_i) are not statistically significant, due to the inherent limitations of satellite imagery in capturing direct human consequences such as injuries or fatalities.

To overcome the weaknesses of remote sensing in assessing human impact, we complement it with social sensing, which can provide granular, on-the-ground information.

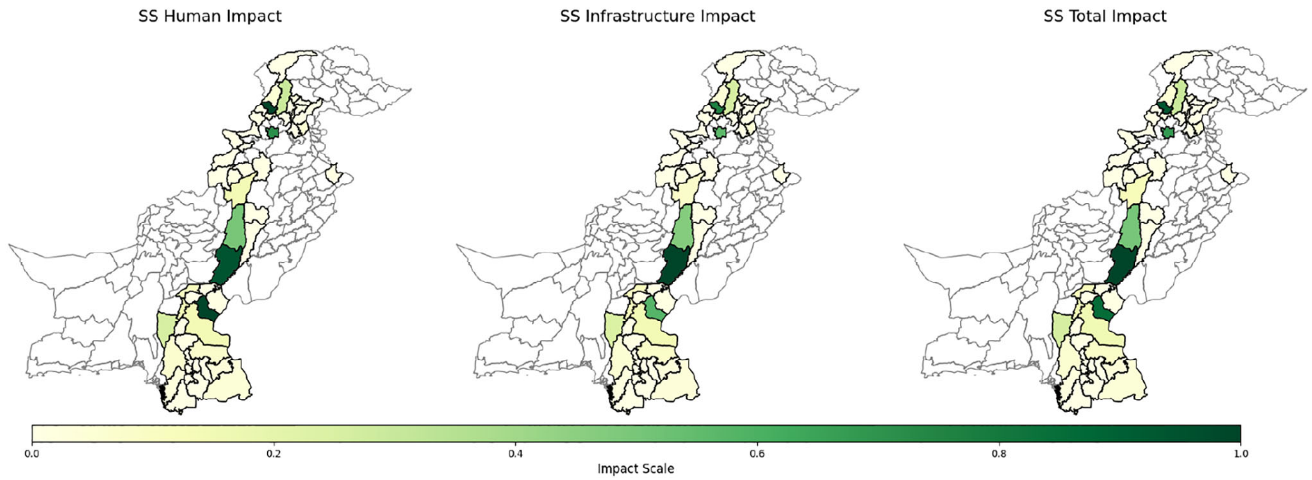

FIGURE 8 Geographic distribution of social sensing human (left), infrastructure (center), and total (right) impact scores.

TABLE 2 Spearman's correlation between remote sensing data (flood extent) and ground truth metrics (53 districts). IV and DV indicate independent and dependent variables, respectively.

IV	DV	Correlation	*p*-value
tF_i	g_i	0.4523	**0.0007**
rF_i		0.4448	**0.0008**
uF_i		0.2760	**0.0454**
tF_i	gI_i	0.4352	**0.0011**
rF_i		0.4280	**0.0014**
uF_i		0.2842	**0.0392**
tF_i	gH_i	0.1368	0.3287
rF_i		0.1319	0.3464
uF_i		0.0085	0.9517

Bold values are statistically significant.

The individual and combined social sensing metrics (sH_i, sI_i, s_i) are correlated with the individual and combined ground truth metrics (gH_i, gIi, g_i). The analysis showed weak positive correlations across all combinations, with *p*-values exceeding 0.05. This prompted us to dig deeper by taking into account the geographic nuances that could influence social sensing data such as variations in internet penetration across geographies. To address this geographic variability, we performed geographically weighted regression (GWR) analysis. Remarkably, the GWR model demonstrated a significant improvement in R^2 values when compared to the global model (ordinary least squares regression) for each variable. This indicates that the relationship between social sensing metrics and ground truth impact scores varies spatially. In particular, the biggest improvement is seen for social sensing infrastructure impact score (sI_i) against ground truth total impact score (g_i), which achieves the highest GWR R^2 of 0.564.

To answer the first question, we find that remote sensing data, particularly total and rural flood extent, as well as social sensing data, when adjusted for geographic variability, can offer valuable insights into ground truth total and infrastructural disaster impacts.

Collective domain effectiveness

Both remote sensing total flood extent percentage (tF_i) and social sensing total impact score (s_i) metrics are then combined to investigate their correlation with ground truth total impact score metric (g_i). We hypothesize that districts with high remote sensing and social sensing will demonstrate a stronger correlation with ground truth impact data. To test this hypothesis, we segmented the dataset into four buckets, each representing a unique combination of remote and social sensing signals:

1. Low remote sensing + Low social sensing (17 districts)
2. Low remote sensing + High social sensing (10 districts)
3. High remote sensing + Low social sensing (10 districts)
4. High remote sensing + High social sensing (16 districts)

To ensure a balanced distribution of districts across these categories, we employed the median value as the differentiating threshold for both metrics. Utilizing Spearman's correlation test, we found a statistically significant positive correlation of $r = 0.5578$ (with a *p*-value of 0.0248, $p < 0.05$) for the fourth bucket encompassing districts with high levels of both remote sensing and social sensing signals, whilst the remaining three buckets were found to be statistically insignificant. This supports the notion that combining data sources provides more accurate insights.

Next, we employed GWR to account for spatial heterogeneity. The correlation analysis using remote sensing impact scores (independent variable) with ground truth impact scores (dependent variable) yields a weak correlation of $R^2 = 0.3483$. Interestingly, using social sensing

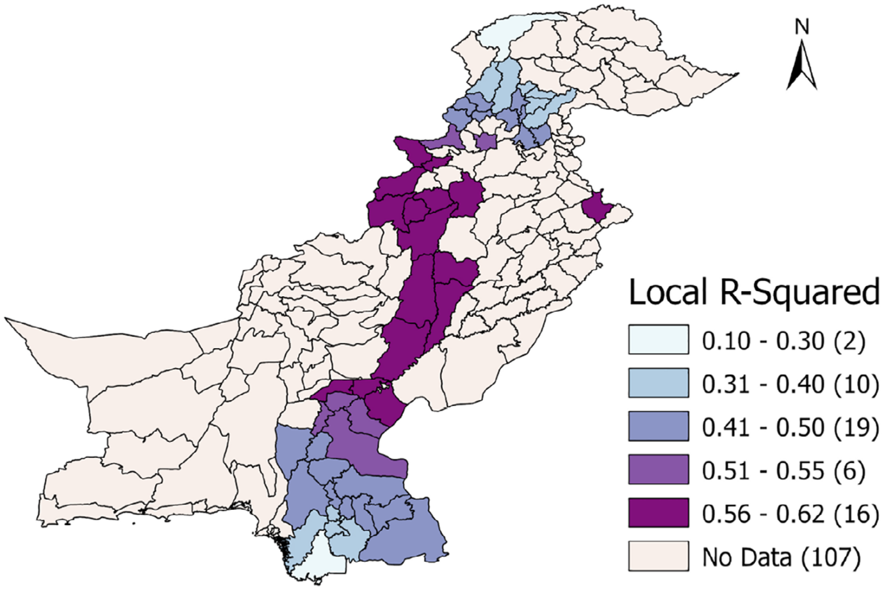

FIGURE 9 Geographically weighted regression modeling of the relationship between the combined remote and social sensing data (tF_i and s_i) versus the ground-truth impact data (g_i) across 53 districts in Pakistan. Numbers in brackets in the legend represent the number of districts in each bucket.

data alone yields a stronger correlation at $R^2 = 0.7744$. The most significant results occurred when combining remote sensing and social sensing, that is, $R^2 = 0.7824$, indicating that information obtained from these data sources (independent variables) has a reasonably good explanatory power and explain a significant portion of the variation in the dependent variable (ground truth total impact score). Next, we used the same combination to run the GWR analysis for individual districts to highlight the distribution of R^2 across all 53 districts as depicted in Figure 9. We observed that most districts ($N = 16$) represent a good fit, that is, $0.56 \leq R^2 \leq 0.62$. Around 25 districts show moderate fit with $0.41 \leq R^2 \leq 0.55$. Whereas only 12 districts fall below $R^2 \leq 0.40$, indicating a weak correlation. Upon further investigation, we observed that of all districts having $R^2 \leq 0.50$, 12 are entirely rural and 17 districts with ≥75% rural population, indicating low internet and social media usage.

To answer the second question, our findings show that districts with high levels of both remote sensing and social sensing data sources have the strongest correlations with ground truth impact data, indicating that a combined approach captures a broader and more accurate picture of disaster effects as compared to using each domain individually.

Qualitative evaluation

As our deployment was activated during the actual flooding event, the data produced by the system were shared with official humanitarian organizations, including UNDP and other local government agencies. The authorities found the system useful for identifying hard-hit districts in real-time. In particular, the stakeholders mentioned that this allows better resource allocation and distribution of funds under time-critical situations.

LESSONS LEARNED

During the development, deployment, and maintenance of the Flood Insights system, there are several lessons learned worth sharing. First, when developing the zero-shot classifier, we followed an iterative approach of honing and expanding prompts within each class of the taxonomy. We found that clear prompts with minimal class overlap are crucial to prevent model confusion. Second, during deployment, we realized that the Flood Insights dashboard is influenced by the chosen AOI during activation. For instance, a province-level boundary restricts insights to that level or broader (i.e., country). Shifting to finer views (i.e., district level) requires time-consuming recomputation of all the satellite and GIS modules. Hence, meticulous granularity selection before activation is advised. Lastly, to maintain the Flood Insights system, three team members monitored the status of all the modules in terms of storage capacity, error handling, and other operational aspects. This monitoring was done on a daily basis to ensure the efficient operation of the Flood Insights system.

CONCLUSIONS

To address the critical information needs of disaster responders, we presented Flood Insights, an on-the-fly system that leverages multisource data and employs state-of-the-art natural language processing and computer vision techniques to surface situational information about flood extent, incurred damages, and population needs. The system relies on remote sensing, social sensing, and geospatial data to provide disaggregated insights at national and subnational levels. We deployed the Flood Insights system during a real-world disaster, the 2022 Pakistan floods, for 1 month and later evaluated the data using official ground truth. Our findings revealed that remote sensing data effectively captures infrastructural disaster impacts, while social sensing data provides valuable insights into human impacts. More importantly, when both domains are combined, it aligns more strongly with the ground truth disaster impacts. The dashboard was conveyed to the UNDP stationed in Pakistan, as well as local authorities, who considered it to be valuable information for enhancing disaster response strategy.

ACKNOWLEDGMENTS

This publication was made possible by BFC Grant #BFC03-0630-190011 from the Qatar National Research Fund (a member of Qatar Foundation). The findings herein reflect the work, and 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 United Kingdom (UKRI), and the United States (NSF).

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict.

ORCID

Zainab Akhtar https://orcid.org/0000-0003-1850-3688
Umair Qazi https://orcid.org/0000-0002-2448-9694
Rizwan Sadiq https://orcid.org/0000-0003-1395-6290
Ferda Ofli https://orcid.org/0000-0003-3918-3230
Muhammad Imran https://orcid.org/0000-0001-7882-5502

ENDNOTES

[1] https://flood-insights.qcri.org/pakistan_floods_2022
[2] https://github.com/CrisisComputing/flood_insights
[3] https://gadm.org/data.html
[4] www.worldpop.org
[5] http://places2.csail.mit.edu/models_places365/resnet50_places365.pth.tar
[6] https://www.pdma.gov.pk/
[7] https://github.com/microsoft/GlobalMLBuildingFootprints

REFERENCES

Akhtar, Z., U. Qazi, R. Sadiq, A. El-Sakka, M. Sajjad, F. Ofli, and M. Imran. 2023. "Mapping Flood Exposure, Damage, and Population Needs Using Remote and Social Sensing: A Case Study of 2022 Pakistan Floods." In *Proceedings of the ACM Web Conference 2023, WWW '23*, 4120–28. NY, USA.

Alam, F., U. Qazi, M. Imran, and F. Ofli. 2021. "HumAID: Human-Annotated Disaster Incidents Data from Twitter with Deep Learning Benchmarks." In *ICWSM*, 933–42. AAAI.

Centre for Peace and Democracy. 2018. "Galgaduud Flooding Assessment Report." https://www.humanitarianresponse.info/fr/operations/somalia/assessment/galgaduud-flooding-assessment-report (accessed November 29, 2023).

Cloud to Street - Microsoft. 2022. "A Global Flood Events and Cloud Cover Dataset (Version 1.0)." https://registry.opendata.aws/c2smsfloods/ (accessed November 29, 2023).

Department of Civil Protection, UN-Agencies, NGOs. 2019. "Inter-Agency Flooding Rapid Assessment Report." https://assessments.hpc.tools/assessment/7aeebd8c-f7f8-4ab0-b767-c7b16a34c410 (accessed November 29, 2023).

FAO Representation in the Sudan. 2020. "The Sudan - 2020 Flood Response Overview." https://www.fao.org/3/cb1386en/cb1386en.pdf (accessed November 29, 2023).

Government of Somalia. 2020. "Joint Rapid Needs Assessment Report on Flood Displaced Population in Afgoye and Surrounding Villages." http://bit.ly/3N7cWme (accessed November 29, 2023).

He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep Residual Learning for Image Recognition." In *CVPR*, 770–78. IEEE.

Kenya National Disaster Operations Centre, UNDP. 2013. "Initial Rapid Assessment Report for Flooding in Narok County." https://www.humanitarianresponse.info/en/operations/kenya/assessment/initial-rapid-assessment-report-flooding-narok-county (accessed November 29, 2023).

Lewis, M., Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer. 2020. "BART: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7871–80, Online. Association for Computational Linguistics.

Macabuag, J., C. Altheim, S. Thorvaldsdottir, and D. Perks. 2022. "Damage Assessments by International Engineers Following the Albania Earthquake of November 2019." *International Journal of Disaster Risk Reduction* 72: 102822.

Malawi Government, The United Nations, The World Bank, GFDRR. 2019. "Malawi 2019 Floods Post-Disaster Needs Assessment Report." https://www.unicef.org/malawi/reports/malawi-2019-floods-post-disaster-needs-assessment-report (accessed November 29, 2023).

Milly, P. C. D., R. T. Wetherald, K. Dunne, and T. L. Delworth. 2002. "Increasing Risk of Great Floods in a Changing Climate." *Nature* 415(6871): 514–17.

Ministry of Agriculture and Irrigation; Ministry of Livestock, Fisheries & Rural Development, FAO, WFP. 2015. "Agriculture and Livelihood Flood Impact Assessment in Myanmar." https://www.fao.org/fileadmin/user_upload/emergencies/docs/Final_Impact_Assessment_Report_final.pdf (accessed November 29, 2023).

NDC Solomon Islands, City of Honiara National Capital. 2018. "Dashboard: Initial Damage Assessment-East Honiara Flooding." https://bit.ly/47ScOi9 (accessed November 29, 2023).

Nemni, E., J. Bullock, S. Belabbes, and L. Bromley. 2020. "Fully Convolutional Neural Network for Rapid Flood Segmentation in Synthetic Aperture Radar Imagery." *Remote Sensing* 12(16): 2532.

Nguyen, D. T., F. Alam, F. Ofli, and M. Imran. 2017. "Automatic Image Filtering on Social Networks Using Deep Learning and Perceptual Hashing During Crises." In *ISCRAM*, 499–511. ISCRAM.

Oman, A. 2011. "Completing the Jigsaw Puzzle: Joint Assessment Missions (JAM)." Field Exchange-Emergency Nutrition Network ENN. 49–51. 40.

Pekel, J.-F., A. Cottam, N. Gorelick, and A. S. Belward. 2016. "High-Resolution Mapping of Global Surface Water and Its Long-Term Changes." *Nature* 540(7633): 418–22.

REACH. 2023. "Multi-Sector Needs Assessment (MSNA) - Global Key Findings, January 2023." https://reliefweb.int/report/world/multi-sector-needs-assessment-msna-global-key-findings-january-2023 (accessed October 9, 2024).

Ronneberger, O., P. Fischer, and T. Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–41. Springer.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, et al. 2015. "ImageNet Large Scale Visual Recognition Challenge." *IJCV* 115(3): 211–52.

Shelter Cluster. 2017. "Afghanisatan (North) Flood Response Evaluation Assessment." https://reliefweb.int/sites/reliefweb.int/files/resources/reach_afg_report_shelter_flood_response_april2017_1.pdf (accessed November 29, 2023).

Suwaileh, R., T. Elsayed, M. Imran, and H. Sajjad. 2022. "When a Disaster Happens, We are Ready: Location Mention Recognition from Crisis Tweets." *International Journal of Disaster Risk Reduction* 78: 103107.

Vieweg, S. E. 2012. "Situational Awareness in Mass Emergency: A Behavioral and Linguistic Analysis of Microblogged Communications." Ph.D. thesis, University of Colorado at Boulder.

Weber, E., N. Marzo, D. P. Papadopoulos, A. Biswas, A. Lapedriza, F. Ofli, M. Imran, and A. Torralba. 2020. "Detecting Natural Disasters, Damage, and Incidents in the Wild." In *ECCV*, 331–50. Springer.

Zhao, M., C. Cheng, Y. Zhou, X. Li, S. Shen, and C. Song. 2022. "A Global Dataset of Annual Urban Extents (1992–2020) from Harmonized Nighttime Lights." *Earth System Science Data* 14(2): 517–34.

Zhou, B., A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. 2017. "Places: A 10 Million Image Database for Scene Recognition." *PAMI* 40(6): 1452–64.

Ziolkovska, A., H. Ali, and B. Qureshi. 2017. "Development of Multi-Cluster Rapid and In-Depth Assessment Methodologies in Afghanistan." Field Exchange. 16. 56.

How to cite this article: Akhtar, Z., U. Qazi, A. El-Sakka, R. Sadiq, F. Ofli, and M. Imran. 2024. "Fusing remote and social sensing data for flood impact mapping." *AI Magazine* 1–16. https://doi.org/10.1002/aaai.12196

AUTHOR BIOGRAPHIES

Zainab Akhtar has experience with the fundamentals of GIS where she is able to analyze and visualize data to uncover relationships that are not readily apparent. Her interest centers around using different research methods to effectively tackle humanitarian issues with the help of social media, machine learning algorithms, and GIS software's.

Umair Qazi is an experienced Software Engineer who holds a degree in Information Systems from Carnegie Mellon University and has contributed to multiple disaster management technologies, combining AI with real-time data systems. With experience in full-stack development, Umair has contributed to the design and implementation of platforms across various industries.

Aya El-Sakka has experience in analyzing complex datasets and developing interactive dashboards that support real-time decision-making. She has contributed to the development of the Flood Insights system to ensure the dashboard operates seamlessly in real time.

Rizwan Sadiq is an AI Specialist focused on developing advanced AI models for disaster management. He has contributed to integrating social and remote sensing data to enhance real-time disaster response and crisis prediction, particularly in flood mapping and damage assessment.

Ferda Ofli is a Senior Scientist at the Qatar Computing Research Institute since 2014. He received the B.Sc. degrees both in Electrical and Electronics Engineering and Computer Engineering, and the Ph.D. degree in Electrical Engineering from Koc University, Istanbul, Turkey, in 2005 and 2010, respectively. From 2010 to 2014, he was a Postdoctoral Researcher at the University of California, Berkeley, CA, USA. His research interests cover computer vision and machine learning with a particular focus on applications in the humanitarian domain. He is a Senior Member of IEEE and

ACM and a member of AAAI with over 80 publications in top-tier conferences and journals including CVPR, ECCV, WWW, AAAI, and PAMI. He won the AAAI Deployed Application Award in 2024, the CVPR Outstanding Reviewer Awards in 2020 and 2021, the ISCRAM Best CoRe Paper Award in 2020, the ISCRAM Best Insight Paper and Best Paper Runner-up Awards in 2019, the ISCRAM Best Paper Runner-up Award in 2017, the Elsevier JVCI Best Paper Award in 2015, and the IEEE SIU Best Student Paper Award in 2011. He also received the Graduate Studies Excellence Award in 2010 for his outstanding academic achievement at Koc University.

Muhammad Imran is a Senior Scientist and the Lead of the Crisis Computing Group at the Qatar Computing Research Institute (QCRI). He completed his Ph.D. in Computer Science at the University of Trento in 2013 and subsequently joined QCRI, where he spent 2 years as a Postdoctoral Researcher and 5 years as a Scientist. His research focuses on using artificial intelligence, machine learning, and big data analytics to improve decision-making in areas such as humanitarian response and resilience. Dr. Imran has authored over 130 peer-reviewed publications in top-tier conferences and journals, including WWW, ACL, AAAI, ECCV, and PAMI. He has received five best paper awards, and his work has been featured in outlets such as the BBC, Wall Street Journal, WIRED Magazine, and New Scientist. He frequently delivers keynotes and invited talks at international events and serves as an Associate Editor for the journal Frontiers in Big Data. Additionally, he contributes as an Editor for special issues in leading journals and is actively involved as a track chair and program committee member for various conferences and journals.