# Automatic analysis of social media images to identify disaster type and infer appropriate emergency response

Amna Asif[1*] , Shaheen Khatoon[1], Md Maruf Hasan[1], Majed A. Alshamari[1], Sherif Abdou[4], Khaled Mostafa Elsayed[2] and Mohsen Rashwan[3]

*Correspondence:
aarkhan@kfu.edu.sa
[1] Information Systems
Department, College
of Computer Sciences &
Information Technology,
King Faisal University (KFU),
AlAhsa, Kingdom of Saudi
Arabia
Full list of author information
is available at the end of the
article

## Abstract

Social media postings are increasingly being used in modern days disaster management. Along with the textual information, the contexts and cues inherent in the images posted on social media play an important role in identifying appropriate emergency responses to a particular disaster. In this paper, we proposed a disaster taxonomy of emergency response and used the same taxonomy with an emergency response pipeline together with deep-learning-based image classification and object identification algorithms to automate the emergency response decision-making process. We used the card sorting method to validate the completeness and correctness of the disaster taxonomy. We also used VGG-16 and You Only Look Once (YOLO) algorithms to analyze disaster-related images and identify disaster types and relevant cues (such as objects that appeared in those images). Furthermore, using decision tables and applied analytic hierarchy processes (AHP), we aligned the intermediate outputs to map a disaster-related image into the disaster taxonomy and determine an appropriate type of emergency response for a given disaster. The proposed approach has been validated using Earthquake, Hurricane, and Typhoon as use cases. The results show that 96% of images were categorized correctly on disaster taxonomy using YOLOv4. The accuracy can be further improved using an incremental training approach. Due to the use of cloud-based deep learning algorithms in image analysis, our approach can potentially be useful to real-time crisis management. The algorithms along with the proposed emergency response pipeline can be further enhanced with other spatiotemporal features extracted from multimedia information posted on social media.

**Keywords:** Convolutional neural networks, Image classification, Object detection, Disaster management

## Introduction

Sometimes disaster comes with little or no warning, arising problems rife with questions and uncertainties and requiring us to make quick decisions based on limited information [1]. In the era of ubiquitous networks and social media, sometimes we are overwhelmed with information that we cannot process or filter within a short time. The onset of disaster demands a fast response in ensuring the safety of affected people

Asif *et al. J Big Data*     (2021) 8:83

Page 2 of 28

and protecting vulnerable places and other artifacts. In such a situation, the information coming instantly and continuously from the affected areas helps the humanitarian organizations reach out to the affected people and specific locations with their services. Numerous studies have shown that online information from social media is useful in crisis management and emergency response operations [1, 2]. The EDM (Emergency Decision Making) [3] plays a significant role in natural disasters to help the responsible bodies' humanitarian efforts with the emergency and rescue operations. The EDM, however, requires a timely collection of relevant information. Most previous studies show that the users' textual data on social media [4] helps with emergency response in all phases of disaster management: the calm before the disaster, during the disaster, the peak, the plateau, the decline, and return to normal [5]. Almost all social media platforms support the facility of uploading images and videos from the users' accounts. Most social media users find it easy to capture and post photos of the crisis instead of writing lengthy texts. The social media platforms such as Twitter also provide API to access, retrieve, and download user-uploaded photos along with metadata. Hence, we are encouraged to use social media images to identify disasters and disaster-related cues in the social media posts. The effective extraction and fusion of information embedded in social media images will be helpful in many aspects of crisis management including fast response, rescue and aid operation, as well as generating timely alerts or warnings that may save lives and property.

It is worth mentioning that social media posts are not always of high quality, accurate and timely; however, the previous studies have proved the usefulness of social media in disaster management and response. For instance, Said et al. [6] justified the role of social media in the communication and dissemination of disaster-related news and relevant approaches for information filtering, events detection, and summarization. Murthy et al. [7] have analyzed the Twitter and Flickr platforms' posts with images of affected areas. They reported that during the disaster, the visual cues of most social media posts are related to food and supplies. Peters and Albuguerque [8] have investigated the image-related geographical information to check if the posted images are relevant to a disaster or not. Similarly, Daly & Thom [9] have analyzed the geographic information of social media images related to fire hazards. Nguyen et al. [10] have investigated social media images to determine the extent of damage during natural disasters using machine learning techniques. The previous studies verify that users' social media photos are a useful and rich source of disaster information.

Many studies [11–13] have contributed to disaster image datasets such as floods, earthquakes, and fire, helpful in the classification of disaster types. However, the visual information embedded in the social media images can be made more useful by recognizing the actual emergency cues as humans naturally perceive by looking at a picture. To do so, we need further annotation of such image datasets. In social media crowdsourcing, generally, the information is available in massive amounts; therefore, it is impossible for a person or a team to manually filter loads of social media photos and find meaningful information (i.e., relevant cues) in them quickly. Therefore, automating the recognition of useful disaster-related objects and their contextual information in the social media images such as damage, fatality, causality, affected people, animals, or supplies, etc., help the rescuers, police, humanitarian organization, and emergency operators in

the EDM. However, none of the previous research contains a detailed investigation of the following: (1) recognizing the emergency cues from the images, (2) automating the decision support for locating the appropriate emergency types and responsible authorities for given image-driven cues, and (3) labeling of emergency cues and disaster objects in the image datasets. In this study, we focus on such aspects of social media image analysis.

This paper aims to address the challenge of processing a huge number of social media images according to disaster type and relevant emergency responses. We have proposed a novel approach to automating emergency information recognition from the social media disaster's images. To automate the social media image analysis in recognizing appropriate emergency response, we developed a disaster taxonomy of emergency response. Previously many studies investigated disaster taxonomies [14, 15] from the perspective of disaster's functional areas and theoretical aspects. However, the existing taxonomies are more general to our requirements as we need lower-level details on the disaster response function to develop our automation pipeline. Therefore, it is required to explore what type of visual cues can accelerate the emergency response. We also developed the labeled images dataset of disaster objects and features that will fit the emergency response taxonomy.

To develop the automated social media image sorting pipeline, we first construct the taxonomy of disaster emergency response suitable for social media visual elements by exploring the literature and augmenting it using a card-sorting activity [16]. Furthermore, we also annotate the dataset according to the proposed taxonomy of emergency response. We used our developed custom images dataset containing labels of disaster-related objects to train deep learning-based models for extracting emergency object labels from the image. Image labels' utilization in a balanced way for correct emergency information recognition is challenging; we employed 'decision tables' and 'analytic hierarchical processes' models for information fusion to find the most suitable place for the image in emergency response taxonomy. The evaluation of the proposed image processing pipeline for emergency recognition is performed on the open-source and custom-developed datasets. In the following, the contributions of this work are summarized:

- Constructing the taxonomy of disaster emergency response for social media images consisting of different categories: *'damage', 'rescue, volunteering, and/or donation', 'food and basic need supply', 'affected individuals', 'caution, warning, advice', and 'affect on social activities'.*
- Developing a labeled image disaster's object dataset to detect disaster information for training CNN-based models.
- Proposing an image processing pipeline that automates social media images' filtering by effective use of disaster emergency response taxonomy and consolidating disaster emergency object labels by decision support modeling technique for information fusion to automate the decision to distinguish the image according to the emergency response disaster category.

The rest of this paper is organized as follows: "Related work" Section provides a comprehensive literature review; "The proposed methodology for automating the emergency

response process" Section elaborates the technical details of each module of the proposed methodology. It presents the disaster taxonomy, dataset development, and CNN-based model used for image classification and disaster-related object detection, followed by the information fusion approach used to map the disaster-related object on the proposed taxonomy. "Concept validation through the image processing pipeline" Section presents an end-to-end image processing pipeline used to evaluate the proposed methodology and report results. We closed this paper with a discussion and conclusion in "Discussion and conclusion" Section.

## Related work

This section reports a comprehensive review of the previous literature on social media image data during the disaster. We start with the studies related to machine learning and deep learning neural network models and algorithms of multimodal social media data in the disaster domain, followed by their evaluation studies. Then we discuss the previous contributions in developing disaster-related image datasets. Lastly, we present the existing work in the disaster classification and highlight the limitations of previous work.

### Machine learning models and algorithms for disaster data crowdsourcing

Francalanci et al. [17] proposed a method to extract the geo-located images from tweets to support the emergency response. They have offered the extraction method of geo-located images comprised of tweet selection, tweet cleaning, geo-localization, selection of tweets with media, image analysis, and fine-grained localization. The authors have developed the system caled Image Extraction from tweets (IMEXT) to validate the proposed method. The IMEXT tool resulted in 14.18% of useful images with location information in the available tweets dataset. This study is limited to one disaster type, and the tweets with location and damage-related keywords filtered manually. Murthy et al. [7] conducted a study on geo-located images taken from the Instagram's app users and posted on Twitter accounts during Hurricane Sandy. The images were posted in three phases of the Hurricane Sandy disaster, i.e., pre-US pre-storm, when Sandy made US landfall, and Sandy's aftermath. The results showed that most photos were about supplies during the time Hurricane Sandy made US landfall, and the most common topic was 'Sandy parties'. Aftermath, users started posting images related to the categories of 'damage' and 'outside' such as 'built environment', 'trees coming down', 'flooding', etc. Regarding images related to 'relief', users captured pictures of the Hurricane Sandy telethon by taking a screenshot of the Red Cross donations posted on Instagram. The authors concluded that Twitter users change posting trends quickly according to 'what's happening' prompts with images posted on Instagram showing identifiable tendencies. The study was limited to pre-defined disaster's motif categories for image coding. Peters and Albuguerque [8] have investigated Flickr and Instagram's tweets and messages during intensive flood events in Saxony, Germany, in 2013. The authors developed a methodological workflow—the dataset consisting of a total of 26,713 messages from all three social media. Every message was timestamped, georeferenced, and retrieved using their public API. The Flickr users posted more messages 15% in the proportion of 2% of Twitter users. Additionally, they found out that the share of messages within the hazard areas was 45%, 40%, and 25% in Flicker, Instagram, and Twitter, respectively. Moreover,

the messages posted closer to disaster events' location comprised of more images. This study showed that social media images with relevant contents are helpful in improving information extraction, consequently enhance situational awareness. The study dataset was collected from a specific location, which was related to a single disaster event; therefore, the study results have limited scope.

Daly & Thom [9] selected Flickr as a social media data source to investigate the visual characteristics and the geospatial characteristics distribution of major fire photos and metadata. The'Bag of Features model' is used for image classification. The authors have compared the four different approaches to train the SVM: SIFT, SURF, ColorSURF, and ColorSIFT. ColorSIFT was found to be the best, with 91% recall and 93% precision. The study is limited to the Flickr platform, and the images were just classified as fire and non-fire events. Nguyen et al. [10] investigated images posted on social media during a natural disaster to determine the level of damage. To achieve that, they employed state-of-the-art machine learning techniques. The authors explored the performance of several image classification techniques, i.e., the traditional technique of Bag-of-Visual-Words (BoVW) and CNN for evaluating the level of damage severity of social media images. Different experiment settings were set for evaluating the performance of machine learning classifiers of BoVW model, VGG16 network, and VGG16 transfer learning approach where the last layer of pre-trained VGG16 was initialized using the new dataset of three damage categories.: (1) severe, (2) mild, and (3) little or no damage. The results showed that VGG16 fine-tuned outperformed other techniques. In this study geo-location information was not explored.

Ahmed et al. [18] integrated information which was retrieved from multiple sources; sensed data were retrieved from satellite images, and visual and meta-data were retrieved from social media images. The challenge was composed of two tasks (i) Disaster Image Retrieval from Social Media (DIRSM) and (ii) Flood Detection in Satellite Images (FDSI). The CNN models are AlexNet, GoogleNet, VGGNet-19, and ResNet (with 50, 101, and 152 layers). The average precision of visual-only they have attended between 89 to 95%. In [19], the authors extracted the visual information from social media images (like Twitter and flicker) about natural disasters along with meta-data such as (users' tags, title, description, and geo-location information) and this information was used for the retrieval of disaster images from social media. The best results were obtained on visual information, only around 95.73%. A technique to identify the disaster from many social media multimedia repositories consisted of early fusion and decision fusion levels. In the early fusion level, data was collected from various sources such as social media, wireless sensor networks, IoT (internet of things), unmanned aerial vehicles (UAV). The acquired data was used to detect the crisis and analyzed for event's severity level. The decision fusion level used the rule-based approach to gather the incident report from the previous layer and combine them. They have used fast R-CNN trained on open image dataset for object detection [21]. The study was limited to the open-image dataset.

Dunnings and Breckon [22, 24] applied AlexNet, InceptionV1, VGG13, InceptionV1-OnFire, and Fire CNN to automatically detect fire pixel regions in video and images within real-time bounds without reliance on temporal scene information. The study resulted in maximal accuracy of 0.93 of whole image binary fire detection. The

**Table 1** Algorithms used for disaster classifications

| Reference | Algorithm | Disaster type |
| --- | --- | --- |
| [10] | Bag-of-Visual-Words (BoVW), SVM, VGG16, VGG15fine-tuned, VGG16fine-tuned | Earthquake, Hurricane, Flood |
| [18] | AlexNet (4096), GoogleNet, VGGNet 19 (4096), ResNet (with 50, 101 and 152 layers) | Flood |
| [22] | AlexNet, InceptionV1, VGG13, InceptionV1-OnFire, FireCNN | Fire |
| [19] | SVM, Sequential CNN | Flood |
| [20] | Faster R-CNN trained on COCO dataset | Fire |

**Table 2** Analysis of related work for disaster image classification models and algorithms

| Reference | Contribution | Case Studies |
| --- | --- | --- |
| [17, 23] | Extracting geo-located images from Tweets to support emergency response. Developed a system called IMEXT (Image Extraction from tweets) to extract the geo-located images from tweets for supporting the emergency response | Tweets posted the two days after the earthquake that occurred in central Italy in August 2016 |
| [7] | Study of how people use social media in disasters situations | Hurricane Sandy images by Instagram and Twitter users |
| [23, 24] | Develop a system (Image4Act) that provides end-to-end social media image processing. The system collects, denoising, and classifying imagery content posted on social media platforms to help humanitarian organizations gain situational awareness and launch relief operations | Cyclone Debbie hit Queensland, Australia in March 2017 |
| [8] | The images on the ground and posted on social media can offer more reliable and valuable information for improving situational awareness | Hurricane Sandy |
| [9] | Build a machine learning model to classifying fire and non-fire images | California Rim Fire |
| [10] | Employ machine learning techniques to analyze images posted on social media platforms during natural disasters to determine the level of damage caused by disasters | Typhoon Ruby typhoon/Hagupit, Nepal Earthquake, Ecuador Earthquake, Hurricane Matthew |
| [25] | Presented a large amount of multimodal Twitter dataset related to natural disasters | Seven disaster events in 2017 |
| [26] | Designed and implemented flood images retrieval pipeline that using deep neural network for comparing various features for image retrieval | No case studies |

study was limited to the fire types of the disaster event. Antzoulatos et al. [20, 21] applied the steps of Image classification, Emergency localization, Object detection, and Severity level estimation to present a warning system in crisis emergencies for detecting people and vehicles in danger. For emergency classification, the authors used fine-tune the pre-trained parameters of the VGG−16 on Places365 datasets. The study was limited to the Common Objects in Context (COCO) dataset, which is not specific to the disaster events.

Table 1 presented the existing work on the deep learning neural network for disaster information classification, especially from the social media platform. Table 2 presents a detailed analysis of related work in disaster management using social media

data. The previous literature is mainly focusing on the problem of image classification as a disaster or non-disaster type. However, the images can provide more detailed information about the objects, people, and the context in emergencies.

### Evaluation studies on images data

Alam et al. [13] evaluated Image4Act that provides end-to-end social media image processing. The authors designed different experiments of relevancy filtering, de-duplication filtering, and damage assessment to assess the performance of image filtering and damage assessment modules. The relevancy filter results in performance achieved 0.99 precision and 0.97 recall. The de-duplication filter removes 58%, 50%, and 30% proportion of images from severe, mild, and non-categories, respectively. Therefore, the image filtering pipeline reduces the size of raw data images by 62%, and the remaining dataset was used for further analysis. The damage assessment experiments scored maximum precision and recall of 0.74 and 0.68, respectively. Nguyen et al. [23] have performed the performance evaluation of the proposed social media image processing pipeline consisting of relevancy and de-duplication modules. A pretrained VGG-16 network was initialized using the custom dataset. The evaluation results show 0.98, 0.99, and 0.97 training accuracies, precision, and recall, respectively. The previous studies showed higher accuracy results of disaster type classification. However, the evaluation studies are mainly focused on image classifications but limited in object detection for emergency response information.

### Datasets development

Alam et al. [13] have developed many datasets (text and images) related to natural disasters, including earthquakes, hurricanes, wildfires, and floods, using crowdsource Twitter data. The authors have sampled 18,126 images from the tweets and manually annotated the sampled data according to informative and non-informative categories, critical and actionable information, and damage severity levels using the services of the CrowdFlower platform. The authors found out that tweet images contained more damage-related information than their corresponding text. Barz et al. [26, 29] have developed the dataset of flood images using two data sources Flickr and Wikimedia Commons, to implement the image pipeline using deep neural networks. The final dataset contained 3,435 flood-related images and 275 water pollution-related images.

Table 3 presents the analysis of previous work in developing disaster image datasets. Mainly, images were collected from social media and internet sites and annotated according to disaster types. Currently, the limited image datasets are available in the domain of disaster emergency response. The availability of more annotated datasets in different formats (such as CNN classification and object detection) and various disaster-related information (such as disaster types, tasks, events, and context, etc.) can benefit researchers' community in developing their models and performing the experiments. In this regard, one part of our study proposed a method of image dataset development for disaster emergency response. Furthermore, we have applied the

**Table 3** The related work on disaster dataset development

| Reference | Data source | Process | Domain |
|---|---|---|---|
| [26] | Flicker100K dataset Wikimedia Common Dataset size (3435 flood-related images and 275) | Collected the images Defined the 3 tasks for relevant images Assigned domain expert for images annotation on defined tasks Validated the quality of annotation | Flood |
| [25] | Twitter Dataset size (16,097 images) | The tweets with only images URL collected also more than one images Discard all non-English language tweets. Discarded single word and single hashtag All duplicate tweets are removed The authors have developed annotation guidelines and, provided them to well-known paid platform CrowdFlower The manual annotation of text and images are attained separately | Earthquakes, Hurricane, Wildfires, Floods |
| [9] | Flickr API | The Flickr API (flickr.photos.search) method is used to collect publicly assessable photos within a time period to avoid repetition in training and test data The images are stored with metadata using MySQL database The image which is retrieved by their geo-spatial data using flickr.photos. GetEXIF API method | Fire |
| [7] | Tweets embedded Instagram images 11,964 Instagram images | Searched 142,768 geo-located tweets sample embedded Instagram photos The images were then manually coded to study separate disaster-related motif categories (ad, animals, damage, drink, food, gear, macro, other, outside, people, and relief). The images may fit multiple codes | Hurricane Sandy |

proposed approach in developing our disaster-related image dataset for proof of the concept.

### Disaster classification

The international disaster database EM-DAT [1] provides a comprehensive classification scheme of disasters by dividing them into multiple levels. The top level of the classification consists of natural and technological disasters. And the bottom levels are divided into disaster types such as earthquakes or floods and their further sub-types. This classification is a valuable source of information for the scientific community's research. Olteanu et al. [27] identified the multiple crisis dimensions (such as affected individuals, infrastructure & utilities, donations & volunteer, caution & advice, and sympathy & emotions) from social media text by analyzing the previous literature. CrisisMMD [13, 25] is a multimodal (text and image) dataset consisting of seven natural disaster types, humanitarian categories (such as affected individuals, infrastructure and utility damage, injured or dead people, missing or found people, rescue, volunteering or donation

---

[1] https://www.emdat.be/index.php.
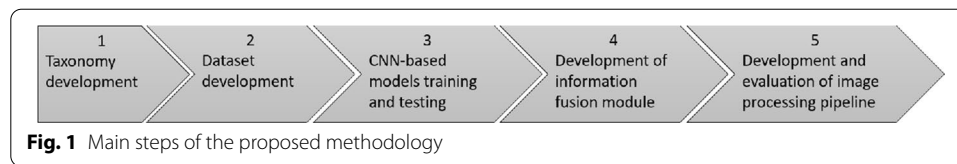
**Fig. 1** Main steps of the proposed methodology

effort, vehicle damage, other relevant information, not relevant or can't judge), and damage severity levels. In [26] dataset was annotated on flooded areas images (flooded and dry), inundation depth estimation, and water pollution categories. In [28], the authors have developed "empathi", the ontology about emergency planning and management during the crisis that contains 423 classes and 338 relations. The most prominent superclasses included hazard type, hazard phases, impact, facility, data modality, place, report, service, status, and surveillance information. Zahra et al. [29] investigated the Twitter data to investigate an eyewitness of disasters and emergencies. They found that direct witnesses have reported the damage level, the impact on social activities, the intensity of the crisis, etc., while indirect witnesses posted prayers, emotions, and thoughts. In this study, we used the information from the previous studies related to disaster data hierarchies, classification, and ontology to develop our disaster taxonomy of emergency response.

The previous literature review on crisis management using social media data helped to understand the contributions and limitations simultaneously. The strengths of previous studies include disaster image dataset development, the disaster image classifications using different machines learning the deep learning neural networks and evaluating proposed models. The previous studies mainly contributed to building social media disaster datasets suitable for CNN classification models to classify disaster types such as fire, earthquake, flood, hurricanes, etc. The model evaluation studies reported good accuracy in identifying different disaster types using CNN algorithms. However, the previous work is limited in such studies that deal with identifying context, activity, and any related artifact in disaster images. Since disaster-related object detection from social media images can help to find further details of a situation (such as any affected human, endanger animal-plant, food, and supplies needs, etc.); consequently, it will be more useful for rescuers or humanitarian organizations.

## The proposed methodology for automating the emergency response process

The comprehensive literature review unveils that there is significant potential towards developing a real-time solution that can seamlessly bring together relevant information to support crisis-related activities amongst the relevant stakeholders. More specifically, we noticed a gap in automatic information processing and delivery to concerned authorities to support rapid response. Motivated by the limitation in the existing studies and recent advancements in neural network architectures for image processing, we proposed an end-to-end methodology to automate the disaster response process. Figure 1 represents the main steps of the proposed methodology to automate the disaster detection and response process.

At first, we developed a taxonomy of disaster emergency response aiming to classify the social media images of disaster based on their labels. Then we built the disaster's image dataset that is labeled according to the proposed taxonomy. Then we proceed to implement CNN-based classification and object detection, models. In developing these models, we first built the CNN-based classification model using the opensource dataset to classify incoming images into the disaster or non-disaster types. Then we trained the CNN-based object detection model for detecting disaster-related labels by utilizing the self-built dataset to identify disaster-related tags from the images previously classified under the disaster category. Finally, we developed an information fusion module that combines the two-level labeled images and maps them to a specific response category inferred by the taxonomy. The following sections provide the technical detail of each step of the proposed methodology we have investigated in the context of the image-based emergency response process.

### The disaster taxonomy

Knowledge of disaster classification is essential to identify major categories of disaster events and appropriate responses. Many existing studies focus on deriving contextual information from a single case of a disaster or crisis based on a specific event (see "Disaster classification" Section). Unfortunately, little attention has been paid to how we broadly classify and respond to any disaster event. We identified the leading crisis dimensions by initially performing the cross-sectional literature analysis to develop an extended scheme for disaster classes in terms of disaster types and relevant responses. The extended set is then validated by conducting the initial pilot testing utilizing the card sorting [16] technique.

The analysis of previous literature has shown various classification schemes to derive contextual information from disaster-related data—however, each study focuses on a specific event covering a few disaster categories. For example, [7–9, 25, 26] focus on identifying images relevant to a particular type of disaster such as fire, flood, hurricane, etc. Similarly, others focus on a piece of specific information need such as severity levels of damage [7, 25, 26], rescue, volunteering, or donation, caution, warning, advice [27], food and basic need supplies, affected individuals (people) [7, 27], and affect on social activities [28]. Motivated by the above studies, we initially prepared six broad categories at the first level and relevant sub-categories at the second level of the classification scheme. The initial categories and sub-categories are then extended by using the card sorting activity. Card sorting is a prevalent method of information structuring in the field of human–computer interaction. Fifteen experts took part in the 'closed card-sorting activity. The objectives of the card-sorting activity were to validate the completeness of the classification hierarchy, the correctness and completeness of labels, and the correctness of information structure. The activity was designed and conducted by using the following steps:

1. The titles of categories and sub-categories were used to develop a Google form to showcase the classification hierarchy.

**Table 4** The proposed disaster taxonomy

| Categories | Labels | Proposed labels |
|---|---|---|
| Damage related information | Accident<br>House damage<br>Car damage and accident<br>Property damage<br>Aeroplan crash<br>Any vehicle damages<br>Missiles or bomb damage<br>Isolated area damages | Car accidents, housebroken or damage, general vehicle accidents, airplane accidents and damages, bomb explosions |
| Rescue, volunteering, and/or dona-tion | Hospital<br>Ambulance<br>First aid<br>Fire-brigade<br>Emergency evacuation<br>NGOs<br>Money and supplies donation<br>Help requests<br>Relief team or person<br>Other resources<br>Medical emergency<br>Relief campaign<br>Land Animals<br>Plants<br>Sea Animals | Hospital staff ward, ambulance, first aid team, laying people or col-lapsed people, fire-brigade, coffin, emergency evacuation people in lines, help banners, animals, plants, aqua animals, medical equipment |
| Food and basic need supplies | Food supplies<br>Basic needs supplies<br>Medical supplies<br>Water shortage | Empty food shelves in hypermarkets<br>Dry water reservoirs, canals, and rivers<br>Empty gallons and people waiting |
| Affected individuals (people) | Fatality<br>Death or causalities<br>Missing goods<br>Missing people<br>Homeless people<br>Other people or a person | Crying people, running people, people or person, children |
| Caution, warning, advice | Caution<br>Advice<br>Warning<br>Government plan or directive<br>Instructions<br>Safety<br>Preparedness<br>Prediction or forecasting | Caution signboard |
| Affect on social activities | Closed business including shops<br>Closed worship centers<br>Closed transportations<br>Closed Institutions (private and public)<br>Lockdown | Empty places such as roads, parks, markets |

2. Participants were provided with three images as a pre-session activity and requested to choose the best-suited first and second-level labels for each image from the given scheme.

3. During the activity, a batch of images was presented to the participants to answer whether each image belongs to a specific category or not and choose appropriate labels from the list of suggested labels for that category. Participants could also sug-gest the new categories and labels for the particular image if it does not belong to pre-defined classes.

**Fig. 2** Main steps of disaster dataset development
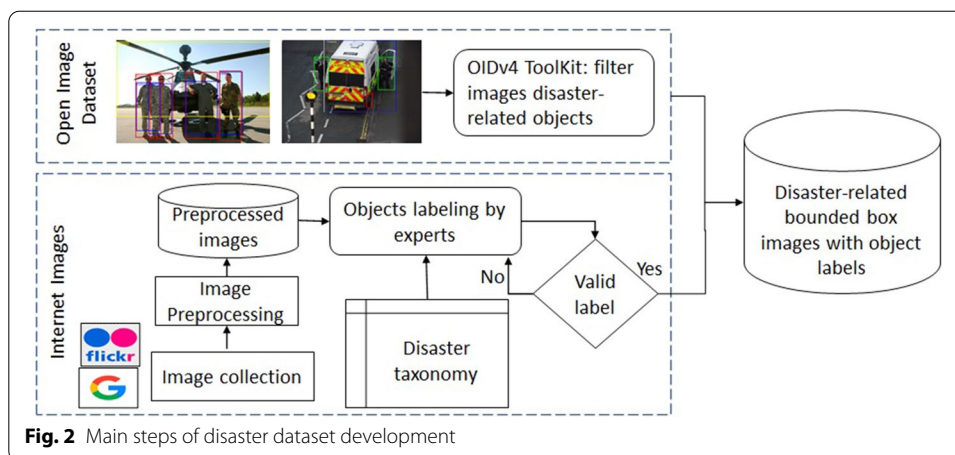
4. At the end of the activity, the participants provided their suggestions and commented on information structure and classification scheme.

Card sorting results were then used to consolidate the taxonomy content and its information structure, as shown in Table 4. We created a fine-grained taxonomy covering six high-level categories such as general type of damage and others' (see Table 4). Human experts then annotated the specific information in each high-level category, which requires human attention, such as car accidents, etc. shown in the second column of Table 4. We then manually pruned the second-level extensive categories by combining similar categories or discarding categories that were hard to visually identify from images, shown in the third column as proposed labels.

**Disaster dataset development**

This section presents the disaster dataset collected to train the CNN-based models to detect disaster types and objects automatically. The image dataset is developed in several stages to ensure an adequate number of images and categories to cover a broad range of disasters.

For high-level disaster classification, we selected CrisisMMD [13, 25] dataset. The dataset consists of 18,126 images collected on seven major natural disasters, including earthquakes, hurricanes, wildfires, and floods. The dataset is annotated on eight humanitarian categories, including affected individuals, infrastructure and utility damage, injured or dead people, missing or found people, rescue, volunteering or donation effort, vehicle damage, and other relevant information. Several studies have used the dataset to train and test several crisis-related CNN-based models [10, 24, 25]. In this study, we have selected the subset of CrisisMMD consisting of 1572 images of three major disasters, i.e., earthquake, hurricane, and typhoon to train the classification model.

For the disaster-related object detection purpose, we developed a customized dataset using Google Open Images Dataset [2](OID) and downloading images from open-source

---

[2] https://storage.googleapis.com/openimages/web/index.html.

**Table 5** Disaster's labels distributions on custom and COCO datasets

| Internet-download images' labels | Open Images Labels | COCO labels |
|---|---|---|
| Accident | Ambulance | Person |
| Casualty | Animal | Dog |
| Crying | Helicopter | Plant |
| Coffin | Bowl | Bottle |
| Damage | Band-aid | Fire hydrant |
| House damage | Ladder | |
| Car damage and accident | Medical equipment | |
| Property damage | | |
| Aeroplan crash | | |
| Any vehicle damages | | |
| Emergency evacuation | | |
| Fire brigade | | |
| First aid | | |
| Lost and found | | |
| Missing people | | |
| Personal protective equipment | | |
| Rescue team | | |

Internet-based resources for disaster-related object detection. Figure 2 shows the steps used in the dataset development. At first, we utilized OID, which contains bounded boxed images with 6000 object classes, including people, animals, ambulance, airplane, plant, building, etc. Then we used OIDv4_ToolKit [30] to download and filter images containing disaster-related objects from OID. We used 7 classes from the OID consisting of 1220 unique images. We also used a few disaster-related labeled images from the coco dataset (see Table 5).

The disaster-related object classes not found in the OID were manually collected and labeled in the next step. We downloaded images from open-source resources, such as Flickr, Google, and other social media platforms, using disaster-related keywords (car accidents, broken houses, rescue team, etc.) specified in the proposed disaster taxonomy. We then performed data pre-processing techniques for duplicates removal and image resizing. For object labeling, we used the opensource tool LabelImg,[3] which allows drawing a bounding box around the object in an image. Object annotation was done using the proposed labeling scheme described in the disaster taxonomy. The bounded box labeled images were then evaluated for compliance with the labeling scheme. Images with bounded-box object labels were stored in the dataset if the labels confirm the proposed labeling scheme; otherwise, either a label was inspected again and improved, or the image was discarded. As a result, 3787 unique images with 19 disaster-related object classes were stored in the disaster dataset. Table 5 represents the distribution of all disaster-related object labels used to annotate the disaster dataset. While annotating the dataset, the internet download images were further inspected for more general class labels on the disaster category such as house damage, property damage, and vehicle damage maps to damage class. Similarly, Aeroplan crash is annotated with the label accident. The consolidated dataset annotated with the proposed labels is made available on GitHub[4] for the fellow research community and for further research and development.

---

[3] https://github.com/tzutalin/labelImg.

[4] https://github.com/amnaasifkhan/RDO-DIsaster-Images-.git.

### CNN-based models' development for disaster classification and object detection

This module consists of deep learning-based models' development for disaster classification and object detection. Previous studies [10, 24, 25] already contributed well to disaster classification; therefore, our focus in this work is a further extension of previous work to objects detection from disaster and identifying specific knowledge in the context of the emergency response, sorting and assigning it to a particular category of crisis. The labeled image datasets are used to train two types of CNN-based algorithms in a single architecture: First, we used VGG-16 as a backbone to classify images to correct disaster types. We then used YOLO to detect disaster-related objects for emergency response. In the following section detail of each model is discussed.

The model VGGNET was initially proposed by Simonyan and Zisserman [31], with $3 \times 3$ convolution layers. Later the depth of the model was expanded to 16–9 layers. We fine-tuned the VGG-16 network for high-level disaster classification due to its better performance on disaster classification [25]. The VGG-16 consists of an input layer followed by pair of $2 \times 2$ and $3 \times 3$ convolutional layers with max pool layers on top of each pair. In this paper, the VGG-16 is fine-tuned by replacing the last softmax layer with three classes (i.e., selected disaster types) instead of the original 1,000-class classification. We used transfer learning to train the model because of the small disaster dataset compared to traditional CNN-based image classification models [32, 33].

You only look once (YOLO) [34] is a state-of-the-art, real-time object detection system. This network divides the whole image into regions to predict bounding boxes and probabilities for each region. Unlike, Region-based convolutional neural network (R-CNN [35] and Faster R-CNN [36]), which identify high scoring regions of interest and then perform prediction multiple time for the various region in an image using CNN-networks for inference, YOLO predicts multiple bounding boxes and class probabilities for these boxes in a full image in one run. YOLO is implemented based on Darknet [37], a highly accurate and fast open-source framework for real-time object detection. YOLO comes with different versions with incremental improvement in previous versions; YOLOv3 and YOLOv4 are widely used in object detection due to their simpler architecture and high accuracy.

YOLOv3 [38] is an end-to-end object detection algorithm implemented in Darknet-53 evolved from YOLOv2 [39]. YOLOv3 runs faster than R-CNN due to its simple architecture. YOLOv3 has one CNN backbone with three object detection layers called heads. It takes full images, divided them into S x S size grid cells. Each cell predicts B bounding boxes along with probabilities and C class probabilities for each object [40]. It also draws on Residual Network (ResNet) [41] with five residual blocks with batch normalization and skips connection concepts. Compared to ResNet and YOLOv2, the YOLOv3 has remarkable performance in terms of detection accuracy, bounding box localization, and detection speed.

The YOLOv4 [42] has CSPDarknet-53 backbone and is designed for large-scale datasets. It combines features from other studies such as CSP [43] connections, SPP (spatial pyramid pooling layer) [44], and PAN (Path Aggregation Network) [34] in its backbone, which has enhanced its performance with the low computational cost of
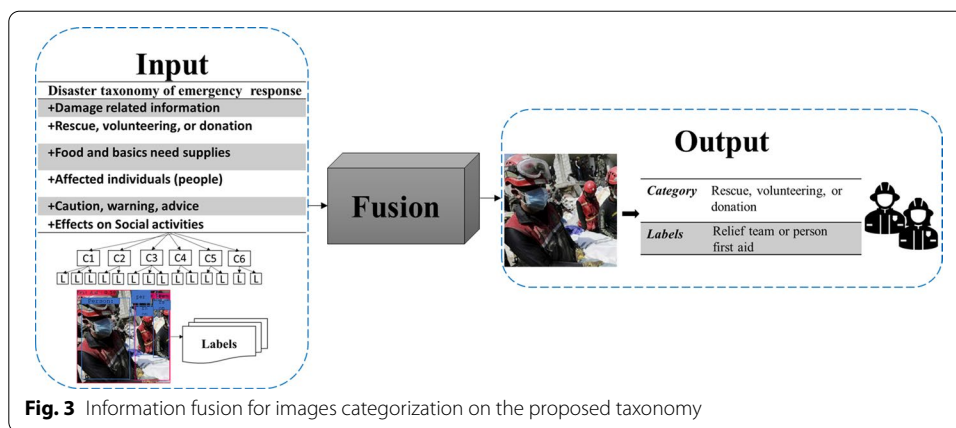
**Fig. 3** Information fusion for images categorization on the proposed taxonomy

inference. The features added in YOLOv4 architecture in both backbone and detector modules are Bag of freebies (Bof) and Bag of specials (BoS). In the BoS part, SPP is tightly coupled with the backbone in the Bof part to improve field sizes of detection layer feature maps. Bof comes from data augmentation techniques.

In the context of this study, we used both YOLOv3 and YOLOv4 for object detection. The size of the disaster dataset was 3787 images with 19 object classes, which is not too large to use YOLOv4; however, we used both versions to observe a significant difference in detection accuracy and speed.

### Information fusion module

The object detection systems (YOLO) used in the previous step predicts labels from the disaster image dataset. To classify each image into an appropriate response category, we need to identify the relevant category of an image according to the labeling scheme (see "The disaster taxonomy" Section). We adopted the information fusion approach to address this problem, motivated by the representative goal-seeking techniques of 'decision tables' and AHP [45]. The information fusion module receives the predicted labels and the labeling scheme as input and calculates each category's local and global weight by following AHP to establish the relevant category of each image based on the predicted labels.

In this process, predicted labels and disaster taxonomy are fed into the information fusion module. If an inferred label matches with a pre-defined label of any taxonomy category, the input image would be sorted in that category. The process of information fusion is shown in Fig. 3 and explained below.

From the figure, it can be observed that the given image has three unique predicted labels: person (6 times), rescue team (1 time), first aid (1 time). These labels match the two categories of response taxonomy, i.e., Rescue, volunteering, or donation (C2), and Affected individuals (C4).

To establish the right category for the given image, we first calculated the global and local weights of each category and then integrated the calculated weights to identify the final category of an input image. The global weight is the ratio of unique labels in

a category to the total number of unique labels for an image with values in the range of 0−1, as follow:

$$C_{i=1:6}\big(wt\,(global)\big) = \frac{\sum_{j=1:m}\big(L_j \in C_i\big)}{X} \tag{1}$$

where Lj is a unique label in a category Ci and X is the number of unique labels in an image. For instance, if C2 has two unique labels and the entire image has three unique labels, the global weight for this category is C2 = (2/3 = 0.67). Based on Eq. 1, the global weight of six categories in the disaster taxonomy is calculated as follow:

$$c\big(wt\,(global)\big) = [C_{i:1:6}(wt(global))] \tag{2}$$

The global weights array for each category from the input image in Fig. 3 is calculated as follow:

$$c(wt(global)) = [C_1(0),\ C_2(0.67),\ C_3(0),\ C_4(0.33),\ C_5(0),\ C_6(0)]$$

To overcome the influence of the same label appearing multiple time in a category, we calculated local weight for each category by considering the ratio of the total number of times (frequency) label in a specific category to the maximum frequency of all the labels in that category, as follow:

$$c_{i=1:6}(wt(local)) = \sum_{j=1:m}\left(\frac{L_j(f)\in C_i}{\max\limits_{j=1:m}\big(L_j(f)\in C_i\big)}\right) \tag{3}$$

where $L_j(f)$ is the frequency of unique label in a category $C_i$. For instance, in Fig. 3, C2 consists of two labels('rescue' and 'first aid' with the frequency of 1 each), and C4 comprised of one label ( 'person' with the frequency of 6). The local weight of these two categories are calculated according to Eq. 3 as follow:

$$C2(wt(local)) = \sum\left[L_{\text{rescue}}\left(\frac{1}{\max(1,1)}\right), L_{\text{firstaid}}\left(\frac{1}{\max(1,1)}\right)\right] = 2$$

$$C4(wt(local)) = \sum\left[L_{\text{person}}\left(\frac{6}{\max(6)}\right)\right] = 1$$

The generalized form of local weight of all categories is calculated as follow:

$$c(wt(local)) = [C_{i:1:6}(wt(local))] \tag{4}$$

The local weight array of all categories in Fig. 3 is calculated according to Eq. 4 as follow:

$$C(wt(local)) = [C1(0), C2(2), C3(0), C4(1), C5(0), C6(0)]$$

The final category of disaster taxonomy of the input image will be calculated by integrating the weight of both groups, as follow.

**Table 6** Detailed procedure of information fusion process

| |
|---|
| **Input: Categories. Labels[][]** |
| **IMGLabels** |
| **Output: Final disaster category** |

1: **Initialize** 2d Matrix Categories.Labels[][]
2: IMGLabels = Image.labels
3:  **FOR** all Labels in Categories.Labels[Rows][Columns] **DO**
4:    **FOR** all Labels in IMGLabels **DO**
5:     **IF** current_label in IMGlabels € Categories.Labels[current_Row]
6:       Matrix_frequency[current_Row].append(SUM(current_label in IMGLabels))
7: X = count_of_unique_labels in IMGLabels
8: **FOR** Rows in Matrix_frequency[][]**DO**
9:   Categories(wt(global))[Current_Row] = $C_{i=1:6}(wt(global))$
10: **FOR** Rows in Matrix_frequency[][] **DO**
11:   **FOR** Columns in Matrix_frequency[][] **DO**
12:     Categories(wt(local)) [Current_Row] = $(C_{i=1:6}(wt(local)))$
13: **FOR** i = 0 to (Number of Rows in Matrix_frequency-1) **DO**
14:   Product[i] = (Categories(wt(global))[i]* Categories (wt(local))[i])
15: Disaster_category = **max**(Product[])
16: **IF** Index(Categories.Labels) = = Index(Disaster_category)
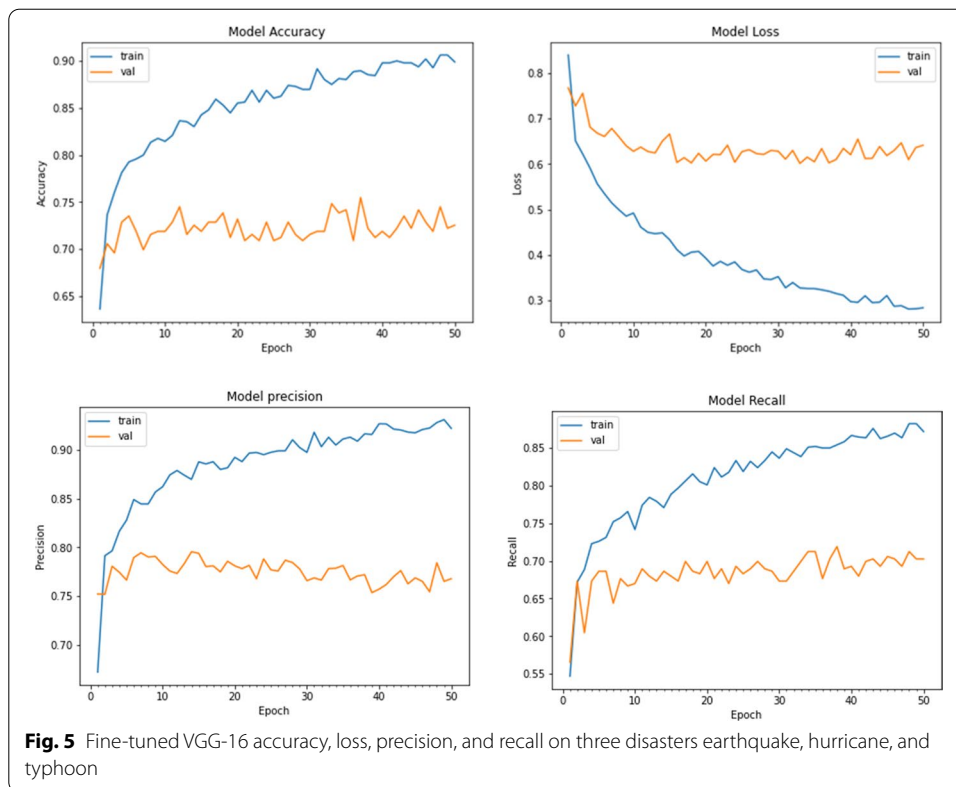
17:    **Output**(Category_Name)
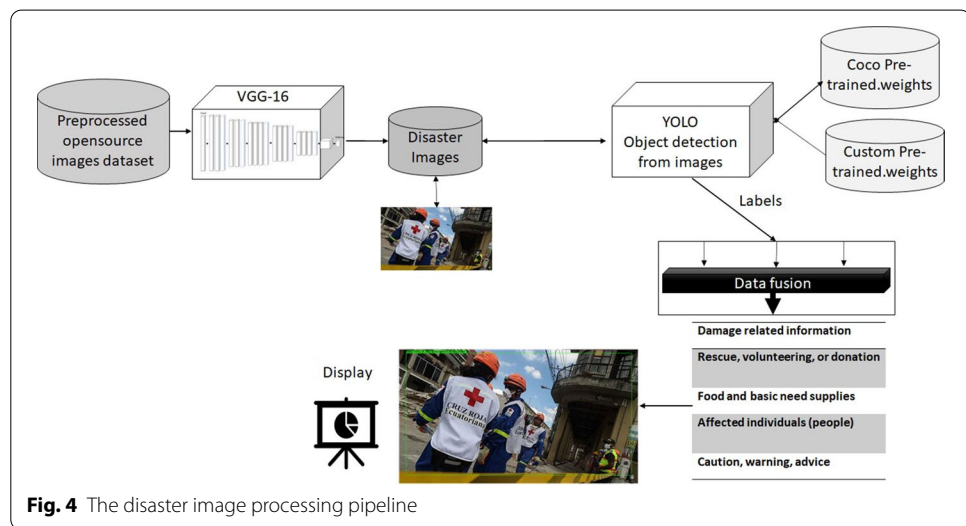


**Fig. 4** The disaster image processing pipeline

$$\text{Disaster\_Category} = \max\left(c\left(wt(global)\right) * c(wt(local))\right) \tag{5}$$

According to Eq. 5, the final disaster category of the input image is calculated as follow:

$$\text{Disaster}_{\text{Category}} = \max([0, 0.67, 0, 0.33, 0, 0] * [0, 2, 0, 1, 0, 0])$$

In the case of multiple maximum values, the input image would be classified into multiple categories in the emergency response taxonomy. If none of the labels of an input image matches the pre-defined labels, the image would not be categories under any of the disaster response categories but would be treated as a disaster image. Table 6 represents the detail of the information fusion process.

**Fig. 5** Fine-tuned VGG-16 accuracy, loss, precision, and recall on three disasters earthquake, hurricane, and typhoon

**Table 7** VGG-16 model accuracy, loss, precision, and recall

| VGG-16 fine-tuned | Accuracy | Loss | Precision | Recall |
| --- | --- | --- | --- | --- |
| Training | 0.94216 | 0.27865 | 0.88229 | 0.91667 |
| Validation | 0.76786 | 0.6411 | 0.70261 | 0.72549 |
| Testing | 0.83674 | 0.51116 | 0.80392 | 0.82026 |

## Concept validation through the image processing pipeline

The proposed methodology is validated by developing an end-to-end image processing pipeline consists of image classification, object detection, and information fusion, as shown in Fig. 4. At first, we used the opensource dataset, CrisisMMD [13], for validating the image classification module. We have selected images related to the earthquake, hurricane, and typhoon, including rescue, people, buildings, vehicles, etc. First, the disaster type classification, the module processes the input image using the fine-tuned VGG-16 weights to identify the type of disaster from the input photo content. If the image is irrelevant to disaster type; then, it is discarded. Once the input image is identified as a relevant image, it is stored in intermediate storage. Later, the input image is transferred to the object detection module trained on YOLO to recognize the image's labels. In this module, the objects/labels present in the picture are extracted and stored in temporary storage. The collection of labels associated with each image are sent to the fusion model. The fusion algorithm is implemented to collect labels and the distinct categories in the

**Fig. 6** YOLOv3 and YOLOv4 training loss on 7000 iterations
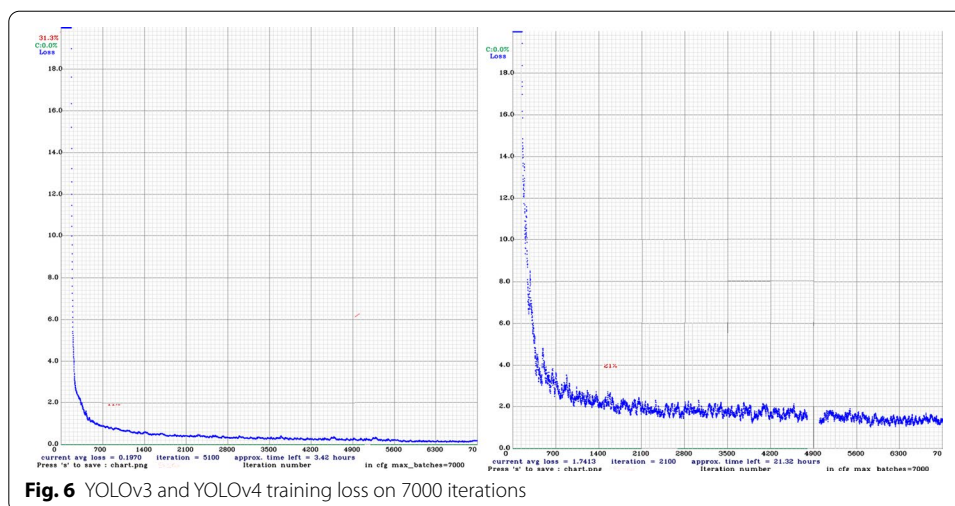
disaster taxonomy to classify input images to their relevant response category. In the following section, the results of the image processing pipeline are discussed.

### Experimental results

In the first phase of the image processing pipeline, the image is classified into different disaster types. To train VGG-16, we used transfer learning by using the 'ImageNet' [46] weights. We selected three different types of disasters from the CrisisMMD dataset, i.e., earthquake, hurricane, and typhoon. The selected open-source images dataset was divided into 60% training, 20% validation, and 20% testing. The model is fine-tuned on the batch size of 32 and 50 epochs.

Figure 5 shows accuracy, loss, precision, and recall on training, validation, and testing data. An average accuracy, loss, precision, and recall for the model on training, validation, and testing datasets are presented in Table 7. The results showed that testing accuracy of 0.83 is comparable with the previous contributors of CrisisMMD [25], which reported average testing accuracy of 0.72. Furthermore, VGG-16 trained on a subset of the CrisisMMD dataset showed a test precision of 0.88 and a recall of 0.92, which is higher than the previous contribution of 0.76 and 0.65 precision and recall.

For object detection, we used both versions of YOLO. We first transformed the labeled disaster images into YOLO format and input parameter of Darknet[5] of recommended size of $416 \times 416$ pixels for training the networks. We run the training session for both YOLOv3 and YOLOv4 up to 7000 iterations with a batch size of 64 by considering the memory limitation of the GPU. We used pre-trained weights "darknet53.conv.74"[6] for YOLOv3 and "YOLOv4.conv.137"[7] weights for YOLOv4. Other parameters such as momentum, initial learning rate, weight decay regularization are the same as in the original versions of both models.

---

[5] https://github.com/pjreddie/darknet.git.

[6] https://pjreddie.com/darknet/yolo/.

[7] https://github.com/AlexeyAB/darknet/releases/download/darknet_yolo_v3_optimal/YOLOv4.conv.137.

**Table 8** AP of YOLOv4 and YOLOv3 on the custom dataset

| Class | YOLOv3 (AP) (%) | YOLOv4 (AP) (%) |
| --- | --- | --- |
| Accident | 48.02 | 95.49 |
| Casualty | 48.57 | 40.00 |
| Crying | 25.00 | 100.00 |
| Coffin | 20.00 | 100.00 |
| Damage | 45.49 | 95.14 |
| Emergency evacuation | 32.45 | 95.59 |
| Fire brigade | 59.29 | 100.00 |
| First aid | 55.89 | 95.69 |
| Lost and Found | 32.16 | 65.18 |
| Mourning people | 28.00 | 53.94 |
| Personal protective equipment | 14.02 | 96.47 |
| Rescue team | 43.87 | 96.47 |
| Ambulance | 54.99 | 98.85 |
| Animal | 30.61 | 93.75 |
| Band-aid | 4.00 | 60.89 |
| Food supply | 1.11 | 50.03 |
| Helicopter | 25.23 | 71.62 |
| Ladder | 41.84 | 81.56 |
| Medical equipment | 15.06 | 26.44 |

**Table 9** Comparison of YOLOv4 and YOLOv3 classification results on emergency response taxonomy on combined custom and COCO dataset

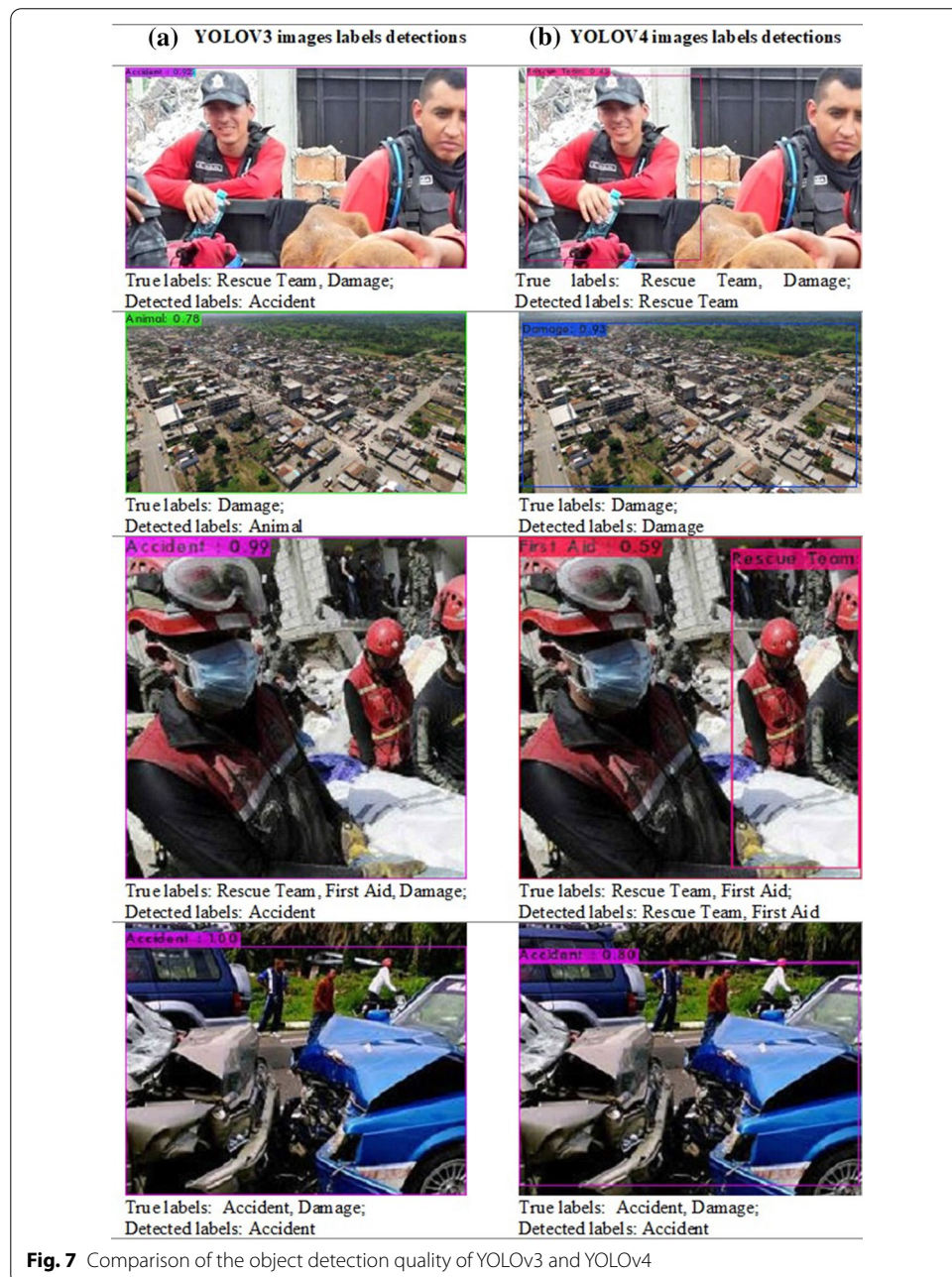| Disaster categories | No. of images classified on disaster taxonomy | |
| --- | --- | --- |
| | YOLOv4 | YOLOv3 |
| Damage related information | 39 | 407 |
| Rescue, volunteering, or donation | 78 | 584 |
| Food and basic need supplies | 4 | 7 |
| Affected individuals (people) | 812 | 687 |
| Caution, warning, advice | 2 | 1 |
| Affect on social activities | 0 | 0 |

Figure 6 reports the training loss while training on the YOLOv3 and YOLOv4 algorithms. We have not observed any significant difference in average loss among the training of both models. The reason might be the size of the dataset, which is not reasonably large to observe the performance difference of YOLOv4.

We measured the performance of the object detection model using AP (average precision) [47], which is a popular method of measuring object detectors' accuracy, and we report the mean overall categories mAP (mean average precision) for object classes. We used IoU (Intersection over Union) to measure the overlap of the predicted boundary over the ground truth boundary. Table 8 presents the AP of YOLOv3 and YOLOv4 on 10% of the dataset's total size. The IoU on 50% threshold performance of YOLOv3 was 32.93% and YOLOv4 was 78.98%. Note that for nearly all incidents classes, AP of YOLOv4 is much higher than YOLOv3.

**Table 10** Comparison of YOLOv4 and YOLOv3 classification results on emergency response taxonomy using only the custom dataset

| Disaster categories | No. of images classified on disaster taxonomy | |
| --- | --- | --- |
| | YOLOv4 | YOLOv3 |
| Damage related information | 38 | 500 |
| Rescue, volunteering, or donation | 46 | 690 |
| Food and basic need supplies | 0 | 5 |
| Affected individuals (people) | 0 | 11 |
| Caution, warning, advice | 0 | 0 |



**Fig. 7** Comparison of the object detection quality of YOLOv3 and YOLOv4

**Table 11** Classification of test dataset on the taxonomy of emergency response using YOLOv4

| Disaster categories | No. of images | | |
|---|---|---|---|
| | Earthquake | Hurricane | Typhoon |
| Damage related information | 9 | 0 | 0 |
| Rescue, volunteering, and/or donation | 22 | 5 | 5 |
| Food and basic need supplies | 1 | 0 | 2 |
| Affected individuals (people) | 160 | 21 | 53 |
| Caution, warning, advice | 0 | 0 | 0 |

We tested the pipeline using the 917 images using a subset of the earthquake disaster type from the Ecuador Earthquake (2016) from CrisisMMD. The evaluation is performed on the dataset with two alternatives: (1) YOLOv3 and YOLOv4 trained on custom dataset + COCO dataset, (2) YOLOv3 and YOLOv4 trained only on the custom dataset. The results showed that the YOLOv4 categorized 876 pictures, and YOLOv3 categorized 907 images in different disaster categories. In general, YOLOv3 detected more labels than YOLOv4. An expert manually evaluated the detected labels by YOLOv3 and YOLOv4 and found that YOLOv4 detected more accurate labels than YOLOv3. Although the quantity of the labels detected by the YOLOv3 is much higher, the quality of the YOLOv4 in terms of the proposed disaster taxonomy was higher. The results presented in Table 9 of the image classification on disaster taxonomy using custom and COCO datasets are significantly different on both YOLOv3 and YOLOv4 algorithms. The manual evaluation of the image classification on disaster taxonomy done by the expert discovered that the images are classified more accurately based on YOLOv4 detection than YOLOv3.

Similarly, Table 10 presents the custom dataset results that show YOLOv4 classified fewer images than YOLOv3. While comparing the accuracy of the classification results on the disaster categories with the human-labeled dataset, we found that YOLOv3 correctly classifies 85% of images, and YOLOv4 accurately categorizes 96% of images. Again, YOLOv4 showed higher accuracy than YOLOv3. The results show that high-quality training data can automatically filter and map millions of social media images to the right disaster category using YOLOv4.

YOLOv4 showed higher detection accuracy, as shown in Fig. 7. From the figure, it can be observed that YOLOv4 detected more specific labels closed to the proposed annotation scheme. For example, the first image was annotated with the labels of 'rescue team' and 'damage'; the YOLOv4 detected it accurately as 'rescue team' whereas YOLOv3 detected it as 'Accident' that is not the actual label of this image.

Based on the higher quality of results, we selected YOLOv4 for further analysis. Although, there is a trade-off in terms of computational requirements since YOLOv4 required more resources than YOLOv3. However, considering the accuracy and criticality of information need for efficient emergency response, it is worth investing in computational resources. For the proof of concept, we selected a 20% test dataset of the earthquake, hurricane, and typhoon from CrisisMMD to evaluate how well YOLOv4 maps these disaster-related images on the proposed emergency response taxonomy. Table 11 presents the results on the number of images correctly classified on the given disaster categories.

## Discussion and conclusion

The power of social media content and crowdsourcing in disaster response has been proven to be extremely effective. However, most previous research tried to utilize text analysis techniques and manual interpretation of social media postings. Automated processing of feature-rich multimedia content is inevitable to fully utilize the benefit of user-generated content in a timely emergency response. In this paper, we proposed a methodology that harnesses cues from information-rich social media images seamlessly. Our work is unique from other research, as, firstly, it uses a comprehensive taxonomy of disaster emergency response utilizing visual contents. Secondly, the disaster-related features (contexts and cues) detected from the images through deep-learning neural network algorithms are automatically mapped to the disaster taxonomy using goal-seeking techniques of 'decision tables' and AHP. We proposed an image processing pipeline to automatically sort disaster-related social media images on the emergency response taxonomy. The pipeline consists of image preprocessing, image classification, and object detection using CNN algorithms, and finally, an information fusion to map the input image to a suitable category of disaster taxonomy of emergency response.

In this paper, we described each module of the social media image processing pipeline. We also validated the effectiveness of the proposed pipeline through a summative evaluation approach. Our results confirm that the image processing pipeline automatically classifies social media images posted by users according to the taxonomy of emergency response with an accuracy of 96%. Due to the CNN-based training algorithm used with the pipeline, the classification accuracy may be further augmented through incremental training with the help of the newly classified images (subject to some manual validation).

The emergency response taxonomy consists of various disaster categories that may help stakeholders be informed in any crisis. This may benefit the users of the information in many ways. First, the millions of images posted on social media or internet platforms can be sorted automatically and more efficiently in a short period of time, avoiding spending days of manual labor work. Secondly, in disasters, the efficient discovery of information can make a massive difference in rescue efforts and significantly improve emergency response efforts if useful information reaches a relevant authority promptly. Social media is the fastest channel to connect with the affected individuals. If the information is available in the form of posted photos, it can help the authorities understand the situations better to prepare for the appropriate emergency response operations accordingly.

In-depth information related to disaster events is required to develop the disaster taxonomy of emergency response. In this paper, we have contributed to the body of knowledge in the direction of using social media images for emergency response by a mixture of different scientific approaches of literature review and card-sorting. As a result, we have proposed the visual disaster taxonomy of emergency response. Previously, Olteanu et al. [27] have investigated information types in crisis from social media's textual information. Barz et al. [26] investigated the types of flood-related information from the visual data. Gaur et al. [28] proposed the ontology related to emergency management

and planning in general. Though, previous contributions are helpful to provide essential disaster information of our proposed visual disaster taxonomy. Our work stands out from previous research as we proposed the first comprehensive visual disaster taxonomy of emergency response that provides clear visual criteria of emergency cues in many catastrophes. The proposed taxonomy will automatically connect a victim to concerned authorities in an efficient manner. This is a task that is otherwise cumbersome to be carried out manually.

The proposed taxonomy of disaster response is used as a source of labels for annotating our custom images dataset[8] as shown in Table 5. The dataset was developed by drawing the bounding box around the objects in images, which may help convey some meaningful message related to accidents, damage, rescue, first-aid, volunteering, food and supplies, affected people and social activities, casualty and death, etc., in a massive emergency. Such messages may help the stakeholders in information seeking and decision making for efficient emergency response. In the previous studies [20, 48], the authors proposed detecting objects such as vehicles and people from the photos related to fire events to estimate emergency severity; however, their contributions are limited to detecting general objects using the COCO dataset. On the contrary, we focused on supporting EDM by exploring the information contained in the images of disaster to find specific knowledge in the context of the emergency response, sorting and assigning it to a particular category of crisis. That may help emergency response in all phases of crisis management.

The image processing pipeline classifies the disaster type (currently earthquake, hurricane, and typhoon) and recognizes the emergency status from the images, e.g., rescue need or work, affected people, or accident damage. The image processing pipeline can be integrated with any social media platform for real-time disaster images collection. Also, it can be easily integrated with any disaster management system, such as Ushahidi [49]. We apply fine-tuned VGG-16 CNN to train the disaster type classification, as it was investigated in [9]; however, our model resulted in improved testing performance. Since the previous studies already contributed well to disaster classification, therefore it is not our paper's primary focus, but our work is a further extension of previous work to objects detection from disaster images. In the direction of disaster object detection from images, we have perceived the IoU 32.93% and 78.98% on YOLOv3 and YOLOv4, respectively, and AP scores on YOLOv4 higher than 90% in many classes. This means that high-quality training data can automatically filter and map millions of social media images to the right disaster category using object detection algorithms like YOLOv4. The trained model helps identify the disaster context and elements from the social media posted videos due to its detection performance in terms of speed. Previously, many studies [8–10, 25, 26] applied machine learning and CNN algorithms to classify the social media images according to various disaster types and severity of damages. Our image processing pipeline is distinctive in terms of information fusion using decision support techniques of 'decision tables' and AHP on the data acquired by CNN classification and detection models to tell the narrative of the disaster situation to the concerned

---

[8] https://github.com/amnaasifkhan/RDO-DIsaster-Images-.git.

stakeholders. The application of the AHP allows avoiding inaccurate mapping to images to disaster taxonomy by normalizing the values of repeated objects labels and prioritizing more distinct labels of the images. The results are validated by the human expert as reported in "Experimental results" Section. Consequently, it allows the acquisition of efficient and reliable disaster information from social media crowdsource images.

Our evaluation setup tested various crisis scenarios consisting of the earthquake, hurricane, and typhoon. The main aim was to test the generality of our proposed taxonomy of emergency responses for multiple types of disasters. The results in Tables 10 and 11 show that the taxonomy of emergency responses from images is suitable for similar natural disasters in dense and semi-populated areas to understand the current situation in the context of crisis. These results provide a significant milestone in automating disaster management systems on image data, consequently improving government performance, as such contribution is missing in the previous studies.

The current study is limited in terms of available computational resources, as the availability of high-performance GPU machines may help improve detection performance. The image mapping results are manually validated by a human expert, which was possible with the given amount of data. However, the manual validations can be cumbersome for big amount of datasets in the future. Therefore, we are planning to enhance our evaluation strategy by improving the evaluation design. Another limitation is that the disaster's objects dataset is small and comprises a small number of images in each category. Furthermore, we have achieved a good accuracy on the custom trained classes compared to annotated data downloaded from the open image. In the future, we are planning to enhance the dataset by increasing the classes and the number of images and make it an opensource. Potential users of such datasets can benefit from using it for their disaster management system.

We would like to highlight that the proposed pipeline can be easily extended or integrated with other challenging features extracted from still images as well as video and other multimedia information. The emerging trend in disaster response-related research will eventually incorporate spatio-temporal features from heterogeneous sources. The proposed pipeline can be integrated with additional information such as geolocation, users' characteristics, etc., to get more insight into the situation to help in EDM. The proposed visual disaster taxonomy of emergency response is helpful to provide fundamental information in developing machine learning and deep learning algorithms of disaster detection and classification. Furthermore, it can be used as criteria for social media data analysis in the emergency response domain. We hope that these contributions will motivate further research on detecting incidents in images, and also promote the development of automatic tools that can be used by humanitarian organizations and emergency response agencies.

**Abbreviations**
SPP: Spatial pyramid pooling layer; PAN: Path Aggregation Network; Bof: Bag of Freebies; Bos: Bag of Specials; AP: Average Precision; mAP: Mean Average Precision; IoU: Intersection Over Union; YOLO: You Only Look Once; AHP: Analytic Hierarchy Processes; EDM: Emergency Decision Making; CNN: Convolutional Neural Networks; IMEXT: Image Extraction from tweets; SVM: Support Vector Machine; BoVW: Bag-of-Visual-Words; COCO: Common Objects in Context; PHOW: Pyramidal Histogram of visual words; EM-DAT: Emergency Events Database; OID: Open Images Dataset.

## Declarations

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare that they have no competing interests.

**Author details**
[1]Information Systems Department, College of Computer Sciences & Information Technology, King Faisal University (KFU), AlAhsa, Kingdom of Saudi Arabia. [2]Faculty of Computing & Information, Cairo University, Giza, Egypt. [3]Faculty of Engineering, Cairo University, Giza, Egypt. [4]Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt.

**References**
1. Imran M, Castillo C, Diaz F, Vieweg S. Processing social media messages in mass emergency: a survey. ACM Comput Surv. 2015;47(4):1–38.
2. Castillo C. Big crisis data: social media in disasters and time-critical situations. Cambridge: Cambridge University Press; 2016.
3. Zhou L, Wu X, Xu Z, Fujita H. Emergency decision making for natural disasters: an overview. IJDRR. 2018;27:567–76.
4. Imran M, Mitra P, Srivastava J. Cross-language domain adaptation for classifying crisis-related short messages. In: Antunes P, Silvera VAB, Albuquerque JPd, Moore KA, Tapia nH, editors. 13th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2016; Rio de Janeiro, Brazil.: Information Systems for Crisis Response and Management, ISCRAM; 2016.
5. Tuech J-J, Gangloff A, Schwarz L. Our challenge is to adapt the organization of our system to the six stages of the epidemic to go beyond the COVID-19 crisis. Br J Surg. 2020;107(7):e189.
6. Said N, et al. Natural disasters detection in social media and satellite imagery: a survey. Multimed Tools Appl . 2019;78(22):31267–302.
7. Murthy D, Gross A, McGarry M. Visual social media and big data. Interpreting Instagram images posted on Twitter. Digital Cult Soc. 2016;2(2):113–34.
8. Peters R, de Albuquerque JP. Investigating images as indicators for relevant social media messages in disaster management. In: Palen L, Buscher M, Comes T, Hughes A, editors. ISCRAM - The 12th International Conference on Information Systems for Crisis Response and Management; 24-27 May; Kristiansand, Norway. 2015.
9. Daly S, Thom JA. Mining and Classifying Image Posts on Social Media to Analyse Fires. In: Tapia A, Antunes P, Bañuls VA, Moore K, Porto J, editors. ISCRAM - The 13th International Conference on Information Systems for Crisis Response and Management; 22-25 May; Rio de Janeiro, Brasil: Federal University of Rio de Janeiro. 2016.
10. Nguyen DT, Ofli F, Imran M, Mitra P. Damage assessment from social media imagery data during disasters. Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; Sydney Australia: Association for Computing Machinery, New York, NY, United States; 2017. p. 569-76. doi:10.1145/3110025.3110109
11. Goodwin M. Fire-Detection-Image-Dataset: Github; 2017 [Available from: https://github.com/cair/Fire-Detection-Image-Dataset.] Saied A. FIRE Dataset: Outdoor-fire images and non-fire images for computer vision tasks.: Kaggle; 2018 [Available from: https://www.kaggle.com/phylake1337/fire-dataset.] Accessed 27 May 2021.
12. Saied A. FIRE Dataset: Outdoor-fire images and non-fire images for computer vision tasks.: Kaggle; 2018 [Available from: https://www.kaggle.com/phylake1337/fire-dataset.] Accessed 27 May 2021.
13. Alam F, Ofli F, Imran M. Crisismmd: Multimodal twitter datasets from natural disasters.  12th International AAAI Conference on Web and Social Media; Palo Alto, California, USA: AAAI Press, Palo Alto, California USA; 2018.
14. Tagarev T, Ratchev V. A taxonomy of crisis management functions. Sustainability. 2020;12(12):5147.
15. Sementelli A. Toward a taxonomy of disaster and crisis theories. ATP. 2007;29(4):497–512.

16. Wood JR, Wood LE. Card sorting: current practices and beyond. J Usability Stud. 2008;4(1):1–6.

17. Francalanci C, Guglielmino P, Montalcini M, Scalia G, Pernici BIMEXT. A method and system to extract geolocated images from Tweets—Analysis of a case study. Brighton, UK: IEEE; 2017. p. 382–90. https://doi.org/10.1109/RCIS.2017.7956563.

18. Ahmad K, Konstantin P, Riegler M, Conci N, Holversen P. CNN and GAN Based Satellite and Social Media Data Fusion for Disaster Detection. In: Gravier G, Bischke B, Demarty C-H, Maia Zaharieva MR, Dellandrea E, Bogdanov D, et al., editors. Working notes proceedings of the Mediaeval workshop; Dublin, Ireland.2017.

19. Ahmad S, Ahmad K, Ahmad N, Conci N. Convolutional Neural Networks for Disaster Images Retrieval. In: Gravier G, Bischke B, Demarty C-H, Maia Zaharieva MR, Dellandrea E, Bogdanov D, et al., editors. Working notes proceedings of the Mediaeval workshop; Dublin, Ireland.2017.

20. Antzoulatos G, Giannakeris P, Koulalis I, Karakostas A, Vrochidis S, Kompatsiaris I. A Multi-Layer Fusion Approach For Real-Time Fire Severity Assessment Based on Multimedia Incidents.  In Proceedings of the 17th International Conference on Information Systems for Crisis Response and Management – ISCRAM 2020 24-27 May; Blacksburg, VA, USA.: ISCRAM Digital Library; 2020.

21. Kuznetsova A, et al. The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale. Int J Comput Vision. 2020;128(7):1956–81.

22. Dunnings AJ, Breckon TP. Experimentally defined convolutional neural network architecture variants for non-temporal real-time fire detection. Athens, Greece: IEEE; 2018. p. 1558–62. https://doi.org/10.1109/ICIP.2018.8451657.

23. Nguyen DT, Alam F, Ofli F, Imran M. Automatic image filtering on social networks using deep learning and perceptual hashing during crises. In: Comes T, Bénaben F, Hanachi C, Lauras M, editors. 14th International Conference on Information Systems for Crisis Response And Management; 21–24 May; Albi, France.: ISCRAM.; 2017.

24. Alam F, Imran M, Ofli F. Image4act: Online social media image processing for disaster response.  2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); 31 July-3 Aug.; Sydney, Australia.: IEEE; 2017. p. 601-4.

25. Alam F, Ofli F, Imran M. Processing social media images by combining human and machine computing during crises. Int J Human Comput Interact. 2018;34(4):311–27.

26. Barz B, Schröter K, Münch M, Yang B, Unger A, Dransch D, et al. Enhancing Flood Impact Analysis using Interactive Retrieval of Social Media Images. arXiv preprint arXiv:190803361. 2019. https://doi.org/10.5445/KSP/1000087327/06

27. Olteanu A, Vieweg S, Castillo C. What to expect when the unexpected happens: Social media communications across crises. CSCW '15: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; Vancouver, BC, Canada.: Association for Computing Machinery, New York, NY, United States.; 2015. p. 994–1009. https://doi.org/10.1145/2675133.2675242

28. Gaur M, Shekarpour S, Gyrard A, Sheth A. Empathi: An Ontology for Emergency Managing and Planning About Hazard Crisis.  2019 IEEE 13th International Conference on Semantic Computing (ICSC); 30 Jan.-1 Feb.; Newport Beach, CA, USA.: IEEE; 2019. p. 396-403. https://doi.org/10.1109/ICOSC.2019.8665539

29. Zahra K, Imran M, Ostermann FO. Automatic identification of eyewitness messages on twitter during disasters. Inf Process Manage. 2020;57(1):102107.

30. Vittorio A. Toolkit to download and visualize single or multiple classes from the huge open images v4 dataset: Github; 2018 [Available from: https://github.com/EscVM/OIDv4_ToolKit.] Accessed 27 May 2021.

31. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Bengio Y, LeCun Y, editors. 3rd International Conference on Learning Representations; 7-9 May.; San Diego, CA, USA.2015.

32. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks?  NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems Montreal, Quebec, Canada.: ACM.; 2014. p. 3320-8.

33. Ozbulak G, Aytar Y, Ekenel HK. How transferable are CNN-based features for age and gender classification?  2016 International Conference of the Biometrics Special Interest Group (BIOSIG); 21-23 Sept. ; Darmstadt, Germany: IEEE; 2016. p. 1-6. https://doi.org/10.1109/BIOSIG.2016.7736925

34. Redmon J, Divvala S, Girshick R, Farhadi A, editors. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016 27-30 June.; Las Vegas, NV, USA: IEEE. https://doi.org/10.1109/CVPR.2016.91

35. Girshick R. Fast R-CNN.  2015 IEEE International Conference on Computer Vision (ICCV); 7-13 Dec.; Santiago, Chile.: IEEE; 2015. p. 1440-8. https://doi.org/10.1109/ICCV.2015.169

36. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell. 2016;39(6):1137–49.

37. Koo Y, Kim S, Ha Y-G. OpenCL-Darknet: implementation and optimization of OpenCL-based deep learning object detection framework. World Wide Web. 2020;1(1):1–21.

38. Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv preprint https://arxiv.org/abs/1804.02767. 2018. Accessed 27 May 2021.

39. Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. Int J Comput Vision. 2010;88(2):303–38.

40. Nie Y, Sommella P, O'Nils M, Liguori C, Lundgren J. Automatic Detection of Melanoma with Yolo Deep Convolutional Neural Networks. Nov.; Iasi, Romania: IEEE; 2019. p. 21–3. https://doi.org/10.1109/EHB47216.2019.8970033.

41. He K, Zhang X, Ren S, Sun J. In: 27–30 June., Las Vegas, NV, editor. Deep residual learning for image recognition. USA: IEEE; 2016. p. 770–8. https://doi.org/10.1109/CVPR.2016.90.

42. Bochkovskiy A, Wang C-Y, Liao H-YM. YOLOv4: Optimal Speed and Accuracy of Object Detection. ArXiv:200410934. 2020.

43. Wang C-Y, Liao H-YM, Wu Y-H, Chen P-Y, Hsieh J-W, Yeh I-H. CSPNet: A new backbone that can enhance learning capability of CNN. Seattle, WA, USA: IEEE; 2020. p. 390–1. https://doi.org/10.1109/CVPRW50498.2020.00203.

44. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell. 2015;37(9):1904–16.

45. Sharda RD. Business intelligence and analytics: systems for decision support. Hoboken: Prentice Hall; 2016.

Asif *et al. J Big Data*        (2021) 8:83

Page 28 of 28

46. Russakovsky O, et al. Imagenet large scale visual recognition challenge. Int J Comput Vision. 2015;115(3):211–52.
47. Hui J. mAP (mean Average Precision) for object detection: Medium; 2018 [Available from: https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173.] Accessed 28 May 2021.
48. Giannakeris P, Avgerinakis K, Karakostas A, Vrochidis S, Kompatsiaris I. People and vehicles in danger-A fire and flood detection system in social media.  2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP); 10-12 June.; Aristi Village, Greece IEEE; 2018. p. 1-5. https://doi.org/10.1109/IVMSPW.2018.8448732
49. Okolloh O. Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. PLA. 2009;59(1):65–70.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.