



A Real-Time System for Detecting Landslide Reports on Social Media Using Artificial Intelligence

Ferda Offi¹ , Umair Qazi¹ , Muhammad Imran¹ , Julien Roch², Catherine Pennington³ , Vanessa Banks³ , and Remy Bossu^{2,4} 

¹ Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
{fofli,uqazi,mimran}@hbku.edu.qa

² European-Mediterranean Seismological Centre, Arpajon, France
{julien.roch,bossu}@emsc-csem.org

³ British Geological Survey, Keyworth, Nottinghamshire, UK
{cpoulton,vbanks}@bgs.ac.uk

⁴ CEA, DAM, DIF, 91297 Arpajon, France

Abstract. This paper presents an online system that leverages social media data in real time to identify landslide-related information automatically using state-of-the-art artificial intelligence techniques. The designed system can (i) reduce the information overload by eliminating duplicate and irrelevant content, (ii) identify landslide images, (iii) infer geolocation of the images, and (iv) categorize the user type (organization or person) of the account sharing the information. The system was deployed in February 2020 online at https://landslide-aidr.qcri.org/landslide_system.php to monitor live Twitter data stream and has been running continuously since then to provide time-critical information to partners such as British Geological Survey and European Mediterranean Seismological Centre. We trust this system can both contribute to harvesting of global landslide data for further research and support global landslide maps to facilitate emergency response and decision making.

Keywords: Landslide detection · Social media · Online system · Real time · Image classification · Computer vision · Artificial intelligence

1 Introduction

Landslides cause thousands of deaths and billions of dollars in infrastructural damage worldwide every year [16]. However, landslide events are often under-reported and insufficiently documented due to their complex natural phenomena oftentimes triggered by earthquakes and tropical storms, which are more conspicuous, and hence, more widely reported [18]. Therefore, any attempt to quantify global landslide hazards and the associated impacts remains an underestimation due to this oversight and lack of global data inventories [8].

Undertaking the challenge of building a global landslide inventory, NASA launched a website¹ in 2018 to allow citizens to report about the regional landslides they see in-person or online [14]. Following a similar Volunteered Geographical Information (VGI) approach, researchers further developed other means such as mobile or web applications to collect citizen-provided data [6, 17]. While VGI-based solutions prove helpful, they are not easily scalable as they require active participation of volunteers that opt-in to use a particular application to collect and share landslide-related data. Furthermore, this means the bulk of data collection and interpretation still involves time consuming work by specialists searching the Internet for news and reports, or directly engaging in communications with those submitting information [14, 17, 28, 36].

To alleviate the need for opt-in participation and manual processing, we developed an online system equipped with state-of-the-art AI models to automatically detect landslide reports posted on social media image streams in real time. The system was developed through an interdisciplinary collaboration between the computer scientists at the Qatar Computing Research Institute (QCRI) and the earthquake and landslide specialists from the European-Mediterranean Seismological Centre (EMSC) and the British Geological Survey (BGS), respectively. The developed system employs several supervised machine learning models to (i) deal with the noisy nature of the social media data by filtering out duplicate and irrelevant images, (ii) detect landslide reports by interpreting the retained images, (iii) infer the location information of the detected landslide reports from the available metadata, and (iv) identify the type of users that have shared the landslide reports. We deployed the system online in February 2020 to monitor live Twitter data stream and it has collected more than 54 million tweets and 15 million image URLs. Only about 2.5 million of these image URLs were deemed unique and downloaded for further analysis. Eventually, the system identified about 38,000 landslide reports worldwide, which corresponds to less than 1% of the collected image URLs and highlights the challenging nature of the problem. Despite this, quantitative verification of the system’s performance during a real-world deployment shows that our system can detect landslide reports with Precision = 76% and Recall = 74% (i.e., F1 = 75%).

2 Related Work

The literature on landslide detection and mapping approaches mainly uses four types of data sources: (i) physical sensors, (ii) remote sensing, (iii) volunteers, and (iv) social networks. Sensor-based approaches rely on land characteristics such as rainfall, altitude, soil type, and slope to detect landslides and develop models to predict future events [19, 31]. While these approaches can be highly accurate at sub-catchment levels, their large-scale deployment is extremely costly.

Earth observation data from high-resolution satellite imagery has been widely used for landslide detection, mapping, and monitoring [37]. Remote sensing techniques either use Synthetic Aperture Radar (SAR) or optical imagery to perform

¹ <https://gpm.nasa.gov/landslides/index.html> (accessed on 12 February 2022).

landslide detection in various formulations including classification, segmentation, object detection, and change detection [4, 13, 20, 29, 30, 35]. While remote sensing through satellites can be useful to monitor landslides globally, their deployment can prove costly and time-consuming.

A few studies demonstrated the use of Volunteered Geographical Information (VGI) as an alternative method to detect landslides [2, 3, 6, 17]. These studies assume active participation of volunteers to collect landslide data where the volunteers opt-in to use a mobile or web application to provide information such as photos, time of occurrence, damage description, and other observations about a landslide event. On the contrary, our work capitalizes on massive social media data without any active participation requirement and with better scalability.

Social media data has been used in many humanitarian contexts ranging from general social analytics [32] and geospatial sentiment analysis [1] to incident detection [38] and rapid damage assessment [10], including multimodal approaches [25]. However, its use for landslide detection has not been explored extensively. The most relevant work by Musaev et al. [21, 23] combines social media text data and physical sensors to detect landslides. In contrast, we focus on analyzing social media images which provide more detailed information about the impact of the landslide event. Therefore, our work complements prior art.

3 System Design

The system is designed to ingest data from an online social media platform (i.e., Twitter), analyze the incoming data, and process relevant information under the condition that all tasks must be performed in a time-sensitive manner. Figure 1 shows a high-level architecture of the system and its various critical components. Data flows from left to right through two types of connections between components. The red lines indicate streaming connections whereas the black lines represent on-demand connections. A streaming connection can be of two types (i) a publisher-subscriber channel, and (ii) a push-pop queue.

3.1 Data Collectors

We have two types of collectors. One collects data (i.e., tweets) directly from Twitter. The other one then downloads images corresponding to collected tweets.

Tweet Collector. This module uses the Twitter Streaming API² to collect live tweets. The Streaming API can provide data in various ways based on (i) a list of keywords, (ii) geographical bounding boxes, or (iii) both. The bounding box approach provides geo-tagged tweets which can be about any topic, thus we use the keyword-based approach through the AIDR system [11]. The keywords are related to multilingual landslide terminology and nomenclature such as landslide,

² <https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/guides/connecting> (accessed on 24 March 2022).

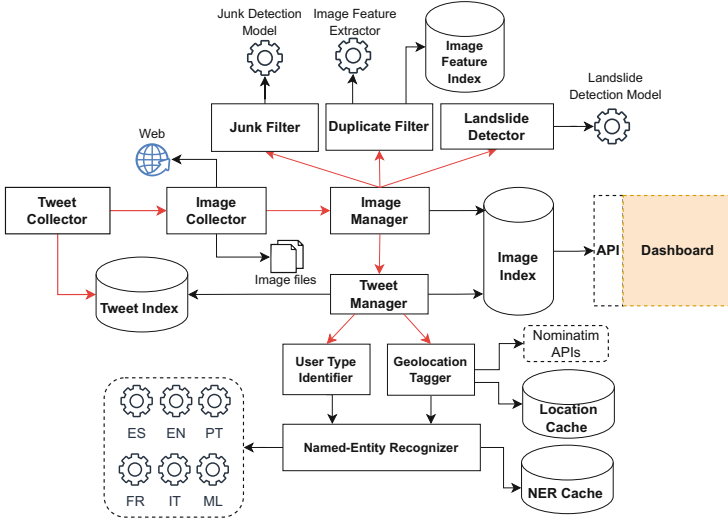


Fig. 1. System architecture with important components and communication flows

landslip, earth slip, mudslide, rockslide, and rock fall (Table 3). Tweets matching with at least one of the pre-specified keywords are acquired from Twitter in the JSON format and persisted into the Tweet Index, which is an Elasticsearch database. If a tweet contains one or more images, its *id* and *URLs* of all images are pushed to the Image Collector through a Redis³ queue.

Image Collector. This module parses image-related attributes dispatched by the Tweet Collector module and extracts image URLs and downloads corresponding images. Due to re-tweets, same image URLs may appear multiple times during the data collection. To avoid redundant downloads, the system keeps track of previously seen image URLs in an in-memory *linked hash map* which has $O(1)$ time complexity for adding and searching an element and $O(n)$ space complexity. The downloaded images are saved on the file system and their paths and tweet ids are pushed to the Image Manager queue for further processing.

3.2 Image Manager

The system has multiple modules that analyze images for different purposes. Two of these modules, namely Duplicate Filter and Junk Filter, are tasked to reduce the data noise by eliminating images that are (i) near-or-exact duplicate and (ii) irrelevant for general disaster response, respectively. The third module, Landslide Detector, is the core module that interprets each image as landslide or not-landslide. All image processor modules are managed by the Image Manager,

³ <https://redis.io/> (accessed on 24 March 2022).

which pops items from its queue and immediately dispatches to the three image processors (i.e., Junk Filter, Duplicate Filter, and Landslide Detector) through their respective queues. The Image Manager also monitors the output of all image processors to persist them into the main Image Index.

Duplicate Filter. Image-level deduplication is important to discard near-or-exact duplicate images that are often due to high retweeting activity. This module identifies duplicate images to prevent further processing as well as information overload on end users. The module acquires images from its input queue and checks whether a given image is near-or-exact duplicate of previously seen images. To this end, it first extracts features from each image using a deep learning model and then compares these features against an Image Feature Index to detect near-or-exact duplicate cases based on a distance threshold. The image feature index keeps a record of all unique image features. If the module identifies a near-or-exact duplicate, it returns the reference image’s id and the computed distance. Otherwise, it tags the image as “not-duplicate”. If the image is “not-duplicate”, then it is also inserted into the Image Feature Index. Section 4.1 presents details of the feature extracting model.

Junk Filter. Even though filtered through landslide-related keywords, the Twitter image stream carries images not pertaining to landslide incidents. Identifying these junk content is important to reduce information overload on end users. To this end, the Junk Filter module pops images from the input queue and processes them through the junk detection model, which outputs a class label (“relevant” or “not-relevant”) and a confidence score. More detailed information about the junk detection model is presented in Sect. 4.2. The processed images are pushed into the output queue of the module.

Landslide Detector. As the main objective of the system is to identify images showing landslide incidents, in this module we perform this task using a deep learning computer vision model. The module first acquires images from its input queue and passes them through the landslide classifier, which outputs a class label (“landslide” or “not-landslide”) and a confidence score. The landslide classifier is a deep learning image classification model that is presented in detail in Sect. 4.3. The classified images are pushed into the module’s output queue.

3.3 Tweet Manager

The system contains three modules, namely Geolocation Tagger, User Type Identifier, and Named-Entity Recognizer, that process textual content for different purposes. Specifically, Geolocation Tagger analyzes various tweet metadata fields to infer geolocation information while User Type Identifier focuses on identifying the type of Twitter account. Both modules use Named-Entity Recognizer to tag text tokens with named-entities.

Geolocation Tagger. Identifying the location of landslide incidents reported on Twitter is an important task. A tweet reporting a landslide with some image content may or may not have an explicit mention of the location in the text where the incident took place. In that case, other meta-data fields are examined to find location cues. These fields include, *GPS-coordinates*, *place*, *user location*, and *user profile description*. To this end, we use our geolocation tagging approach presented in [12] with a different field priority order. We observed that most tweets with landslide reporting images contain location cues in their text content. Therefore, if a tweet does not contain GPS-coordinates, we give high priority to the location names mentioned in the text. Place, user location, and user profile description come later in the order, respectively. The geolocation tagger uses the named-entity recognizer to get named-entities for tweet text and user profile description fields. The geolocation tagger uses Nominatim geocoding and reverse geocoding APIs and tags each tweet with country, state, county, and city information, when possible. More details of the geotagging approach can be found in [12]. The module maintains a cache of processed locations to increase its efficiency for recurring requests.

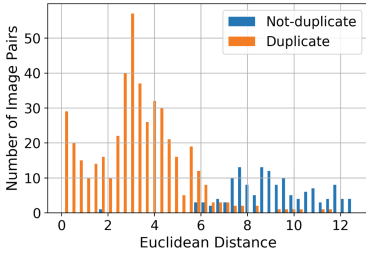
User Type Identifier. This module uses the name of the tweet author to determine whether the account is of type person or organization. Landslide incidents reported by personal accounts are more important for our end users than those reported by organizational accounts. For this purpose, we use the English NER model through the Named-Entity Recognizer module, which tags name tokens with one of the several predefined named-entities, including PERSON.

Named-Entity Recognizer. As described above, both Geolocation Tagger and User Type Identifier modules use Named-Entity Recognizer to perform their operation. To support these operations for multilingual tweets, we use five NER models representing five international languages, including English, French, Spanish, Portuguese, and Italian. Additionally, we use a multilingual NER model (denoted as ML) for all other languages. All of these multilingual models and their performance scores are publicly available at spaCy website.⁴ This module also maintains a cache of processed NER requests to increase its efficiency for recurring requests.

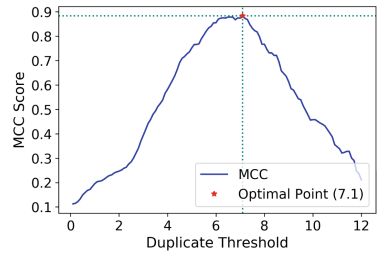
4 Experiments

In this section, we first describe the design and development of our image models and present experimental results. Then, we present performance evaluation and benchmarking results for the most critical components of the system. For image models, we follow the popular transfer learning approach based on convolutional neural networks (CNNs) as many studies have shown that features learned by CNNs are effectively transferable between different visual recognition tasks [7, 27, 34], particularly when training samples are limited.

⁴ <https://spacy.io/usage/models> (accessed on 24 March 2022).



(a) Histogram of pairwise distances



(b) MCC vs. duplicate threshold

Fig. 2. Optimal duplicate distance threshold determination: (a) Distribution of the Euclidean distances between the image pairs in the duplicate test set. (b) MCC performance as a function of distance threshold.

4.1 Duplicate Filtering

The Duplicate Filter is responsible for extracting a feature vector from a given image using a state-of-the-art deep learning model and comparing this feature vector with the feature vectors of previously seen images based on a pre-defined distance threshold d . For this purpose, we extract deep features from the penultimate layer of a ResNet-50 model [9] pre-trained on the Places data set [39], which comprises 10 million images collected for scene recognition.⁵ Each feature vector has a size of 2,048. To determine the optimal distance threshold d , we performed experiments on a manually annotated set of 600 image pairs including 460 duplicate and 140 non-duplicate cases with varying pairwise distances (Fig. 2a). We used Euclidean distance metric (i.e., L2 norm) to measure the distance between two image feature vectors. Note that image pairs with a distance greater than 12.5 looked trivially distinct, and hence, we did not include them in our experiments. We then performed a grid search over a range of threshold values from 0 to 12 with a step size of 0.1 and measured the performance of each threshold value by computing the Matthew’s Correlation Coefficient (MCC), which is regarded as a balanced measure for imbalanced classification problems [5]. As depicted in Fig. 2b, the optimal performance is achieved when the duplicate distance threshold is $d = 7.1$.

4.2 Junk Classification

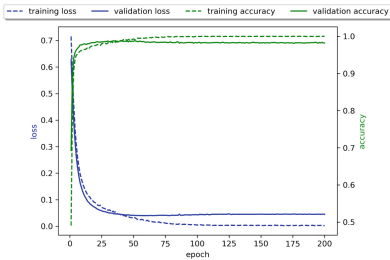
The Junk Filter employs a CNN model to determine whether an image is relevant or not for general emergency management and response. To this end, we took a ResNet-50 model [9] pre-trained on ImageNet [33], adopted its final layer to binary classification task, and fine-tuned it on a custom data set introduced by Nguyen et al. [24]. We merged the validation set with the training set, and used the test set to evaluate the performance of the model as summarized in

⁵ The pre-trained model is available at http://places2.csail.mit.edu/models_places365/resnet50_places365.pth.tar (accessed on Jan 23, 2022).

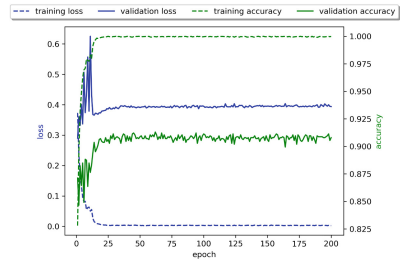
Table 1a. We used Adam optimizer [15] with an initial learning rate of 10^{-6} and configured the `ReduceLROnPlateau` scheduler to decay the learning rate by 0.1 with a patience of 50 epochs. We trained the model for a total of 200 epochs. The training process of the junk classification model is plotted in Fig. 3a and its performance evaluation is presented in Table 1b. The model achieves almost perfect performance in all measures due to the distinct features between relevant and not-relevant images in the training data set. We note that early stopping the model training after 100 epochs would yield the same model performance.

Table 1. Details of the data set used for training the junk classification model and the performance of the trained model on the test set.

(a) Training data set				(b) Model performance (Acc: 98.79)			
Class	Train	Test	Total	Class	Precision	Recall	F1
Relevant	2,814	704	3,518	Relevant	98.31	99.29	98.80
Not-relevant	2,814	704	3,518	Not-relevant	99.28	98.30	98.79
Total	5,628	1,408	7,036	Macro avg.	98.80	98.79	98.79



(a) Junk classification



(b) Landslide classification

Fig. 3. Model training progress in terms of accuracy and loss achieved on the training and validation sets

4.3 Landslide Classification

The Landslide Detector is the most important component of the proposed system. Therefore, we performed a separate, comprehensive study to identify the optimal configuration for the landslide classification model [26]. To recap, we first created a large landslide image data set labeled by landslide specialists, who are also co-authors of this paper. The data set contains 11,737 images,

which are split into training, validation, and test sets as shown in Table 2a. Then, adopting a transfer learning approach, we conducted an extensive set of experiments using various CNN architectures with different optimizers, learning rates, weight decays, and class balancing strategies. The winning model configuration is a ResNet-50 architecture trained using Adam optimizer with an initial learning rate of 10^{-4} , a weight decay of 10^{-3} , and without a class balancing strategy. Figure 3b displays the training progress of the best performing landslide classification model, which is also integrated into our system, whereas Table 2b summarizes the performance of the model on the test set. As with the junk classification model, the plot indicates that the model training process could be early stopped after 100 epochs without affecting the model performance.

Table 2. Details of the data set used for training the landslide classification model and the performance of the trained model on the test set.

(a) Training data set					(b) Model performance (Acc: 86.97)			
Class	Train	Val	Test	Total	Class	Precision	Recall	F1
Landslide	1,883	271	536	2,690	Landslide	73.66	66.79	70.06
Not-landslide	6,332	902	1,813	9,047	Not-landslide	90.45	92.94	91.68
Total	8,215	1,173	2,349	11,737	Macro avg.	82.05	79.87	80.87

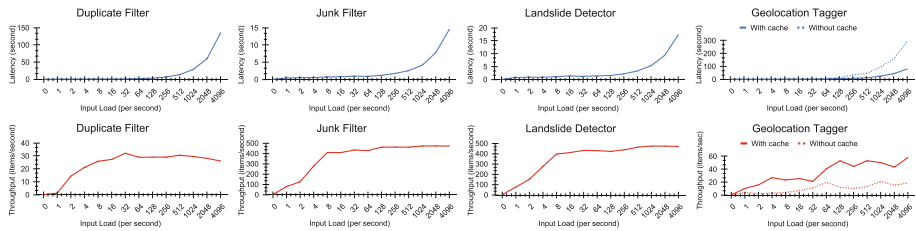


Fig. 4. Latency (top) and throughput (bottom) of the Junk Filter, Duplicate Filter, Landslide Detector, and Geolocation Tagger (left to right).

4.4 Performance Evaluation and Benchmarking

To stress-test the system and understand its scalability, we conducted performance experiments on four critical modules, i.e., Duplicate Filter, Junk Filter, Landslide Detector, and Geolocation Tagger. We use latency and throughput, as they are considered reliable measures to test a system’s performance. In our case, the latency is the time taken by a module to process a given input load consisting of images. Whereas, the throughput is the number of images processed in a unit time (one second) given an input load. The experiments were

conducted using a pool of 50,000 images. We developed a simulator to mimic the functionality of the Image Collector. The simulator pushed varying amounts of input loads to Redis channels, which were then consumed by modules. Based on the real-world deployment, we observed that the input load reaches a maximum of 0.08 images per second (on average). Therefore, we tested a range of input loads defined as 2^n , $n \in \{0, 1, \dots, 12\}$. We performed the tests on a Linux server with 256GB RAM, 2.2GHz processor with 32 cores and two Tesla V100 GPUs with 16 GB.

Figure 4 shows the performance results. The latency for all modules follows the same pattern, i.e., as the input load (per second) increases, the latency also increases. However, as the computational responsibilities of each module differ, so do their latencies at different input loads. For instance, both Relevancy Filter and Landslide Detector show a decent latency of around five seconds even at 1024 input load. The Duplicate Filter, however, exhibits high latency (i.e., 29s) at the same load. The latency for Geolocation Tagger is measured with and without cache, which makes a significant difference. The cache keeps a record of all existing unique requests and hence, on average, the latency of the cached version is about four times less.

In terms of throughput, Relevancy Filter and Landslide Detector maintain a high throughput of more than 400 images/second, even at the maximum input load. Throughput for Junk Filter reaches module capacity at 467 images/second on average and for Landslide Detector it goes up to 457 images/second on average. For Duplicate Filter, the throughput initially increases but then starts decreasing as the size of Image Feature Index grows. The throughput is also about 4 times higher on average with cache compared to without cache. Geolocation Tagger reaches its capacity at about 50 images per second with cache. With an empty cache, it goes as high as 21 images per second on average with cache and is expected to increase as the cache grows in size.

5 Real-World Deployment

Here we present details about our real-world deployment including data collection and statistics, quantitative verification of the detected landslide reports, and a comparison with a text-based approach.

5.1 Data Collection and Statistics

In February 2020, we launched the system online at https://landslide-aidr.qcri.org/landslide_system.php to monitor live Twitter stream for landslide-related reports. Figure 5 shows a snapshot of the system dashboard. We note that, by landslide, we refer to all downward and outward movement of loosened slope materials such as landslip, debris flows, mudslides, rockfalls, earthflows, and other mass movements. As mentioned in Sect. 3.1, the system follows a keyword-based data collection strategy. Hence, we curated a list of 339 multilingual keywords covering all types of landslides in 32 languages including English, Albanian, Arabic, Basque, Bengali, Bosnian, Catalan, Chinese, Croatian, Dutch,



AIDR Twitter Image Analysis: Detecting Landslides, Mudslides, Rockslides.

Total images collected: **2,527,175**

Data available from 2020-02-04 to 2021-12-29
 Start date: [12/29/2021, 12:00:00 AM] End date: [12/29/2021, 11:59:59 PM] [Apply]
 Remove duplicates: YES NO
 Remove irrelevant: YES NO
 Show landslide images: YES NO
 User type: Person Organization All
 Location filter:
 Country: [All] State: [Not available] City: [Not available]
 Include non-geotagged images: YES NO

Total images: 1,886

Filtered images: 7

*All times are shown in UTC.

Images Layout Map Layout

(=) : Inferred location
 (o) : Exact location

Page 1 of 1 prev next

LANDSLIDE	LANDSLIDE	LANDSLIDE	LANDSLIDE	LANDSLIDE
<p>Time: 29-Dec-2021 13:04:34 From: Servicio de Información Pública Text: C12 Un deslizamiento de tierra a causa de las lluvias mantiene cerrada parte de la calzada en la carretera local 004, específicamente en la quebrada La Boba en Campo Elias, #Mérida. Autoridades atienden la situación para despejar la vía. :OC @ivanive3 Location (=): Kinshasa, Capital District, Congo Tweet link</p>	<p>Time: 29-Dec-2021 13:04:34 From: Servicio de Información Pública Text: C12 Un deslizamiento de tierra a causa de las lluvias mantiene cerrada parte de la calzada en la carretera local 004, específicamente en la quebrada La Boba en Campo Elias, #Mérida. Autoridades atienden la situación para despejar la vía. :OC @ivanive3 Location (=): Kinshasa, Capital District, Congo Tweet link</p>	<p>Time: 29-Dec-2021 11:00:25 From: penerangan 071/WK Text: Hujan deras, Dua Rumah Warga Rusak Terkena Tanah Longsor Location: N/A Tweet link</p>	<p>Time: 29-Dec-2021 05:56:14 From: Ingeniería en la Red @pantoja_niza #PC-142#E #2#B #0#2#0#E, un tipo de deslizamiento traslacional que está controlado por dos o más discontinuidades (estratificación, esquistosidad, diaclasas, fallas). Location (=): Andalusia, Spain Tweet link</p>	<p>Time: 29-Dec-2021 01:15:16 From: Obras Quito Text: 69 Trabajamos en la limpieza de tierra y escombros por deslizamiento de talud sobre la vía en la av. Simón Bolívar, en el sector del Mirador de Cumbayá. #PortoQuitoDigno Location (=): Pichincha, Ecuador Tweet link</p>

Fig. 5. Snapshot of the online system dashboard

French, Georgian, German, Greek, Hindi, Hungarian, Indonesia, Iranian, Italian, Japanese, Korean, Malaysia, Philippines, Polish, Portuguese, Romanian, Russian, Sesotho, Slovenian, Spanish, Swedish, and Turkish (Table 3).

Since its deployment until December 31, 2021, the system has collected more than 54 million tweets and 15 million image URLs, out of which about 2.5 million were deemed unique and downloaded for further analysis. Figure 6 depicts the weekly volume of raw tweets and images collected during this time period as well as the distributions of images filtered by the Junk Filter, Duplicate Filter, and Landslide Detector. The data do not show any gaps, which is an important factor for robust monitoring of real-world events continuously. On average, the Junk Filter eliminates around 76% of the collected images, the Duplicate Filter further reduces the redundancy by an additional 9%, and finally, the Landslide Detector classifies only 0.84% of the remaining 15% images as landslides. This corresponds to a significant (i.e., more than 99%) reduction of information overload for our end users. 6,523 of all the detected landslide reports were shared by personal accounts and 4,553 by organizational accounts. Figure 7 shows the worldwide

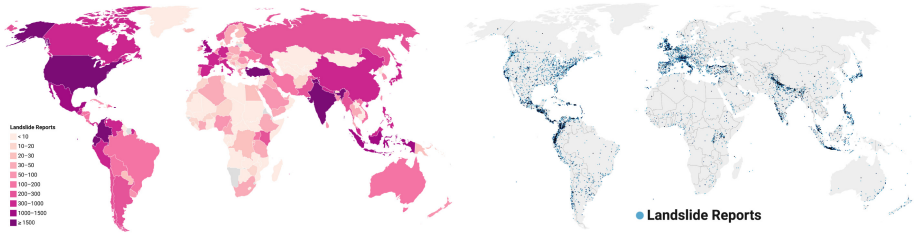


Fig. 7. Worldwide distribution of the collected landslide reports

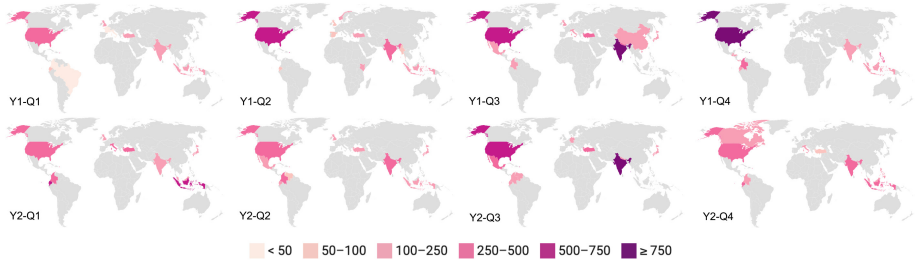


Fig. 8. Top-10 countries with the highest landslide reports in each quarter

5.2 Validation of the Landslide Model Predictions

Although the system has collected more than 2.5 million images since its deployment in February 2020, there are only about 17,000 images labeled as landslide (or 38,000 images including near-and-exact duplicates), which corresponds to less than 1% of the total volume. This highlights the difficulty of the task even though a carefully curated set of landslide-related keywords has been used to collect data from Twitter. To validate the performance of the landslide model in the real-world deployment, we sampled $N = 3,600$ tweets with images collected by our system. To avoid overburdening our landslide specialists with noisy data as well as to warrant robust statistics, we sampled only from the subset of tweets with images labeled as non-duplicate and relevant. Our landslide specialists then reviewed these images and annotated them with ground-truth landslide/not-landslide labels. Eventually, we compared the machine-predicted labels with expert annotations to evaluate the performance of the landslide model in a real-world scenario. Table 4 summarizes the number of correct (i.e., True Positive (TP) and True Negative (TN)) and incorrect (i.e., False Positive (FP) and False Negative (FN)) predictions as well as the performance scores such as accuracy, precision, recall, F1, and MCC. Overall, we see that the performance of the model in a real-world scenario is comparable to the experimental results (Sect. 4.3).

Table 4. Validation of landslide model predictions

TP	FP	FN	TN	Total
123	39	43	3,395	3,600
Accuracy	Precision	Recall	F1	MCC
97.72	75.93	74.10	75.00	73.81

5.3 Comparison with a Text-Based Approach

Text-based landslide detection is a nascent problem that only a couple of studies have addressed so far [22,23]. Since these studies did not share their data sets and models, we do not have any off-the-shelf text-based landslide classification model to use as a baseline in our study. Therefore, we consider an alternative scenario with a *proxy* text classification model based on lexicon (i.e., keyword) matching, which is already implemented in our system. That is, we assume all the retrieved tweets are already labeled as landslide by a hypothetical model. We then use the previously sampled set of tweets with their expert annotations to compute the precision of a lexicon-based text model. Unsurprisingly, we found that the lexicon-based text model achieved only about 5% precision (i.e., only about 5% of the tweets retrieved were indeed related to landslides) while the image classification model achieved 76% precision as reported before.

6 Conclusion

In this paper, we presented a system that was developed through an interdisciplinary collaboration between the computer scientists at the Qatar Computing Research Institute (QCRI) and the earthquake and landslide specialists from the European-Mediterranean Seismological Centre (EMSC) and the British Geological Survey (BGS), respectively. The developed system leverages online social media data in real time to identify landslide-related information automatically using state-of-the-art artificial intelligence techniques. The designed system (i) reduces the information overload by eliminating duplicate and irrelevant content, (ii) identifies landslide images, (iii) infers their geolocation, and (iv) categorizes the user type (organization or person) of the account sharing the information. We presented results of our model development as well as system performance evaluation and benchmarking experiments. We demonstrated the system’s success with a real-world deployment. We note that some of the system components, such as the Geolocation Tagger, can be further improved for better efficiency. Online dashboard can also be enriched with a mechanism to collect user feedback. Besides, the landslide reports detected by the system can be verified against “ground truth” from official reports (when/if available), which we plan as future work. Consequently, we believe that our system can contribute to harvesting of global landslide data and facilitate further landslide research. Furthermore, it can support global landslide susceptibility maps to provide situational awareness and improve emergency response and decision making.

Acknowledgments. The British Geological Survey (UK Research and Innovation) granted supporting research funding through National Capability (Shallow Geohazards) and Innovation funding streams. European-Mediterranean Seismological Centre was partially funded by the European Union's (EU) Horizon 2020 Research and Innovation Program under Grant Agreement RISE Number 821115. Opinions expressed in this article solely reflect the authors' views; the EU is not responsible for any use that may be made of information it contains.

References

1. Alfarrarjeh, A., Agrawal, S., Kim, S.H., Shahabi, C.: Geo-spatial multimedia sentiment analysis in disasters. In: DSAA, pp. 193–202 (2017)
2. Can, R., Kocaman, S., Gokceoglu, C.: A convolutional neural network architecture for auto-detection of landslide photographs to assess citizen science and volunteered geographic information data quality. *ISPRS Int. J. Geo-Inf.* **8**(7), 300 (2019)
3. Can, R., Kocaman, S., Gokceoglu, C.: Development of a CitSci and artificial intelligence supported GIS platform for landslide data collection. *Int. Arch. Photogramm.* **43**, 43–50 (2020)
4. Cheng, G., Guo, L., Zhao, T., Han, J., Li, H., Fang, J.: Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA. *Int. J. Remote Sens.* **34**(1), 45–59 (2013)
5. Chicco, D., Jurman, G.: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **21**(1), 1–13 (2020)
6. Choi, C.E., Cui, Y., Zhou, G.G.: Utilizing crowdsourcing to enhance the mitigation and management of landslides. *Landslides* **15**(9), 1889–1899 (2018)
7. Donahue, J., et al.: DeCAF: a deep convolutional activation feature for generic visual recognition. In: ICML, pp. 647–655 (2014)
8. Froude, M.J., Petley, D.N.: Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazard Earth Syst.* **18**(8), 2161–2181 (2018)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
10. Imran, M., Alam, F., Qazi, U., Peterson, S., Ofli, F.: Rapid damage assessment using social media images by combining human and machine intelligence. In: ISCRAM, pp. 1–13, May 2020
11. Imran, M., Castillo, C., Lucas, J., Meier, P., Vieweg, S.: AIDR: artificial intelligence for disaster response. In: WWW, pp. 159–162 (2014)
12. Imran, M., Qazi, U., Ofli, F.: TBCOV: two billion multilingual COVID-19 tweets with sentiment, entity, geo, and gender labels. *Data* **7**(1), 8 (2022)
13. Ji, S., Yu, D., Shen, C., Li, W., Xu, Q.: Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides* **17**(6), 1337–1352 (2020). <https://doi.org/10.1007/s10346-020-01353-2>
14. Juang, C.S., Stanley, T.A., Kirschbaum, D.B.: Using citizen science to expand the global map of landslides: introducing the Cooperative Open Online Landslide Repository (COOLR). *PLoS ONE* **14**(7), e0218657 (2019)
15. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: ICLR (2015)
16. Kjekstad, O., Highland, L.: Economic and social impacts of landslides. In: Sassa, K., Canuti, P. (eds.) *Landslides – Disaster Risk Reduction*. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-69970-5_30

17. Kocaman, S., Gokceoglu, C.: A CitSci app for landslide data collection. *Landslides* **16**(3), 611–615 (2018). <https://doi.org/10.1007/s10346-018-1101-2>
18. Lee, E.M., Jones, D.K.: *Landslide Risk Assessment*, vol. 10. Thomas Telford, London (2004)
19. Merghadi, A., et al.: Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance. *Earth-Sci. Rev.* **207**, 103225 (2020)
20. Mohan, A., Singh, A.K., Kumar, B., Dwivedi, R.: Review on remote sensing methods for landslide detection using machine and deep learning. *Trans. Emerg. Telecommun. Technol.* **32**(7), e3998 (2021)
21. Musaev, A., Wang, D., Pu, C.: LITMUS: landslide detection by integrating multiple sources. In: ISCRAM (2014)
22. Musaev, A., Wang, D., Shridhar, S., Pu, C.: Fast text classification using randomized explicit semantic analysis. In: IRI, pp. 364–371. IEEE (2015)
23. Musaev, A., Wang, D., Xie, J., Pu, C.: REX: rapid ensemble classification system for landslide detection using social media. In: ICDCS, pp. 1240–1249. IEEE (2017)
24. Nguyen, D.T., Alam, F., Ofli, F., Imran, M.: Automatic image filtering on social networks using deep learning and perceptual hashing during crises. In: ISCRAM, pp. 499–511, May 2017
25. Ofli, F., Alam, F., Imran, M.: Analysis of social media data using multimodal deep learning for disaster response. In: ISCRAM, pp. 1–10, May 2020
26. Ofli, F., et al.: Landslide detection in real-time social media image streams. arXiv preprint [arXiv:2110.04080](https://arxiv.org/abs/2110.04080) (2021)
27. Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Learning and transferring mid-level image representations using convolutional neural networks. In: CVPR, pp. 1717–1724 (2014)
28. Pennington, C., Freeborough, K., Dashwood, C., Dijkstra, T., Lawrie, K.: The national landslide database of great Britain: acquisition, communication and the role of social media. *Geomorphology* **249**, 44–51 (2015)
29. Prakash, N., Manconi, A., Loew, S.: Mapping landslides on EO data: performance of deep learning models vs. traditional machine learning models. *Remote Sens.* **12**(3), 346 (2020)
30. Prakash, N., Manconi, A., Loew, S.: A new strategy to map landslides with a generalized convolutional neural network. *Sci. Rep.* **11**(1), 1–15 (2021)
31. Ramesh, M.V., Kumar, S., Rangan, P.V.: Wireless sensor network for landslide detection. In: ICWN, pp. 89–95 (2009)
32. Razis, G., Theofilou, G., Anagnostopoulos, I.: Latent twitter image information for social analytics. *Information* **12**(2), 49 (2021)
33. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**(3), 211–252 (2015). <https://doi.org/10.1007/s11263-015-0816-y>
34. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: OverFeat: integrated recognition, localization and detection using convolutional networks. In: ICLR, April 2014
35. Tavakkoli Piralilou, S., et al.: Landslide detection using multi-scale image segmentation and different machine learning models in the higher Himalayas. *Remote Sens.* **11**(21), 2575 (2019)
36. Taylor, F.E., Malamud, B.D., Freeborough, K., Demeritt, D.: Enriching great Britain’s national landslide database by searching newspaper archives. *Geomorphology* **249**, 52–68 (2015)
37. Tofani, V., Segoni, S., Agostini, A., Catani, F., Casagli, N.: Use of remote sensing for landslide studies in Europe. *Nat. Hazard Earth Syst.* **13**(2), 299–309 (2013)

38. Weber, E., et al.: Detecting natural disasters, damage, and incidents in the wild. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12364, pp. 331–350. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58529-7_20
39. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: a 10 million image database for scene recognition. PAMI **40**(6), 1452–1464 (2017)