Deliverable

D5.7

Detection of landslides and fires from Twitter monitoring

Deliverable information	
Work package	[D5.7 WP5]
Lead	[EMSC]
Authors	[Julien Roch, EMSC; Remy Bossu, EMSC]
Reviewers	[Gianfranco Vanucci, UNIBO]
Approval	[Management Board]
Status	[final]
Dissemination level	[Public]
Delivery deadline	[30.04.2022]
Submission date	[17.05.2022]
Intranet path	[DOCUMENTS/DELIVERABLES/File Name]

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821115

Table of Contents

Summary

Indirect effects (tsunami, fires, and landslides) may significantly contribute to the overall impact of a given earthquake. The specify of landslides is that even when they do not directly impact population they can still have adverse effects by blocking roads and hampering rescue operations. In this research, we aim at detecting possible earthquake triggered landslides in order to improve response efficiency.

In practice, we develop a methodology which first harvest tweets containing both keywords related to landslides and a picture, we then trained an artificial intelligence (AI) system to select the small proportion of pictures actually related to landslides, the final analysis and validation being done by human.

The AI system reduces the amount of images to analyze by 99%. To classify the images, a training model has been developed by a multidisciplinary team involving the Euro-Mediterranean Seismological Center (EMSC), the Qatar Computing Research Institute (QCRI) and the landslide team of the British Geophysical Survey (BGS). Training the system has been a huge part of the work. Real-time performance evaluation of the model shows that the system can detect a landslide with a precision of 76%.

This work initially developed for earthquake-triggered landslides only has attracted interest from the landslide community with the involvement of the BGS landslide team. Thanks to them, the scope has been widened to a prototype service now called "Global Landslide Reporter", 2 papers have been submitted for publication (one accepted, one in review) and once they are both published a post on the reference landslide blog will be released. The review process has already initiate a discussion to link our prototype service to landslide satellite observations. More generally, the purpose of the blog post will be to invite the landslide community to test the prototype service, hopefully improve it and turn it to an operational and widely used service.

The demonstrator is currently running in real-time and is available to the community (https://landslide-aidr.qcri.org/service.php).

1. Introduction

Landslides are responsible of thousands of death and can cost up to 1 billion every year. Landslides may hamper rescue operation by blocking roads (Figure 1), especially when it happens in mountain when only one road is available (Figure 1 - right). Having information quickly after landslides occur can contribute to save many lives and improve response efficiency. A small number of landslides are permanently monitored otherwise they are the subject of direct or satellite observations.

Figure 1: Examples of roads blocked by a landslide

Smartphone applications or web services have been developed to allow citizen to report a landslide. They have to use a specific application and landslide specialists are still needed to analyze and search for more information on the web. Some agencies also manually harvest landslide observations on social media but it is a labor extensive process.

Our approach is based on a machine learning system that collect and classify landslide-related images from twitter in real-time and automatically. In 2019, 550 million tweets were send Twitter user every day (5787 tweets/sec) (Phengsuwan et al. 2021). While Twitter is already in use at EMSC to detect felt earthquakes, finding information about landslides on twitter is like finding a needle in haystack. There is lots of noisy information which makes the extraction of relevant information difficult. While an earthquake can be felt hundreds of kilometers away from the epicenter location, a landslide is very local phenomenon which makes it difficult to witness. The Figure 2 show a comparison of the number of tweets containing the keyword “earthquake” after an earthquake in California in this example (left) and the number of tweets containing the keywords landslide (red curve) over 7 month during 6 hours window (right) superimpose with earthquake magnitudes > 5 (blue dots). It shows that while after an earthquake, the number tweets increase really fast after the quake (the vertical red line on the left) and the number of tweets is very important. Figure 2 (right) shows that the number of tweets remains very small (~5 tweets/min) even when 32 languages are monitored.

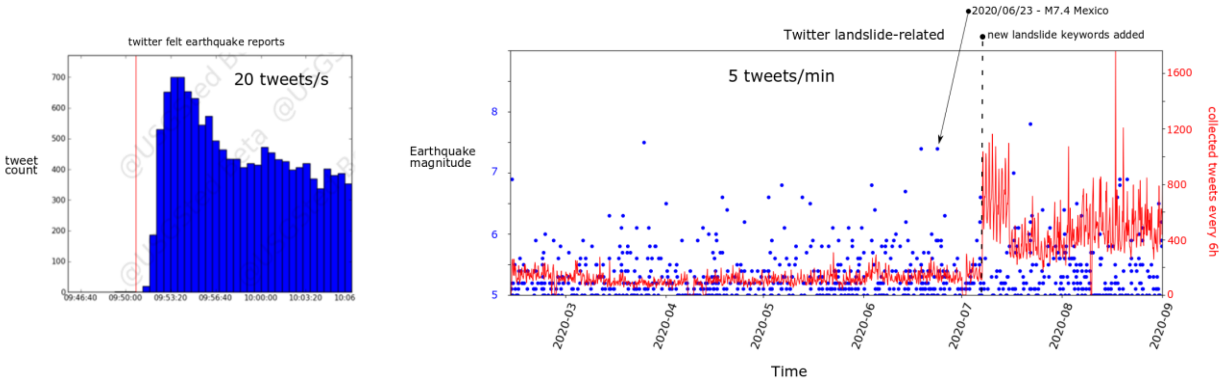


Figure 2: Comparison of the number of tweet containing the keyword "earthquake" after an earthquake in California in this example (left) and the number of tweet containing the keyword landslide (right) over 6 hour window.

In this project, involving three teams, the Euro-Mediterranean Seismological Center (EMSC), the Qatar Computing Research Institute (QCRI) and the landslide team of the British Geophysical Survey (BGS), we will present a new method based on artificial intelligence using the Artificial Intelligence for Disaster Reduction (AIDR) developed by QCRI. The system collects tweets based on landslide-related keywords, it filters out irrelevant and duplicate images to reduce the amount of data and classify the remaining image into landslide and not landslide (Figure 3).

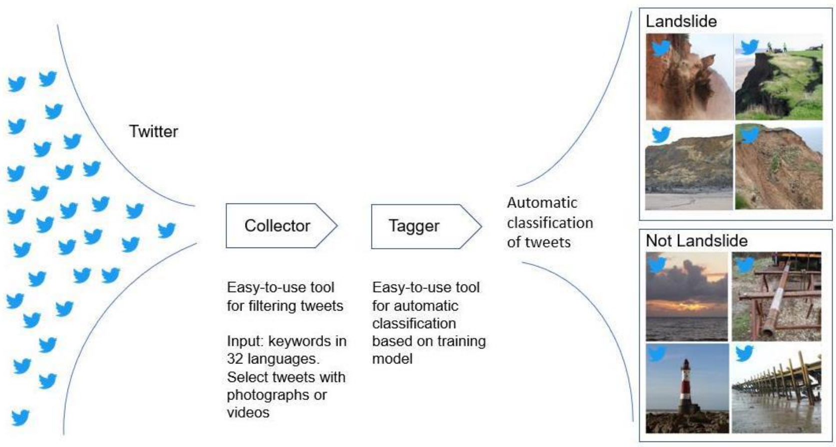


Figure 3: Graphical representation illustrating the workflow involved in collecting, tagging and *classifying images from Tweets as 'landslide' and 'not landslide' (from Pennington et al. 2022). Photographs BGS © UKRI [2022].*

The initial goal was to detect earthquake-triggered landslides and fires by real time monitoring of Twitter. No tweets related to triggered fires have been collected, so we only focused on landslides, and as shown on Figure 2, the number of tweet reporting landslide is very small, so we decided to extend our search to all landslides regardless of the trigger.

In the next section, we will present our methodology, and discuss our results. This project has led to 2 submitted papers in appendix and a future one in Dave Petley's blog (https://blogs.agu.org/landslideblog/).

2. Methodology

Using artificial intelligence (AI) requires 2 main steps. First, create a large dataset of manually annotated images of landslide and no landslide and then train the AI based on the manually annotated dataset.

We have created a large dataset of landslides images gathered from google (6284), twitter (1153) using landslide keywords and the GeoScenic BGS database (4300) (BGS, 2021). In total, we have collected 11737 landslides images. All images have been independently labelled as landslide and not landslide by 3 persons from EMSC and BGS based on a methodology defined in Pennington et al. 2022 (in appendix). To ensure a good agreement between the landslides specialists, two statistical measures have been carried out based, one based on the Fleiss' Kappa (Fleiss, 1971) and a percentage agreement. The results show a good agreement between the specialists with 76% agreement and Fleiss' Kappa score of 0.58 which correspond to almost "substantial agreement". The distribution of the images in the training dataset is summarized in Table 1. 23% of images dataset correspond to landslide. This low percentage show why we can't only rely on landslide keywords to collect landslide data.

Table 1: Distribution of the images across data sources.

	Google	Twitter	GeoScenic	Total
Landslide	1240	598	852	2690
Not landslide	5044	555	3448	9047
Total	6284	1153	4300	11737

The manual labelling methodology is detailed with examples in Pennington et al. 2022 in appendix. The system has been trained using a convolutional neural network with our training dataset. The Figure 4 shows on few examples how the AI is interpreting the raw images collected using a heat map which show the zone where the AI is analyzing.

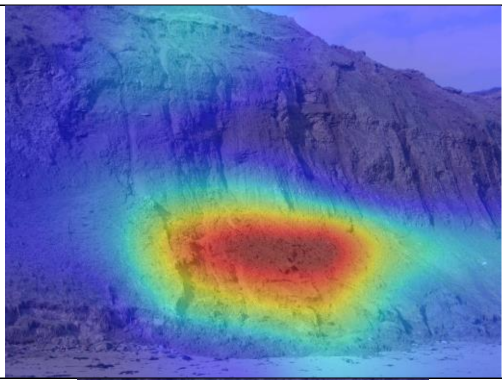
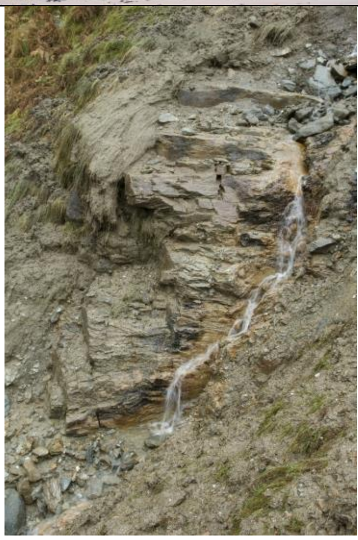
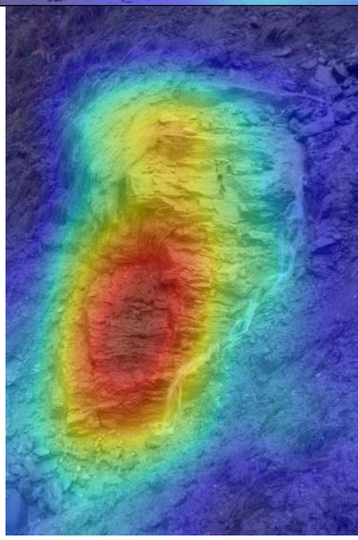
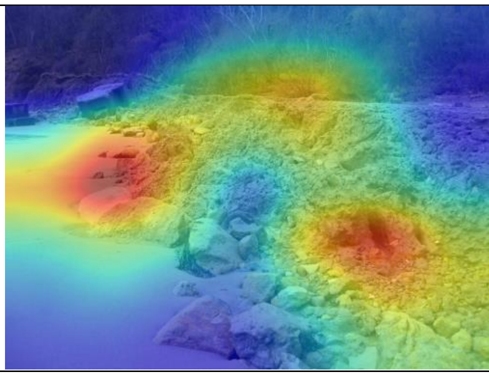
	Original images	Heat map prediction	Model prediction
a			Landslide 99%
b			Not a landslide 100%
c			Not a landslide 100%

Figure 4: Examples of the raw images collected on the left and the prediction associated, by the AI with the help of a heat map which show where the AI is looking at.

3. Real-time landslide classification

The real time landslide AIDR platform (https://landslide-aidr.qcri.org/landslide_system.php) has been monitoring twitter data between February 2020 and December the 29th 2021. Images are first collected from a text-based approach. We developed a multi-lingual list of 339 landslide-related keywords in 32 languages (see appendix 1). Keywords such as landslip, debris flow, mudslides, rockfalls, avalanche… are also searched.

During this period, 2.5 million unique landslide-related images have been analyzed and only 17000 images were labeled as landslide, corresponding to only 1% of the collected images. To validate the landslide model in real-time and because of the huge amount of data, we sampled 3600 tweets with images collected that have been labelled by the system. These 3600 images have been reviewed by the 3 landslide specialists and compared by the machine-predicted labels. The results are summarized in Table 2.

Table 2: Validation of landslide model predictions.

True positive (TP)	False Positive (FP)	False Negative (FN)	True Negative (TN)	Total
123	39	43	3395	3600

True positive and True Negative correspond to correct label by the system, while false positive (the model incorrectly classifies a not-landslide image as landslide) and false negative (the model incorrectly classifies a landslide image as not-landslide).

The performance of the model shows that the AI is able to detect a landslide with a precision of 76%, where the precision is defined as the ratio TP/(TP+FP) and it is a better statistics when dealing with imbalance distribution.

4. Examples

Even though we extend our search to all landslides regardless of the trigger, we present 2 examples of landslide triggered by earthquakes. The Figure 5 shows a landslide/rockfalls triggered by a magnitude 7.4 in Oaxaca Province in (Mexico) the 2020-06-23T15:29:04 UTC. The rockfalls completely blocked a road, and a car has been crushed by boulders (Figure 5-c). It is also a typical example where a road is closed and could impede any rescue operations. The two first images have been collected by the AIDR platform 5h45 after the earthquake origin time, while the Figure 5-c has been collected 7h20 after the earthquake origin time. The earthquake occurred in the morning local time.

The Figure 6 shows different landslides collected automatically by the AIDR platform after a magnitude 7.1 in Japan the 2021-02-13 14:07:51 UTC. The earthquake occur during night time in Japan (23:07 Japan time), thus only few images were collected in the first hours. The first image (Figure 6a) has been collected 1h45 after the earthquake and is a photo taken on a TV showing the landslide on the news. The three other images have collected the day after, around 10 hours after the earthquake origin time.

Figure 5: Examples of landslide images collected automatically by the AIDR platform few hours after the Oaxaca M7.4 earthquake the 2020-06-23T15:29:04 UTC.

Figure 6: Examples of images harvested automatically by the AIDR platform few hours after a Japanese M7.1 earthquake the 2021-02-13 14:07:51 UTC.

5. Discussion

The initial objective of the project was to detect earthquake-triggered landslide on Twitter. It has been extended to a "Global Landslide Reporter" prototype. The system collect images based on keywords, then filters out irrelevant and duplicate images using an AI approach.

The demonstrator used during this study is available at https://landslide-aidr.qcri.org/landslide_system.php. A new interface is now publicly available at https://landslide-aidr.qcri.org/service.php which includes the latest updates. It allows the user to check live system, to filter the data by date, country and explore all data on a map. The user can also give feedback on the system. 2 papers have been submitted (appendix 2) and in order to gain visibility from the landslide community, we will post on Dave Petley's blog which most people in the landslide community read.

Tweets are no longer geolocated by default, which mean that except if a user chooses to share its location or if a user write his location on the tweet, the location is no longer available. The current version inferred the location based on the texts within the tweet when no location is explicitly available (Imran et al. 2022). Although not perfect, it allows displaying the landslide on the map but as for all harvesting results from Twitter, the location has to be manually checked.

As shown in Figure 2 (right) adding more keywords in various languages strongly increase the number of images collected. We still need to add more languages and invite through our publications potential users to share their feedback and turn this prototype in an operational service.

More details about this work can be found in our 2 submitted papers in appendix 2.

6. Conclusions and Perspectives

In this study, a model has been developed to detect automatically and in real time landslide images from twitter. This work is a collaboration between computer scientists (QCRI), earthquake (EMSC) and landslide (BGS) specialists which has led to 2 submitted articles. A future post will be written on Dave Petley's blog (https://blogs.agu.org/landslideblog) once Pennington et al. 2022 will be published.

This work shows how artificial intelligence can help us sort out irrelevant images collected from Twitter. The AI decreases by more than 99% the pictures requiring a visual inspection!

Perhaps more importantly, this work which was initially focused on earthquake-triggered landslides is turning, thanks to the interest of the landslide community, to a global landslide reporter service. This interest has been further demonstrated by a recent discussion to join it with landslide imagery services, a combination which could prove crucial to improve the location of the observations. These are strong indications that this project development will be maintained far beyond the end of the RISE project and will serve a large community of users.

References

British Geological Survey (2021) Report a landslide to the BGS landslides team, https://www2.bgs.ac.uk/reportalandslide/reportForm.html [Accessed 30th March 2021]

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological bulletin, 76(5), p 378. https://doi.org/10.1037/h0031619

Imran, M., Castillo, C., Lucas, J., Patrick, M., Vieweg, S. (2014). AIDR: Artificial Intelligence for Disaster Response. WWW '14 Companion: Proceedings of the 23rd International Conference on World Wide Web, p 159–162, https://doi.org/10.1145/2567948.2577034.

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).

Ofli, F., Qazi, U., Imran, M., Roch, J., Pennington, C., Banks, V., Bossu, R. (2022). A Real-time System for Detecting Landslide Reports on Social Media using Artificial Intelligence. Preprint arXiv:2202.07475.

Pennington, C., Bossu, R., Olfi, F., Imran, M., Qazi, U., Roch, J., Banks, V. (2022). A near-real-time global landslide incident reporting tool demonstrator using social media and artificial intelligence. IJDRR-D-22-00635.

Phengsuwan, J., Shah, T., Thekkummal, N.B., Wen, Z., Sun, R., Pullarkatt, D., Thirugnanam, H., Ramesh, M.V., Morgan, G.; James, P. and Ranjan, R. (2021) Use of Social Media Data in Disaster Management: A Survey. Future Internet, v13, 46. https://doi.org/10.3390/fi13020046

Appendix 1 – List of all keywords

landslide, landslides, rockfall, rock-fall, rockslide, rockslides, mudslide, mudslides, mudflow, mudflows, landslip, earthslip, Sturzstrom, avalanche, glissement de terrain, glissements de terrain, chute de pierres, coulée de boue, effondrement, avalanche, frana, frane, crollo di roccia, crolli di roccia, caduta massi, cadute massi, smottamento, smottamenti, slavina, slavine, lliscament de terra, esllavissada de terra, Despreniments de roques, colada de terra, corriment de terra, allau, Esllavissaments superficials, deslizamiento, deslizamiento de tierra, caída de roca, desprendimientos de rocas, deslizamiento de rocas, avalancha de barro, deslizamiento de barro, deslizamiento de lodo, Colapso, tanah runtuh, kejatuhan batu, tanah runtuh, lumpur, tanah longsor, batu jatuh, guguran, lahar, longsoran besar, انهيار, يديالجل أرضي , انزلاق أرضي , سقوط صخري, انهيار صخري , الانهيارات الصخرية, الإنزلاق الطيني , الإنزلاقات الطينية, الانزلاق الأرضي ، انزلاق الأرض, الانهيار, оползень, оползни, зсув, зсув грунту, зсув ґрунту, обвал , обвал скель , падіння скель, камнепад , грязьовий потік , селеві потоки , обрушение, лавина, 산사태, 낙석, 암반사태, 진흙사태, 이류, 사태, 눈사태, heyelan, toprak kayması, heyelanlar, toprak kaymaları, kaya düşmesi, kaya kayması, kaya göçmesi, kaya akıntısı, moloz akıntısı, moloz akışı, moloz akması, moloz kayması, kaya düşmeleri, kaya kaymaları, kaya göçmeleri, kaya akıntıları, moloz akıntıları, moloz kaymaları, çamur akıntısı, çamur akışı, çamur kayması, çamur akması, tortu akıntısı, tortu akışı, tortu akması, tortu kayması, döküntü akıntısı, döküntü akışı, döküntü akması, döküntü kayması, lahar, çamur akıntıları, çamur kaymaları, tortu akıntıları, tortu kaymaları, döküntü akıntıları, döküntü kaymaları, laharlar, çığ, çığlar, Erdrutsch, Erdrutsche, Bergrutsch, Bergrutsche, Hangrutsch, Hangrutsche, Hangrutschung, Hangrutschungen, Abrutschung, Abrutschungen, Steinrutschung, Steinrutschung, Hangmure, Hangmuren, Steinschlag, Steinschläge, Murgang, Murgänge, Mure, Muren, Schlammlawine, Schlammlawinen, Murenabgang, Murenabgänge, Bergsturz, Bergstürze, Lawine, Lawinen, Schneelawine, Schneelawinen, Eislawine, Eislawinen, Staublawine, Staublawinen, ভূমিস্খলন, भूस्खलन, 地すべり, 土砂災害, 土砂崩れ, 山津波, 地滑り, 山崩れ, 山, 坍, 土石流, 雪崩, 表層雪崩, 滑り, κατολίσθηση, καθίζηση εδάφους, πτώση βράχου, βραχολίσθηση, ολίσθηση λάσπης, λασπολίσθηση, καθίζηση έδαφους, χιονοστιβάδα, pagguho ng lupa, pagkahulog ng bato, putik sa lupa, 滑坡 , 山体滑坡, 岩崩, 岩滑, 泥石流, 山体塌方 , 地崩, 雪崩, deslizamento de terras, Queda de rochas, deslizamento de rochas, Queda de blocos, deslizamento de lamas, lamas, Movimentos de massa, Avalanche, lur-irristatzea, harri-jausia , harri erorketa, arroka-irristatzea, Lur-kolada, azal-irristatzea , gainazaleko-irristatzea, higakin-korrontea, alunecare de teren, alunecare de teren cu caderi de roci, caderi de roci, alunecare de noroi, alunecare de pamant, avalansa, Rrëshqitje toke, Rrëshqitje shkëmbore, Rrëshqitjet shkëmbore, Rrëshqitje e tipit “rënie e coprave dhe blloqeve shkembore, Rrjedhje balte, Rrjedhjet balte, Rrëshqitje dheu, Rrëshqitje toke, Rrëshqitje e tipit ortekë, zemeljski plaz, skalni podor, skalni zdrs, skalni zdrsi, blatni tok, blatni tokovi, zdrs pobočja, zdrs zemljine, plaz, skred, skreden, jordskred, jordskreden, bergskred, lerskred, släntstabilitet, kvicklera, snöskred, lavin, odron, klizište, lavina, kőlavina, hólavina, földcsuszamlás, talajcsúszás, talajcsuszamlás, sárfolyás, talajfolyás, iszapfolyás, sárlavina, talajkúszás, suvadás, kőomlás, hegyomlás, hegyomlások, iszapár, földcsuszamlás, földcsuszamlások, sárlavina, sárlavinák, törmeléklavina, lavina, lavinák, alunecare de teren, alunecare alunecări de stănci , alunecări de stănci, alunecare de noroi, alunecări de noroi, deplasări de teren, alunecare de pământ, avalanșă, Aardverschuiving, aardverschuivingen, Bergstorting , Rotslawine, steenlawine, puinlawine, Modderlawine, Modderstroom, modderstromen, Lawine, kliziste, klizista, odron, odroni, blatni tok, blatni tokovi, zemljani tok, zemljani tokovi, lavina, lavine, osuwisko, osuwiska, obryw skalny, obrywy skale, osuwisko skalne, osuwiska skalne, lawina błotna, lawiny błotne, osuwisko, lawina, lawiny, بهمن , رانش زمین , زمین لغزه, ریزش سنگ, سنگ بارش, سنگ لغزه, گل لغزه, Ho heleha hoa mobu, Ho theteha hoa mafika, Seretse se phallang, Ho hlefoha hoa lefats'e, Ho heleha, le ho phalla hoa lehloa, მეწყერი, ქვათაცვენა, კლდეზვავი, კლდეზვავები, ღვარცოფი, ღვარცოფები, ზვავი , maanvyöry, putoavia kiviä, kivivyöry, kivivyöryt, mutavyöry, mutavyöryt, sortuma, lumivyöry, snowmelt, snow melt, debris, flow, cliff fall, cliff collapse, landslips.

Appendix 2 – submitted papers

A Real-time System for Detecting Landslide Reports on Social Media using Artificial Intelligence

Ferda Ofli[1][0000−0003−3918−3230], Umair Qazi[1][0000−0002−2448−9694], Muhammad Imran[1][0000−0001−7882−5502], Julien Roch[2], Catherine Pennington[3][0000−0002−3560−9030], Vanessa Banks[3], and Remy Bossu[2][0000−0002−9927−9122]

[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, United Kingdom
{cpoulton,vbanks}@bgs.ac.uk

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 [13]. However, landslide events are often underreported 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 [15]. Therefore, any attempt to quantify global landslide hazards and the associated impacts remains to be an underestimation due to this oversight and lack of global data inventories [7].

Undertaking the challenge of building a global landslide inventory, NASA launched a website[4] in 2018 to allow citizens to report about the regional landslides they see in-person or online [11]. Following a similar Volunteered Geographical Information (VGI) approach, researchers further developed other means such as mobile or web applications to collect citizen-provided data [5, 14]. 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, 11, 24, 31].

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 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 the challenging nature of the problem, 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 [16, 27]. 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 [32]. Remote sensing techniques either use Synthetic Aperture Radar (SAR) or optical imagery to perform

[4] `https://gpm.nasa.gov/landslides/index.html`

landslide detection in various formulations including classification, segmentation, object detection, and change detection [17, 3, 30, 10, 25, 26]. 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 [5, 14, 1, 2]. 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.

Use of social media data for landslide detection has not been explored extensively. The most relevant work by Musaev et al. [18, 20] combines social media text data and physical sensors to detect landslides. In contrast, we focus on analyzing social media images which can provide more detailed information about the impact of the landslide event. To this end, our work complements prior art.

3 System Design

The system is designed to ingest data from an online social media platform (i.e., Twitter), process and analyze the incoming data, and persist relevant information under the condition that all tasks must be performed in a time-sensitive manner. Fig. 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 Twitter Streaming API[5] 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. Our system employs only the keyword-based data collection approach since the bounding box approach provides only geo-tagged tweets which can be about any topic. Tweets matching with at least one pre-specified keywords are acquired from Twitter in 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 Redis[6] queue.

[5] `https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/guides/connecting`

[6] `https://redis.io/`

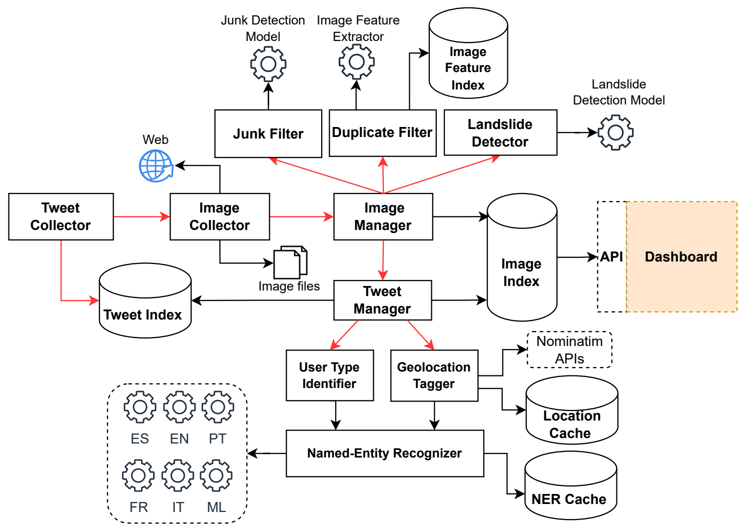


Fig. 1: System architecture with important components and communication flows

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 Junk Filter and Duplicate 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, 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 images. 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 Section 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 Section 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 [9] 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 [9]. 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 operation 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 are publicly available at spaCy[7]. This module also maintains a cache of processed NER requests to increase its efficiency for recurring requests.

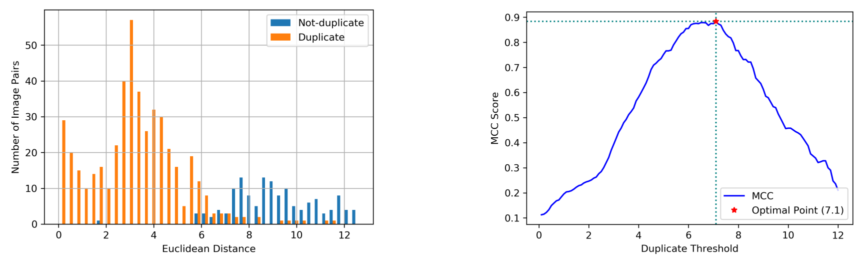
4 Experiments

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 [6, 29, 23], particularly when training samples are limited.

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 [8] pre-trained on the Places data set [33], which

[7] `https://spacy.io/usage/models`



(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.

comprises 10 million images collected for scene recognition.[8] 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 [4]. 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 [8] pre-trained on ImageNet [28], adopted its final layer to binary classification task, and fine-tuned it on a custom data set introduced by Nguyen et al. [21]. We merged the validation set with the training set, and used the test set to evaluate the performance of the model as summarized in Table 1a. We used Adam optimizer [12] 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.

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

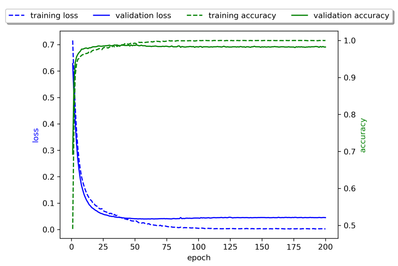
Table 1: Details of the data set used for training the junk classification model and the performance of the trained model on the test set.

(a) Training data set

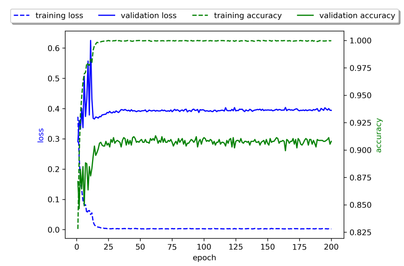
Class	Train	Test	Total
Relevant	2,814	704	3,518
Not-relevant	2,814	704	3,518
Total	5,628	1,408	7,036

(b) Model performance (Acc: 98.79)

Class	Precision	Recall	F1
Relevant	98.31	99.29	98.80
Not-relevant	99.28	98.30	98.79
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 [22]. 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. Fig. 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.

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,

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

Class	Train	Val	Test	Total
Landslide	1,883	271	536	2,690
Not-landslide	6,332	902	1,813	9,047
Total	8,215	1,173	2,349	11,737

(b) Model performance (Acc: 86.97)

Class	Precision	Recall	F1
Landslide	73.66	66.79	70.06
Not-landslide	90.45	92.94	91.68
Macro avg.	82.05	79.87	80.87

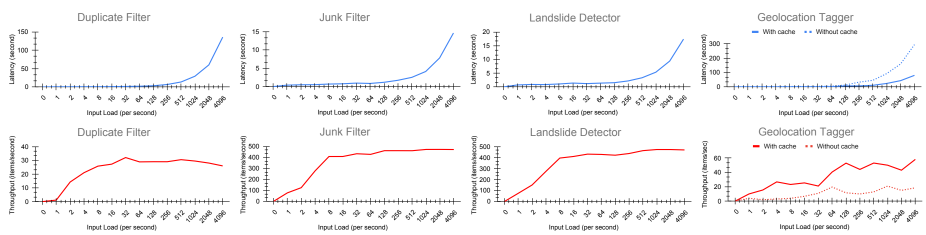


Fig. 4: Latency (top) and throughput (bottom) of the Junk Filter, Duplicate Filter, Landslide Detector, and Geolocation Tagger (left to right).

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 $2^n, n \in \{0, 1, ..., 12\}$. We performed the tests on a Linux server with 256GB RAM, 2.2 GHz processor with 32 cores and two Tesla V100 GPUs with 16GB.

Fig. 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., 29 seconds) 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.

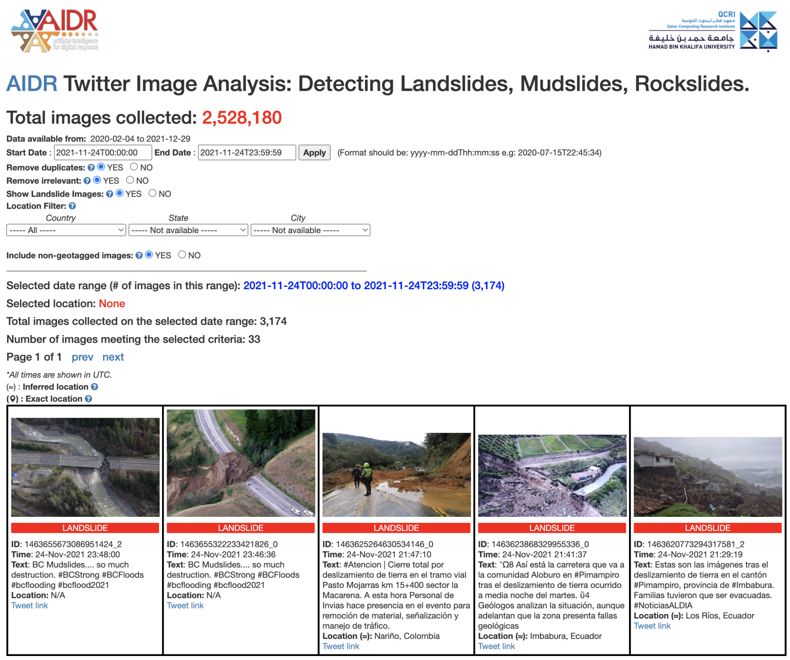


Fig. 5: Snapshot of the online system

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. Fig. 5 shows a snapshot of the system dashboard. It is important

Table 3: List of all keywords in 32 languages used for data collection.

landslide, landslides, rockfall, rock-fall, rockslide, rockslides, mudslide, mudslides, mudflow, mudflows, landslip, earthslip, Sturzstrom, avalanche, glissement de terrain, glissements de terrain, chute de pierres, coulée de boue, effondrement, avalanche, frana, frane, crollo di roccia, crolli di roccia, caduta massi, cadute massi, smottamento, smottamenti, slavina, slavine, lliscament de terra, esllavissada de terra, Despreniments de roques, colada de terra, corriment de terra, allau, Esllavissaments superficials, Deslizamiento, deslizamiento de tierra, caída de roca, desprendimientos de rocas, deslizamiento de rocas, avalancha de barro, deslizamiento de barro, deslizamiento de lodo, Colapso, tanah runtuh, kejatuhan batu, tanah runtuh, lumpur, tanah longsor, batu jatuh, guguran, lahar, longsoran besar, , انهيار أرضي , انزلاق أرضي , سقوط صخري, انهيار صخري , الانهيارات الصخرية, الإنزلاق الطيني , الإنزلاقات الطينية, الانزلاق الأرضي ، انزلاق الأرض, الانهيار الجليدي, оползень, оползни, зсув, зсув грунту, зсув ґрунту, обвал , обвал скель , падіння скель, камнепад , грязьовий потік , селеві потоки , обрушение, лавина, 산사태, 낙석, 암반사태, 진흙사태, 이류, 사태, 눈사태, heyelan, toprak kayması, heyelanlar, toprak kaymaları, kaya düşmesi, kaya kayması, kaya göçmesi, kaya akıntısı, moloz akıntısı, moloz akışı, moloz akması, moloz kayması, kaya düşmeleri, kaya kaymaları, kaya göçmeleri, kaya akıntıları, moloz akıntıları, moloz kaymaları, çamur akıntısı, çamur akışı, çamur kayması, çamur akması, tortu akıntısı, tortu akışı, tortu akması, tortu kayması, döküntü akıntısı, döküntü akışı, döküntü akması, döküntü kayması, lahar, çamur akıntıları, çamur kaymaları, tortu akıntıları, tortu kaymaları, döküntü akıntıları, döküntü kaymaları, laharlar, çığ, çığlar, Erdrutsch, Erdrutsche, Bergrutsch, Bergrutsche, Hangrutsch, Hangrutsche, Hangrutschung, Hangrutschungen, Abrutschung, Abrutschungen, Steinrutschung, Steinrutschung, Hangmure, Hangmuren, Steinschlag, Steinschläge, Murgang, Murgänge, Mure, Muren, Schlammlawine, Schlammlawinen, Murenabgang, Murenabgänge, Bergsturz, Bergstürze, Lawine, Lawinen, Schneelawine, Schneelawinen, Eislawine, Eislawinen, Staublawine, Staublawinen, ভূমিস্খলন, , भूस्खलन, 地すべり, 土砂災害, 土砂崩れ, 山津波, 地滑り, 山崩れ, 山, 坍, 土石流, 雪崩, 表層雪崩, 滑り, κατολίσθηση, καθίζηση εδάφους, πτώση βράχου, βραχολίσθηση, ολίσθηση λάσπης, λασπολίσθηση, καθίζηση έδαφους, χιονοστιβάδα, pagguho ng lupa, pagkahulog ng bato, putik sa lupa, 滑坡 , 山体滑坡, 岩崩, 岩滑, 泥石流, 山体塌方 , 地崩, 雪崩, deslizamento de terras, Queda de rochas, deslizamento de rochas, Queda de blocos, deslizamento de lamas, lamas, Movimentos de massa, Avalanche, lur-irristatzea, harri-jausia , harri erorketa, arroka-irristatzea, Lur-kolada, azal-irristatzea , gainazaleko-irristatzea, higakin-korrontea, alunecare de teren, alunecare de teren cu caderi de roci, caderi de roci, alunecare de noroi, alunecare de pamant, avalansa, Rrëshqitje toke, Rrëshqitje shkëmbore, Rrëshqitjet shkëmbore, Rrëshqitje e tipit “rënie e coprave dhe blloqeve shkembore, Rrjedhje balte, Rrjedhjet balte, Rrëshqitje dheu, Rrëshqitje toke, Rrëshqitje e tipit ortekë, zemeljski plaz, skalni podor, skalni zdrs, skalni zdrsi, blatni tok, blatni tokovi, zdrs pobočja, zdrs zemljine, plaz, skred, skreden, jordskred, jordskreden, bergskred, lerskred, släntstabilitet, kvicklera, snöskred, lavin, odron, klizište, lavina, kőlavina, hólavina, földcsuszamlás, talajcsúszás, talajcsuszamlás, sárfolyás, talajfolyás, iszapfolyás, sárlavina, talajkúszás, suvadás, kőomlás, hegyomlás, hegyomlások, iszapár, földcsuszamlás, földcsuszamlások, sárlavina, sárlavinák, törmeléklavina, lavina, lavinák, alunecare de teren, alunecare alunecări de stănci , alunecări de stănci, alunecare de noroi , alunecări de noroi, deplasări de teren, alunecare de pămănt, avalanșă, Aardverschuiving, aardverschuivingen, Bergstorting , Rotslawine, steenlawine, puinlawine, Modderlawine, Modderstroom, modderstromen, Lawine, kliziste, klizista, odron, odroni, blatni tok, blatni tokovi, zemljani tok, zemljani tokovi, lavina, lavine, osuwisko, osuwiska, obryw skalny, obrywy skale, osuwisko skalne, osuwiska skalne, lawina błotna, lawiny błotne, osuwisko, lawina, lawiny, , رانش زمین, زمین لغزه, ریزش سنگ, سنگ بارش, سنگ لغزه, گل لغزه , بهمن, Ho heleha hoa mobu, Ho theteha hoa mafika, Seretse se phallang, Ho hlefoha hoa lefats'e, Ho heleha, le ho phalla hoa lehloa, მეწყერი, ქვათაცვენა, კლდეზვავი, კლდეზვავები, ღვარცოფი, ღვარცოფები, ზვავი , maanvyöry, putoavia kiviä, kivivyöry, kivivyöryt, mutavyöry, mutavyöryt, sortuma, lumivyöry, snowmelt, snow melt, debris flow, cliff fall, cliff collapse, landslips

to note that, by landslide, we refer to all downward and outward movement of loosen slope materials such as landslip, debris flows, mudslides, rockfalls, earthflows, and other mass movements. As mentioned in Section 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, 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. Fig. 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 about 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. Of all the detected landslide reports, 6,523 were shared by personal accounts and 4,553 by organizational accounts. Fig. 7 shows the worldwide distribution of the detected landslide reports while Fig. 8 highlights the top-10 countries with the highest number of landslide reports in each quarter.

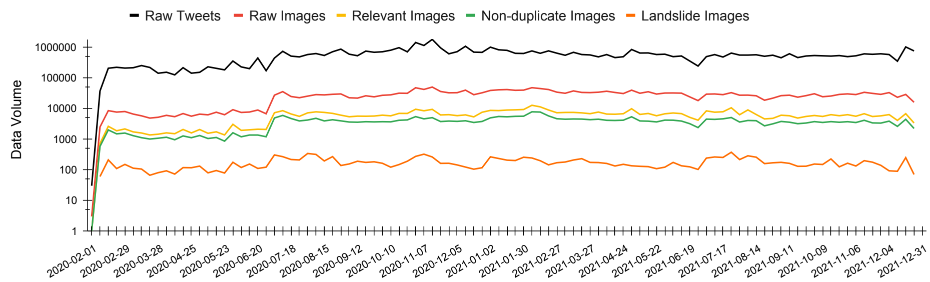


Fig. 6: Weekly distributions of raw tweets and images as well as the relevant, non-duplicate, and landslide images (y-axis is in log scale).

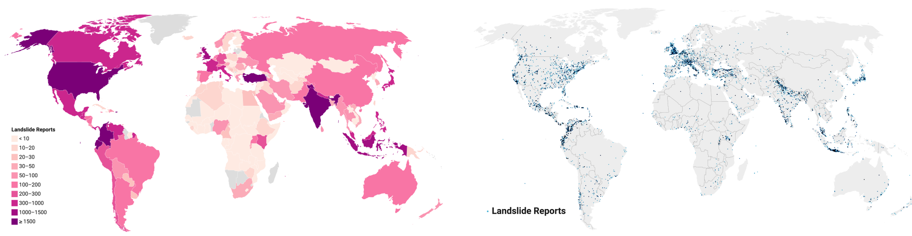


Fig. 7: Worldwide distribution of the collected landslide reports

We see that US, Ecuador, Colombia, and India experience significant landslides all year round. For India, landslides become even more prevalent in Q3. Likewise, Mexico experiences a significant increase in Q3. In contrast, prominent landslides in Indonesia and Malaysia happen in Q1 and Q4 whereas in the UK they occur more in Q1 and Q2. Turkey experiences most landslides in Q1 through Q3.

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

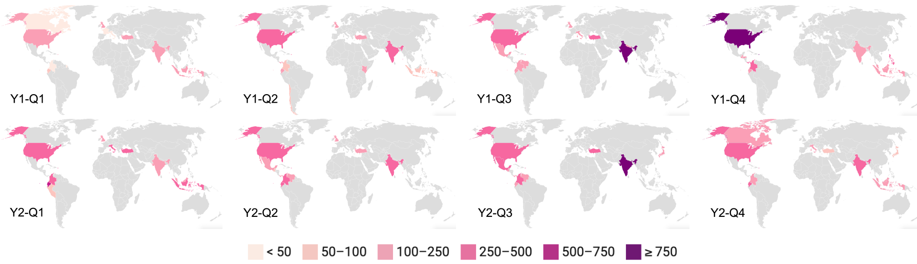


Fig. 8: Top-10 countries with the highest landslide reports in each quarter

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

Negative (FN)) predictions together with the corresponding 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 results achieved in our experiments (Section 4.3).

5.3 Comparison with a Text-based Approach

Text-based landslide detection is a nascent problem where only a couple of studies have addressed so far [19, 20]. 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 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.

References

1. 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)
2. 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)
3. 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)
4. Chicco, D., Jurman, G.: The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. BMC genomics **21**(1), 1–13 (2020)
5. Choi, C.E., Cui, Y., Zhou, G.G.: Utilizing crowdsourcing to enhance the mitigation and management of landslides. Landslides **15**(9), 1889–1899 (2018)
6. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. In: ICML. pp. 647–655 (2014)
7. Froude, M.J., Petley, D.N.: Global fatal landslide occurrence from 2004 to 2016. Nat Hazard Earth Sys **18**(8), 2161–2181 (2018)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)
9. 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)
10. 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)
11. 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)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015)
13. Kjekstad, O., Highland, L.: Economic and social impacts of landslides. In: Landslides–disaster risk reduction, pp. 573–587. Springer (2009)
14. Kocaman, S., Gokceoglu, C.: A CitSci app for landslide data collection. Landslides **16**(3), 611–615 (2019)

15. Lee, E.M., Jones, D.K.: Landslide risk assessment, vol. 10. Thomas Telford London (2004)
16. Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R., Abderrahmane, B.: Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Sci Rev **207**, 103225 (2020)
17. Mohan, A., Singh, A.K., Kumar, B., Dwivedi, R.: Review on remote sensing methods for landslide detection using machine and deep learning. T Emerg Telecommun T **32**(7), e3998 (2021)
18. Musaev, A., Wang, D., Pu, C.: Litmus: Landslide detection by integrating multiple sources. In: ISCRAM (2014)
19. Musaev, A., Wang, D., Shridhar, S., Pu, C.: Fast text classification using randomized explicit semantic analysis. In: IRI. pp. 364–371. IEEE (2015)
20. 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)
21. 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)
22. Ofli, F., Imran, M., Qazi, U., Roch, J., Pennington, C., Banks, V.J., Bossu, R.: Landslide detection in real-time social media image streams. arXiv preprint arXiv:2110.04080 (2021)
23. 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)
24. 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)
25. Prakash, N., Manconi, A., Loew, S.: Mapping landslides on eo data: Performance of deep learning models vs. traditional machine learning models. Remote Sensing **12**(3), 346 (2020)
26. Prakash, N., Manconi, A., Loew, S.: A new strategy to map landslides with a generalized convolutional neural network. Scientific reports **11**(1), 1–15 (2021)
27. Ramesh, M.V., Kumar, S., Rangan, P.V.: Wireless sensor network for landslide detection. In: ICWN. pp. 89–95 (2009)
28. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. IJCV **115**(3), 211–252 (2015)
29. 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)
30. Tavakkoli Piralilou, S., Shahabi, H., Jarihani, B., Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S.R., Aryal, J.: Landslide detection using multi-scale image segmentation and different machine learning models in the higher himalayas. Remote Sensing **11**(21), 2575 (2019)
31. 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)
32. Tofani, V., Segoni, S., Agostini, A., Catani, F., Casagli, N.: Use of remote sensing for landslide studies in europe. Nat Hazard Earth Sys **13**(2), 299–309 (2013)
33. 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)

International Journal of Disaster Risk Reduction

A near-real-time global landslide incident reporting tool demonstrator using social media and artificial intelligence

--Manuscript Draft--

Manuscript Number:	IJDRR-D-22-00635
Article Type:	Research Paper
Keywords:	Landslides; triggered-landslides; image-labelling; Artificial Intelligence; database
Corresponding Author:	Catherine Pennington British Geological Survey Keyworth, Nottinghamshire UNITED KINGDOM
First Author:	Catherine Pennington
Order of Authors:	Catherine Pennington
	Rémy Bossu
	Ferda Ofli
	Muhammad Imran
	Umair W. Qazi
	Julien Roch
	Vanessa J. Banks
Abstract:	The development of a system that monitors social media continuously for general landslide-related content using a landslide classification model to identify and retain the most relevant information is described and validated. The system harvests photographs in real-time from these data and tags each image as landslide or not-landslide. A training model was developed with input from computer scientists, geologists (landslide specialists) and social media specialists to establish a large image dataset that has then been applied to the live Twitter data stream. The preliminary model was developed by training a convolutional neural network on the dataset. Quantitative verification of the system's performance during a real-world deployment shows that the system can detect landslide reports with Precision=76%. The demonstrator model is currently running live; the next stage of development will incorporate stakeholder and user feedback.

A near-real-time global landslide incident reporting tool demonstrator using social media and artificial intelligence

Catherine V. L. Pennington[a*], Rémy Bossu[b,c], Ferda Ofli[d], Muhammad Imran[d], Umair W. Qazi[d], Julien Roch[b], Vanessa J. Banks[a]

[a] British Geological Survey, Keyworth, Nottinghamshire, NG12 5GG, United Kingdom

[b] EMSC, European-Mediterranean Seismological Centre, Arpajon, France

[c] CEA, DAM, DIF, F-91297 Arpajon, France

[d] Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

Abstract

The development of a system that monitors social media continuously for general landslide-related content using a landslide classification model to identify and retain the most relevant information is described and validated. The system harvests photographs in real-time from these data and tags each image as landslide or not-landslide. A training model was developed with input from computer scientists, geologists (landslide specialists) and social media specialists to establish a large image dataset that has then been applied to the live Twitter data stream. The preliminary model was developed by training a convolutional neural network on the dataset. Quantitative verification of the system's performance during a real-world deployment shows that the system can detect landslide reports with Precision=76%. The demonstrator model is currently running live; the next stage of development will incorporate stakeholder and user feedback.

Keywords (6 max)

Landslides, triggered-landslides, image-labelling, Artificial Intelligence, database

Introduction

The reporting of landslides and their impacts (damage and loss) varies widely across the globe reflecting a range of physical and socio-economic drivers and contexts. This means that any attempt to quantify global landslide hazards and the associated impacts is an underestimation (Froude and Petley, 2018). Landslides often occur in a multi-hazard cascading environment triggered by other more conspicuous, and therefore more widely reported, hazards such as earthquakes and tropical storms (Lee and Jones, 2004). Consequently, impacts such as the number of fatalities caused by landslides themselves are underestimated because they can be incorrectly reported as being the result of the trigger event, e.g. earthquake (Kjekstad and Highland, 2008). Further, global studies have confirmed that fatalities attributed to non-seismically induced landslides were underestimated in the International Disaster Database (EM-DAT). Between 2004 and 2010, Petley (2012) estimates that the EM-DAT database under-reported the number of fatalities by 2000% whilst between 2007 and 2013, Kirschbaum *et al*. (2015) found this under-representation to be by 1400%. These numbers are an order of magnitude greater than previous studies had indicated and highlight inaccurate quantification and thus appreciation of the true impacts of landslides, resulting in poor prioritisation of global-scale landslide research and mitigation (Petley, 2012).

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

National and regional landslide databases have been established in many countries to document and map hazard events and their associated damage and losses over time e.g., in Europe (Haque *et al*., 2016; Herrera *et al*., 2018), India (National Disaster Management Guidelines, 2011), China (Shi *et al*., 2000), Japan (National Research Institute for Earth Science and Disaster Prevention, 2021), Africa (Broeckx *et al*., 2018), Canada (Public Safety Canada, 2013), the Caribbean (van Western *et al*., 2016), the United States of America (Jones *et al*., 2019), New Zealand (Rosser *et al*., 2017), Australia (Geoscience Australia, 2012), globally (Froude and Petley, 2018; Juang *et al*., 2019; NASA Landslide Reporter, 2018) and globally for non-specialists (*ThinkHazard!*, 2020). These databases have various applications including scientific research, the creation of landslide susceptibility maps (e.g., Foster et al., 2012; Damm and Klose, 2015), Disaster Risk Reduction (e.g., Han *et al*., 2021), planning (e.g., Gibson *et al*., 2012), landslide forecasting models (e.g., Kirschbaum and Stanley, 2018) and the building of resilience to or documenting impacts of climate change (e.g., Andersson-Sköld *et al*., 2013; Wood *et al*., 2020). Databases vary depending on the states of wealth, politics and governance, education, insurance and the availability of institutions willing and able to maintain such databases, as well as the landslide strategies adopted by the host nation or region (Herrera et al., 2018). Whereas the physical location and dimensions of landslides form the backbone to many landslide databases, the associated impacts (i.e., damage and loss) are much less catalogued despite this being a main component of Disaster Risk Reduction (Corominas *et al*., 2014; Herrera *et al*., 2018). This is due to multiple factors and challenges associated with, and varying priorities for, capturing data as well as difficulties in using international standards at different scales (Gunawan and Aldridge, 2018; United Nations Development Programme, 2013).

Historically, national and regional landslide databases have required substantial investment to enable the manual trawling of maps, aerial photographs, scientific papers, reports and the printed news media for data population (e.g., Geomorphological Services Limited, 1986/7; 1989). The British Geological Survey (BGS) estimates that an average year may require 180 staff hours spent manually trawling the news and social media for data on UK landslides. This work feeds into the UK National Landslide Database (Foster et al., 2012), which underpins much of the landslides research carried out by this national geological survey such as the Daily Landslide Hazard Assessments for the Natural Hazard Partnership (http://naturalhazardspartnership.org).

Landslide data-gathering processes have changed considerably over the last two decades as digital technology, data availability, earth observation techniques, database standards and software interoperability have improved (e.g., European Commission, 2021a). The use of smartphones has increased the incidence, detail and speed of data reported (Niles *et al*., 2019) where, in general, information (including photographs) of landslides are published on social media inherently because they have had an impact for humans of some kind (Pennington et al., 2015). Indeed, dedicated citizen science smartphone applications have been developed specifically for hazard data capture by non-specialists, the benefits and challenges of which are discussed in Lee *et al*. (2020) and Bossu et al. (2015; 2016). LastQuake (https://www.emsc-csem.org; Bossu et al., 2018; Steed *et al*., 2019) is a successful example for earthquakes where eyewitnesses can share their felt experiences as well as pictures and videos, data which aim to improve rapid situational awareness (Bossu et al., 2020). myHaz-VCT (https://oda.bgs.ac.uk) is focused geographically on St Vincent and the

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Grenadines and collects photographs, videos and free-text descriptions on a range of natural hazards including flooding, storms, landslides, earthquakes, volcanoes, tsunamis and other environmental phenomena such as drought, ground subsidence and changes in water levels (Duncan *et al.*, 2019).

At the time of writing, there are two known smartphone applications for the specific acquisition of landslide data, both of which require active participation by the user: Landslide Monitoring App (LaMA) in Turkey (Kocaman and Gokceoglu, 2019) and the Landslide Information System (LIS) in Hong Kong (Choi et al., 2018). Collecting data on landslides through such applications is challenging in terms of user engagement and retention; it is far less common to witness a landslide than an earthquake or a flood, for example, due to them being highly localised.

Elsewhere, the use of citizen science has mostly extended to inviting the public to contact the researchers through online portals, web forms or via email. Examples include 'Report a Landslide' by the BGS (2021a), as well as their engagement via social media, and the 'Report a landslide' and 'Did You See It?' public engagement (Baum *et al.*, 2014) carried out by the United States Geological Survey, now superseded by the NASA Landslide Reporter (Juang et al., 2019). An example of a regional study is Kostelnik et al. (2021) where non-specialists are invited to 'Report an event' for the Bond Fire Debris Flows in California.

Perceptions of what constitutes reliable information are evolving to include unstructured data, such as that published on social media. This is now becoming more valued as a tool to record hazard and hazard impact information, particularly as it can include eye-witness accounts and facilitate the reconstruction of events (Kocaman and Gokceoglu, 2019; Cieslik et al., 2019). To add weight to this, it has become increasingly recognised that news media sources have reporting biases, such as factual accuracy or not reporting at all due to prioritisation of other news (Guzzetti and Tonelli, 2004; Moeller, 2006; Pennington and Harrison, 2013). Despite this however, the bulk of data collection and interpretation still involves time consuming work by specialists searching the Internet for news and social media reports, directly engaging in communications with those submitting information and then interpreting the data received (Kocaman and Gokceoglu, 2019; Juang et al., 2019; Pennington et al., 2015; Taylor et al., 2015).

Under-representation in landslide databases can feed through to emergency planning and preparedness for landslide response when natural disasters occur, particularly in areas for which regional landslide susceptibility mapping has not been completed. In such regions, especially if they are remote with poor access to communication technologies, international responses to natural disasters are supported by attempts to better understand the distribution of triggered landslides, e.g. Nepal where over 4,000 landslides were mapped using satellite imagery after the Gorkha earthquake in 2015 (Lacroix, 2016). It is important to understand both where landslides may have impacted communities with potential damage or loss of life, and where they may affect transport routes and impede emergency response activities. For the latter, even small landslide events can block major transport routes so these are particularly important to identify. In multi-hazard scenarios timely understanding of impacts such as damming of rivers may be important in terms of protecting communities from consequential flooding for example.

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

When natural disasters occur, their impacts are usually not discovered beyond the attention of first responders or government agencies until the news media are able to attend the scene or, for example in remote areas, once satellites have been able to collect imagery and their responding communities have activated the Disaster Charter (https://disasterscharter.org) and processed the data. There is currently an estimated time lag, or data latency, ranging from several hours to several days from when a disaster happens and reliable spatial data becoming available to users, particularly with respect to satellite data (NASA, 2020; Copernicus, 2020; Kaku, 2019; Voigt et al., 2016). Data latency is associated with the satellite return path and the route that it takes, image quality and processing time. Landslides can be associated with rainfall or volcanoes meaning satellite data acquisition can be delayed due to poor image quality caused by cloud cover or whether the satellite passes the area in the day or night (Santangelo et al., 2022). Interpretation of these images also requires considerable effort by specialists although recent work aims to speed this process up using automatic image recognition (Mondini et al; Yi et al., 2020; Yang et al., 2019; Ji et al., 2020; Gudžius et al., 2021).

Aims

Social media data allow access to a rich source of human information such as text, videos, photographs, timestamps and coordinates (e.g., Lacassin et al., 2020; Alam et al., 2018). In 2021, there were 3.78 billion social media users worldwide (Mohsin, 2021) and acquiring disaster data through these platforms has gathered pace, particularly over the last decade. This is, however, an inherently imperfect information source when compared to conventional sensors, aerial imagery or expert interpretation, but it provides large quantities of data, in near-real-time and at spatial densities that may exceed conventional sensor networks and this can complement data from other sources (Li et al., 2021). While these data have great potential for disaster management, they are noisy and it is difficult for disaster managers to extract relevant and timely information (Phengsuwan et al., 2021; Alam et al., 2017).

With an aim to tackle the aforementioned issues, this paper explores a well-known microblogging platform, Twitter, to identify landslide-related posts, specifically those with images containing landslides. Twitter allows users to read and post short messages called 'Tweets'. Tweets are limited to 280 characters and photos or short videos can be included. Tweets are posted to a publicly available profile or can be sent as direct messages to other users. In 2019, Twitter had 330 million monthly active users and 145 million daily active users; a total of 500 million Tweets were sent by Twitter users every day, equivalent to 5787 Tweets per second (Phengsuwan et al., 2021).

In this paper, a new methodology is presented that harvests landslide photographs from Tweets automatically and in real-time. To do this, different types of noise and irrelevant content that that can be associated with landslide-related social media imagery data are identified. Moreover, a further aim is the annotation and release of a dataset for the community to develop image filtering and landslide detection tools. The specific objectives of this paper are:

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

1. Novel qualitative analysis of a non-traditional data source (Twitter) for capturing landslide reports
2. Image labelling methodology for landslide classification
3. Expert-labelled dataset consisting of 11,737 images

A similar study was undertaken by Can et al. (2019) who used a smaller image dataset from different sources and recommend further work with a larger dataset before their algorithm can be used without manual intervention. The work presented here involves more extensive model training experiments and a larger dataset.

We suggest that the methodology and labelled dataset will help the disaster management community build tools to detect landslide images automatically from social media, with potential for incorporation in multi-hazard impact assessment workflows alongside other established methods. Moreover, we anticipate that such a tool will improve response times for first responders. This will enable responders to add information pertaining to their understanding of what is happening on the ground in near-real-time providing data from those affected as soon as it is published on social media.

This interdisciplinary work is the result of the collaboration between computer scientists, earthquake-, social media- and landslide hazard specialists. The initial objective was for the earthquake-triggered landslides to be reported to the European Civil Protection Unit as this hazard can hamper rescue operations. The objective was then extended to incorporate all landslides regardless of their trigger. This tool will be open for any institute wanting to speed up social media harvesting on this topic.

Supervised Machine Learning Approach

The process of using Artificial Intelligence (AI) or Machine Learning (ML) for the identification of landslides in photographs typically requires two steps: (1) create a large, labelled dataset for the task at hand, and (2) train a ML model to achieve the desired classification task. Figure 1 shows a graphical representation of the workflow. This training dataset contains a collection of photographs showing particular characteristics associated with landslides. To create a diverse dataset, we curated a total of 11,737 images from three data sources: Google, Twitter and BGS's image database: GeoScenic (BGS, 2021b). 6,284 images were downloaded from Google by querying landslide-related keywords such as landslide, landslip, earth slip, mudslide, rockslide and rock fall. We developed a multi-lingual list currently comprising 339 keywords in 32 languages: English, Albanian, Arabic, Basque, Bengali, Bosnian, Catalan, Chinese, Croatian, Dutch, French, Georgian, German, Greek, Hindi, Hungarian, Indonesia, Iranian, Italian, Japanese, Korean, Malaysia, Philippines, Polish, Portuguese, Romanian, Russian, Sesotho, Slovenian, Spanish, Swedish, and Turkish (Appendix 1). A total of 1,153 images were collected from Twitter through its Streaming API using the same keywords. In addition, 4,300 photographs were donated by the GeoScenic database that were known to be associated with fieldtrips involving landslides. Three landslide specialists, co-authors of this paper, then carried out an independent yes/no landslide interpretation on the 11,737 photographs using the methodology described below. Figure 2 shows examples of collected photographs divided into 'landslides' and 'not landslides' that

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

demonstrates the kind of noise associated with image harvesting.

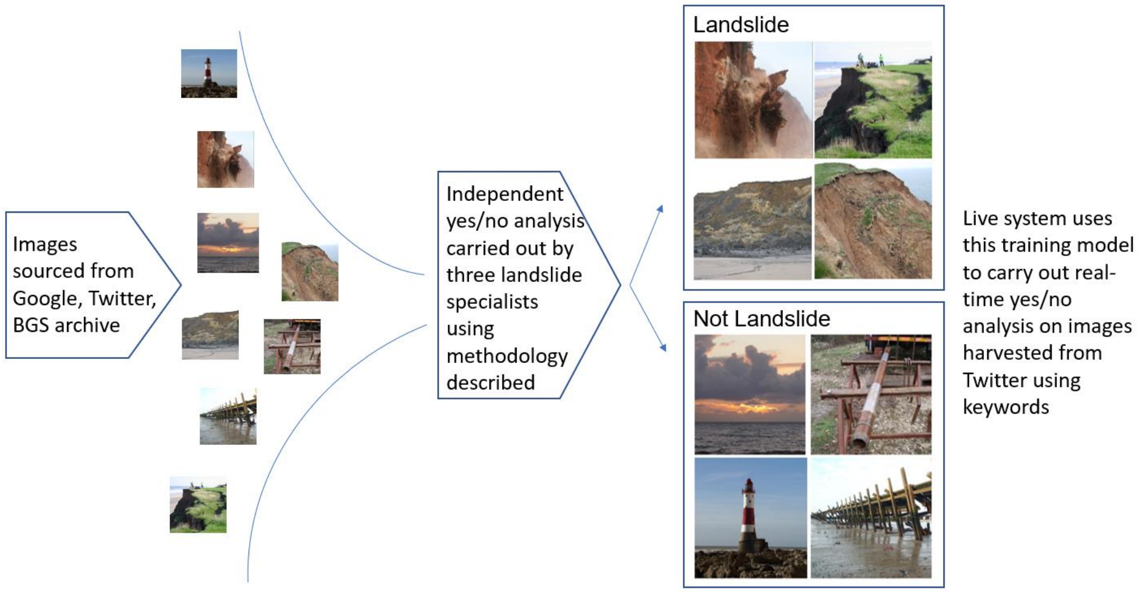


Figure 1 Graphical representation showing the workflow for the development of the training model. Photographs BGS © UKRI [2022].

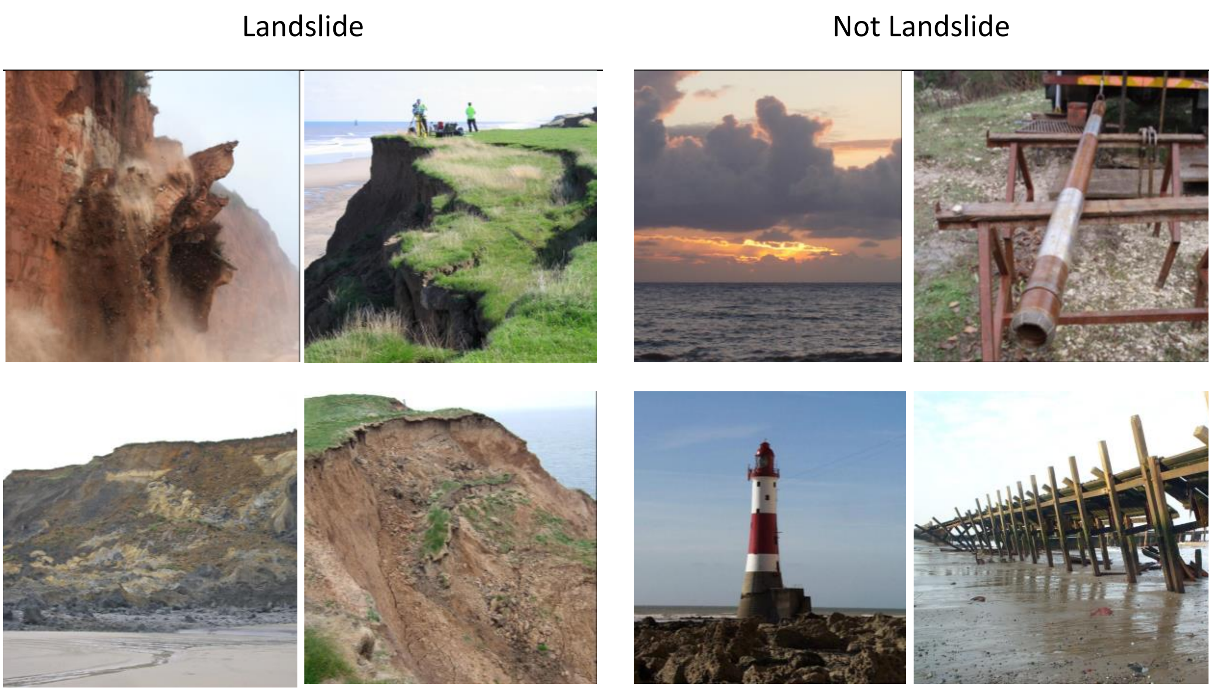


Figure 2 Examples of images collected showing landslides and examples of noise (not landslides) Photographs BGS © UKRI [2022]

Although manually curated, keywords were used to acquire images from Twitter and Google; the resultant images are not always related to landslides and often contained irrelevant and noisy content. This demonstrates why the use of text-based data collection alone is not enough to gather landslide-related reports from social media or the Internet. While the

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

images from the GeoScenic database were known to be associated with fieldtrips involving landslides, the set included both landslide and non-landslide photographs. Therefore, the collected images needed to be evaluated manually by the landslide specialists. Since the AI task is “given an image, recognise landslide” without any other external information or expert knowledge available to the AI model, the landslide specialists were tasked to devise a labelling methodology while keeping this “computer vision” perspective in mind.

Expert-labelling methodology

The decision-making process carried out for the purpose of training the computer model to identify landslide features in photographs differs from conventional desk- or field-based landslide identification familiar to the geologist. Expert assessment of photographs involved the application of several assumptions as outlined in the following methodology.

1. There is no contextual knowledge or previous understanding of the landslide. A data-gathering exercise would usually be carried out by landslide specialists to gain as much ground information as possible before any interpretations are made. This requires a different approach. Here, information such as any landslide nomenclature, ground conditions, antecedent meteorological context or geographic region are excluded from the decision-making process.

2. Each photograph must be treated in isolation. This may show all or part of a landslide and is confined to one viewpoint. Ideally, conventional landslide analysis involves viewing the landslide from several different perspectives and scales before an interpretation is made.

3. The model does not discriminate landslide ‘type’ (i.e. Hungr et al., 2014; Cruden and Varnes 1996), but aims to recognise zones of depletion (where the material has come from) and accretion (where it has been deposited). This excludes, therefore, events where the landslide debris has been removed by coastal or fluvial erosion or where a landslide has been remediated.

4. The model aims to show contemporary landslides. This means older but perhaps still active or dormant landslides are omitted from the model. Examples of this may include landslides that are slow moving or cyclic but are nonetheless active. Fully vegetated landslides may also fall into this category if there is no exposure of geological materials (e.g., rock or earth; Figure 3).

5. In order to train the model there was a requirement for a clear representation of a landslide as the major component of the image (e.g., Figure 4).

6. Where representation was borderline, consideration was given to whether the end user would be concerned by the image being returned as a landslide, e.g., in the situation where another geomorphological feature such as a retaining wall or a sinkhole might be returned as a landslide. Borderline cases are broadly grouped as (Figure 5A) backscarps and extensional that could be faults, (Figure 5B) material engulfing buildings that could be the landslide deposit but could also have formed through other natural or manmade processes, (Figure 5C) debris falling onto roads that could be a landslide deposit or vegetation or mixed debris not

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

associated with landsliding and (Figure 5D) rivers in flow or flood channels that have a similar appearance to debris flow channels.

Figure 3 Example of a completely vegetated landslide (flow) that would be excluded from the dataset. BGS © UKRI [2022].

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Figure 4 An example of where there is a clear representation of a landslide as a major component of the photograph BGS © UKRI [2022].

Figure 5 Examples of borderline classes. A: This image could be a landslide backscarp or a fault; B: Material engulfing buildings; C: Debris falling onto a road; D: Rivers in flow or channels that have a similar form to debris flow type landslides. Photographs BGS © UKRI [2022].

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Once the dataset creation and model training stages were completed, the demonstrator model was run using Twitter images in real-time. Figure 6 illustrates the workflow involved in collecting, tagging and classifying images as 'landslide' and 'not landslide'.

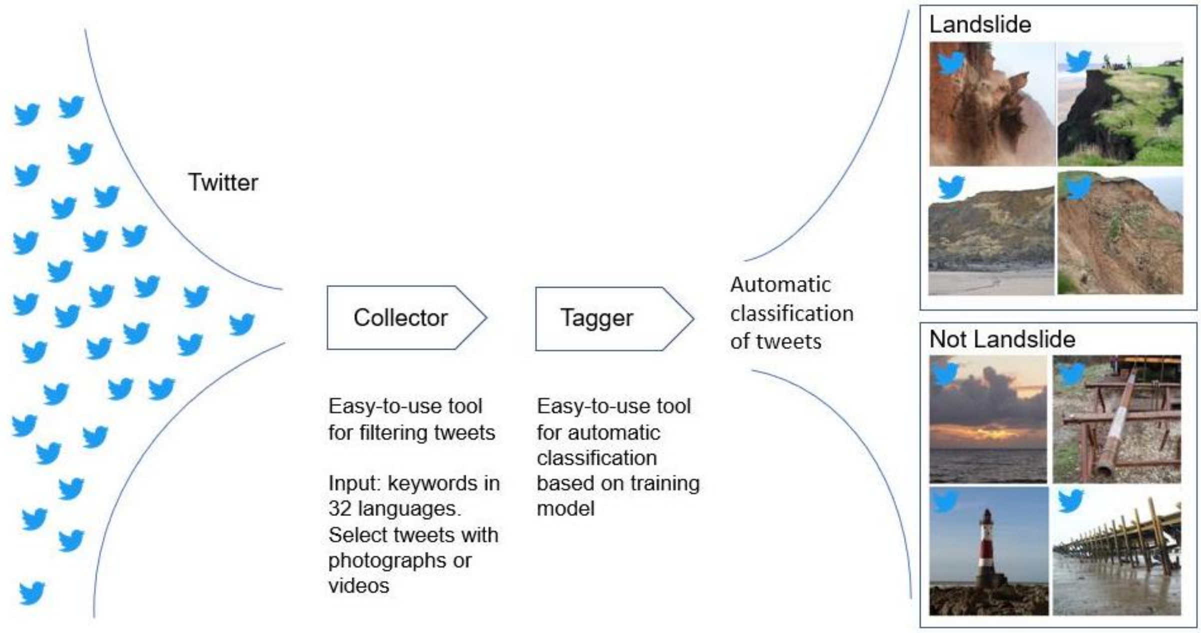


Figure 6 Graphical representation illustrating the workflow involved in collecting, tagging and classifying images from Tweets as 'landslide' and 'not landslide'. Photographs BGS © UKRI [2022].

Expert-labelling results

Using the methodology outlined above, the three landslide specialists carried out independent yes/no interpretations of 11,737 photographs. In order to ensure reliability of the final labels, an analysis was carried out to measure their agreement using two statistical measures: Fleiss' Kappa (Fleiss, 1971) and percentage agreement (observer agreement). Despite the inherent difficulty of the labelling task, the three landslide specialists achieved good overall agreement. An overall Fleiss' Kappa score of 0.58 was achieved, which indicates an almost 'substantial' inter-annotator agreement between the three landslide specialists. The percentage agreement is 76%, which is only slightly below the 80% mark set as a rule-of-thumb by Bayerl and Paul (2011).

Since the ultimate goal is to develop a system that will monitor the noisy social media streams continuously to detect landslide reports in real-time, negative (i.e., not-landslide) images were also retained in the dataset to represent completely irrelevant cases (e.g., cartoons, advertisements, selfies) as well as difficult scenarios (i.e., those which may look similar to landslides) such as post-disaster images from earthquakes and floods in addition to other natural scenes without landslides for model training purposes. The distribution of the images in the final dataset across different categories and data sources are summarised in Table 1.

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

As suggested by the table, only about 23% of the images are labelled as landslide in the final dataset. This shows an imbalanced class distribution, which presents a challenge in model training simply because the model may decide to always predict not-landslide and achieve 77% accuracy (because of the skew in the distribution) but this would not be useful at all. Solutions to problems like this (i.e., finding a needle in the haystack) do always need to deal with the class imbalance issue meaning the training set presented here reflects this realistic scenario.

Table 1 Distribution of the images across different categories and data sources

	Google	Twitter	BGS	Total
Landslide	1,240	598	852	2,690
Not-landslide	5,044	555	3,448	9,047
Total	6,284	1,153	4,300	11,737

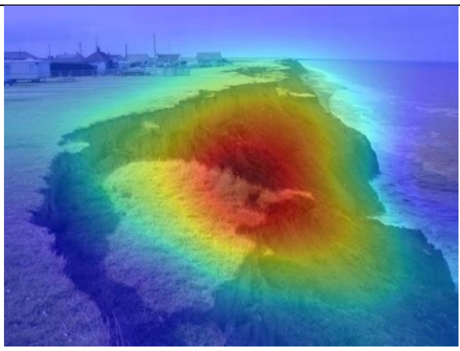
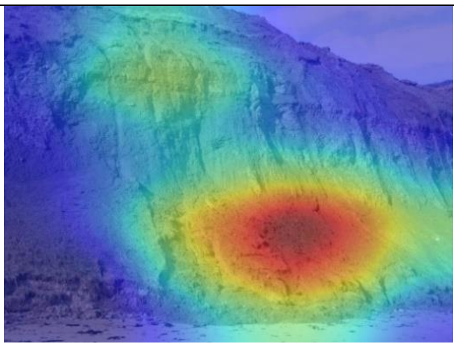
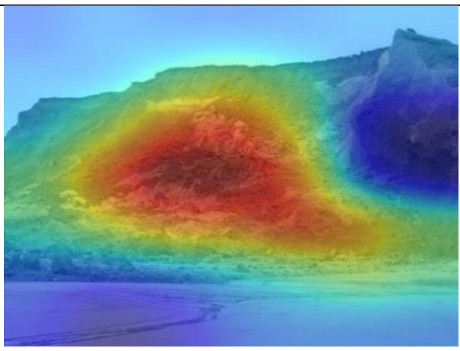
Demonstrator model results

Since the focus of this study is to establish a methodology for the landslide dataset creation, a technical paper, conducted in conjunction with this study, describes the underpinning ML theory and presents a detailed experimental approach to the model development step (Ofli et al., 2021). The demonstrator model presented here is developed by training a convolutional neural network on the dataset introduced in this paper. Quantitative verification of the system's performance during a real-world deployment shows that our system can detect landslide reports with Precision=76%.

We deployed the system online in February 2020 to monitor the 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. 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. More details about this system deployment can be found in Ofli et al. (2022). Below, we evaluate the performance of this demonstrator model on a few example images with the help of heat maps or class activation maps (Zhou, 2016), which highlight the discriminative parts of a photograph that the model is paying attention to (Tables 2 and 3).

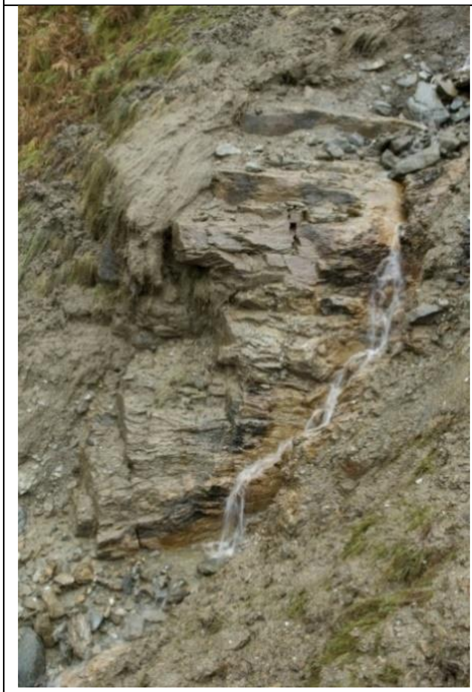
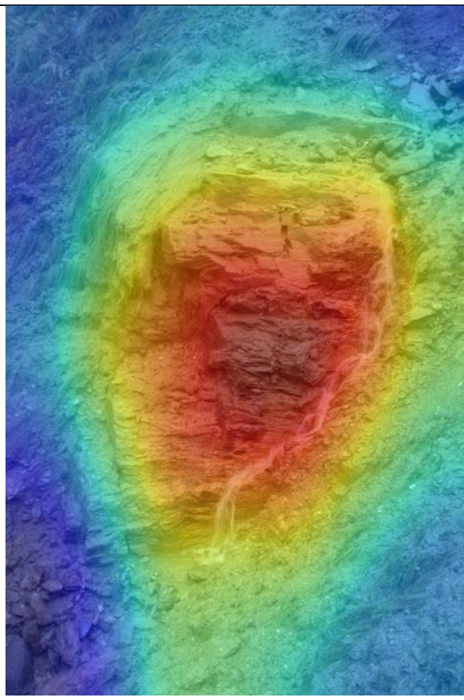
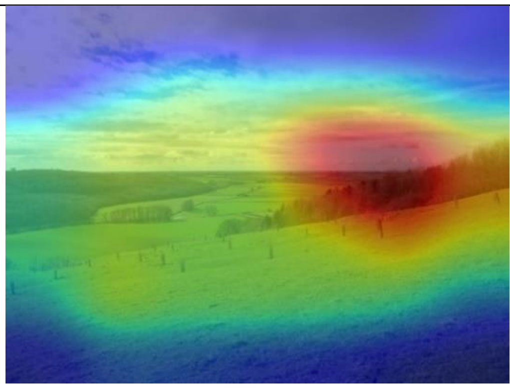
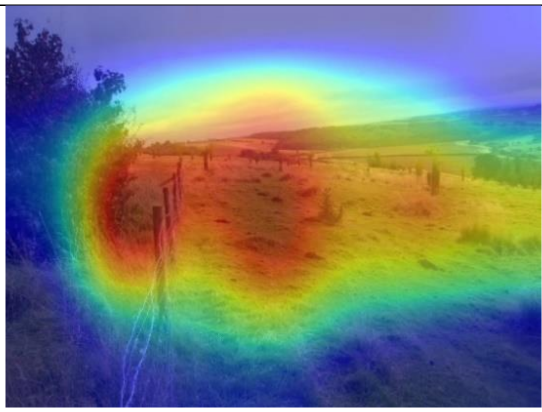
*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Table 2 Photographs identified as containing a landslide showing the class activation map interpretation generated by the system. [Class activation map produced by QCRI Contains BGS © UKRI]

Original photograph	Class Activation Map interpretation	Confidence
		100%
		100%
		98.7%

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Table 3 Photographs identified as NOT containing a landslide. [Class activation map produced by QCRI Contains BGS © UKRI]

Original photograph	Heat map interpretation	Confidence
		99.9% This is a rock exposure, not landslide
		99.9% A field, not landslide
g 		99.9% A field, not landslide

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Discussion

The aim of this work was to develop a system that monitors social media continuously and in real-time for general landslide-related content, using the landslide classification model to identify and retain the most relevant information. The system harvests photographs from these data and tags each image as landslide or not-landslide. A training model was developed through interdisciplinary working by the authors to establish a large image dataset that has then been applied to the live Twitter data stream.

The demonstrator model is currently running live and landslide images are being harvested in real-time (https://landslide-aidr.qcri.org/landslide_system.php). This is publicly available and users can filter by date and country as well as being able to explore data spatially via a map interface. The map interface uses a range of factors to geolocate data markers. The current version prioritises location-based text within Tweets over geolocation data or stated location of the user. While not perfect, this allows the map to display landslide sites and prevents it from becoming purely a representation of user locations. If geolocation data are used, the location is downgraded to adhere to rules around viewing geodata (Twitter, 2022a, b) and to protect user privacy. Future improvements to locating data are discussed below.

Also available on the demonstrator model website is the list of keywords from Appendix 1 used to initially extract Tweets. We invite users to provide feedback on both the demonstrator itself and the list of keywords via the link above. Once feedback has been collated, we plan to carry out future iterations to move this work from a demonstrator model to an operational service.

The image interpretation process used by the three landslide specialists was iterative in the initial phase of work. To maintain consistency of agreement, the methodology described above was established through much interdisciplinary discussion, which led to a phase of reinterpretation. While this put demands on the landslide specialists, the combined understanding produced this novel methodology with high levels of agreement.

The methodology aimed to identify landslide features, but the task was not to discriminate scale, meaning that images labelled as landslides may be very small (<1m and not strictly a landslide) and aerial photographs including multiple landslide events are not captured by the model (e.g., Figure 7). Further iterations of this work could use more sophisticated object detection or image segmentation techniques to solve this issue.

Future work will include a Geolocation Inference module that will use Tweet metadata to geolocate images following the approach used by Imran et al. (2022) for spatial analysis of various factors associated with the COVID-19 pandemic. An automated real-time geographic representation of landslide locations will be developed. Understanding the location of landslides is an important element of this work as there may not be the magnitude of data compared to other hazards such as earthquakes. However, there are ethics to be considered as part of this location-based work such as that adopted by the UK government through the Data Ethics Framework (UK Government, 2020) and the Locus Charter (2021). The work

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

described in this paper could also be adapted to complement other hazard inventories, such as snow avalanches.

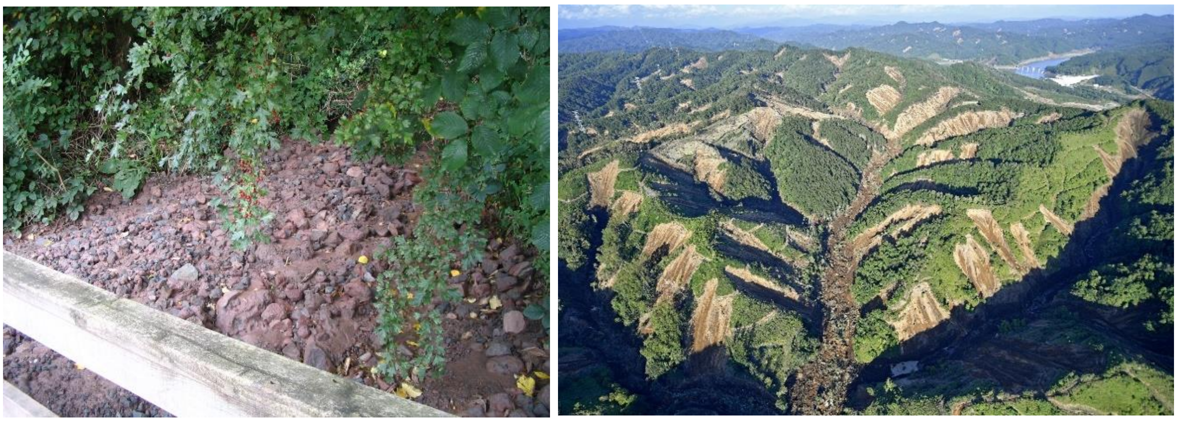


Figure 7 Examples of photographs that demonstrate the issue of scale. (Left) a small area (<1m) of mixed debris, BGS © UKRI [2022] (Right) an aerial photograph showing multiple landslide deposits across a large area [Photograph supplied under licence © KYODO Kyodo/Reuters Pictures].

It is important to reiterate that this work is not intended to be used in isolation during a disaster scenario. As well as the inherent noise within the data content itself, there are inaccuracies that could, for example in the worst case, hinder rescue operations if not combined with other data sources. Disaster managers should note that this work does not take into account:

1. Areas without mobile or internet coverage (even if temporary). As natural hazards cause damage to infrastructure, this may lead to mobile phone or internet outages meaning information cannot be published to social media.
2. The geographical variation in population density. In densely populated areas, there are likely to be more relevant Tweets due to numbers of people that could skew the data away from less densely populated areas that may have suffered greater damage.
3. Variations in use of social media (i.e. Twitter) as a result of trends in national or regional uptake or demographics.
4. Photographs that are embedded as thumbnails in web page links in Tweets. For example, an article published by the news media with photographs that was included in a Tweet is currently excluded.

For these reasons, the authors recommend that this work is used as a tool to provide additional information to established workflows for disaster management.

For landslides research, such as that involving national or regional landslide databases, it is hoped that this work will introduce considerable efficiency savings for institutions responsible for maintaining this workflow. Images of landslide events and impacts will be available automatically and social media is trawled in a systematic and continuous way. This has been adapted to the terminologies used in different countries through the list of keywords. The authors would like to improve this list to make the operational model more accurate.

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Conclusion

This paper demonstrates the potential application of artificial intelligence for landslide recognition in images harvested from social media. In this study, we aimed to develop a model that can detect landslides in social media image streams automatically and in real-time. For this purpose, we created a large image collection from multiple sources with different characteristics to ensure data diversity. The collected images were assessed by three landslide specialists independently to attain high quality labels with almost substantial inter-annotator agreement. The assessment methodology is described and is the result of interdisciplinary working between geologists, computer scientists and social media specialists. The resulting model achieved high performance in terms of accuracy scores, which can be deemed sufficient for the purpose. The demonstrator model is publicly available and running in real-time and the authors invite feedback. There are a number of potential applications for this research. In this account image processing has been focused on "fresh" landslides as evidenced by the exposure of geological materials, which lends itself to the focus on the potential for Disaster Risk and Resilience. This paper is published in association with a technical paper that describes the model in detail.

Funding

This article 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.

The British Geological Survey (UK Research and Innovation) granted supporting research funding through National Capability (Shallow Geohazards) and Innovation funding streams.

Declaration of competing interest

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

Acknowledgments

This paper presents results from interdisciplinary research carried out by computer scientists at the Qatar Computing Research Institute (QCRI), earthquakes and social media specialists at the European-Mediterranean Seismological Centre (EMSC) and landslide hazard expertise from the British Geological Survey (BGS). The authors would like to thank those who donated photographs from their archives for the training model and for discussions around the applicability of this tool: Professor Dave Petley at Sheffield University for advice; Jo Walsh and Brian McIntyre at the British Geological Survey for collating photographs from the UK National Landslide Database and GeoScenic photograph archive. This work is about Twitter, not by Twitter; the terms 'Twitter', 'Tweet' and the Twitter Bird logo featured in Figure 6 are trademarks of Twitter Inc. or its affiliates.

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

References

Alam, F., Imran, M., and Ofli, F. (2017) Image4Act: Online social media image processing for disaster response. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp 601–604. https://doi.org/10.1145/3110025.3110164

Alam, F., Ofli, F., and Imran, M. (2018) Processing social media images by combining human and machine computing during crises. International Journal of Human–Computer Interaction, 34(4): pp 311–327. https://doi.org/10.1080/10447318.2018.1427831

Andersson-Sköld, Y., Bergman, R., Johansson, M., Persson, E. and Nyberg, L. (2013) Landslide risk management—A brief overview and example from Sweden of current situation and climate change, International Journal of Disaster Risk Reduction, Volume 3, pp 44-61, https://doi.org/10.1016/j.ijdrr.2012.11.002

Baum, R.L., Highland, L.M., Lyttle, P.T., Fee, J.M., Martinez, E.M. and Wald, L.A. (2014) “Report a Landslide” A Website to Engage the Public in Identifying Geologic Hazards. In: Sassa K., Canuti P., Yin Y. (eds) Landslide Science for a Safer Geoenvironment. Springer, Cham. https://doi.org/10.1007/978-3-319-04999-1_8

Bayerl, P. S., & Paul, K. I. (2011). What determines inter-coder agreement in manual annotations? A meta-analytic investigation. Computational Linguistics, 37(4), 699-725. https://doi.org/10.1162/COLI_a_00074

BGS (2021a) Report a landslide, https://www2.bgs.ac.uk/reportalandslide/reportForm.html [Accessed 29th September 2021]

BGS (2021b) GeoScenic, http://geoscenic.bgs.ac.uk [Accessed 29th September 2021]

Bossu, R., Laurin, M., Mazet-Roux, G., Roussel, F., & Steed, R. (2015). The importance of smartphones as public earthquake-information tools and tools for the rapid engagement with eyewitnesses: A case study of the 2015 Nepal earthquake sequence. Seismological Research Letters, v86(6), pp 1587-1592. https://doi.org/10.1785/0220150147

Bossu, R., Steed, R., Mazet-Roux, G., Roussel, F., Etivant, C.,Frobert, L., & Godey, S. (2016). The key role of eyewitnesses in rapid impact assessment of global earthquakes. In Earthquakes and their Impact on Society (pp 601-618). Springer, Cham. https://doi.org/10.1007/978-3-319-21753-6_25

Bossu, R., Roussel, F., Fallou, L., Landès, M., Steed, R., Mazet-Roux, G., Dupont, A., Frobert, L. and Petersen, L. (2018). "LastQuake: From rapid information to global seismic risk reduction." International journal of disaster risk reduction v28, pp 32-42. https://doi.org/10.1016/j.ijdrr.2018.02.024

Bossu, R., Fallou, L., Landès, M., Roussel, F., Julien-Laferrière, S., Roch, J., and Steed, R. (2020). Rapid public information and situational awareness after the November 26, 2019, Albania earthquake: lessons learned from the LastQuake system. Frontiers in Earth Science, v8, p 235. https://doi.org/10.3389/feart.2020.00235

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

British Geological Survey (2021) Report a landslide to the BGS landslides team, https://www2.bgs.ac.uk/reportalandslide/reportForm.html [Accessed 30th March 2021]

Broeckx, J., Vanmaercke, M., Duchateau, R. and Poesen, J. (2018). A data-based landslide susceptibility map of Africa. Earth-Science Reviews, v185, pp 102-121. https://doi.org/10.1016/j.earscirev.2018.05.002

Can, R., Kocaman, S. and Gokceoglu, C. (2019) A Convolutional Neural Network Architecture for Auto-Detection of Landslide Photographs to Assess Citizen Science and Volunteered Geographic Information Data Quality. ISPRS International Journal of Geo-Information. v8(7): p 300. https://doi.org/10.3390/ijgi8070300

Choi, C.E., Cui, Y. and Zhou, G.G.D. (2018) Utilizing crowdsourcing to enhance the mitigation and management of landslides. Landslides, v 15, pp 1889–1899. https://doi.org/10.1007/s10346-018-1034-9

Cieslik, K., Puja, S., Madhab, U., Art, D., Russell, C., Clark, J., Raj, D.M. and Amrit C. (2019) Building Resilience to Chronic Landslide Hazard Through Citizen Science. Frontiers in Earth Science, v 7, p278, https://doi.org/10.3389/feart.2019.00278

Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J.P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M.G., Pastor, M., Ferlisi, S., Tofarni, V., Hervás, J. and Smith, J. T. (2014) Recommendations for the quantitative analysis of landslide risk. Bulletin of Engineering Geology and the Environment, v73, pp 209–263. https://doi.org/10.1007/s10064-013-0538-8

Copernicus (2021) Linking with early warning systems. https://emergency.copernicus.eu/mapping/ems/linking-early-warning-systems [Accessed 26/09/2021]

Cruden D.M. and Varnes D.J. (1996) Landslide types and processes. In: Turner AK, Schuster RL (eds) Landslides investigation and mitigation. Transportation research board, US National Research Council. Special Report 247, Washington, DC, Chapter 3, pp 36–75.

Damm, B. and Klose, M. (2015) The landslide database for Germany: Closing the gap at national level. Geomorphology, v249, pp 82-93, https://doi.org/10.1016/j.geomorph.2015.03.021

Duncan, M., Mee, K., Hicks, A., Richardson, S., Robertson, R.E.A., Forbes, M., Ferdinand, I., Jordan, C. and Loughlin, S. (2019) Citizen science using mobile phone technology in St Vincent & the Grenadines to facilitate near-real time multi-hazard observations. Geophysical Research Abstracts, v 21, p1. https://meetingorganizer.copernicus.org/EGU2019/EGU2019-14536.pdf

European Commission (2021a). INSPIRE Principles https://inspire.ec.europa.eu/inspire-principles/9. Accessed 27/01/2021, last updated 27/01/2021.

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological bulletin, 76(5), p 378. https://doi.org/10.1037/h0031619

Foster, C, Pennington, C V L, Culshaw, M G, and Lawrie, K. (2012). The National Landslide Database of Great Britain: development, evolution and applications. Environmental Earth Sciences, 66(3), pp 941-953, https://doi.org/10.1007/s12665-011-1304-5

Froude, M. J. and Petley, D. N. (2018). Global fatal landslide occurrence from 2004 to 2016. Natural Hazards and Earth System Science, 18, pp 2161-2181, https://doi.org/10.5194/nhess-18-2161-2018.

Geomorphological Services Limited (GSL). 1986/7. Review of research into landsliding in Great Britain. Report to the Department of the Environment, 15 volumes. Geomorphological Services (Publications and Reprographics) Limited, Newport Pagnell.

Geomorphological Services Limited (GSL). 1989. Landslide data bank for Britain: user manual version 1.0. Geomorphological Services (Publications and Reprographics) Limited, Newport Pagnell.

Geoscience Australia (2012) Landslide Search. Software, http://pid.geoscience.gov.au/dataset/ga/74273

Gibson, A.D., Culshaw, M.G., Dashwood, C. and Pennington, C.V.L. (2012) Landslide management in the UK—the problem of managing hazards in a 'low-risk' environment. Landslides 10, pp 599–610. https://doi.org/10.1007/s10346-012-0346-4

Gudžius, P., Kurasova, O., Darulis, V. and Filatovas E. (2021) Deep learning-based object recognition in multispectral satellite imagery for real-time applications. Machine Vision and Applications, v32(98). https://doi.org/10.1007/s00138-021-01209-2

Gunawan, O. and Aldridge, T. (2018) Disaster Loss Data Management in Scotland. Technical Report DMS/2018/01 for the Scottish National Centre for Resilience. https://core.ac.uk/download/pdf/296209205.pdf

Guzzetti, F. and Tonelli, G. (2004) Information system on hydrological and geomorphological catastrophes in Italy (SICI): a tool for managing landslide and flood hazards. Natural Hazards and Earth Systems Science, v4, pp 213–232. https://doi.org/10.5194/nhess-4-213-2004

Han, Q., Sassa, K. and Mikoš, M. (2021) International Programme on Landslides (IPL): A Programme of the ICL for Landslide Disaster Risk Reduction. In: Sassa K., Mikoš M., Sassa S., Bobrowsky P.T., Takara K., Dang K. (eds) Understanding and Reducing Landslide Disaster Risk. WLF 2020. ICL Contribution to Landslide Disaster Risk Reduction. Springer, Cham. https://doi.org/10.1007/978-3-030-60196-6_11

Haque, U., Blum, P., da Silva, P.F., Andersen, P., Pilz, J., Chalov, S.R., Malet, J.P., Auflič, M.J., Andres, N., Poyiadji, E., Lamas, P.C., Zhang, W., Pesevski, I., Pétursson, H.G., Kurt, T., Dobrev, N., Davalillo, J.C.G., Halkia, M., Ferri, S., Gaprindashvili, G., Engström, J. and Keellings, D.

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

(2016) Fatal landslides in Europe. Landslides, v13, pp 1545–1554. https://doi.org/10.1007/s10346-016-0689-3

Herrera, G., Mateos, R.M., García-Davalillo, J.C., Grandjean, G., Poyiadji, E., Maftei, R., Filipciuc, T-C., Auflič, M.J., Jež, J., Podolszki, L., Trigila, A., Iadanza, C., Raetzo, H., Kociu, A., Przyłucka, M., Kułak, M., Sheehy, M., Pellicer, X.M., McKeown, C., Ryan, G., Kopačková, V., Frei, M., Kuhn, D., Hermanns, R.L., Koulermou, N., Smith, C.A., Engdahl, M., Buxó, P., Gonzalez, M., Dashwood, C., Reeves, H., Cigna, F., Liščák, P., Pauditš, P., Mikulėnas, V., Demir, V., Raha, M., Quental, L., Sandić, C., Fusi B. and Jensen, O. A. (2018) Landslide databases in the Geological Surveys of Europe. Landslides, v15, pp 359–379. https://doi.org/10.1007/s10346-017-0902-z

Hough, S. (2013). Spatial Variability of "Did You Feel It?" Intensity Data: Insights into Sampling Biases in Historical Earthquake Intensity Distributions. Bulletin of the Seismological Society of America, 103, pp 2767-2781. https://doi.org/10.1785/0120120285

Hungr, O., Leroueil, S. and Picarelli, L. (2014) The Varnes classification of landslide types, an update. Landslides, v11, pp 167-194. https://doi.org/10.1007/s10346-013-0436-y

Ji, S., Yu, D., Shen, C., Li, W. and Xu, Q. (2020) Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. Landslides, v17, pp 1337–1352 https://doi.org/10.1007/s10346-020-01353-2

Jones, E.S., Mirus, B.B., Schmitt, R.G., Baum, R.L., Burns, W.J., Crawford, M., Godt, J.W., Kirschbaum, D.B., Lancaster, J.T., Lindsey, K.O., McCoy, K.E., Slaughter, S. and Stanley, T.A. (2019) Summary Metadata – Landslide Inventories across the United States: U.S. Geological Survey data release, https://doi.org/10.5066/P9E2A37P.

Juang, C.S., Stanley, T.A., and Kirschbaum, D.B. (2019) Using citizen science to expand the global map of landslides: Introducing the Cooperative Open Online Landslide Repository (COOLR). PLOS ONE v14(7): e0218657. https://doi.org/10.1371/journal.pone.0218657

Locus Charter (2021) Locus Charter Report https://ethicalgeo.org/wp-content/uploads/2021/03/Locus_Charter_March21.pdf Last updated February 2021, [Accessed 29th September 2021].

Kazuya K. (2019) Satellite remote sensing for disaster management support: A holistic and staged approach based on case studies in Sentinel Asia. International Journal of Disaster Risk Reduction, v13, pp 417-432, https://doi.org/10.1016/j.ijdrr.2018.09.015

Kirschbaum, D.B., Adler, R., Hong, Y., Hill, S., and Lerner-Lam, A.L. (2010) A global landslide catalog for hazard applications—Method, results, and limitations. Natural Hazards, v. 52, pp 561–575, https://doi.org/10.1007/s11069-009-9401-4

Kirschbaum, D.B. and Stanley, T. (2018) Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness. Earth's Future, v6 (3), pp 297-616, https://doi.org/10.1002/2017EF000715

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Kirschbaum, D., Stanley, T. and Zhou, Y. (2015): Spatial and temporal analysis of a global landslide catalog. Geomorphology, v249, pp 4–15, https://doi.org/10.1016/j.geomorph.2015.03.016

Kjekstad O. and Highland L.M., 2008, Economic and social impacts of landslides, in Sassa D. Canuti P., eds., Landslides—Disaster risk reduction: Berlin, Springer-Verlag, pp 573–587. https://doi.org/10.1007/978-3-540-69970-5_30

Kocaman, S. and Gokceoglu, C.A. (2019). CitSci app for landslide data collection. Landslides, v16, pp 611–615. https://doi.org/10.1007/s10346-018-1101-2

Kostelnik, J., Wald, L., Schmitt, R. and Rengers, F.K. (2021) Bond fire debris flows, California. https://landslides.usgs.gov/storymap/bond/ [Accessed 16th September 2021].

Lacroix, P. (2016) Landslides triggered by the Gorkha earthquake in the Langtang valley, volumes and initiation processes. Earth Planets Space v68, pp46. https://doi.org/10.1186/s40623-016-0423-3

Lacassin, R., Devès, M., Hicks, S. P., Ampuero, J.-P., Bossu, R., Bruhat, L., Daryono, Wibisono, D. F., Fallou, L., Fielding, E. J., Gabriel, A.-A., Gurney, J., Krippner, J., Lomax, A., Sudibyo, Muh. M., Pamumpuni, A., Patton, J. R., Robinson, H., Tingay, M., and Valkaniotis, S. (2020) Rapid collaborative knowledge building via Twitter after significant geohazard events, Geoscience Communication, v3, pp 129–146, https://doi.org/10.5194/gc-3-129-2020

Lee E. M. and Jones D.K. (2004) Background to landslide Assessment. Thomas Telford Ltd, London, UK.

Lee, K.A., Lee, J.R. and Bell, P. (2020) A review of Citizen Science within the Earth Sciences: potential benefits and obstacles, Proceedings of the Geologists' Association, v131(6), pp605-617. https://doi.org/10.1016/j.pgeola.2020.07.010

Li, L., Bensi, M., Cui, Q., Baecher, G.B. and Huang, Y. (2021) Social media crowdsourcing for rapid damage assessment following a sudden-onset natural hazard event. International Journal of Information Management, v60, 102378, https://doi.org/10.1016/j.ijinfomgt.2021.102378

Moeller, S. D. (2006) “Regarding the pain of others": media, bias and the coverage of international disasters. Journal of International Affairs, v59(2) The Globalization of Disaster (SPRING/SUMMER 2006), pp 173-196. https://www.jstor.org/stable/24358432

Mondini, A. C., Guzzetti, F., Reichenbach, P., Rossi, M., Cardinali, M. and Ardizzone, F. (2011) Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. Remote Sensing of Environment, v115, pp 1743-1757. https://doi.org/10.1016/j.rse.2011.03.006

Mohsin, M.Y.S. (2021) 10 Social Media Statistics You Need to Know in 2021 [Infographic]. Available online: https://www.oberlo.com/blog/social-media-marketing-statistics [Accessed 29th September 2021].

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Napolitano, E., Marchesini, I., Salvati, P., Donnini, M., Bianchi, C. and Guzzetti, F. (2018) LAND-deFeND - An innovative database structure for landslides and floods and their consequences. Journal of Environmental Management, v207, pp 203-218. https://doi.org/10.1016/j.jenvman.2017.11.022

NASA Landslide Reporter (2018) https://pmm.nasa.gov/landslides/report.html (21.08.2018)

NASA (2020). What is data latency? https://earthdata.nasa.gov/learn/backgrounders/data-latency [Accessed 27/09/2021].

National Disaster Management Guidelines (2011) National Disaster Management Information and Communication System (NDMICS). A publication of the National Disaster Management Authority, Government of India ISBN: 978-93-80440-12-5, February 2012, New Delhi.

National Research Institute for Earth Science and Disaster Prevention (2021) Landslide Disasters, https://www.bosai.go.jp/e/research/database/earth-and-sand.html [Accessed 4th November 2021].

Niles, M.T., Emery, B.F., Reagan, A.J., Dodds, P.S. and Danforth, C.M. (2019) Social media usage patterns during natural hazards. PLoS ONE 14(2): e0210484. https://doi.org/10.1371/journal.pone.0210484

Ofli, F., Imran, M., Qazi, U., Roch, J., Pennington, C., Banks, V. J., & Bossu, R. (2021). Landslide Detection in Real-Time Social Media Image Streams. arXiv preprint arXiv:2110.04080.

Ofli, F., Qazi, U., Imran, M., Roch, J., Pennington, C., Banks, V., & Bossu, R. (2022). A Real-time System for Detecting Landslide Reports on Social Media using Artificial Intelligence. arXiv preprint arXiv:2202.07475.

Osuchowski, M. and Atkinson, R. (2008) Connecting diverse landslide inventories for improved information in Australia. In: Casagli N, Fanti R, Tofani V (eds) Proceedings of the First World Landslide Forum, Tokyo, Japan, 67–70. (http://www.150.217.73.85/wlfpdf/02_Osuchowski.pdf), [accessed 29 November 2009]

Pennington, C.V.L., Freeborough, K. A., Dashwood, C., Dijkstra, T. J. and Lawrie, K. (2015) The National Landslide Database of Great Britain: Acquisition, communication and the role of social media, Geomorphology, v249, pp 44-51. https://doi.org/10.1016/j.geomorph.2015.03.013

Pennington, C.V.L. and Harrison, A.M. (2013). 2012: Landslide Year? Geoscience Magazine, v23, pp 10–15. https://www.geolsoc.org.uk/Geoscientist/Archive/June-2013/2012-Landslide-year

Petley, D. (2012) Global patterns of loss of life from landslides. Geology, v40 (10), pp 927–930.

Phengsuwan, J., Shah, T., Thekkummal, N.B., Wen, Z., Sun, R., Pullarkatt, D., Thirugnanam, H., Ramesh, M.V., Morgan, G.; James, P. and Ranjan, R. (2021) Use of Social Media Data in

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Disaster Management: A Survey. Future Internet, v13, 46. https://doi.org/10.3390/fi13020046

Public Safety Canada (2013). Canada Disaster Database, https://cdd.publicsafety.gc.ca/srchpg-eng.aspx Accessed March 2021, Last modified 12th September 2013.

Imran M, Qazi U, Ofli F. TBCOV: Two Billion Multilingual COVID-19 Tweets with Sentiment, Entity, Geo, and Gender Labels. Data. 2022; 7(1):8. https://doi.org/10.3390/data7010008

Rosser, B., Dellow, S., Haubrock, S. and Glasset, P. (2017) New Zealand's National Landslide Database. Landslides, published online open access 30 May 2017. https://link.springer.com/content/pdf/10.1007/s10346-017-0843-6.pdf [accessed Mar 29 2021].

Santangelo, M., Cardinali, M., Bucci, F., Fiorucci, F., Mondini, A.C., (2022) Exploring event landslide mapping using Sentinel-1 SAR backscatter products. Geomorphology, v397, 10802. https://doi.org/10.1016/j.geomorph.2021.108021

Shi, P., Wang, J., Yang, M., Chen, J., Pan, Y., Wang, P., Fang, W., Zhou, W., Wang, Y., Yuan, Y., Xu, X. and Luo, W. (2000) Understanding of natural disaster database design and compilation of digital atlas of natural disasters in China. Geographic Information Sciences, v6(2), pp 153-158. https://doi.org/10.1080/10824000009480544

Steed, R.J., Fuenzalida, A., Bossu, R., Bondár, I., Heinloo, A., Dupont, A., Saul, J. and Strollo, A. (2019). Crowdsourcing triggers rapid, reliable earthquake locations. Science Advances, https://www.science.org/doi/pdf/10.1126/sciadv.aau9824

Taylor, F.E., Malamud, B. D., Freeborough, K.A. and Demeritt, D. (2015) Enriching Great Britain's National Landslide Database by searching newspaper archives. Geomorphology, v249, pp 52-68, https://doi.org/10.1016/j.geomorph.2015.05.019

ThinkHazard! (2020) https://thinkhazard.org, last updated 30th June 2020. [Accessed 4th November 2021].

Twitter (2022a) Developer Terms: Geo guidelines, https://developer.twitter.com/en/developer-terms/geo-guidelines, last updated 22nd October 2014. [Accessed 14th March 2022].

Twitter (2022b) Developer Policy, Section Using geo-data, https://developer.twitter.com/en/developer-terms/policy. [Accessed 14th March 2022].

UK Government (2020) Data Ethics Framework: legislation and codes of practice for use of data. https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-legislation-and-codes-of-practice-for-use-of-data Last updated 16 September 2020, [Accessed 29th September 2021].

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

United Nations Development Programme (2013) A Comparative Review of country-level and regional disaster loss and damage databases, Bureau for Crisis Prevention and Recovery. https://www.preventionweb.net/publications/view/33067 [Accessed: 4th November 2021]

van Western, C., Jetten, V. and Alkema, D. (2016) Validating national landslide susceptibility and hazard maps for Caribbean island countries: the case of Dominica and tropical storm Erika. EGU General Assembly 2016, held 17-22 April, 2016 in Vienna, Austria. https://meetingorganizer.copernicus.org/EGU2016/EGU2016-4334.pdf

Varnes, D.J. (1978) Slope movement types and processes. In: Schuster RL, Krizek RJ (eds)Landslides, analysis and control, special report 176: Transportation research board, National Academy of Sciences, Washington, DC., pp 11–33. https://onlinepubs.trb.org/Onlinepubs/sr/sr176/176-002.pdf

Voigt, S., Giulio-Tonolo, F., Lyons, J., Kučera, J., Jones, B., Schneiderhan, T., Platzeck, G., Kaku, K., Kumar Hazarika, M., Czaran, L., Li, S., Pedersen, W., Kadiri James, G., Proy, C., Macharia Muthike, D., Bequignon, J. and Guha-Sapir, D. (2016) Global trends in satellite-based emergency mapping. Science, v353(6296), pp 247-252, https://doi.org/10.1126/science.aad8728

Wood, J.L., Harrison, S., Reinhardt, L. and Taylor, F.E. (2020) Landslide databases for climate change detection and attribution. Geomorphology, v355, https://doi.org/10.1016/j.geomorph.2020.107061

Yang Y., Song, S., Yue, F., He, W., Shao, W., Zhao, K. and Nie, W. (2019) Superpixel-based automatic image recognition for landslide deformation areas. Engineering Geology, v 259, 105166, https://doi.org/10.1016/j.enggeo.2019.105166

Yi, Y. and Zhang, W. (2020) A new deep-learning-based approach for earthquake-triggered landslide detection from single-temporal RapidEye satellite imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, v13, pp 6166-6176, https://doi.org/10.1109/JSTARS.2020.3028855

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. and Torralba, A. (2016) Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2921-2929, https://doi.org/10.1109/CVPR.2016.319

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Appendix 1

List of all keywords in 32 languages used for data collection

landslide, landslides, rockfall, rock-fall, rockslide, rockslides, mudslide, mudslides, mudflow, mudflows, landslip, earthslip, Sturzstrom, avalanche, glissement de terrain, glissements de terrain, chute de pierres, coulée de boue, effondrement, avalanche, frana, frane, crollo di roccia, crolli di roccia, caduta massi, cadute massi, smottamento, smottamenti, slavina, slavine, lliscament de terra, esllavissada de terra, Despreniments de roques, colada de terra, corriment de terra, allau, Esllavissaments superficials, Deslizamiento, deslizamiento de tierra, caída de roca, desprendimientos de rocas, deslizamiento de rocas, avalancha de barro, deslizamiento de barro, deslizamiento de lodo, Colapso, tanah runtuh, kejatuhan batu, tanah runtuh, lumpur, tanah longsor, batu jatuh, guguran, lahar, longsoran besar, انهيار أرضي , انزلاق أرضي , سقوط صخري, انهيار صخري , الانهيارات الصخرية, الإنزلاق الطيني , الإنزلاقات الطينية, الانزلاق الأرضي ، انزلاق الأرض, الانهيار الجليدي, оползень, оползни, зсув, зсув грунту, зсув ґрунту, обвал , обвал скель , падіння скель, камнепад , грязьовий потік , селеві потоки , обрушение, лавина, 산사태, 낙석, 암반사태, 진흙사태, 이류, 사태, 눈사태, heyelan, toprak kayması, heyelanlar, toprak kaymaları, kaya düşmesi, kaya kayması, kaya göçmesi, kaya akıntısı, moloz akıntısı, moloz akışı, moloz akması, moloz kayması, kaya düşmeleri, kaya kaymaları, kaya göçmeleri, kaya akıntıları, moloz akıntıları, moloz kaymaları, çamur akıntısı, çamur akışı, çamur kayması, çamur akması, tortu akıntısı, tortu akışı, tortu akması, tortu kayması, döküntü akıntısı, döküntü akışı, döküntü akması, döküntü kayması, lahar, çamur akıntıları, çamur kaymaları, tortu akıntıları, tortu kaymaları, döküntü akıntıları, döküntü kaymaları, laharlar, çığ, çığlar, Erdrutsch, Erdrutsche, Bergrutsch, Bergrutsche, Hangrutsch, Hangrutsche, Hangrutschung, Hangrutschungen, Abrutschung, Abrutschungen, Steinrutschung, Steinrutschung, Hangmure, Hangmuren, Steinschlag, Steinschläge, Murgang, Murgänge, Mure, Muren, Schlammlawine, Schlammlawinen, Murenabgang, Murenabgänge, Bergsturz, Bergstürze, Lawine, Lawinen, Schneelawine, Schneelawinen, Eislawine, Eislawinen, Staublawine, Staublawinen, ভূমিস্খলন, भूस्खलन, 地すべり, 土砂災害, 土砂崩れ, 山津波, 地滑り, 山崩れ, 山, 坍, 土石流, 雪崩, 表層雪崩, 滑り, κατολίσθηση, καθίζηση εδάφους, πτώση βράχου, βραχολίσθηση, ολίσθηση λάσπης, λασπολίσθηση, καθίζηση έδαφους, χιονοστιβάδα, pagguho ng lupa, pagkahulog ng bato, putik sa lupa, 滑坡 , 山体滑坡, 岩崩, 岩滑, 泥石流, 山体塌方 , 地崩, 雪崩, deslizamento de terras, Queda de rochas, deslizamento de rochas, Queda de blocos, deslizamento de lamas, lamas, Movimentos de massa, Avalanche, lur-irristatzea, harri-jausia , harri erorketa, arroka-irristatzea, Lur-kolada, azal-irristatzea , gainazaleko-irristatzea, higakin-korrontea, alunecare de teren, alunecare de teren cu caderi de roci, caderi de roci, alunecare de noroi, alunecare de pamant, avalansa, Rrëshqitje toke, Rrëshqitje shkëmbore, Rrëshqitjet shkëmbore, Rrëshqitje e tipit "rënie e coprave dhe blloqeve shkembore, Rrjedhje balte, Rrjedhjet balte, Rrëshqitje dheu, Rrëshqitje toke, Rrëshqitje e tipit ortekë, zemeljski plaz, skalni podor, skalni zdrs, skalni zdrsi, blatni tok, blatni tokovi, zdrs pobočja, zdrs zemljine, plaz, skred, skreden, jordskred, jordskreden, bergskred, lerskred, släntstabilitet, kvicklera, snöskred, lavin, odron, klizište, lavina, kőlavina, hólavina, földcsuszamlás, talajcsúszás, talajcsuszamlás, sárfolyás, talajfolyás, iszapfolyás, sárlavina, talajkúszás, suvadás, kőomlás, hegyomlás, hegyomlások, iszapár, földcsuszamlás, földcsuszamlások, sárlavina, sárlavinák, törmeléklavina, lavina, lavinák, alunecare de teren, alunecare alunecări de stănci , alunecări de stănci, alunecare de noroi , alunecări de noroi, deplasări de teren, alunecare de pământ, avalanșă, Aardverschuiving, aardverschuivingen, Bergstorting , Rotslawine, steenlawine, puinlawine, Modderlawine, Modderstroom, modderstromen, Lawine, kliziste, klizista, odron, odroni, blatni tok, blatni tokovi, zemljani tok, zemljani tokovi, lavina, lavine, osuwisko, osuwiska, obryw skalny, obrywy skale, osuwisko skalne, osuwiska skalne, lawina błotna, lawiny błotne, osuwisko, lawina, lawiny, رانش زمین , زمین لغزه, ریزش سنگ, سنگ بارش, سنگ لغزه, گل لغزه , بهمن, Ho heleha hoa mobu, Ho theteha hoa mafika, Seretse se phallang, Ho hlefoha hoa lefats'e, Ho heleha, le ho phalla hoa lehloa, მეწყერი, ქვათაცვენა, კლდეზვავი, კლდეზვავები, ღვარცოფი, ღვარცოფები, ზვავი , maanvyöry, putoavia kiviä, kivivyöry, kivivyöryt, mutavyöry, mutavyöryt, sortuma, lumivyöry, snowmelt, snow melt, debris flow, cliff fall, cliff collapse, landslips

*Corresponding author.
Email addresses: cpoulton@bgs.ac.uk (C.V.L. Pennington), bossu@emsc-csem.org (R. Bossu), fofli@hbku.edu.qa (F. Ofli), mimran@hbku.edu.qa (M. Imran), uqazi@hbku.edu.qa (U. W. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V. J. Banks)

Declaration of interests

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

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Declarations of interest: none