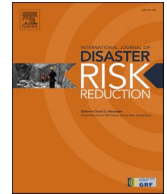




ELSEVIER

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

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

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 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

ARTICLE INFO

Keywords:

6 max)

Landslides

Triggered-landslides

Image-labelling

Artificial intelligence

Database

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 <https://landslide-aidr.qcri.org/service.php>; the next stage of development will incorporate stakeholder and user feedback.

1. Introduction

When landslides 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 [1–4]. 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 [5,6]. Interpretation of these images also requires considerable effort by specialists although recent work aims to speed this process up using automatic image recognition [7–12]. This means that locating and assessing the disaster can take considerable time and most studies are based on post-event analysis.

Social media data, while inherently imperfect, provide near-real-time information in large quantities and at spatial densities that

* Corresponding author. (C.V.L. Pennington).

E-mail 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. Qazi), julien.roch@emsc-csem.org (J. Roch), vbanks@bgs.ac.uk (V.J. Banks).

<https://doi.org/10.1016/j.ijdrr.2022.103089>

Received 30 March 2022; Received in revised form 24 May 2022; Accepted 26 May 2022

Available online 31 May 2022

2212-4209/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

may exceed conventional sensor networks that can complement data from other sources [13]. These ‘social sensor’ data allow access to a rich source of human information such as text, videos, photographs, timestamps and coordinates [14,15] and can report disaster information quicker than the observatories [16]. In 2021, there were 3.78 billion social media users worldwide [17] and acquiring disaster data through these platforms has gathered pace, particularly over the last decade. While these data have great potential for near-real-time reporting of landslide events and their associated impacts, they are noisy and it is difficult for disaster managers to extract relevant and timely information [18,19].

2. Background

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 and the attention of the mass media is not uniformly distributed across disaster-affected areas [20]. This means that any attempt to quantify global landslide hazards and the associated impacts is an underestimation [21]. Few studies exist at present that provide a globally consistent set of estimates for landslide hazard [22]. 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 [23]. 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 [24]. 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 [25], estimates that the EM-DAT database under-reported the number of fatalities by 2000% whilst between 2007 and 2013 [26], 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 [25].

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 [27,28], India [29], China [30], Japan [31], Africa [32], Canada [33], the Caribbean [34], the United States of America [35], New Zealand [36], Australia [37], globally [21,38,39] and globally for non-specialists [40]. These databases have various applications including scientific research, the creation of landslide susceptibility maps (e.g. Refs. [41,42], Disaster Risk Reduction (e.g. Ref. [43], planning (e.g. Ref. [44], landslide forecasting models (e.g. Ref. [45], and the building of resilience to or documenting impacts of climate change (e.g., Refs. [46,47]. 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 [28]. 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 [28,48]. 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 [49,50].

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., Refs. [51,52]. 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 [41], 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://naturalhazardpartnership.org>). This time-consuming work has been addressed to an extent by using automated procedures to harvest landslide-related news media [26,53–56] although this still requires manual interpretation to extract details such as location or impacts [57]. have taken this a step further by developing an automated system that analyses the content of news media, assigning the information to different classes according to ‘news relevance’ and creating a landslide inventory.

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., Ref. [58]. The use of smartphones has increased the incidence, detail and speed of data reported [59] 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 [60]. 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 Refs. [61–63]. LastQuake (<https://www.emsc-csem.org> [64,65]; 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 [66]. myHaz-VCT (<https://oda.bgs.ac.uk>) is focused geographically on St Vincent and the 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 [67].

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 [68] and the Landslide Information System (LIS) in Hong Kong [69]. 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 [70]; as well as their engagement via social media, and the ‘Report a landslide’ and ‘Did You See It?’ public engagement [71] carried out by the United States Geological Survey, now superseded by the NASA Landslide Reporter [38]. An example of a regional study is [72] 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 [68,73]. 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 [74–76]. 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 [38,54,60,68].

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 [77]. 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.

3. Aims

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 5,787 Tweets per second [18].

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 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. 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 theory and presents a detailed experimental approach to the model development step [78]. The specific objectives of this paper are:

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

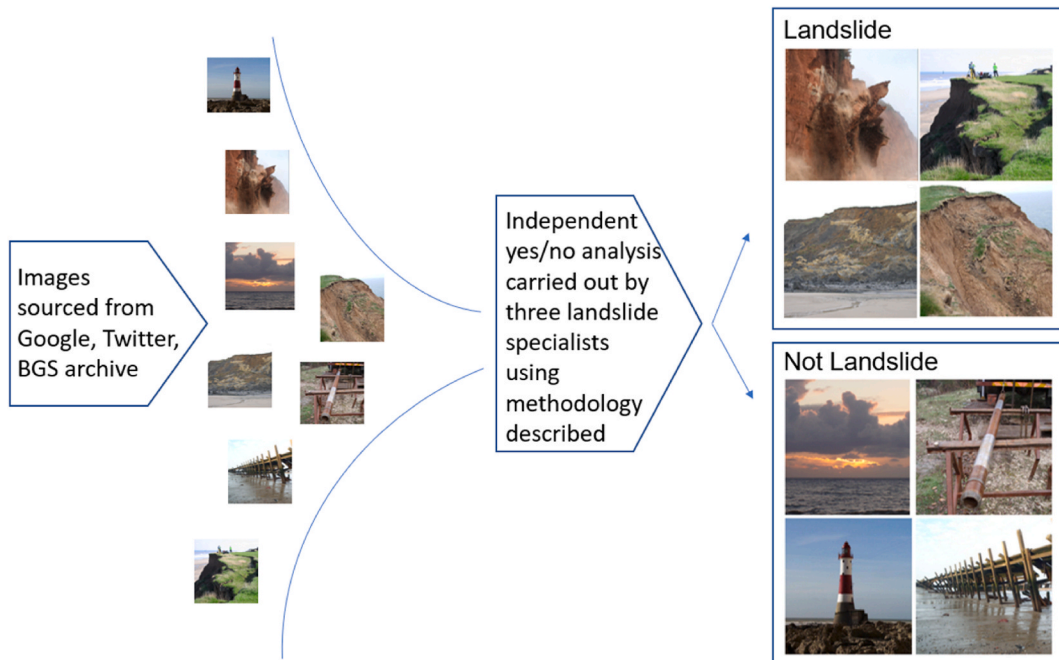


Fig. 1. Graphical representation showing the workflow for the development of the training model. The live system uses this training model to carry out real-time yes/no analysis on images harvested from Twitter using keywords. Photographs BGS © UKRI [2022].

A similar study was undertaken by Ref. [79] 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.

4. 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. Fig. 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 [80] all of which contained images of landslides, and non-landslides. 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. Fig. 2 shows examples of collected photographs divided into 'landslides' and 'not landslides' that demonstrates the kind of noise associated with image harvesting.

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 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.



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

4.1. 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. [81,82]), 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; Fig. 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., Fig. 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 (Fig. 5A) backscarps and extensional that could be faults (Fig. 5B), material engulfing buildings that could be the landslide deposit but could also have formed through other natural or manmade processes (Fig. 5C), debris falling onto roads that could be a landslide deposit or vegetation or mixed debris not associated with landsliding and (Fig. 5D) rivers in flow or flood channels that have a similar appearance to debris flow channels.

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

4.2. 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 [83] 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 Ref. [84].

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.

As suggested by the table, only about 23% of the images are labelled as landslide in the final dataset. This shows an imbalanced



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



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



Fig. 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].

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.

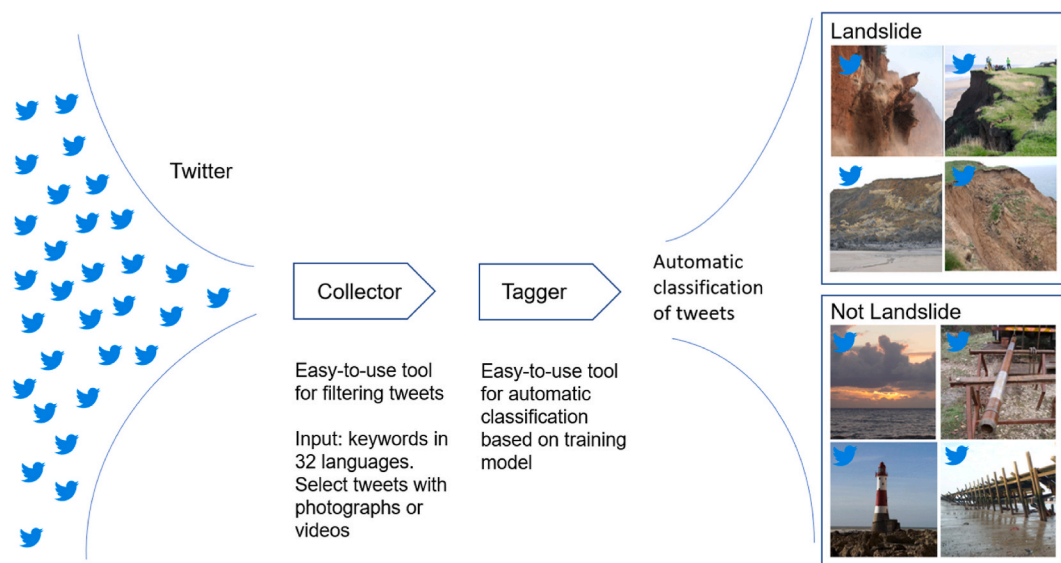


Fig. 6. Graphical representation illustrating the workflow involved in collecting, tagging and classifying images from Tweets as 'landslide' and 'not landslide'. Photographs BGS © UKRI [2022].

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

4.3. Demonstrator model results

The demonstrator model presented here is developed by training a convolutional neural network on the dataset introduced in this paper. Specifically, the dataset was split into training, validation, and test sets following a 70%, 10%, and 20% ratio, respectively. Then, a model based on ResNet-50 architecture was trained using Adam optimizer with an initial learning rate of 10^{-4} and a weight decay of 10^{-3} . On the test set, the trained model achieves an overall Accuracy = 87% and Precision = 74%, Recall (Sensitivity) = 67%, and F1-score = 70% for the landslide class. For more details about the extensive model development experiments, we refer the reader to Ref. [78].

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 (including near-and-exact duplicates) 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 Ref. [85]. To verify the system performance in the real-world, 3,600 images that were deemed relevant and non-duplicate by the system were randomly sampled and labelled by the system as landslide and non-landslide images. System-predicted labels were then compared with expert annotations to evaluate the performance of the demonstrator model. There were 123 images correctly labelled by the system as landslides (true positives) and 3,395 images correctly labelled as non-landslides (true negatives). On the other hand, there were 39 images that were incorrectly labelled as landslides (false positives) and 43 images that were incorrectly labelled as non-landslides (false negatives). This quantitative verification exercise showed that the demonstrator model can detect landslide reports with Accuracy = 98%, Precision = 76%, Recall (Sensitivity) = 74%, and F1-score = 75%. Below, we further evaluate the performance of this demonstrator model qualitatively on a few example images with the help of heat maps or class activation maps [86], which highlight the discriminative parts of a photograph that the model is paying attention to (Tables 2 and 3). The confidence scores are computed by the demonstrator model using the SoftMax function.


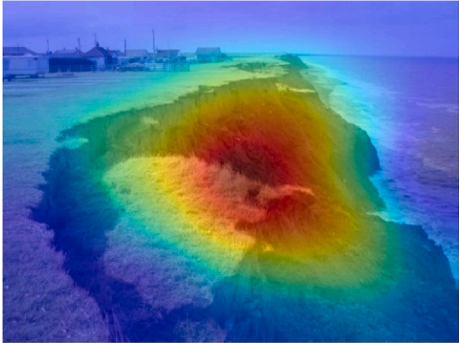

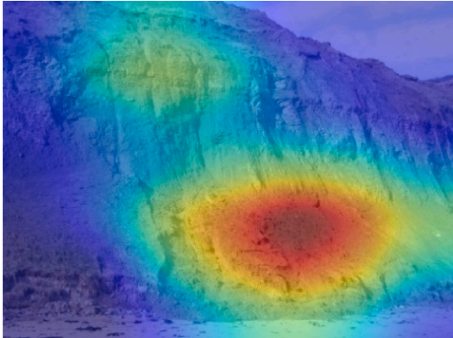

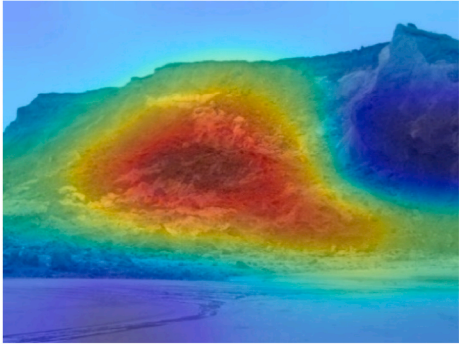
5. Discussion

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>).

Table 2

Photographs identified as containing a landslide showing the class activation map interpretation generated by the system. The confidence scores are computed by the demonstrator model using the SoftMax function. [Class activation map produced by QCRI Contains BGS © UKRI].

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

qcri.org/service.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 [87,88] 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 (<1 m and not strictly a landslide) and aerial photographs including multiple landslide events are not captured by the model (e.g., Fig. 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

Table 3

Photographs identified as NOT containing a landslide. The confidence scores are computed by the demonstrator model using the SoftMax function. [Class activation map produced by QCRI Contains BGS © UKRI].


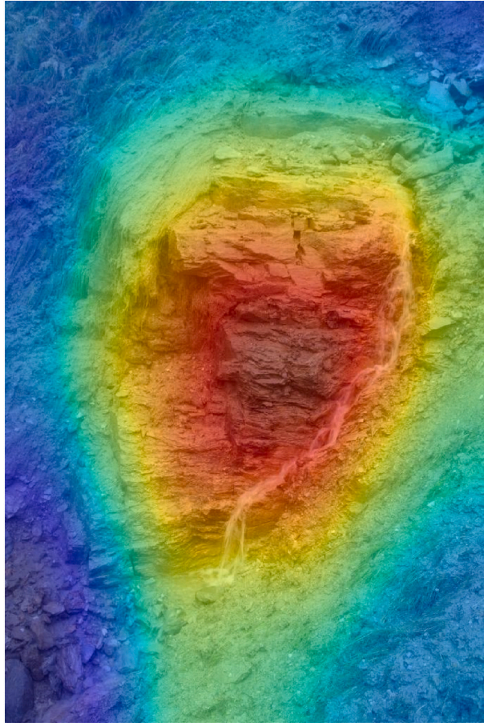

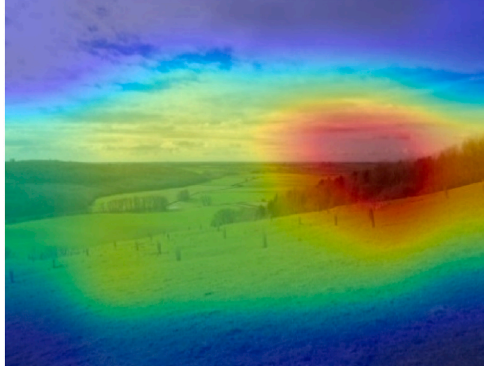

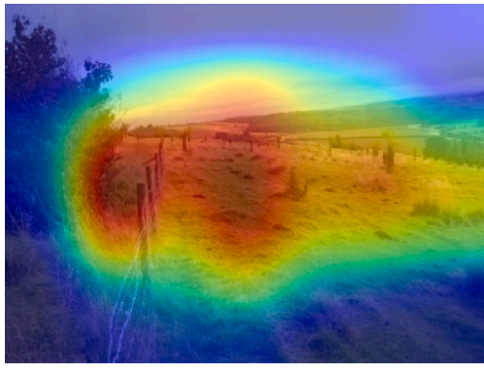
Original photograph	Heat map interpretation	Confidence
		<p>99.9% This is a rock exposure, not landslide</p>
		<p>99.9% A field, not landslide</p>
		<p>99.9% A field, not landslide</p>



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

used by Ref. [89] 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 [90] and the [91]. The work described in this paper could also be adapted to complement other hazard inventories, such as snow avalanches.

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.

6. 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.

(continued)

kiviä, kivivyöry, kivivyöryt, mutavyöry, mutavyöryt, sortuma, lumivyöry, snowmelt, snow melt, debris flow, cliff fall, cliff collapse, landslips

References

- [1] NASA, What is data latency?. <https://earthdata.nasa.gov/learn/backgrounders/data-latency>, 2020. (Accessed 27 September 2021).
- [2] S. Voigt, F. Giulio-Tonolo, J. Lyons, J. Kučera, B. Jones, T. Schneiderhan, G. Platzcek, K. Kaku, M. Kumar Hazarika, L. Czarán, S. Li, W. Pedersen, G. Kadiri James, C. Proy, D. Macharia Muthike, J. Bequignon, D. Guha-Sapir, Global trends in satellite-based emergency mapping, *Science* v353 (6296) (2016) 247–252, <https://doi.org/10.1126/science.aad8728>.
- [3] Copernicus, Linking with Early Warning Systems, 2021. <https://emergency.copernicus.eu/mapping/ems/linking-early-warning-systems>. (Accessed 26 September 2021).
- [4] K. Kaku, Satellite remote sensing for disaster management support: A holistic and staged approach based on case studies in Sentinel Asia, *Int. J. Disaster Risk Reduc.* 33 (2019) 417–432, <https://doi.org/10.1016/j.ijdrr.2018.09.015>.
- [5] M. Santangelo, M. Cardinali, F. Bucci, F. Fiorucci, A.C. Mondini, Exploring event landslide mapping using Sentinel-1 SAR backscatter products, *Geomorphology* v397 (2022) 10802, <https://doi.org/10.1016/j.geomorph.2021.108021>.
- [6] K. Burrows, O. Marc, D. Remy, Establishing the timings of individual rainfall-triggered landslides using Sentinel-1 satellite radar data, *Nat. Hazards Earth Syst. Sci.* (2022), <https://doi.org/10.5194/nhess-2022-21>. In press.
- [7] J.V. Fayne, A. Ahamed, J. Roberts-Pierel, A.C. Rumsey, D. Kirschbaum, Automated satellite-based landslide identification product for Nepal, *Earth Interact.* v23 (3) (2018) p1–21, <https://doi.org/10.1175/EI-D-17-0022.1>.
- [8] Y. Yang, S. Song, F. Yue, W. He, W. Shao, K. Zhao, W. Nie, Superpixel-based automatic image recognition for landslide deformation areas, *Eng. Geol.* 259 (2019) 105166, <https://doi.org/10.1016/j.enggeo.2019.105166>.
- [9] S. Ji, D. Yu, C. Shen, W. Li, Q. Xu, Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks, *Landslides* v17 (2020) 1337–1352, <https://doi.org/10.1007/s10346-020-01353-2>.
- [10] P. Gudzius, O. Kurasova, V. Darulis, E. Filatovas, Deep learning-based object recognition in multispectral satellite imagery for real-time applications, *Mach. Vis. Appl.* v32 (98) (2021), <https://doi.org/10.1007/s00138-021-01209-2>.
- [11] A.C. Mondini, F. Guzzetti, P. Reichenbach, M. Rossi, M. Cardinali, F. Ardizzone, Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images, *Remote Sensing of Environment* v115 (2011) 1743–1757, <https://doi.org/10.1016/j.rse.2011.03.006>.
- [12] Y. Yi, W. Zhang, A new deep-learning-based approach for earthquake-triggered landslide detection from single-temporal RapidEye satellite imagery, *IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens.* v13 (2020) 6166–6176, <https://doi.org/10.1109/JSTARS.2020.3028855>.
- [13] L. Li, M. Bensi, Q. Cui, G.B. Baecher, Y. Huang, Social media crowdsourcing for rapid damage assessment following a sudden-onset natural hazard event, *Int. J. Inf. Manag.* v60 (2021) 102378, <https://doi.org/10.1016/j.ijinfomgt.2021.102378>.
- [14] R. Laccasin, M. Devès, S.P. Hicks, J.-P. Ampuero, R. Bossu, L. Bruhat, Daryono, D.F. Wibisono, L. Fallou, E.J. Fielding, A.-A. Gabriel, J. Gurney, J. Krippner, A. Lomax, Muh M. Sudibyo, A. Pamumpuni, J.R. Patton, H. Robinson, M. Tingay, S. Valkaniotis, Rapid collaborative knowledge building via Twitter after significant geohazard events, *Geoscience Communication* v3 (2020) 129–146, <https://doi.org/10.5194/gc-3-129-2020>.
- [15] F. Alam, F. Ofli, M. Imran, Processing social media images by combining human and machine computing during crises, *Int. J. Hum. Comput. Interact.* v34 (4) (2018) 311–327, <https://doi.org/10.1080/10447318.2018.1427831>.
- [16] S. Goswami, S. Chakraborty, S. Ghosh, A. Chakrabarti, B. Chakraborty, A review on application of data mining techniques to combat natural disasters, *Ain Shams Eng. J.* 9 (2018) 365–378, <https://doi.org/10.1016/j.asej.2016.01.012>.
- [17] M.Y.S. Mohsin, 10 Social Media Statistics You Need to Know in 2021 [Infographic], 2021, 29th September 2021, <https://www.oberlo.com/blog/social-media-marketing-statistics>.
- [18] J. Phengsuwan, T. Shah, N.B. Thekkummal, Z. Wen, R. Sun, D. Pullarkatt, H. Thirugnanam, M.V. Ramesh, G. Morgan, P. James, R. Ranjan, Use of social media data in disaster management: a survey, *Future Internet* v13 (2021) 46, <https://doi.org/10.3390/fi13020046>.
- [19] F. Alam, M. Imran, F. Ofli, 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*, 2017, pp. 601–604, <https://doi.org/10.1145/3110025.3110164>.
- [20] C. Fan, F. Wu, A. Mostafavi, A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters, *IEEE Access* v8 (2020) 10478–10490, <https://doi.org/10.1109/ACCESS.2020.2965550>.
- [21] M.J. Froude, D.N. Petley, Global fatal landslide occurrence from 2004 to 2016, *Nat. Hazards Earth Syst. Sci.* v18 (2018) 2161–2181, <https://doi.org/10.5194/nhess-18-2161-2018>.
- [22] R. Emberson, D. Kirschbaum, T. Stanley, New global characterization of landslide exposure, *Nat. Hazards Earth Syst. Sci.* v20 (2020) 3413–3424, <https://doi.org/10.5194/nhess-20-3413-2020>.
- [23] E.M. Lee, D.K. Jones, *Background to Landslide Assessment*, Thomas Telford Ltd, London, UK, 2004.
- [24] O. Kjekstad, L.M. Highland, Economic and social impacts of landslides, in: D. Sassa, P. Canuti (Eds.), *Landslides—Disaster Risk Reduction*, Springer-Verlag, Berlin, 2008, pp. 573–587, https://doi.org/10.1007/978-3-540-69970-5_30.
- [25] D. Petley, *Global patterns of loss of life from landslides*, *Geology* v40 (10) (2012) 927–930.
- [26] D. Kirschbaum, T. Stanley, Y. Zhou, Spatial and temporal analysis of a global landslide catalog, *Geomorphology* v249 (2015) 4–15, <https://doi.org/10.1016/j.geomorph.2015.03.016>.
- [27] U. Haque, P. Blum, P.F. da Silva, P. Andersen, J. Pilz, S.R. Chalov, J.P. Malet, M.J. Aulfić, N. Andres, E. Poyiadji, P.C. Lamas, W. Zhang, I. Pesevski, H. G. Pétursson, T. Kurt, N. Dobrev, J.C.G. Davalillo, M. Halkia, S. Ferri, G. Gaprindashvili, J. Engström, D. Keellings, Fatal landslides in Europe, *Landslides* v13 (2016) 1545–1554, <https://doi.org/10.1007/s10346-016-0689-3>.
- [28] G. Herrera, R.M. Mateos, J.C. García-Davalillo, G. Grandjean, E. Poyiadji, R. Maftai, T.-C. Filipciuc, M.J. Aulfić, J. Jež, L. Podolszki, A. Trigila, C. Iadanza, H. Raetzko, A. Kociu, M. Przytucka, M. Kutak, M. Sheehy, X.M. Pellicer, C. McKeown, G. Ryan, V. Kopačková, M. Frei, D. Kuhn, R.L. Hermanns, N. Koulermou, C. A. Smith, M. Engdahl, P. Buxó, M. Gonzalez, C. Dashwood, H. Reeves, F. Cigna, P. Liščák, P. Paudits, V. Mikulénas, V. Demir, M. Raha, L. Quental, C. Sandić, B. Fusi, O.A. Jensen, Landslide databases in the geological surveys of Europe, *Landslides* v15 (2018) 359–379, <https://doi.org/10.1007/s10346-017-0902-z>.
- [29] *National Disaster Management Guidelines, National Disaster Management Information and Communication System (NDMICS)*, A publication of the National Disaster Management Authority, Government of India, New Delhi, 2011, ISBN 978-93-80440-12-5. February 2012.
- [30] P. Shi, J. Wang, M. Yang, J. Chen, Y. Pan, P. Wang, W. Fang, W. Zhou, Y. Wang, Y. Yuan, X. Xu, W. Luo, Understanding of natural disaster database design and compilation of digital atlas of natural disasters in China, *Geographic Information Sciences* v6 (2) (2000) 153–158, <https://doi.org/10.1080/10824000009480544>.
- [31] National Research Institute for Earth Science and Disaster Prevention, *Landslide Disasters*, 2021, 4th November 2021, <https://www.bosai.go.jp/e/research/database/earth-and-sand.html>.
- [32] J. Broeckx, M. Vanmaercke, R. Duchateau, J. Poesen, A data-based landslide susceptibility map of Africa, *Earth Sci. Rev.* v185 (2018) 102–121, <https://doi.org/10.1016/j.earscirev.2018.05.002>.
- [33] Public Safety Canada, *Canada Disaster Database*, March 2021, Last modified, 2013, 12th September 2013, <https://cdd.publicsafety.gc.ca/srchpg-eng.aspx>.

- [34] C. van Western, V. Jetten, D. Alkema, Validating National Landslide Susceptibility and Hazard Maps for Caribbean Island Countries: the Case of Dominica and Tropical Storm Erika, in: EGU General Assembly 2016, Held 17-22 April, 2016 in Vienna, Austria, 2016. <https://meetingorganizer.copernicus.org/EGU2016/EGU2016-4334.pdf>.
- [35] E.S. Jones, B.B. Mirus, R.G. Schmitt, R.L. Baum, W.J. Burns, M. Crawford, J.W. Godt, D.B. Kirschbaum, J.T. Lancaster, K.O. Lindsey, K.E. McCoy, S. Slaughter, T. A. Stanley, Summary Metadata – Landslide Inventories across the United States: U.S. Geological Survey data release, 2019, <https://doi.org/10.5066/P9E2A37P>.
- [36] B. Rosser, S. Dellow, S. Haubrock, P. Glasset, 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>, 2017. (Accessed 29 March 2021).
- [37] Geoscience Australia, Landslide Search. Software, 2012. <http://pid.geoscience.gov.au/dataset/ga/74273>.
- [38] C.S. Juang, T.A. Stanley, D.B. Kirschbaum, Using citizen science to expand the global map of landslides: introducing the cooperative open online landslide repository (COOLR), PLoS One v14 (7) (2019), e0218657, <https://doi.org/10.1371/journal.pone.0218657>.
- [39] NASA Landslide Reporter (2018) <https://pmm.nasa.gov/landslides/report.html> (21.08.2018).
- [40] ThinkHazard! (2020) <https://thinkhazard.org>, last updated 30th June 2020. [Accessed 4th November 2021].
- [41] C. Foster, C.V.L. Pennington, M.G. Culshaw, K. Lawrie, The national landslide database of great Britain: development, evolution and applications, Environ. Earth Sci. v66 (3) (2012) 941–953, <https://doi.org/10.1007/s12665-011-1304-5>.
- [42] B. Damm, M. Klose, The landslide database for Germany: closing the gap at national level, Geomorphology v249 (2015) 82–93, <https://doi.org/10.1016/j.geomorph.2015.03.021>.
- [43] Q. Han, K. Sassa, M. Mikoš, International programme on landslides (ipl): a programme of the icl for landslide disaster risk reduction, in: K. Sassa, M. Mikoš, S. Sassa, P.T. Bobrowsky, K. Takara, K. Dang (Eds.), Understanding and Reducing Landslide Disaster Risk. WLF 2020. ICL Contribution to Landslide Disaster Risk Reduction, Springer, Cham, 2021, https://doi.org/10.1007/978-3-030-60196-6_11.
- [44] A.D. Gibson, M.G. Culshaw, C. Dashwood, C.V.L. Pennington, Landslide management in the UK—the problem of managing hazards in a 'low-risk' environment, Landslides v10 (2012) 599–610, <https://doi.org/10.1007/s10346-012-0346-4>.
- [45] D.B. Kirschbaum, T. Stanley, Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness, Earth's Future v6 (3) (2018) 297–616, <https://doi.org/10.1002/2017EF000715>.
- [46] Y. Andersson-Sköld, R. Bergman, M. Johansson, E. Persson, L. Nyberg, Landslide risk management—a brief overview and example from Sweden of current situation and climate change, Int. J. Disaster Risk Reduc. 3 (2013) 44–61, <https://doi.org/10.1016/j.ijdr.2012.11.002>.
- [47] J.L. Wood, S. Harrison, L. Reinhardt, F.E. Taylor, Landslide databases for climate change detection and attribution, Geomorphology v355 (2020), <https://doi.org/10.1016/j.geomorph.2020.107061>.
- [48] J. Corominas, C. van Westen, P. Frattini, L. Cascini, J.P. Malet, S. Fotopoulou, F. Catani, M. Van Den Eeckhaut, O. Mavrouli, F. Agliardi, K. Pitilakis, M.G. Winter, M. Pastor, S. Ferlisi, V. Tofarni, J. Hervás, J.T. Smith, Recommendations for the quantitative analysis of landslide risk, Bull. Eng. Geol. Environ. v73 (2014) 209–263, <https://doi.org/10.1007/s10064-013-0538-8>.
- [49] O. Gunawan, T. Aldridge, Disaster Loss Data Management in Scotland, 2018. Technical Report DMS/2018/01 for the Scottish National Centre for Resilience, <https://core.ac.uk/download/pdf/296209205.pdf>.
- [50] United Nations Development Programme, A Comparative Review of Country-Level and Regional Disaster Loss and Damage Databases, Bureau for Crisis Prevention and Recovery, 2013, 4th November 2021, <https://www.preventionweb.net/publications/view/33067>.
- [51] Geomorphological Services Limited (GSL), Review of Research into Landsliding in Great Britain, Report to the Department of the Environment, Newport Pagnell, 1986/7, 15 volumes. Geomorphological Services (Publications and Reprographics) Limited.
- [52] Geomorphological Services Limited (GSL), Landslide Data Bank for Britain: User Manual Version 1.0, Geomorphological Services (Publications and Reprographics) Limited, Newport Pagnell, 1989.
- [53] D.B. Kirschbaum, R. Adler, Y. Hong, S. Hill, A.L. Lerner-Lam, A global landslide catalog for hazard applications—method, results, and limitations, Nat. Hazards v52 (2010) 561–575, <https://doi.org/10.1007/s11069-009-9401-4>.
- [54] F.E. Taylor, B.D. Malamud, K.A. Freeborough, D. Demerit, Enriching great Britain's national landslide database by searching newspaper archives, Geomorphology v249 (2015) 52–68, <https://doi.org/10.1016/j.geomorph.2015.05.019>.
- [55] J. Klimeš, J. Stemberk, J. Blahut, V. Krejčí, O. Krejčí, F. Hartvich, P. Kysel, Challenges for landslide hazard and risk management in 'low-risk' regions, Czech Republic-landslide occurrences and related costs (IPL project no. 197), Landslides v14 (2017) 771–780, <https://doi.org/10.1007/s10346-017-0798-7>.
- [56] T. Görüm, S. Fidan, Spatiotemporal variations of fatal landslides in Turkey, Landslides v18 (2021) 1691–1705, <https://doi.org/10.1007/s10346-020-01580-7>.
- [57] R. Franceschini, R. Ascanio, F. Catani, N. Casagli, Exploring a landslide inventory created by automated web data mining: the case of Italy, Landslides v19 (2022) 841–853.
- [58] European Commission, INSPIRE Principles, 2021 last updated 27/01/2021, <https://inspire.ec.europa.eu/inspire-principles/9>. (Accessed 27 January 2021).
- [59] M.T. Niles, B.F. Emery, A.J. Reagan, P.S. Dodds, C.M. Danforth, Social media usage patterns during natural hazards, PLoS One v14 (2) (2019), e0210484, <https://doi.org/10.1371/journal.pone.0210484>.
- [60] C.V.L. Pennington, K.A. Freeborough, C. Dashwood, T.J. Dijkstra, K. Lawrie, The national landslide database of great Britain: acquisition, communication and the role of social media, Geomorphology v249 (2015) 44–51, <https://doi.org/10.1016/j.geomorph.2015.03.013>.
- [61] K.A. Lee, J.R. Lee, P. Bell, A review of Citizen Science within the Earth Sciences: potential benefits and obstacles, Proc. Geologists' Assoc. v131 (6) (2020) 605–617, <https://doi.org/10.1016/j.pgeola.2020.07.010>.
- [62] R. Bossu, M. Laurin, G. Mazet-Roux, F. Roussel, R. Steed, 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, Seismol. Res. Lett. v86 (6) (2015) 1587–1592, <https://doi.org/10.1785/0220150147>.
- [63] R. Bossu, R. Steed, G. Mazet-Roux, F. Roussel, C. Etivant, L. Frobert, S. Godey, The key role of eyewitnesses in rapid impact assessment of global earthquakes, in: Earthquakes and Their Impact on Society, Springer, Cham, 2016, pp. 601–618, https://doi.org/10.1007/978-3-319-21753-6_25.
- [64] R. Bossu, F. Roussel, L. Fallou, M. Landès, R. Steed, G. Mazet-Roux, A. Dupont, L. Frobert, L. Petersen, LastQuake: from rapid information to global seismic risk reduction, Int. J. Disaster Risk Reduc. v28 (2018) 32–42, <https://doi.org/10.1016/j.ijdr.2018.02.024>.
- [65] R.J. Steed, A. Fuenzalida, R. Bossu, I. Bondár, A. Heinloo, A. Dupont, J. Saul, A. Strollo, Crowdsourcing triggers rapid, reliable earthquake locations, Sci. Adv. (2019), <https://www.science.org/doi/pdf/10.1126/sciadv.aau9824>.
- [66] R. Bossu, L. Fallou, M. Landès, F. Roussel, S. Julien-Laferrrière, J. Roch, R. Steed, Rapid public information and situational awareness after the November 26, 2019, Albania earthquake: lessons learned from the LastQuake system, Front. Earth Sci. v8 (2020) 235, <https://doi.org/10.3389/feart.2020.00235>.
- [67] M. Duncan, K. Mee, A. Hicks, S. Richardson, R.E.A. Robertson, M. Forbes, I. Ferdinand, C. Jordan, S. Loughlin, Citizen science using mobile phone technology in St Vincent & the Grenadines to facilitate near-real time multi-hazard observations, Geophys. Res. Abstr. v21 (2019) 1, <https://meetingorganizer.copernicus.org/EGU2019/EGU2019-14536.pdf>.
- [68] S. Kocaman, C.A. Gokceoglu, CitSci app for landslide data collection, Landslides v16 (2019) 611–615, <https://doi.org/10.1007/s10346-018-1101-2>.
- [69] C.E. Choi, Y. Cui, G.G.D. Zhou, Utilizing crowdsourcing to enhance the mitigation and management of landslides, Landslides v15 (2018) 1889–1899, <https://doi.org/10.1007/s10346-018-1034-9>.
- [70] BGS, Report a Landslide, 2021, 29th September 2021, <https://www2.bgs.ac.uk/reportlandslide/reportForm.html>.
- [71] R.L. Baum, L.M. Highland, P.T. Lyttle, J.M. Fee, E.M. Martinez, L.A. Wald, "Report a landslide" A website to engage the public in identifying geologic hazards, in: K. Sassa, P. Canuti, Y. Yin (Eds.), Landslide Science for a Safer Geoenvironment, Springer, Cham, 2014, https://doi.org/10.1007/978-3-319-04999-1_8.
- [72] J. Kostelnik, L. Wald, R. Schmitt, F.K. Rengers, Bond Fire Debris Flows, California, 2021, 16th September 2021, <https://landslides.usgs.gov/storymap/bond/>.
- [73] K. Cieslik, S. Puja, U. Madhab, D. Art, C. Russell, J. Clark, D.M. Raj, C. Amrit, Building resilience to chronic landslide hazard through citizen science, Front. Earth Sci. 7 (2019) p278, <https://doi.org/10.3389/feart.2019.00278>.
- [74] F. Guzzetti, G. Tonelli, Information system on hydrological and geomorphological catastrophes in Italy (SICD): a tool for managing landslide and flood hazards, Nat. Hazards Earth Syst. Sci. v4 (2004) 213–232, <https://doi.org/10.5194/nhess-4-213-2004>.

- [75] S.D. Moeller, "Regarding the pain of others": media, bias and the coverage of international disasters, *J. Int. Aff.* v59 (2) (2006) 173–196. The Globalization of Disaster (SPRING/SUMMER 2006), <https://www.jstor.org/stable/24358432>.
- [76] C.V.L. Pennington, A.M. Harrison, Landslide year?, 2012, *Geoscience Magazine* v23 (2013) 10–15, <https://www.geolsoc.org.uk/Geoscientist/Archive/June-2013/2012-Landslide-year>.
- [77] P. Lacroix, Landslides triggered by the Gorkha earthquake in the Langtang valley, volumes and initiation processes, *Earth Planets Space* v68 (2016) 46, <https://doi.org/10.1186/s40623-016-0423-3>.
- [78] F. Ofli, M. Imran, U. Qazi, J. Roch, C. Pennington, V.J. Banks, R. Bossu, Landslide Detection in Real-Time Social Media Image Streams, 2021, <https://doi.org/10.48550/arXiv.2110.04080>.
- [79] R. Can, S. Kocaman, C. Gokceoglu, 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.* v8 (7) (2019) 300, <https://doi.org/10.3390/ijgi8070300>.
- [80] BGS, GeoScenic, 2021, 29th September 2021, <http://geoscenic.bgs.ac.uk>.
- [81] O. Hungr, S. Leroueil, L. Picarelli, The Varnes classification of landslide types, an update, *Landslides* v11 (2014) 167–194, <https://doi.org/10.1007/s10346-013-0436-y>.
- [82] D.M. Cruden, D.J. Varnes, Landslide types and processes, in: A.K. Turner, R.L. Schuster (Eds.), *Landslides Investigation and Mitigation*, Transportation research board, US National Research Council. Special Report 247, Washington, DC, 1996, pp. 36–75 (Chapter 3).
- [83] J.L. Fleiss, Measuring nominal scale agreement among many raters, *Psychol. Bull.* 76 (5) (1971) 378, <https://doi.org/10.1037/h0031619>.
- [84] P.S. Bayerl, K.I. Paul, What determines inter-coder agreement in manual annotations? A meta-analytic investigation, *Comput. Ling.* 37 (4) (2011) 699–725, https://doi.org/10.1162/COLI_a_00074.
- [85] F. Ofli, U. Qazi, M. Imran, J. Roch, C. Pennington, V. Banks, R. Bossu, A Real-Time System for Detecting Landslide Reports on Social Media Using Artificial Intelligence, 2022. <https://arxiv.org/pdf/2202.07475v1.pdf>.
- [86] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2921–2929, <https://doi.org/10.1109/CVPR.2016.319>.
- [87] Twitter, Developer Terms: Geo Guidelines, last updated 22nd October 2014, 2022, 14th March 2022, <https://developer.twitter.com/en/developer-terms/geo-guidelines>.
- [88] Twitter, Developer Policy, Section Using Geo-Data, 2022, 14th March 2022, <https://developer.twitter.com/en/developer-terms/policy>.
- [89] M. Imran, U. Qazi, F. Ofli, TBCOV: two billion multilingual COVID-19 tweets with sentiment, entity, geo, and gender labels, *Data* 7 (1) (2022) 8, <https://doi.org/10.3390/data7010008>.
- [90] UK Government, Data Ethics Framework: Legislation and Codes of Practice for Use of Data, 2020. Last updated 16 September 2020, <https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-legislation-and-codes-of-practice-for-use-of-data>, 29th September 2021.
- [91] Locus Charter, Locus Charter Report, 2021. Last updated February 2021, https://ethicalgeo.org/wp-content/uploads/2021/03/Locus_Charter_March21.pdf, 29th September 2021.