LANDSATQUAKE: A LARGE-SCALE DATASET FOR PRACTICAL LANDSLIDE DETECTION

Vihaan Akshaay Rajendiran[1]*, **Amanda Roeliza Guilalas Hunt**[1]*, **Gen K. Li**[1]†, **Lei Li**[2]†
[1]University of California, Santa Barbara [2]Carnegie Mellon University
{vihaanakshaay,amandaroelizahunt,ligen}@ucsb.edu, leili@cs.cmu.edu

ABSTRACT

Identifying landslides from remote imagery is critical for rapid responses after landslide hazards and for assessing their environmental impacts. Existing datasets for landslide detection models are mostly based on multi-sourced, high-resolution (e.g., 1-5 m) satellite imagery from commercial companies (e.g., Planet Labs) and ultra-high-resolution (e.g., <1m) photos from unmanned aerial vehicle (UAV) surveys. However, obtaining those data is often economically expensive and labor-intensive, limiting their applicability. Here we present 'LandsatQuake,' a benchmark dataset composed of 31 landslide inventories from 21 earthquake-prone regions across the world covering a total area of 5.56×10^7 km^2 and spanning the last 40 years. This dataset emphasizes practicality by using satellite images acquired by Landsat, the only satellite system that has recorded Earth's land surface for >40 years. The public availability, high coverage of the world, and longevity make the Landsat data ideal for developing historical and recent landslide inventories caused by known triggers (e.g., earthquakes or rainstorms). Additionally, we demonstrate the challenges of applying existing computer vision algorithms to practical landslide detection problems by evaluating several baselines. The processing codes and dataset are publicly available at GitHub and Hugging Face, respectively.

1 INTRODUCTION

Landslides represent a severe natural hazard. In the United States, landslides are estimated to cause 25-50 deaths annually and more than one billion dollars in damages (Froude & Petley, 2018; Schuster & Highland, 2001). Most high-quality landslide inventories, which are critical for assessing damage, are produced from manual mapping, which is labor intensive and time consuming (Galli et al., 2008). We still lack ready open-source algorithms that allow efficient and rapid mapping of landslides after catastrophic landslide-triggering events. Automated landslide detection using machine learning is a rapidly growing field and provides a much faster alternative to manual mapping (Milledge et al., 2021).

Studies have shown that, in comparison to traditional machine learning methods, deep learning techniques can achieve superior performance when handling large-scale remote sensing data (qiang Yang et al., 2024). Although there have been various attempts to provide new architectures and techniques for landslide detection ((Ghorbanzadeh et al., 2019), (Su et al., 2021), Ullo et al. (2020)), most of these focus only on specific regions and fail to generalize to a broader distribution of satellite images or regions. In addition, current landslide detection models mainly rely on high-resolution imagery, which only cover recent time spans and are expensive. Additionally, current landslide detection models are unable to separate amalgamated landslides, which heavily skews volume calculations and gives an inaccurate estimate of landslide hazard magnitudes (Larsen et al., 2010). To address these challenges, we present **LandsatQuake (LQ)**, which offers:

- **Broad coverage and historical depth**: four decades of Landsat imagery across 21 earthquake-prone mountain ranges spanning 5.56×10^7 km^2.

*Equal contribution
†Corresponding author

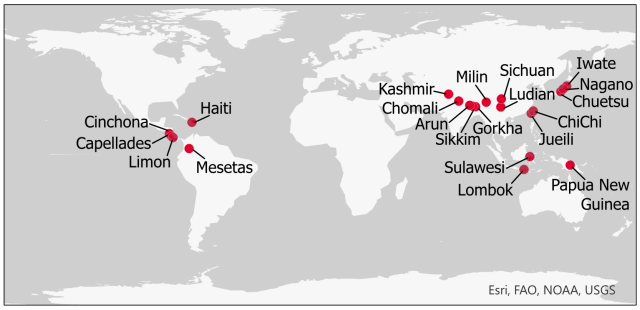

Figure 1: Map showing locations of all landslide inventories path of LandsatQuake

- **Practical relevance**: moderate 30 m resolution suitable for large, deep-seated landslides.
- **Integration with DEM data**: slope-based enhancements for improved detection.
- **Cost-effective and accessible data**: open-source, no licensing constraints.

2 DATASETS

In recent years, efforts have focused on developing benchmark datasets for landslide detection, exemplified by three recent contributions: **(A) CAS Landslide Dataset**, which uses multi-sensor RGB images (5 m or finer) and UAV drone surveys with polygons from open-source repositories (Xu et al., 2024); **(B) Landslide4Sense**, a competition dataset integrating 14 bands from Sentinel-2A and DEM/slope layers from ALOS (Ghorbanzadeh et al., 2022); and **(C) HR-GLDD**, a global dataset composed of 3 m PlanetScope imagery (RGB + NIR) covering various landslide triggers (Meena et al., 2023). While these datasets provide high-resolution (<10 m) images mostly from 2015 onward, they often lack global distribution and omit topographic information, limiting their applicability to moderate-resolution imagery and ignoring the valuable slope/elevation context crucial for capturing real-world landslide patterns (Guzzetti et al., 2012).

2.1 LANDSATQUAKE DATASET CONSTRUCTION

Satellite imagery for LandsatQuake was acquired from EarthExplorer (usg), focusing on scenes within ± 2 years of each earthquake event, with less than 20% cloud cover. Downloaded images spanned approximately $2^\circ \times 2^\circ$ in local UTM coordinate zones, drawing from Landsat 4, 5, and 8 missions at 30 m ground resolution. Landsat 7 was excluded due to its faulty scan line corrector (Storey et al., 2003). Landsat 4 and 5 images were formatted at 8-bit, while Landsat 8 was provided at 16-bit and subsequently converted to 8-bit to maintain consistency. Digital elevation model (DEM) data were sourced from the ALOS Global Digital Surface Model (**?**), downloaded in $1^\circ \times 1^\circ$ tiles at 30 m resolution. These DEM tiles were re-projected from EPSG 4326 to local UTM coordinate zones.

The LandsatQuake dataset itself incorporates 31 landslide inventories from 21 major earthquake-impacted mountain regions (Table 1), including polygons from the USGS earthquake-triggered landslide repository and prior studies (Marc et al., 2016; Schmitt et al., 2017; Li et al., 2022a). Given the 30 m resolution of Landsat, landslides under 10,000 m^2 were excluded because they are too small for reliable detection, and larger, deep-seated failures have greater geomorphic and socioeconomic impacts. Data processing involved mosaicking DEM tiles, compositing Landsat bands into multi-band images, and overlaying the landslide boundaries before segmenting the data into patches of 224×224 pixels. We then eliminated patches containing artifacts or high cloud cover, ensuring high-quality samples for model training.

In total, we included six Landsat reflectance bands (1, 2, 3, 4, 5, and 7), omitting the thermal band (6) due to its unavailability in Landsat 4 and 5. Natural-color (RGB) images were created by assigning band 3 to the red channel, band 2 to green, and band 1 to blue. Each RGB image underwent a contrast stretch to enhance visibility, followed by a slight intensity dampening (Figure 2). Afterwards, we

appended near-infrared (band 4), shortwave infrared (bands 5 and 7), the DEM, and a slope map (derived from the DEM) as additional channels, leveraging the known importance of topography for landslide detection (Guzzetti et al., 2012).

3 EXPERIMENTS

Experts: Identifying and quantifying every factor influencing landslide detection in satellite imagery can be both challenging and impractical. To circumvent this abstraction, we directly engaged experts by presenting them with both raw and annotated images, ensuring that any key challenges are reflected in the feedback provided. Specifically, we randomly sampled 100 images each from L4S, CAS, and LandsatQuake, annotated them using reference labels, then asked three experts to rate the ease of identifying landslides (1 = hardest/most likely to miss, 10 = easiest/most likely to detect).

As summarized in Table 1, LandsatQuake images received noticeably lower scores, highlighting the increased difficulty of detecting landslides using open-source Landsat data compared to higher-resolution imagery in existing datasets. This highlights the challenges associated with using open-source Landsat imagery for landslide detection in contrast to high-resolution images from previous deep learning-based datasets.

Score	**Expert A**	**Expert B**	**Expert C**	**Average**
CAS	5.61	7.60	6.03	6.41
Landslide4Sense	6.9	8.39	4.78	6.69
LandsatQuake	**2.0**	**1.81**	**1.12**	**1.64**

Table 1: Scores of three GeoScience experts and the average scores for each dataset

Models: Acknowledging the difficulty in detecting landslides from Landsat imagery, we measured the performance of various existing computer vision models to assess how current CV methods struggle with realistic, moderate-resolution datasets such as LandsatQuake.

Model (Backbone)	**mAP (Average)**	**mAP @ 50% IoU (mAP50)**
ConvNeXt Small (Liu et al., 2022)	0.000018	0.000090
ConvNeXt Tiny (Liu et al., 2022)	0.000064	0.000319
EfficientNet (Tan & Le, 2020)	0.000081	0.000202
Darknet (Redmon, 2013–2016)	0.000990	0.004950
ResNet18 (He et al., 2015)	0.000021	0.000140
ResNet50 FPN (He et al., 2015)	**0.002975**	**0.010006**
ResNet101 (He et al., 2015)	0.000033	0.000083
ResNet152 (He et al., 2015)	0.000024	0.000066
SqueezeNet 1.0 (Iandola et al., 2016)	0.000073	0.000163
SqueezeNet 1.1 (Iandola et al., 2016)	0.000043	0.000163
Swin (Liu et al., 2021b)	0.000040	0.000146
ViTDet Tiny (Li et al., 2022b)	0.000017	0.000086
ViTDet (Li et al., 2022b)	0.000043	0.000143

Table 2: Comparison of mAP scores for Faster R-CNN with different backbones evaluated on LQ.

To quantify these challenges, we trained Faster R-CNN (Ren et al., 2016) object detection models using PyTorch with various feature-extracting backbones. The dataset was prepared by segregating images and labels into training, validation, and test datasets with a split of 80%, 10%, and 10%, respectively, with data configurations loaded from YAML files. Since our data is very different in distribution from most standard pre-training datasets such as ImageNet (Krizhevsky et al., 2012) or Common Objects in Context (Lin et al., 2015), we evaluated our models by using backbones trained from scratch. We chose a batch size of 8 and 20 training epochs across all different configurations (backbones), and optimization was performed using the Stochastic Gradient Descent (SGD) (LeCun et al., 1998) optimizer with momentum 0.9 and a learning rate of 0.001. Model performance was evaluated based on mean Average Precision (mAP) at different Intersection-over-Union (IoU) (Everingham et al., 2010) thresholds (specifically from 0.5 to 0.95 in steps of 0.05), and models were

saved according to the best validation mAP observed. We have reported the mAP average and mAP @ IoU=50 values for various transformer-based and convolution-based backbones in Table 3.

The results show that traditional computer vision models, which are optimized and improved iteratively for standard deep learning benchmarks, perform poorly on LandsatQuake. Thus, applying existing models to Landsat-based data remains challenging.

Effect of Extra Spectral Bands: All experiments so far have used only RGB bands to ensure consistent evaluation across different satellite sources and alignment with models typically designed for RGB inputs. In this section, we explore the benefits of adding Landsat Bands 4 (NIR), 5 (SWIR 1), and 7 (SWIR 2) by training a Faster R-CNN model (ResNet-50 backbone) modified for six input channels. We trained for 20 epochs with an Adam (Kingma & Ba, 2017) optimizer (learning rate = 0.0005, weight decay enabled), employing a ReduceLROnPlateau scheduler (patience = 5, factor = 0.1) and custom normalization (mean = 0.0, std = 1.0). We initially masked the additional three channels and then progressively unmasked them to gauge their impact on mAP.

#Bands	**mAP (Average)**	**mAP @ 50% IoU (mAP50)**
3 bands	0.0009197	0.003822
4 bands	0.0010123	0.001817
5 bands	**0.0077761**	0.022527
6 bands	0.0053710	0.012381
DEM (8 bands)	0.0034852	**0.034486**

Table 3: Comparison of Faster R-CNN (ResNet-50) with different number of input bands.

Given the known utility of DEM and slope for landslide detection (Guzzetti et al., 2012; Wang et al., 2021), we appended these as seventh and eighth input channels for a parallel experiment. As shown in Table 4, while the overall mAP average is not always higher, the mAP at 50% IoU improves with each added band, indicating that DEM and slope in particular provide valuable ancillary information for more accurate landslide detection.

4 Limitations, Future Work, and Conclusion

First, due to the dataset's complexity, most out-of-the-box computer vision models struggle to learn effectively from LandsatQuake. This difficulty hampers the ability to conduct analysis of learned models (Grün et al., 2016) for insights, making it challenging to understand the models and visualize the internal representation of features.

Second, despite improvements in data quality and satellite technology over time, our dataset relies on shapefiles for incidents from various years in the past. Consequently, the only available source of satellite images for those old events is often Landsat (which is open-source and has a timespan of over 40 years (Wulder et al., 2022)), either due to licensing restrictions or because it is the sole source available for older events.

Third, the continuing evolution of satellite technology (García-Arenal & Fraile, 2017) makes it challenging to provide a universal format that can integrate our Landsat-based dataset with those using images from other sources.

Finally, although the inclusion of DEM bands improves performance, the typical mAP scores for small object detection problems (Liu et al., 2021a) and sparse datasets, including those in aerial imagery (Koyun et al., 2022), can be up to an order of magnitude higher. This emphasizes the need for more specific methods for landslide detection.

In conclusion, we present **LandsatQuake**, a large-scale dataset for landslide detection that covers 21 landslide-active regions and 31 landslide inventories spanning the last 40 years. By leveraging open-access Landsat imagery, this dataset provides historical data from multiple global locations, mirroring real-world conditions where high-resolution data are not always available. Our evaluations with standard computer vision models demonstrate that existing methods underperform in this setting, highlighting the need for targeted approaches. This initiative underlines the importance of methodologies and datasets that not only optimize traditional performance metrics but also address practical and realistic challenges in landslide detection.

REFERENCES

Landsat-4, Landsat-5, and Landsat-8 images courtesy of the U.S. Geological Survey.

Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88 (2):303–338, 2010. doi: 10.1007/s11263-009-0275-4. URL https://doi.org/10.1007/s11263-009-0275-4.

M. J. Froude and D. N. Petley. Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, 18:2161–2181, 2018. doi: 10.5194/nhess-18-2161-2018.

Mirco Galli, Francesca Ardizzone, Mauro Cardinali, Fausto Guzzetti, and Paola Reichenbach. Comparing landslide inventory maps. *Geomorphology*, 94(3):268–289, 2008. ISSN 0169-555X. doi: https://doi.org/10.1016/j.geomorph.2006.09.023. GIS technology and models for assessing landslide hazard and risk.

Fernando García-Arenal and Aurora Fraile. Chapter 56 - origin and evolution of satellites. In Ahmed Hadidi, Ricardo Flores, John W. Randles, and Peter Palukaitis (eds.), *Viroids and Satellites*, pp. 605–614. Academic Press, Boston, 2017. ISBN 978-0-12-801498-1. doi: https://doi.org/10.1016/B978-0-12-801498-1.00056-5. URL https://www.sciencedirect.com/science/article/pii/B9780128014981000565.

Omid Ghorbanzadeh, Thomas Blaschke, Khalil Gholamnia, Sansar Raj Meena, Dirk Tiede, and Jagannath Aryal. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sensing*, 11(2), 2019. ISSN 2072-4292. doi: 10.3390/rs11020196. URL https://www.mdpi.com/2072-4292/11/2/196.

Omid Ghorbanzadeh, Yonghao Xu, Pedram Ghamisi, Michael Kopp, and David Kreil. Landslide4sense: Reference benchmark data and deep learning models for landslide detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022. doi: 10.1109/TGRS.2022.3215209.

Felix Grün, Christian Rupprecht, Nassir Navab, and Federico Tombari. A taxonomy and library for visualizing learned features in convolutional neural networks, 2016.

Fausto Guzzetti, Alessandro Cesare Mondini, Mauro Cardinali, Federica Fiorucci, Michele Santangelo, and Kang-Tsung Chang. Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, 112(1):42–66, 2012. ISSN 0012-8252. doi: https://doi.org/10.1016/j.earscirev.2012.02.001. URL https://www.sciencedirect.com/science/article/pii/S0012825212000128.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size, 2016.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

Onur Can Koyun, Reyhan Kevser Keser, İbrahim Batuhan Akkaya, and Behçet Uğur Töreyin. Focus-and-detect: A small object detection framework for aerial images. *Signal Processing: Image Communication*, 104:116675, May 2022. ISSN 0923-5965. doi: 10.1016/j.image.2022.116675. URL http://dx.doi.org/10.1016/j.image.2022.116675.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In Peter L. Bartlett, Fernando C. N. Pereira, Christopher J. C. Burges, Léon Bottou, and Kilian Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, 2012.

I. Larsen, D. Montgomery, and O. Korup. Landslide erosion controlled by hillslope material. *Nature Geoscience*, 3:247–251, 2010. doi: 10.1038/ngeo776. URL https://doi.org/10.1038/ngeo776.

Yann LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pp. 9–50. Springer, 1998.

G. K. Li, S. Moon, and J. T. Higa. Residence time of over-steepened rock masses in an active mountain range. *Geophysical Research Letters*, 49(7):e2021GL097319, 2022a. doi: 10.1029/2021GL097319.

Yanghao Li, Hanzi Mao, Ross Girshick, and Kaiming He. Exploring plain vision transformer backbones for object detection, 2022b.

Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015.

Yang Liu, Peng Sun, Nickolas Wergeles, and Yi Shang. A survey and performance evaluation of deep learning methods for small object detection. *Expert Systems with Applications*, 172:114602, 2021a. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2021.114602. URL https://www.sciencedirect.com/science/article/pii/S0957417421000439.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows, 2021b.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. *CoRR*, abs/2201.03545, 2022. URL https://arxiv.org/abs/2201.03545.

Odin Marc, Niels Hovius, Patrick Meunier, Tolga Gorum, and Taro Uchida. A seismologically consistent expression for the total area and volume of earthquake-triggered landsliding. *Journal of Geophysical Research: Earth Surface*, 121(4):640–663, 2016. doi: 10.1002/2015JF003732.

S. R. Meena, L. Nava, K. Bhuyan, S. Puliero, L. P. Soares, H. C. Dias, M. Floris, and F. Catani. Hr-gldd: a globally distributed dataset using generalized deep learning (dl) for rapid landslide mapping on high-resolution (hr) satellite imagery. *Earth System Science Data*, 15(7):3283–3298, 2023. doi: 10.5194/essd-15-3283-2023. URL https://essd.copernicus.org/articles/15/3283/2023/.

David G. Milledge, Dino G. Bellugi, Jack Watt, and Alexander L. Densmore. Automated landslide detection outperforms manual mapping for several recent large earthquakes. *Natural Hazards and Earth System Sciences*, 2021. doi: 10.5194/nhess-2021-168. URL https://doi.org/10.5194/nhess-2021-168.

Zhi qiang Yang, Wen wen Qi, Chong Xu, and Xiao yi Shao. Exploring deep learning for landslide mapping: A comprehensive review. *China Geology*, 7(2):330–350, 2024. ISSN 2096-5192. doi: https://doi.org/10.31035/cg2024032. URL https://www.sciencedirect.com/science/article/pii/S2096519224001137. Special issue on Landslide Monitoring, Early Warning, and Risk Assessment.

Joseph Redmon. Darknet: Open source neural networks in c. http://pjreddie.com/darknet/, 2013–2016.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks, 2016.

Robert G. Schmitt, Hakan Tanyas, M. Anna Nowicki Jessee, Jing Zhu, Katherine M. Biegel, Kate E. Allstadt, Randall W. Jibson, Eric M. Thompson, Cees J. van Westen, Hiroshi P. Sato, David J. Wald, Jonathan W. Godt, Tolga Gorum, Chong Xu, Ellen M. Rathje, and Keith L. Knudsen. An open repository of earthquake-triggered ground-failure inventories. Data Series 1064, US Geological Survey, Reston, VA, 2017. URL https://doi.org/10.3133/ds1064.

Robert L. Schuster and Lynn M. Highland. Socioeconomic and environmental impacts of landslides in the western hemisphere. Open-File Report 2001-276, US Geological Survey, 2001. URL https://doi.org/10.3133/ofr01276.

James Storey, Pasquale Scaramuzza, Gail Schmidt, and Julia Barsi. LANDSAT 7 SCAN LINE CORRECTOR-OFF GAP-FILLED PRODUCT DEVELOPMENT, 2003. Work performed under USGS contract 03CRCN0001.

Zhi Su, Joe K. Chow, Ping-Seng Tan, and et al. Deep convolutional neural network–based pixel-wise landslide inventory mapping. *Landslides*, 18:1421–1443, 2021. doi: 10.1007/s10346-020-01557-6. Received 02 December 2019; Accepted 06 October 2020; Published 28 October 2020; Issue Date April 2021.

Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020.

Silvia Liberata Ullo, Amrita Mohan, Alessandro Sebastianelli, Shaik Ejaz Ahamed, Basant Kumar, Ramji Dwivedi, and G. R. Sinha. A new mask r-cnn based method for improved landslide detection, 2020.

Haojie Wang, Limin Zhang, Kesheng Yin, Hongyu Luo, and Jinhui Li. Landslide identification using machine learning. *Geoscience Frontiers*, 12(1):351–364, 2021. ISSN 1674-9871. doi: https://doi.org/10.1016/j.gsf.2020.02.012. URL https://www.sciencedirect.com/science/article/pii/S1674987120300542.

Michael A. Wulder, David P. Roy, Volker C. Radeloff, Thomas R. Loveland, Martha C. Anderson, David M. Johnson, Sean Healey, Zhe Zhu, Theodore A. Scambos, Nima Pahlevan, Matthew Hansen, Noel Gorelick, Christopher J. Crawford, Jeffrey G. Masek, Txomin Hermosilla, Joanne C. White, Alan S. Belward, Crystal Schaaf, Curtis E. Woodcock, Justin L. Huntington, Leo Lymburner, Patrick Hostert, Feng Gao, Alexei Lyapustin, Jean-Francois Pekel, Peter Strobl, and Bruce D. Cook. Fifty years of landsat science and impacts. *Remote Sensing of Environment*, 280:113195, 2022. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2022.113195. URL https://www.sciencedirect.com/science/article/pii/S0034425722003054.

Yezhou Xu, Chen Ouyang, Qiang Xu, et al. Cas landslide dataset: A large-scale and multisensor dataset for deep learning-based landslide detection. *Scientific Data*, 11:12, 2024. doi: 10.1038/s41597-023-02847-z. Received 11 August 2023; Accepted 11 December 2023; Published 02 January 2024.