https://doi.org/10.5194/egusphere-2026-6509
Preprint. Discussion started: 5 November 2026
© Author(s) 2026. CC BY 4.0 License.

Brief communication: Landsifier 2 – Python-based Software for Automated Landslide Type Classification

Nico Schlauf[1,*], Kushanav Bhuyan[2,*], Manan Kapoor[3,4], Lorenzo Nava[5,6], Kamal Rana[3], Atakan Aral[1], and Ugur Ozturk[7]

[1]Faculty of Computer Science, University of Vienna, Vienna, Austria
[2]State Key Laboratory of Geohazard Prevention and Geoenvironment Protection (SKLGP), Chengdu University of Technology, Chengdu, China
[3]Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, 14623, NY, USA
[4]Office of Research, University of Cincinnati, Cincinnati, USA
[5]Department of Geography, King's College London, London, UK
[6]King's Institute for Artificial Intelligence, King's College London, London, UK
[7]Department of Geography and Regional Research, University of Vienna, Vienna, Austria
[*]These authors contributed equally to this work.

Correspondence: Kushanav Bhuyan (kushanav.bhuyan@cdut.edu.cn)

Abstract.

Most mapped landslide inventories lack movement-type and trigger information, limiting their value for process-based research and hazard assessment. Here, we present Landsifier 2, an open-source Python library that enriches polygon inventories using three-dimensional geometry, topological descriptors, and configurable classification. Building on our existing trigger identification, this revamped library adds movement-type prediction, customizable classes, and geospatial outputs with class probabilities. Applications to independently labelled and previously undocumented inventories demonstrate its flexibility. Landsifier 2 provides a reproducible means of transforming landslide inventories into informative datasets for comparative geomorphology and hazard assessment.

1 Introduction

Landslide hazard models are typically informed by analysing past landslide events and their causative factors, including seismic activity (Xiao et al., 2024) and precipitation events (e.g., Strauch et al., 2018). Here, systematic landslide inventories becomes important as they provide the essential information for understanding landslide occurrences and supporting hazard assessments (Ozturk et al., 2021). With the advances in remote sensing and automated mapping (Behling and Roessner, 2020; Bhuyan et al., 2023), landslides commonly presented to the landslide research community in bulk as landslide polygons have become increasingly available across different geographical and climatic settings. A persistent limitation of such inventories is the absence of detailed information on triggering mechanisms and landslide typology, unless the data are event-based (Nava et al., 2022) or derived from field mapping (e.g., Görüm et al., 2023). This gap is not trivial: without this information, hazard

https://doi.org/10.5194/egusphere-2026-6509
Preprint. Discussion started: 5 November 2026
© Author(s) 2026. CC BY 4.0 License.

models risk misattributing landslide triggers, mischaracterising propagation processes, and ultimately producing less reliable risk assessments.

The geometry of landslides offers an alternative means of retrieving information about their characteristics. Landslides are inherently three-dimensional phenomena, with distinct shapes linked to both their triggering processes (Taylor et al., 2018) and their movement types (Cruden and Varnes, 1996). The morphological signatures of landslides encoded in the topography offer valuable clues to a landslide's origin and dynamics. Our recent work demonstrates a clear relationship between the three-dimensional geometry of landslides and the geomorphological responses visible in the terrain (Bhuyan et al., 2024). Building upon the former development, Rana et al. (2022) introduced Landsifier, an open-source Python library for classifying landslide triggering mechanisms from mapped inventories. Nevertheless, incorporating movement-type classification into a readily accessible and configurable workflow has remained a practical challenge, particularly when users seek to train models using their own inventory classifications. This lack hinders the operational use of large-scale, remotely mapped inventories (e.g., Behling et al., 2014) for hazard modelling, as manual type classification is resource-intensive and often unfeasible as we are talking about hundreds of thousands of landslide polygons across regional or global datasets.

To address this gap, we present Landsifier 2, an extended and streamlined version of the original library (Rana et al., 2022) that integrates automated landslide movement-type classification while retaining its existing trigger-classification functionality. Its initial functions focused on planform analysis to estimate whether landslides were earthquake-triggered or rainfall-induced (Rana et al., 2021). In our latest development, we have expanded Landsifier to classify landslides into the main types—flows, falls, slides, and complex—using only their mapped polygons. This new functionality is now fully integrated into Landsifier version 2.0 (https://pypi.org/project/Landsifier/), providing the landslide research and hazard modelling communities with a much-needed, automated solution for type classification at scale.

2 Landsifier 2: implementation and functionality

2.1 Geometric reconstruction and topological feature extraction

Landsifier 2 provides an automated framework for extracting topological descriptors from mapped landslide geometries. The `extract_features` function integrates two-dimensional polygon inventories with elevation information obtained from either a user-provided digital elevation model (DEM) or automatically retrieved Shuttle Radar Topography Mission (SRTM) data. Landslide planforms are reconstructed as three-dimensional point representations by interpolating elevation values at polygon boundary vertices. Users may optionally incorporate regularly spaced interior points to capture additional information on the topographic structure within individual landslides. Boundary-only sampling provides a computationally simpler representation, whereas interior sampling offers a more detailed characterization of the landslide surface, potentially benefiting classification schemes that distinguish specific movement types. To standardize the three-dimensional representations, the spatial coordinates (x, y, z) are uniformly normalized by the elevation difference between the highest and lowest points within each landslide polygon. This landslide-specific relief factor preserves the relative geometric proportions of the three-dimensional point configuration while expressing spatial distances relative to the landslide's vertical relief. The normalized point repre-

https://doi.org/10.5194/egusphere-2026-6509
Preprint. Discussion started: 5 November 2026
© Author(s) 2026. CC BY 4.0 License.

sentations are subsequently processed using persistent homology, implemented through the giotto-tda library (Tauzin et al., 2021). Vietoris–Rips (Zomorodian, 2010) filtrations characterize the emergence and disappearance of topological features across increasing spatial scales, considering homology dimensions 0–2. The resulting persistence information is summarized into numerical descriptors, including persistence entropy and topological feature counts, which form the feature space used for classification. By integrating geometric reconstruction, normalization, and topological feature extraction into a common workflow, Landsifier 2 eliminates the need for users to independently implement these computational procedures.

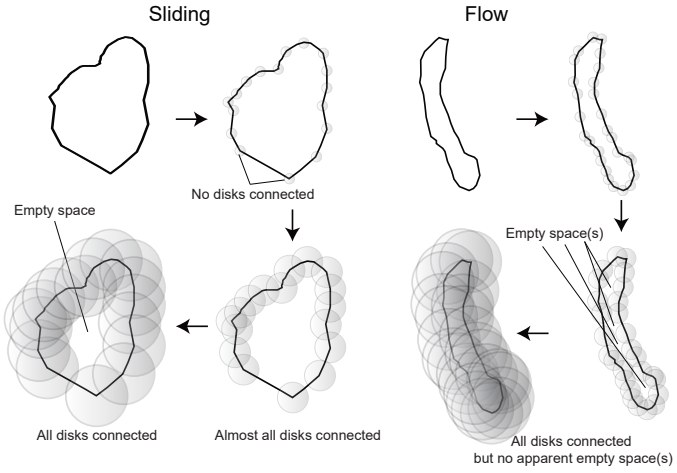


Figure 1. Example process to extract features from a 2D landslide planform. In each point along the landslide polygon, a buffer zone is introduced, and its size is gradually increased to identify different properties. In the figure, we show an example where we follow the empty spaces that appear and vanish as the buffer area increases. One can see in a flow process more voids appear and vanish, when compared to a sliding process.

2.2 Configurable classification and inventory enrichment

The extended library enables user-defined landslide classification through the `train_classifier` function, which trains Random Forest models using the extracted topological descriptors and corresponding movement-type labels. Rather than imposing a fixed classification scheme, Landsifier allows users to define categories according to their inventory nomenclature and desired level of detail, ranging from broad movement types (e.g., slides, flows, and falls) to more specific classifications (e.g., rockslides, debris flows, and earthflows), provided suitable labelled training data are available. To accommodate differences in the predictive relevance of topological descriptors, the framework supports both full-feature and optimized-feature classifiers. The former retains the complete set of extracted descriptors, whereas the latter uses feature-importance ranking to select a reduced subset of informative features. Model evaluation supports stratified cross-validation and standard classification metrics, allowing users to compare feature configurations and assess model performance. In the experiments presented here, five-fold cross-validation was adopted, with macro-F1 scores averaged across folds and feature selection performed independently within each training fold. Trained classifiers are saved for subsequent application to unlabelled inventories using the

https://doi.org/10.5194/egusphere-2026-6509
Preprint. Discussion started: 5 November 2026
© Author(s) 2026. CC BY 4.0 License.

`predict_inventory` function. During prediction, the target polygons undergo the same geometric reconstruction, normalization, and topological feature extraction procedures applied during training, ensuring consistency between the training and prediction feature spaces. The resulting movement-type predictions and associated class probabilities are appended to the original landslide geometries and returned as a GeoDataFrame, facilitating visualization, export, and integration into existing geospatial workflows. Together, these functionalities extend Landsifier beyond its original trigger-classification application, providing an accessible and configurable framework for enriching landslide inventories with movement-type information.

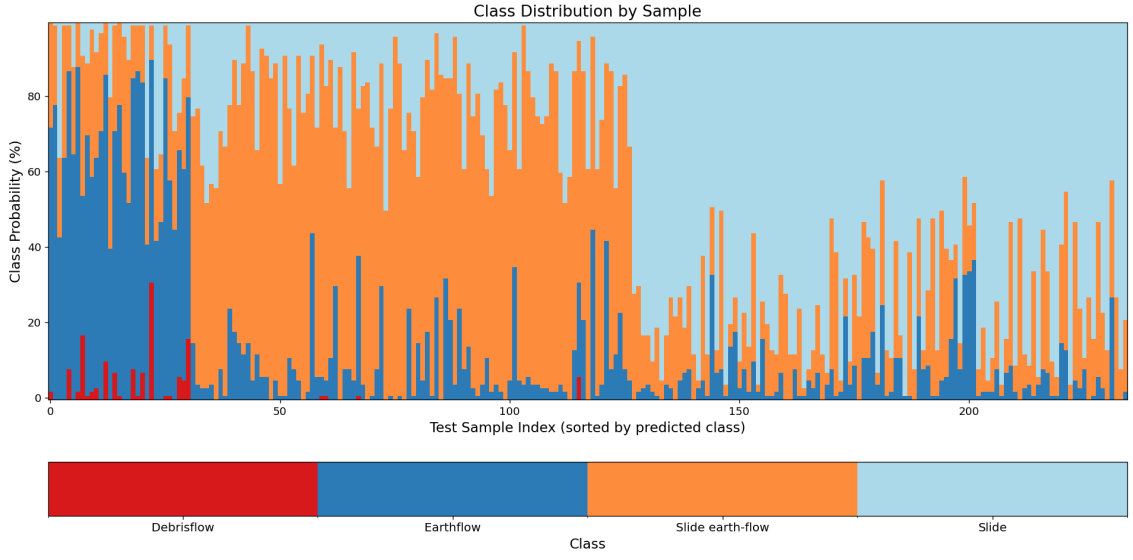


Figure 2. An example of visualization with predictions denoted for different samples across the test inventory sorted by the class of failure typology/movement.

3 Summary of the workflow

The library provides an end-to-end, modular workflow (Figure 3) for landslide classification using topological data analysis (TDA), persistent homology, and machine learning. Persistent homology-based features are extracted and classified via a Random Forest model, with each stage of the pipeline designed to be integrated into external geospatial or ML frameworks. In this section, we walk through the complete process of installing Landsifier, generating a model based on a labelled landslide inventory, and using it to classify an inventory of unlabelled landslides. The results will be presented the next section.

For feature generation, the library uses the GDAL API for geospatial handling and Giotto-TDA for topological analysis. On macOS, installing both can be challenging due to limited installation wheels for recent Python versions, whereas Windows users can typically install them using available wheels. Our testing shows GDAL installs most reliably on Anaconda Python 3.8 or earlier, using the `conda-forge` channel. Giotto-TDA is also installed exclusively via `conda-forge`, which may be in-

https://doi.org/10.5194/egusphere-2026-6509
Preprint. Discussion started: 5 November 2026
© Author(s) 2026. CC BY 4.0 License.

compatible with some Homebrew-based setups. To ensure smooth installation of `Landsifier` v2.0 across platforms, we recommend creating an Anaconda environment with Python 3.8 or lower, which satisfies both GDAL and Giotto-TDA dependencies efficiently.

The workflow begins with the `extract_features` function, which overlays landslide inventory shapefiles on a digital elevation model (DEM) to extract topological features, saving them as organized `.npy` files (Figure 3). The `train_classifier` function then trains a Random Forest Classifier, producing both an all-features model and an optimized-features model based on feature importance analysis. Predictions on new data are generated with the `predict_inventory` function, outputting a GeoDataFrame of classified landslides for further geospatial analysis. Model confidence can be visualized using the `get_image_data` function, which produces a visualization of class probabilities. Additional features—such as interior point sampling, cross-validation, and model persistence—support both research and operational landslide mapping.

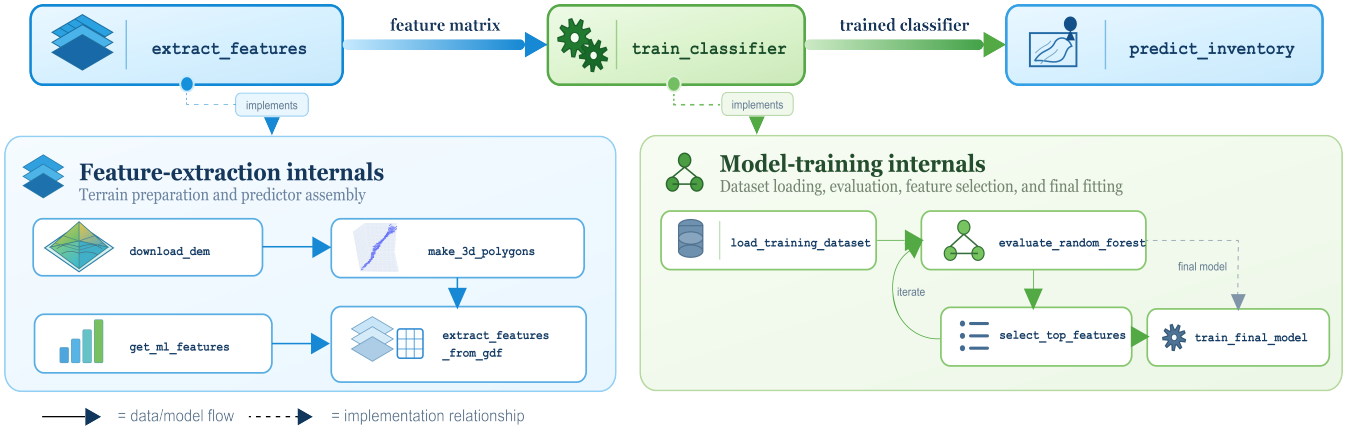


Figure 3. A workflow of the different library modules and functions of Landsifier 2.0

https://doi.org/10.5194/egusphere-2026-6509
Preprint. Discussion started: 5 November 2026
© Author(s) 2026. CC BY 4.0 License.

4 Experimental demonstration and geographical application

We evaluated Landsifier 2 using three independently labelled landslide inventories from Liguria (Italy; approximately 1200 landslides), Denmark (approximately 3000), and Washington State (USA; approximately 13 000). The inventories use different movement-type classifications: slides, flows, falls, and complex landslides in Liguria; slides, flows, falls, and spreads in Denmark; and slides, flows, and falls in Washington. We trained separate classifiers for each inventory to demonstrate the framework's adaptability to different movement-type nomenclatures. For each region, we evaluated four configurations combining boundary-only or boundary-and-interior point sampling with either the complete or optimized feature set.

For each inventory and configuration, we randomly partitioned the data into training and test subsets for multiple iterations, using a training-test split (and stratification, if and where applicable. For the optimized-feature configurations, feature importances from the training runs were averaged across iterations to rank the descriptors. We selected the optimized feature set using (selection rule, i.e., the top k features). Macro-F1 was calculated on the test subset in each iteration and summarized in Table 1, Figure 2.

Classification performance varied across inventories and feature configurations (Table 1). With boundary-only sampling and optimized features, the reported macro-F1 scores were 0.95, 0.87, and 0.85 for Liguria, Denmark, and Washington, respectively. Including interior points changed these scores to 0.77 and 0.67 for Liguria and Denmark, and 0.90 for Washington. Feature optimization maintained or improved performance relative to the complete descriptor set. These results indicate that a more detailed geometric representation does not necessarily improve movement-type discrimination. The suitability of boundary and interior sampling appears to depend on inventory characteristics, supporting configurable feature extraction rather than a single prescribed representation.

To demonstrate applications beyond the labelled inventories, we applied the independently trained classifiers to undocumented landslide inventories from Beijing and Hunan (China) and Japan (Fig. 4). Because independently verified movement-type labels were unavailable, we evaluated agreement among the three classifiers using an ensemble consensus derived from their hard class predictions. The consensus was the most frequent predicted class; when all three predictions differed, we selected the lexicographically first class label. Thus, a three-way disagreement among flow, slide, and fall predictions was assigned to fall. We quantified each classifier's agreement with these consensus labels using macro-F1 and selected the highest-scoring classifier for each application region. We retained the original movement-type labels without prior harmonization to represent the heterogeneity encountered when applying models trained on independently developed inventories. These scores therefore measure agreement with an ensemble-derived reference, not independently validated classification accuracy. The tie-breaking rule may affect model rankings where predictions are highly discordant.

The cross-regional applications also highlight a broader challenge in automated inventory enrichment. Movement-type classifications reflect institutional mapping protocols, expert interpretation, inventory objectives, and the level of taxonomic detail used during mapping. Differences among regional classifiers may therefore reflect both geographical variation in landslide characteristics and inconsistencies in movement-type definitions; these experiments do not distinguish their individual contributions. A more standardized, hierarchical nomenclature, progressing from broad movement categories to specific subclasses,

https://doi.org/10.5194/egusphere-2026-6509
Preprint. Discussion started: 5 November 2026
© Author(s) 2026. CC BY 4.0 License.

could improve the comparability of training inventories and resulting predictions. Landsifier 2 accommodates existing classification diversity through user-defined labels while providing a common computational framework for landslide inventory enrichment.

Landslide inventory	AF w/o IP	OF w/o IP	AF w/ IP	OF w/ IP
Liguria	0.95	0.95	0.69	0.77
Denmark	0.84	0.87	0.53	0.67
Washington	0.84	0.85	0.86	0.90

AF All features **w/o IP** Without interior points
OF Optimized features **w/ IP** With interior points

Table 1. F1 scores for the twelve Landsifier models trained using different configurations.

5 Conclusions

Landsifier 2 extends the original trigger-classification library into an accessible and configurable framework for enriching polygon-based landslide inventories with movement-type information. By integrating automated topological feature extraction, configurable geometric sampling, feature optimization, user-defined classifier training, and geospatial prediction outputs, the extended library facilitates the application of previously developed topology-based classification methods without requiring users to independently implement the underlying computational procedures.

Experiments involving three independently labelled regional inventories demonstrate that classification performance depends on the choice of geometric sampling strategy and feature configuration. While feature optimization consistently maintained or improved classification performance, the inclusion of interior points produced contrasting outcomes among inventories, highlighting the importance of adapting feature representations to the available data and classification objectives. Applications to undocumented inventories further demonstrate the framework's operational flexibility but also expose inconsistencies among independently trained classifiers. These differences may arise from regional variations in landslide characteristics and heterogeneous movement-type definitions, while ensemble agreement alone cannot establish predictive accuracy.

The flexibility to accommodate user-defined movement types provides opportunities for enriching existing inventories at different levels of classification detail. Nevertheless, broader applications would benefit from more consistent, hierarchical movement-type nomenclatures and independent validation across geographical settings. Landsifier 2 thus provides a common computational framework for advancing automated landslide inventory characterization while retaining the adaptability required by diverse mapping conventions and research objectives.

https://doi.org/10.5194/egusphere-2026-6509
Preprint. Discussion started: 5 November 2026
© Author(s) 2026. CC BY 4.0 License.

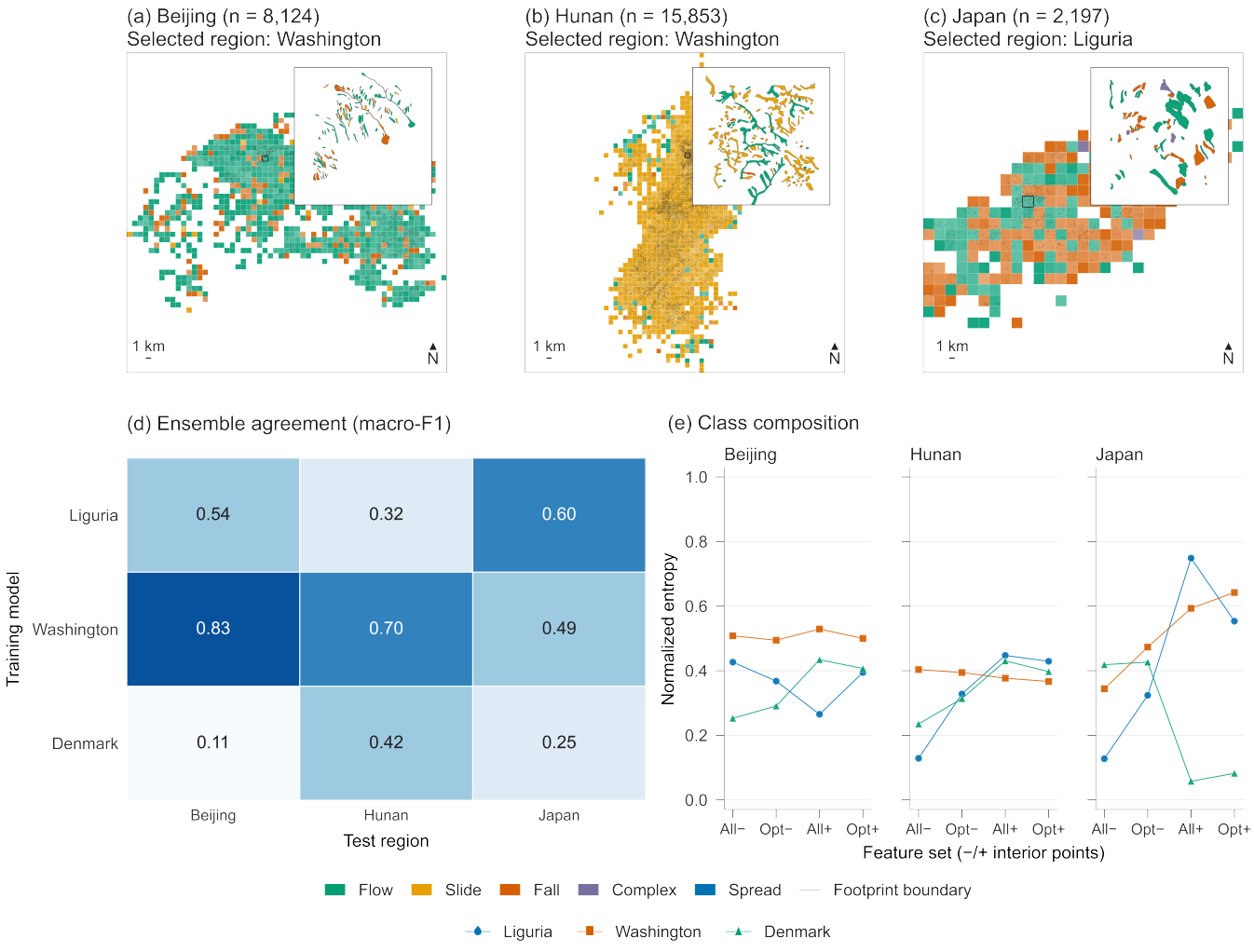


Figure 4. Cross-regional application of independently trained Landsifier classifiers to landslide inventories without documented movement types in Beijing, Hunan, and Japan. Agreement was evaluated using the original movement-type labels from each training inventory without prior taxonomic harmonization. Reported agreement measures characterize consistency among model predictions rather than independently validated classification accuracy.

https://doi.org/10.5194/egusphere-2026-6509
Preprint. Discussion started: 5 November 2026
© Author(s) 2026. CC BY 4.0 License.

Code and data availability. We performed all data processing and analysis in Python using its standard libraries. The analysis scripts are openly available in https://pypi.org/project/Landsifier/

The Italian dataset was sourced from the Inventario dei Fenomeni Franosi in Italia (IFFI). Established in 1999 to map and catalog landslides across Italy, the IFFI database (www.progettoiffi.isprambiente.it) contains information for over 250,000 usable landslide polygons. Entries were compiled and verified through aerial photo interpretation, historical records, and field surveys, and are classified following the schemes of Cruden et al. (1996). For this study, we selected the polygonal landslide records and post-processed them to match the spatial extent and 25-m resolution of the EU-DEM.

The US dataset draws on the inventories from the Washington State Landslide Inventory Database (WASLID updated 2018/08/01; Slaughter et al. (2017)), mapped by the Department of Natural Resources, Washington Geological Survey (WGS).

Author contributions. NS, MK, KB, KR developed the library and contributed to the writing of the manuscript. KB and KR developed the tools that are used in the library. LN tested the functions and ensured an error-free workflow. KB, KR, and UO conceptualized the research. UO and KB drafted the first version of the manuscript with review and editing provided by LN and AA.

Competing interests. One author is a member of the editorial board of Natural Hazards and Earth System Sciences.

Acknowledgements. KB is funded by the National Natural Science Foundation of China - Research Fund for International Scientists (NSFC-RFIS, Grant No. W2633127) and the National Foreign Experts Program, China (Grant No. Y20250027).

UO is funded by the European Union (ERC, UrbanSlide, Grant 101160656, Doi: 10.3030/101160656). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

https://doi.org/10.5194/egusphere-2026-6509
Preprint. Discussion started: 5 November 2026
© Author(s) 2026. CC BY 4.0 License.

References

Behling, R. and Roessner, S.: Multi-temporal landslide inventory for a study area in Southern Kyrgyzstan derived from multi-sensor optical satellite time series data (1986 – 2013), https://doi.org/10.5880/GFZ.1.4.2020.002, artwork Size: 2 Files Medium: application/octet-stream,application/octet-stream Pages: 2 Files, 2020.

Behling, R., Roessner, S., Kaufmann, H., and Kleinschmit, B.: Automated Spatiotemporal Landslide Mapping over Large Areas Using RapidEye Time Series Data, Remote Sensing, 6, 8026–8055, https://doi.org/10.3390/rs6098026, 2014.

Bhuyan, K., Tanyaş, H., Nava, L., Puliero, S., Meena, S. R., Floris, M., van Westen, C., and Catani, F.: Generating multi-temporal landslide inventories through a general deep transfer learning strategy using HR EO data, Scientific Reports, 13, 162, https://doi.org/10.1038/s41598-022-27352-y, 2023.

Bhuyan, K., Rana, K., Ferrer, J. V., Cotton, F., Ozturk, U., Catani, F., and Malik, N.: Landslide topology uncovers failure movements, Nat Commun, 15, 2633, https://doi.org/10.1038/s41467-024-46741-7, 2024.

Cruden, D. M., Varnes, D. J., Turner, A., and Schuster, R.: Landslides: investigation and mitigation, Transportation research board special report, 247, 36–75, 1996.

Cruden, M. D. and Varnes, J. D.: Landslide Types and Processes, in: Landslides: Investigation and Mitigation, no. 247 in Special Report(National Research Council (U.S.) Transportation Research Board), pp. 36–75, National Academy Press, Washington, D.C, ISBN 0-309-06208-X, 1996.

Görüm, T., Tanyas, H., Karabacak, F., Yılmaz, A., Girgin, S., Allstadt, K. E., Süzen, M. L., and Burgi, P.: Preliminary documentation of coseismic ground failure triggered by the February 6, 2023 Türkiye earthquake sequence, Engineering Geology, 327, 107 315, https://doi.org/10.1016/j.enggeo.2023.107315, 2023.

Nava, L., Bhuyan, K., Meena, S. R., Monserrat, O., and Catani, F.: Rapid Mapping of Landslides on SAR Data by Attention U-Net, Remote Sensing, 14, https://doi.org/10.3390/rs14061449, 2022.

Ozturk, U., Pittore, M., Behling, R., Roessner, S., Andreani, L., and Korup, O.: How robust are landslide susceptibility estimates?, Landslides, 18, 681–695, https://doi.org/10.1007/s10346-020-01485-5, 2021.

Rana, K., Ozturk, U., and Malik, N.: Landslide Geometry Reveals its Trigger, Geophys Res Lett, https://doi.org/10.1029/2020GL090848, 2021.

Rana, K., Malik, N., and Ozturk, U.: Landsifier v1.0: a Python library to estimate likely triggers of mapped landslides, Nat. Hazards Earth Syst. Sci., 22, 3751–3764, https://doi.org/10.5194/nhess-22-3751-2022, 2022.

Slaughter, S., Burns, W., Mickelson, K., Jacobacci, K., Biel, A., and Contreras, T.: Protocol for landslide inventory mapping from lidar data in Washington State, Wash. Geol. Surv. Bull, 82, 2017.

Strauch, R., Istanbulluoglu, E., Nudurupati, S. S., Bandaragoda, C., Gasparini, N. M., and Tucker, G. E.: A hydroclimatological approach to predicting regional landslide probability using Landlab, Earth Surf. Dynam., 6, 49–75, https://doi.org/10.5194/esurf-6-49-2018, 2018.

Tauzin, G., Lupo, U., Tunstall, L., Pérez, J. B., Caorsi, M., Medina-Mardones, A. M., Dassatti, A., and Hess, K.: giotto-tda:: A topological data analysis toolkit for machine learning and data exploration, Journal of Machine Learning Research, 22, 1–6, 2021.

Taylor, F. E., Malamud, B. D., Witt, A., and Guzzetti, F.: Landslide shape, ellipticity and length-to-width ratios, Earth Surf Processes Landf, 43, 3164–3189, https://doi.org/10.1002/esp.4479, 2018.

Xiao, S., Xiao, T., Jiang, R., Wang, H., Ju, L., and Zhang, L.: Two-phase strategy for rapid and unbiased assessment of earthquake-induced landslides, Engineering Geology, 336, 107 562, https://doi.org/10.1016/j.enggeo.2024.107562, 2024.

https://doi.org/10.5194/egusphere-2026-6509
Preprint. Discussion started: 5 November 2026
© Author(s) 2026. CC BY 4.0 License.



205 Zomorodian, A.: Fast construction of the Vietoris-Rips complex, Computers & Graphics, 34, 263–271, 2010.