

Rana, K., Ozturk, U., Malik, N. (2021): Landslide
Geometry Reveals its Trigger. - Geophysical Research
Letters, 48, e2020GL090848.

<https://doi.org/10.1029/2020GL090848>

Geophysical Research Letters

RESEARCH LETTER

10.1029/2020GL090848

Key Points:

- We present a method to identify the trigger of landslides based on their geometries
- Our method could improve the utility of the existing landslide databases for hazard modeling
- We show that geometric features of a landslide reflect its trigger

Supporting Information:

- Supporting Information S1

Correspondence to:

N. Malik,
nxmsma@rit.edu



Citation:

Rana, K., Ozturk, U., & Malik, N. (2021). Landslide geometry reveals its trigger. *Geophysical Research Letters*, 48, e2020GL090848. <https://doi.org/10.1029/2020GL090848>

Received 16 SEP 2020

Accepted 5 JAN 2021

Landslide Geometry Reveals its Trigger

Kamal Rana¹, Ugur Ozturk^{2,3} , and Nishant Malik⁴ 

¹Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, USA, ²Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany, ³Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany, ⁴School of Mathematical Sciences, Rochester Institute of Technology, Rochester, NY, USA

Abstract Electronic databases of landslides seldom include the triggering mechanisms, rendering these inventories unusable for landslide hazard modeling. We present a method for classifying the triggering mechanisms of landslides in existing inventories, thus, allowing these inventories to aid in landslide hazard modeling corresponding to the correct event chain. Our method uses various geometric characteristics of landslides as the feature space for the machine-learning classifier *random forest*, resulting in accurate and robust classifications of landslide triggers. We applied the method to six landslide inventories spread over the Japanese archipelago in several different tests and training configurations to demonstrate the effectiveness of our approach. We achieved mean accuracy ranging from 67% to 92%. We also provide an illustrative example of a real-world usage scenario for our method using an additional inventory with unknown ground truth. Furthermore, our feature importance analysis indicates that landslides having identical trigger mechanisms exhibit similar geometric properties.

Plain Language Summary There is a general shortage of high-quality spatiotemporal data sets on landslides. Even the existing inventories often lack critical information, such as the absence of details about the landslide-triggering mechanisms. These missing pieces of information render these databases useless for landslide modeling and analysis. Here, we propose a method that can estimate the likely triggering mechanism of landslides using the geometric properties of landslides' physical outlines provided in the existing inventories. Our method uses a machine-learning algorithm known as random forest, which is computationally simple to implement but, at the same time, highly robust and accurate. Our method's success indicates that landslides with similar triggers display similar geometric properties. A long-suspected fact that we anticipate landslide modelers will find useful.

1. Introduction

Landslides pose a constant threat to human life, individual property, and infrastructure in rugged terrains globally. Thus, algorithms that can learn patterns from past landslides and provide early-warning signals are highly sought-after (Osanaï et al., 2010). However, the performance of such algorithms critically depends on the quality of the existing landslide inventories. Some of these inventories are compiled by mapping landslides that cause damages, e.g., along roads (Pittore et al., 2018) and merging these mappings with existing data sets to create a knowledge base (Havenith et al., 2015). More contemporary mapping includes repeated satellite monitoring of landslide-prone regions to create complete landslide databases (Behling et al., 2014; Tanyas et al., 2017). Although event-based inventories include clear traces of the triggering mechanisms of landslides (von Specht et al., 2019), many other inventories, such as satellite-based, lack crucial information linking a given landslide to a specific triggering mechanism (Behling et al., 2014). The missing information about triggering mechanisms decreases the efficacy of these inventories in landslide hazard analyses, as this could introduce biases, for instance, inadvertently using earthquake-triggered landslides to assess landslide hazard for extreme rainfall (Ozturk et al., 2020). Hence, there is a need to identify triggers of landslides in existing databases to make them usable in hazard models. Furthermore, as the trigger mechanisms get engraved into the landslide polygons' geometry (e.g., the outline of debris field), identifying geometric features that best classify a particular landslide will provide physical insights into the trigger mechanisms.

Geometric features of landslide polygons are suspected to reflect their trigger (Varnes, 1996). Coseismic landslides tend to have a lower length-to-width ratio than the rainfall-induced landslides due to their extended accumulation zone (Taylor et al., 2018). Topographic site effects amplify seismic signals at higher altitudes

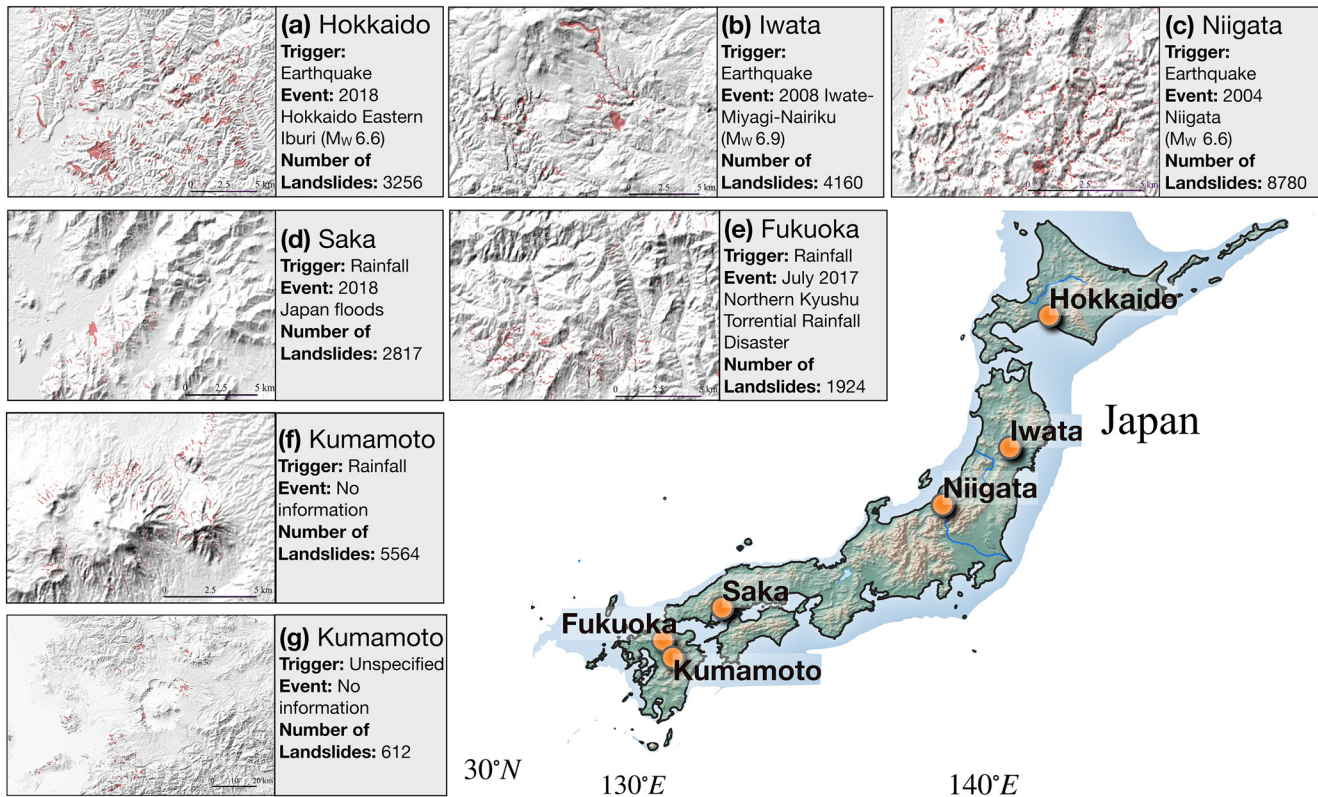


Figure 1. The map of Japan shows the geographical locations of the seven landslide inventories used in this work. (a)–(g) The Digital Elevation Model (DEM) of these inventories, and the adjoining panels list the region, trigger type, the origin of the event, and the number of landslides. The red color overlays on the DEM are a subset of landslide polygons of each inventory. Japanese Geospatial Information Authority (GSI) is the source of data in (a), (d), and (e); National Research Institute for Earth Science and Disaster Resilience (NIED) is the source of (f) and (g) and data in (c) and (b) is from Schmitt et al. (2018).

causing coseismic landslides to cluster around mountain ridges (Meunier et al., 2008; Rault et al., 2019). On the contrary, rainfall-induced landslides happen to connect to the local drainage network, which could be deceptive considering their extended accumulation (Marc et al., 2018; Ozturk et al., 2018). Accordingly, existing studies try to categorize landslide based on their triggers using (a) perimeter area indices (Pourghasemi et al., 2014); (b) similarity of a landslide planform to a circle or an ellipsoid (Samia et al., 2017; Taylor et al., 2018); and (c) scaling relationships between landslide dimensions and area (Milledge et al., 2014). However, it is challenging to apply these metrics for automatic classification of landslides due to the irregularities in landslide polygons and amalgamation of several types of landslides in the databases (e.g., Marc & Hovius, 2015).

Here, we present a machine-learning-based approach that uses the geometric features of landslide polygons to identify the underlying triggering mechanisms of landslides. To illustrate the effectiveness and accuracy of our approach, we apply it to several inventories spread over the Japanese archipelago with known triggers. We anticipate that our robust yet straightforward approach will be transferable to build a landslide knowledge base. For future practical deployment of the approach, we also demonstrate the applicability of our method on a landslide inventory without any triggering information. Apart from providing a useful technique for landslide-trigger classification, our study also highlights that the information of triggering mechanisms is embedded in the geometric features of landslides.

2. Data

In this work, we analyze seven landslide inventories that belong to six different regions of Japan (see Figure 1). We know the trigger mechanisms in six of the seven inventories, and we employ these inventories with known triggers to test the efficiency of our method. For the seventh inventory, the triggering

mechanism is unknown, and we use it to demonstrate the practical implementation of the method as this scenario represents the most probable real-world usage of our method. We used the coseismic landslides associated with following earthquakes: the 2018 Hokkaido Eastern Iburi (M_w 6.6); the 2008 Iwate-Miyagi Nairiku (M_w 6.9), and the 2004 Niigata (M_w 6.6). The Geospatial Information Authority of Japan (GSI) is the source of landslide inventory from the Hokkaido Eastern Iburi earthquake, while the source of the other two coseismic inventory is the repository created by Schmitt et al. (2017). GSI also provides rainfall-induced landslide inventories of the Fukuoka (July 2017) and Saka (July 2018) regions. We employ two more inventories from the Kumamoto region provided by the National Research Institute for Earth Science and Disaster Resilience (NIED) of Japan. In one of these inventories, the underlying trigger is documented as rainfall; however, the second inventory lacks any triggering information. We will refer to this second inventory as “unspecified.”

3. Method

The landslide planforms (polygons) are one of the primary information in landslide inventories, and geometric features of these polygons are a rich source for understanding physical mechanisms underlying a particular landslide (Kasai & Yamada, 2019; Milledge et al., 2014; Pourghasemi et al., 2014; Samia et al., 2017; Taylor et al., 2018). Therefore, we explored various geometric properties of landslide polygons and identified a subset to form the feature space for machine-learning-based automatized classification of landslides into two categories: earthquake-induced and rainfall-induced landslides.

We started by exploring a broad set of measures to quantify the geometric shapes of two-dimensional polygons extracted from the landslide planforms. In Table S1, we provide a complete list of these geometric measures. Using a combination of feature selection approaches (described in detail in Sections S2 and S3 of the Supporting Information), we were able to select seven features that lead to the best classification accuracy (Ambrose & McLachlan, 2002; Chandrashekar & Sahin, 2014; Friedman et al., 2009; Scott, 1992). These seven geometric features were area A , perimeter P , convex hull-based measure $C_h = \frac{A}{A_c}$ (A_c is the area of the convex hull fitted to the polygon and hereafter, we will refer C_h as convex hull measure), the ratio of area and perimeter $\frac{A}{P}$, width of the minimum area bounding box W , minor axis s_m , and eccentricity of the fitted ellipse e (Figure 2). As each feature has a different range of measurement values, we standardized the data by calculating z-scores of each feature.

For classifying landslides, we employ *random forest*, an ensemble-based learning method, which is known to be a highly robust and accurate for tasks such as classification and regression (Barnett et al., 2019; Biau, 2012; Biau & Scornet, 2016; Breiman, 2001; Kursa, 2014; Liaw et al., 2002; Rodriguez-Galiano et al., 2012; Roy & Larocque, 2012). Unlike the standard tree-based methods, where all the attributes are used for the best split of a node, in a random forest, only a random subset are used, and each decision tree is constructed using different bootstrap samples of the data (Breiman, 2001; Liaw et al., 2002). For the testing sample, each tree predicts the class independently, and the class having the majority vote is the class prediction of the sample. Below we briefly describe the steps in implementing random forest for our binary classification problem (Breiman, 2001; Friedman et al., 2009; Liaw et al., 2002; Zhang & Ma, 2012).

We build a decision tree T_b for each bootstrap sample b of the training set; let $\mathbf{X}_i = [x_{ij}]_{j=1}^p$ represent a p -dimensional feature vector for data point i . Next, we recursively repeat the following two steps for each node of the tree: (a) select m variables among p and (b) split the node into daughter nodes that best separate the classes. As we carry out binary classification, i.e., classifying landslides into earthquake or rainfall triggered, the parent node q is split into two daughter nodes left l and right r . For this splitting, we employ the Gini index method, where Gini index for l and r are $G_l = 1 - p_{l1}^2 - p_{l2}^2$ and $G_r = 1 - p_{r1}^2 - p_{r2}^2$ respectively. Note p_{lj} is the probability of data points of class j in node l . For each split of a node, the Gini index of the subnodes should be less than the parent node, and can be achieved for a split s_q of node q if we maximize the decrease in the quantity $\Delta\theta(s_q) = G_q - \rho_{rq}G_r - \rho_{lq}G_l$; where ρ_{rq} (ρ_{lq}) are the ratio of the number of data points in daughter nodes r (l) to the total number of points in the parent node q (Kuhn & Johnson, 2013; Zhang &

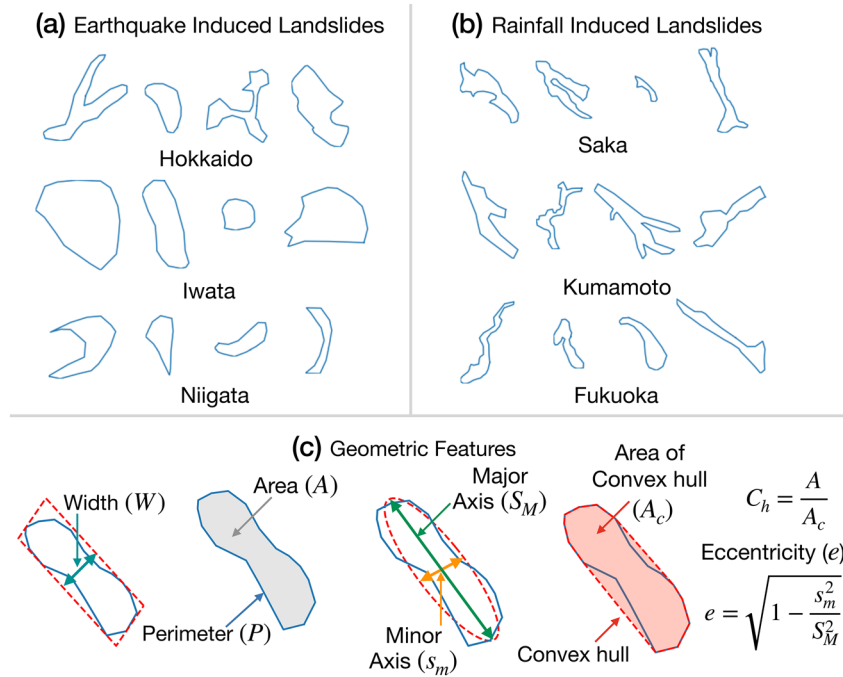


Figure 2. Sample landslide planforms from (a) Earthquake-triggered, (b) rainfall-triggered inventories, (c) geometric features (left to right): width (W) of minimum area bounding box fitted to the landslide polygon, area (A) and perimeter (P) of the landslide polygon, minor (s_m) and major axis (S_M) lengths of an ellipse fitted to the polygon and convex hull-based measure $C_h = \frac{A}{A_c}$ where A_c is the area of convex hull fitted to the polygon.

Ma, 2012). The two recursive steps (a) and (b) continue until a predefined criterion is satisfied. For example, we are left with one data point each in the daughter nodes, and no further splitting is possible.

Now if $C_b(\mathbf{Y}_i)$ is the class prediction of the b th random forest tree for \mathbf{Y}_i point in the training set, then the class prediction of \mathbf{Y}_i is given by the majority vote in the set $\{C_b(\mathbf{Y}_i)\}_1^B$; where B is the total number of bootstrap samples. In this setting, the importance of the k th feature in predicting the training set is given by

$I(k) = \frac{1}{B} \sum_{T_b} \sum_{q \in T_b: D(s_q)=k} \rho_q \theta(s_q)$; where ρ_q is ratio of number of points in the q th node to total number of points in training data and $D(s_q) = k$ implies that the feature involved in the split s_q is k th feature. The $I(k)$ in above formulation measures the average of weighted impurity decrease $\rho_q \theta(s_q)$ over all the splits in the ensemble of random forest trees. As $\sum_k I(k) = 1$, we express the feature importance in percentage as $\frac{I(k)}{\sum_k I(k)} \times 100$.

4. Results

For the first numerical experiment, we constructed testing and training set by combining the six inventories with known triggers to validate our approach (Figure 3). In this experiment, the total number of samples in the combined data set were $n_{\text{total}} = 26,501$, with $n_{\text{rainfall}} = 10,305$ and $n_{\text{earthquake}} = 16,196$. We randomly resample the data to take an equal number of the earthquake and rainfall samples ($n_{\text{rainfall}} = n_{\text{earthquake}} = 10,305$) to avoid any class imbalance. Thus, we apply the algorithm on $n = 2 \times 10,305 = 20,610$ samples. We also employed 1,000 runs of 10-fold cross-validation to swap training and testing sets to avoid the likelihood of results influenced by overfitting and smaller standard deviation in each case indicate more stable class performances. Note that splitting 20,610 landslide samples into 10-folds with an equal number of landslide samples in each fold leads to uneven numbers of earthquake and rainfall samples ($20,610/10 = 2,061$; an odd number). To tackle this issue, the 5-folds out of 10-folds have 2,062 samples (both earthquake and rainfall have 1,031 samples), and the remaining 5-folds have 2,060 landslide samples (each earthquake and rainfall have 1,030 samples). Thus, each iteration of 10-fold cross-validation has either $n_{\text{train}} = 18,548$, $n_{\text{test}} = 2,062$

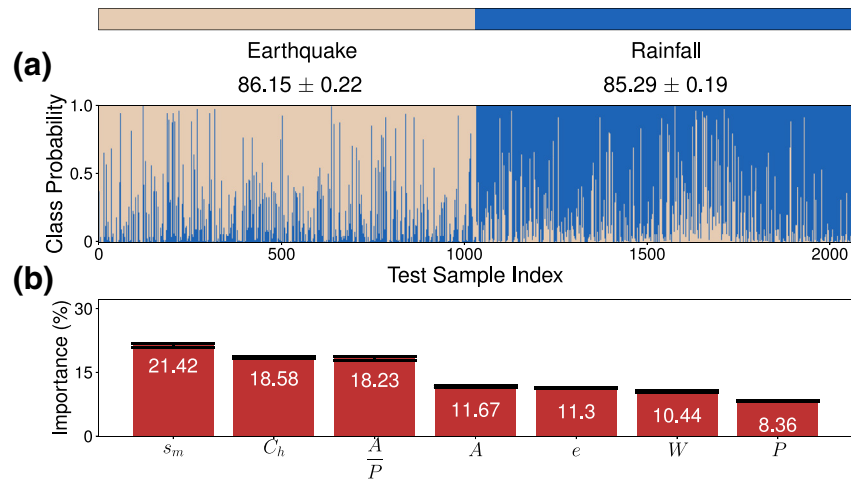


Figure 3. (a) Earthquake- and rainfall-triggered class accuracies using random forest classifier on geometric features of landslide polygons. The underlying data are a concoction of all the six inventories with available ground truth, where the training set has 18,548 and the test set has 2,062 samples. Within the test set, 1,031 samples are earthquake driven and 1,031 are rainfall driven. Using 1,000 runs of 10-fold cross-validation, we identified $86.15 \pm 0.22\%$ earthquake-triggered landslides and $85.29 \pm 0.19\%$ rainfall-triggered landslides correctly. The plot in (a) is the output from one of the runs of the random forest classifier. The class probability represents the proportion of votes for a class in the ensemble of trees. The x axis is the index of the sample in the test set. (b) The importance of geometric features used in (a). The percentage corresponding to each feature represents the mean decrease in the tree leaf impurity over the full random forest such that the total percentage sums to 100.

(1,031 each rainfall-induced and coseismic) or $n_{\text{train}} = 18,550$, $n_{\text{test}} = 2,060$ (1,030 each rainfall-induced and coseismic). Here, n_{train} and n_{test} are the number of training and testing samples.

In the experiment with combined data, using $\mathbf{X} = [A, P, C_h, \frac{A}{P}, W, s_m, e]$ as the feature vector, we achieved the mean classification accuracy of $85.73 \pm 0.16\%$, where $86.15 \pm 0.22\%$ coseismic and $85.29 \pm 0.19\%$ rainfall-induced events were classified correctly (see Figure 3a). Among all the geometric features, minor axis length s_m has the highest feature importance of 21.42%, followed by the convex hull measure C_h (18.58%) (see Figure 3b). In Section S5, we provide further detailed analysis of these results, including additional metrics evaluating the algorithm's performance. Also, see Figures S6 and S7 and Table S2 in the SI. During the feature selection process, we removed highly correlated features (see Section S2). However, there is still a possibility of bias in feature importance due to multicollinearity in the geometric predictors discussed above. In Section S6, we present some additional analysis of bias in feature importance, and it indicates that the presence of collinearities (if any) does not impact the relative order of features in terms of their importance.

In the second experiment, we applied to our approach to individual inventories; we train the algorithm on five of the six inventories and predict the trigger of landslides in the sixth inventory. Hence, the training data has no information on the test data inventory—a situation similar to the one we anticipate this method will be used in the real world. We use the same set of geometric features as in Figure 3, and to avoid class imbalance, we keep the number of rainfall and earthquake samples the same by resampling the data. The method achieved over 85% classification accuracy for the Saka ($n_{\text{train}} = 14,976$, $n_{\text{test}} = 2,817$), and Niigata ($n_{\text{train}} = 14,832$, $n_{\text{test}} = 8,780$) region. 83.63% accuracy for the Kumamoto region ($n_{\text{train}} = 9,482$, $n_{\text{test}} = 5,564$), 75.59% for the Iwata region ($n_{\text{train}} = 20,610$, $n_{\text{test}} = 4,160$), and 66.62% for the Hokkaido ($n_{\text{train}} = 20,610$, $n_{\text{test}} = 3,256$), and 69.40% for Fukuoka ($n_{\text{train}} = 16,762$, $n_{\text{test}} = 1,924$) regions (see Figure 4). In this experiment, the model performed better classifying the rainfall-triggered inventories than the earthquake-triggered inventories. Performance drops as low as 67% (approximately) in the case of Hokkaido. We repeat the run 1,000 times with a random selection of an equal number of earthquake- and rainfall-triggered landslide samples for training. The results are stable to changes in training samples, as the standard error is small ($<0.65\%$). For further analysis of this experiment (see Section S5).

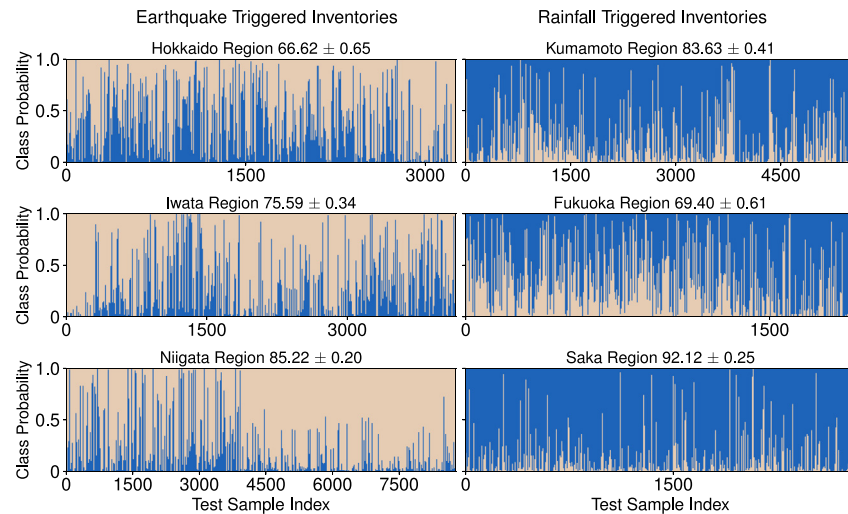


Figure 4. Predicting one of the six inventories while the random forest classifier was trained on the rest (five inventories), i.e., the classifier has no information on the data being predicted. The geometric features used are the same as in Figure 3. The deviation quoted in the accuracy percentage was calculated by running the classifier algorithm 1,000 times (note cross-validation is not possible to implement in this configuration of testing and training data). For further analysis of these results see Figures S6 and S8.

In the third experiment, we applied our method to the unspecified trigger inventory from the Kumamoto region (inventory (g) in Figure 1). Out of 612 landslides in the test case, 604 were classified as earthquake-induced and 8 as rainfall-induced. We present a detailed analysis and discussion on this inventory and our results in Section S4.

5. Discussion

We showed that using the random forest algorithm in conjunction with geometric features of landslide planforms is a robust technique for classifying landslides trigger mechanisms, and it can achieve excellent classification rates in a variety of settings. For example, while using the aggregate data from different regions of Japan with varied geology and topography, this scheme achieved an accuracy of 85%. Whereas in the individual analysis of different regions, the classification accuracy went up as high as 92%. Moreover, we identified seven geometric features of landslide polygons that appear to be the best predictors of the underlying trigger mechanisms. Our results indicate that although there is diversity in the physical mechanisms producing landslide events, there is also a universality in these mechanisms that get embedded in the geometry of landslide planforms.

A possible reason for the emergence of this universality could be that both rainfall-induced and coseismic landslides in our databases consist of a mostly shallow landslide, e.g., debris flows (Kasai & Yamada, 2019; von Specht et al., 2019; Watakabe & Matsushi, 2019). Although coseismic rockslides tend to cut the mountain ridge and slide through concave slopes likely on a lithological failure plain with relatively short accumulation zones (Havenith, 2002), the number of such failures is limited in our databases. Shallow landslides, tend to flow through convex slopes, following the local morphology with extended accumulation that could increase the source area 10-fold (Hungur et al., 2014a; Uchida et al., 2013; Wang et al., 2015). Additionally, debris flows are prone to form concave planforms due to flow divergence along valley bottoms. Hence, more extended accumulation areas will result in elevated convex hull and area-perimeter ratio (Figure 2c) in our analyses. Although these shallow slope failures could be classified into the same landslide type (Hungur et al., 2014b; Varnes, 1996), our results indicate that the coseismic and rainfall-induced landslides form divergent planforms (Taylor et al., 2018). Hence convex hull based measure and area-perimeter ratios, together with the minor axis, contribute the most to the classification (see Figure 3), highlighting that rainfall and earthquake driven landslides have distinct geometry.

To further demonstrate the practical aspects of our approach, we also applied it to an unspecified trigger inventory from the Kumamoto region (inventory (g) in Figure 1). The landslides in this inventory are mapped along the rims of the Aso Caldera; the active volcano Mount Aso shakes the surrounding area, frequently triggering landslides within its vicinity (Saito et al., 2018). Therefore, the majority of the unspecified landslides may be seismically triggered. Accordingly, our algorithm classified 604 as earthquake-induced and 8 as rainfall-induced out of 612 landslides. Given that most of the landslides are classified as coseismic, we are confident in flagging this inventory as seismically triggered due to the volcanic activity.

The size of the databases we used is relatively small, i.e., it is challenging to apply more sophisticated image classification algorithms, such as the convolutional neural networks (CNN) to the problem at hand, as these algorithms require extensive training sets. In contrast, as illustrated above, the random forest achieved good enough accuracy to apply it to real-world settings even when training sets are relatively small. The random forest is comparatively computationally inexpensive, its complexity in our case was $\mathcal{O}(n \log n)$, where n is the number of samples. Furthermore, the random forest is highly portable, as nowadays many machine-learning packages include an elaborate implementation of this algorithm (Pedregosa et al., 2011). Given these advantages, we anticipate that the landslide modeling community will find our scheme useful.

6. Conclusion

Historic landslide inventories rarely include the triggering mechanisms of the observed landslides, a critical piece of information for landslide hazard and susceptibility models. We developed a method to fill this missing information by classifying existing landslides in digital databases of landslides. Our approach uses geometric characteristics of landslide polygons as features for the random forest classifier. The resulting algorithm is highly portable and accurate and can be applied to any region of interest with adequate training data from areas with similar tectonic and climatic features. We also identified seven geometric features of landslide planforms that appear to capture some universal patterns in the landslide-trigger mechanisms.

Furthermore, we applied our scheme to several different tests and training set configurations of the available data from Japan, and our results indicate this method is versatile, robust, and can classify landslide triggers with high accuracy. Envisioning that this method will be applied to individual inventories with unknown triggers in practice, we prepared such a test application and demonstrated that the model classifies it as an earthquake driven inventory, a highly plausible classification based on the geographic location of the inventory.

Although our study is limited to the Japanese archipelago consisting of landslide samples triggered by either earthquake or rainfall, our method is computationally simple, portable, and robust enough to anticipate that the landslide research community will find it useful for classifying landslides in inventories from various other geographical regions and triggering mechanisms.

Acknowledgments

The authors thank the Geospatial Information Authority (GSI) and the National Research Institute for Earth Science and Disaster Resilience (NIED) of Japan for providing the landslide data freely and the ALOS for 30 m DEM (<https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm>). N. Malik and K. Rana were supported by RIT's College of Science DRIG Grant. U. Ozturk was supported by the German Academic Exchange Service (DAAD) within the Co-PREPARE project of the German-Indian Partnerships Support Program (DIP Project No. 57553291). This research work has been partly supported financially by the Federal Ministry of Education and Research of Germany (BMBF) within the project CLIENT II—CaTeNA (FKZ 03G0878A). In memory of Prof. Hiroshi Fukuoka (Sassa et al., 2018).

References

- Ambrose, C., & McLachlan, G. J. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences*, 99(10), 6562–6566. <https://doi.org/10.1073/pnas.102102699>
- Barnett, I., Malik, N., Kuijjer, M. L., Mucha, P. J., & Onnela, J.-P. (2019). Endnote: Feature-based classification of networks. *Network Science*, 7(3), 438–444.
- Behling, R., Roessner, S., Segl, K., Kleinschmit, B., & Kaufmann, H. (2014). Robust automated image co-registration of optical multi-sensor time series data: Database generation for multi-temporal landslide detection. *Remote Sensing*, 3, 2572–2600.
- Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research*, 13(1), 1063–1095.
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, 40(1), 16–28.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). New York, NY: Springer-Verlag. <https://doi.org/10.1007/978-0-387-84858-7>
- Havenith, H.-B. (2002). Site effect analysis around the seismically induced Ananevo Rockslide, Kyrgyzstan. *Bulletin of the Seismological Society of America*, 92(8), 3190.
- Havenith, H. B., Torgoev, A., Braun, R. S. A., Torgoev, I., & Ischuk, A. (2015). Tien Shan geohazards database: Landslide susceptibility analysis. *Geomorphology*, 249(1), 32–43.
- Hungr, O., Leroueil, S., & Picarelli, L. (2014). The Varnes classification of landslide types, an update. *Landslides*, 11(2), 167–194. <https://doi.org/10.1007/s10346-013-0436-y>

- Kasai, M., & Yamada, T. (2019). Topographic effects on frequency-size distribution of landslides triggered by the Hokkaido Eastern Iwate earthquake in 2018. *Earth, Planets and Space*, *71*(1), 89. <https://doi.org/10.1186/s40623-019-1069-8>
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (1st ed., Vol. 26). New York, NY: Springer-Verlag. <https://doi.org/10.1007/978-1-4614-6849-3>
- Kursa, M. B. (2014). Robustness of random forest-based gene selection methods. *BMC Bioinformatics*, *15*(1), 8.
- Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*(3), 18–22.
- Marc, O., & Hovius, N. (2015). Amalgamation in landslide maps: Effects and automatic detection. *Natural Hazards and Earth System Sciences*, *15*, 723–733.
- Marc, O., Stumpf, A., Malet, J. P., Gosset, M., Uchida, T., & Chiang, S. (2018). Initial insights from a global database of rainfall-induced landslide inventories: The weak influence of slope and strong influence of total storm rainfall. *Earth Surface Dynamics*, *4*(6), 903–922.
- Meunier, P., Hovius, N., & Haines, J. (2008). Topographic site effects and the location of earthquake induced landslides. *Earth and Planetary Science Letters*, *275*(3–4), 221–232. <https://doi.org/10.1016/j.epsl.2008.07.020>
- Milledge, D., Bellugi, D., McKean, J., Densmore, A. L., & Dietrich, W. E. (2014). A multidimensional stability model for predicting shallow landslide size and shape across landscapes. *Journal of Geophysical Research: Earth Surface*, *119*, 2481–2504. <https://doi.org/10.1002/2014JF003135>
- Osanai, N., Shimizu, T., Kuramoto, K., Kojima, S., & Noro, T. (2010). Japanese early-warning for debris flows and slope failures using rainfall indices with radial basis function network. *Landslides*, *7*(3), 325–338.
- Ozturk, U., Pittore, M., Behling, R., Roessner, S., Andreani, L., & Korup, O. (2020). How robust are landslide susceptibility estimates? *Landslides*. <https://doi.org/10.1007/s10346-020-01485-5>
- Ozturk, U., Wendi, D., Crisologo, I., Riemer, A., Agarwal, A., Vogel, K., et al. (2018). Rare flash floods and debris flows in Southern Germany. *Science of the Total Environment*, *626*, 941–952.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
- Pourghasemi, H. R., Moradi, H. R., Aghda, S. F., Sezer, E. A., Jirandeh, A., & Pradhan, B. (2014). Assessment of fractal dimension and geometrical characteristics of the landslides identified in north of Tehran, Iran. *Environmental Earth Sciences*, *71*(1), 3617–3626. <https://doi.org/10.1007/s12665-013-2753-9>
- Rault, C., Robert, A., Marc, O., Hovius, N., & Meunier, P. (2019). Seismic and geologic controls on spatial clustering of landslides in three large earthquakes. *Earth Surface Dynamics*, *3*(7), 829–839.
- Rodríguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sánchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 93–104.
- Roy, M.-H., & Larocque, D. (2012). Robustness of random forests for regression. *Journal of Nonparametric Statistics*, *24*(4), 993–1006.
- Saito, H., Uchiyama, S., Hayakawa, Y. S., & Obanawa, H. (2018). Landslides triggered by an earthquake and heavy rainfalls at Aso Volcano, Japan, detected by UAS and SFM-MVS photogrammetry. *Progress in Earth and Planetary Science*, *5*(1), 15. <https://doi.org/10.1186/s40645-018-0169-6>
- Samia, J., Temme, A., Bregt, A., Wallinga, J., Guzzetti, F., Ardizzone, F., & Rossi, M. (2017). Do landslides follow landslides? Insights in path dependency from a multi-temporal landslide inventory. *Landslides*, *14*(2), 547–558.
- Sassa, K., Canuti, P., Takara, K., & Watanabe, N. (2018). In memoriam—Hiroshi Fukuoka (1961–2018). *Landslides*, *15*(9), 1905–1906. <https://doi.org/10.1007/s10346-018-1048-3>
- Schmitt, R., Tanyas, H., Jessee, A., Zhu, J., Biegel, K., Allstadt, K., et al. (2018). An open repository of earthquake-triggered ground-failure inventories (17 p.). U.S. Geological Survey Data Series 1064. <https://doi.org/10.3133/ds1064>
- Schmitt, R. G., Tanyas, H., Nowicki Jessee, M. A., Zhu, J., Biegel, K. M., Allstadt, K. E., et al. (2017). *An open repository of earthquake-triggered ground-failure inventories (ver 2.0, December 2018)*. U.S. Geological Survey data release collection. <https://doi.org/10.5066/F7H70DB4>
- Scott, D. W. (1992). *Multivariate density estimation: Theory, practice, and visualization*. Hoboken, NJ: John Wiley & Sons. <https://doi.org/10.1002/9780470316849>
- Tanyas, H., van Westen, C. J., Allstadt, K. E., Anna Nowicki Jessee, M., Görüm, T., Jibson, R. W., et al. (2017). Presentation and analysis of a worldwide database of earthquake-induced landslide inventories. *Journal of Geophysical Research: Earth Surface*, *122*, 1991–2015. <https://doi.org/10.1002/2017JF004236>
- Taylor, F., Malamud, B., Witt, A., & Guzzetti, F. (2018). Landslide shape, ellipticity and length-to-width ratios. *Earth Surface Processes and Landforms*, *43*(15), 3164–3189.
- Uchida, T., Nishiguguchi, Y., Nakatani, K., Satofuka, Y., Yamakoshi, T., Okamoto, A., & Mizuyama, T. (2013). New numerical simulation procedure for large-scale debris flows (Kanako-LS). *International Journal of Erosion Control Engineering*, *6*(2), 58–67. <https://doi.org/10.13101/ijece.6.58>
- Varnes, D. C. D. (1996). *Landslides: Investigation and mitigation chapter 3-landslide types and processes* (pp. 247). Transportation Research Board Special Report. <https://trid.trb.org/view/462501>
- von Specht, S., Ozturk, U., Veh, G., Cotton, F., & Korup, O. (2019). Effects of finite source rupture on landslide triggering: The 2016 Mw 7.1 Kumamoto earthquake. *Solid Earth*, *10*(2), 463–486.
- Wang, Z.-Y., Lee, J. H., & Melching, C. S. (2015). Debris flows and landslides. In *River dynamics and integrated river management* (pp. 193–264). New York, NY: Springer.
- Watakabe, T., & Matsushi, Y. (2019). Lithological controls on hydrological processes that trigger shallow landslides: Observations from granite and hornfels hillslopes in Hiroshima, Japan. *Catena*, *180*, 55–68. <https://doi.org/10.1016/j.catena.2019.04.010>
- Zhang, C., & Ma, Y. (2012). *Ensemble machine learning: Methods and applications* (Vol. 1). New York, NY: Springer-Verlag. <https://doi.org/10.1007/978-1-4419-9326-7>