Multitemporal landslide inventory and susceptibility map for the Arun River Basin, Nepal.

Pukar Amatya[1,2], Robert Emberson[1,2] and Dalia Kirschbaum[3]

[1]GESTAR II, University of Maryland Baltimore County, Baltimore, MD, USA
[2]Hydrological Sciences laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[3]Earth Science Division, NASA Goddard Space Flight Center, Greenbelt, MD, USA

* Correspondence: pukar.m.amatya@nasa.gov

Running Title: Landslide dataset for the Arun River Basin

Funding information: Funding for this project was provided under NASA's High Mountain Asia Science Team (NNH19ZDA001N-HMA) and the World Bank.

Data availability statement: The landslide inventory is available via NASA National Snow and Ice Data Center Distributed Active Archive Center (https://doi.org/10.5067/E4M8GRXKLCRC). Landslide susceptibility map is available in Figshare (https://doi.org/10.6084/m9.figshare.20986888).

Dataset Details:
Landslide Inventory
Identifier: https://doi.org/10.5067/E4M8GRXKLCRC
Creator: Pukar Amatya
Dataset correspondence: pukar.m.amatya@nasa.gov
Title: High Mountain Asia Multitemporal Landslide Inventory for the Pumqu/Arun River Basin, Version 1 (HMA2_MTLI)
Publisher: NASA National Snow and Ice Data Center Distributed Active Archive Center
Publication year: 2022
Resource type: Landslide inventory
Version: 1

Landslide Susceptibility
Identifier: https://doi.org/10.6084/m9.figshare.20986888
Creator: Robert Emberson
Dataset correspondence: robert.a.emberson@nasa.gov
Title: Landslide susceptibility map for the Arun River Basin
Publisher: Figshare
Publication year: 2022
Resource type: Landslide susceptibility
Version: 1

Abstract
The transboundary Arun River Basin (ARB) spreads across Nepal and Tibet. Nearly 95% of the basin lies in Tibet through which the Pumqu River flows, forming the Arun River once it enters Nepal. The ARB has five large hydropower projects undergoing construction or planned for the future. Rainfall and earthquake-induced landslides, landslide dammed lakes and landslide-induced glacial lake outburst floods pose major risks to smooth operation of these projects. To safeguard upcoming hydropower projects, areas susceptible to landslides in the ARB must be identified. We used high-resolution satellite imagery and open-source tools to generate a multitemporal landslide inventory for the basin. The rigorously quality-controlled inventory represents a yearly record of landslides from 2011 to 2020. A data-driven approach was used to map areas susceptible to landslides within the ARB. The multitemporal landslide inventory combined with other readily available Earth observation-based variables were used to create a landslide susceptibility map. The susceptibility analysis provides a valuable initial estimate of where landslides are likely to initiate. These landslide products could form the basis of more comprehensive local studies to inform hydropower project development.

Keywords: Arun River Basin, Landslides, Inventory, Susceptibility, Hydropower

1. Introduction

Nepal's domestic electricity generation is dependent on hydropower. With a growing economy, the power demand is forecasted to more than double by 2025 compared to 2018 (International Hydropower Association, 2020). As such, new hydropower projects are a key national need. The Arun River Basin (ARB) in eastern Nepal (Figure 1) has a constant source of water from the Pumqu (in Tibet)/Arun (in Nepal) rivers and has a cascade of five large hydropower projects that are under construction or planned (Table 1). Rainfall and earthquake-induced landslides, landslide dammed lakes and landslide-induced glacial lake outburst floods pose major risks to smooth operation of these projects. On April 20, 2017, flooding and debris from a rock avalanche that impacted a small moraine-dammed lake resulted in formation of a temporary lake that was 2 to 3 km long approximately 2 km upstream of the Upper Arun hydropower site (Byers *et al.*, 2019). On June 15, 2018, landslides induced by heavy rainfall blocked an entrance to a tunnel, trapping workers at Arun 3 hydropower plant (Reynolds, 2018). On June 21, 2020, a landslide dammed the Arun River creating flood hazards downstream (Shakya, 2020). To safeguard upcoming hydropower projects, it is necessary to characterize the likely source areas for landslide materials (susceptibility).

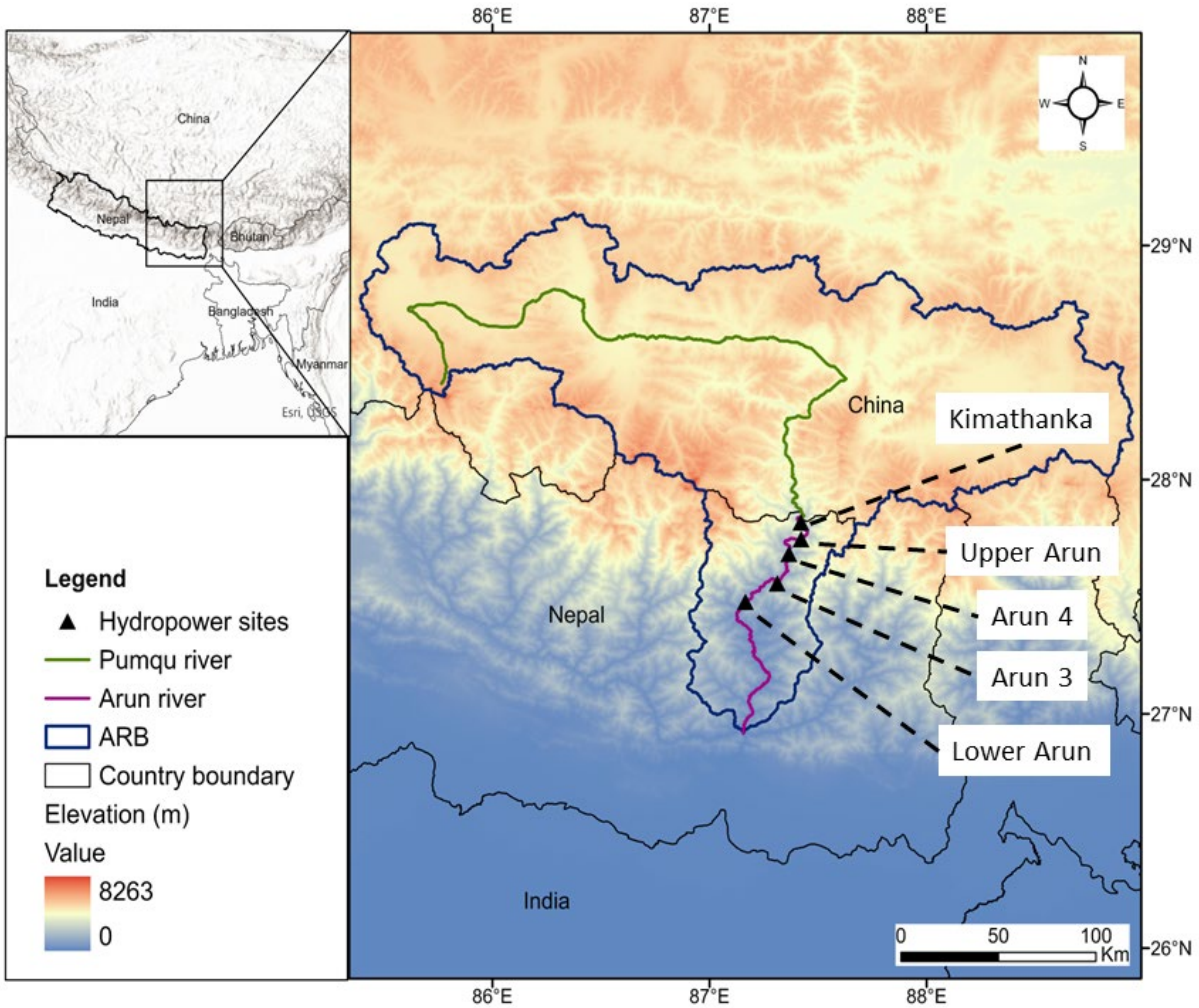

Figure 1. Location of the transboundary ARB and hydropower projects along the Arun River.

Table 1. Hydropower Projects in the ARB and their status

Hydropower Project	Capacity (MW)	Status
Kimathanka	500	Under study
Upper Arun	1,060	Under study
Arun 4	303	Under study
Arun 3	900	Under construction
Lower Arun	400	Under study

In this effort, we used open-source methods to automatically generate a multitemporal landslide inventory, and susceptibility map for the ARB. The first and most important data needed for conducting landslide susceptibility analysis are comprehensive landslide inventories of the basin. Different landslide inventories are available for this region, which were produced as part of a much larger mapping effort. Landslide inventories for year 1992 (ICIMOD, 2017a) and 2010 (ICIMOD, 2017b) for the Koshi River Basin, which encompasses the ARB are available from ICIMOD's regional database system. These inventories were created by manually mapping landslides using topographic maps, imagery from the Landsat, Gaofen-1 and Gaofen-2 satellites, and Google Earth (Zhang et al. 2016, 2019). Recently, a multitemporal manually mapped inventory based on Landsat imagery for eastern Nepal was released by Roberts et al. (2021) and Jones et al. (2021). However due to the 30 m resolution of Landsat, small landslides were missing from these inventories. In order to generate an inventory which includes smaller landslides and at increased temporal scale, we utilized high-resolution imagery from PlanetScope and RapidEye (Planet Team, 2017) and the Semi-Automatic Landslide Detection (SALaD) system (Amatya *et al.*, 2021) to generate a multitemporal (2011 – 2020) landslide inventory for the ARB.

Landslide susceptibility maps provide an estimate of where landslides might occur, based on knowledge of landslides that have occurred in the past (Guzzetti *et al.*, 2005). Landslide inventory, in combination with readily available Earth observation (EO)-based variables commonly used to explain landslide position (e.g. slope, surface roughness, elevation, land cover) and other global datasets such as lithology, can be used to model regional perspective of landslide susceptibility. Statistical and machine learning methods have been gaining popularity in the field of landslide susceptibility modeling (Reichenbach *et al.*, 2018; Merghadi *et al.*, 2020). Goetz et al. (2015) showed random forests (Breiman, 2001) provided marginally better performance in terms of predictive skill in comparison with other susceptibility models. The susceptibility map was generated using the random forests-based approach described by Emberson et al. (2021).

These datasets were developed to support the Risk Assessment of Natural Disasters in Upper Arun Hydropower Project. Beyond this project, these basin wide datasets will be valuable for advancing landslide studies in the basin and upcoming hydropower projects.

2. Methodology

2.1. Landslide Mapping

2.1.1. Data

Imagery from RapidEye and PlanetScope satellites were used to generate the multitemporal landslide inventory. The imagery was made available through NASA's Commercial Smallsat Data Acquisition Program (https://earthdata.nasa.gov/esds/csdap). RapidEye was a constellation of 5 satellites launched in 2008 and retired in 2020. PlanetScope is a constellation of > 180 CubeSats that started imaging the earth's surface in 2016. We used the RapidEye Ortho Tile and the PlanetScope Ortho Tile to produce the multitemporal landslide inventory.

We selected imagery acquired in December of each year from 2010 to 2020 to generate yearly inventories. When imagery for December was not available, early January imagery from subsequent year was used as substitute. Imagery with less than 20 % cloud cover was used. RapidEye imagery was used for production of inventories from 2010-2018. Inventories for 2019 and 2020 were produced using imagery from PlanetScope due to the lack of imagery from RapidEye. The dominant rainfall regime in the Arun valley is the Indian Summer Monsoon, and so while small date differences in imagery may exist year-to-year the vast majority of landsliding occurs in the summer interval that is clearly bracketed by imagery for each year.

In addition to the optical data, we also utilized the 30 m NASADEM (NASA JPL, 2020), a reprocessed Shuttle Radar Topography Mission (SRTM) dataset with improved height accuracy and void filling.

2.1.2. Semi-Automatic Landslide Detection (SALaD) system

The open-source Python-based Semi-Automatic Landslide Detection (SALaD) system (Amatya *et al.*, 2021) combines Object-based Image Analysis (OBIA) and machine learning to map landslides (Figure 2).This system was deployed to process 234 RapidEye and 58 PlanetScope tiles in the Advanced Data Analytics Platform at NASA Goddard Space Flight Center (GSFC)'s NASA Center for Climate Simulation (NCCS) (http://www.nccs.nasa.gov/services/adapt).

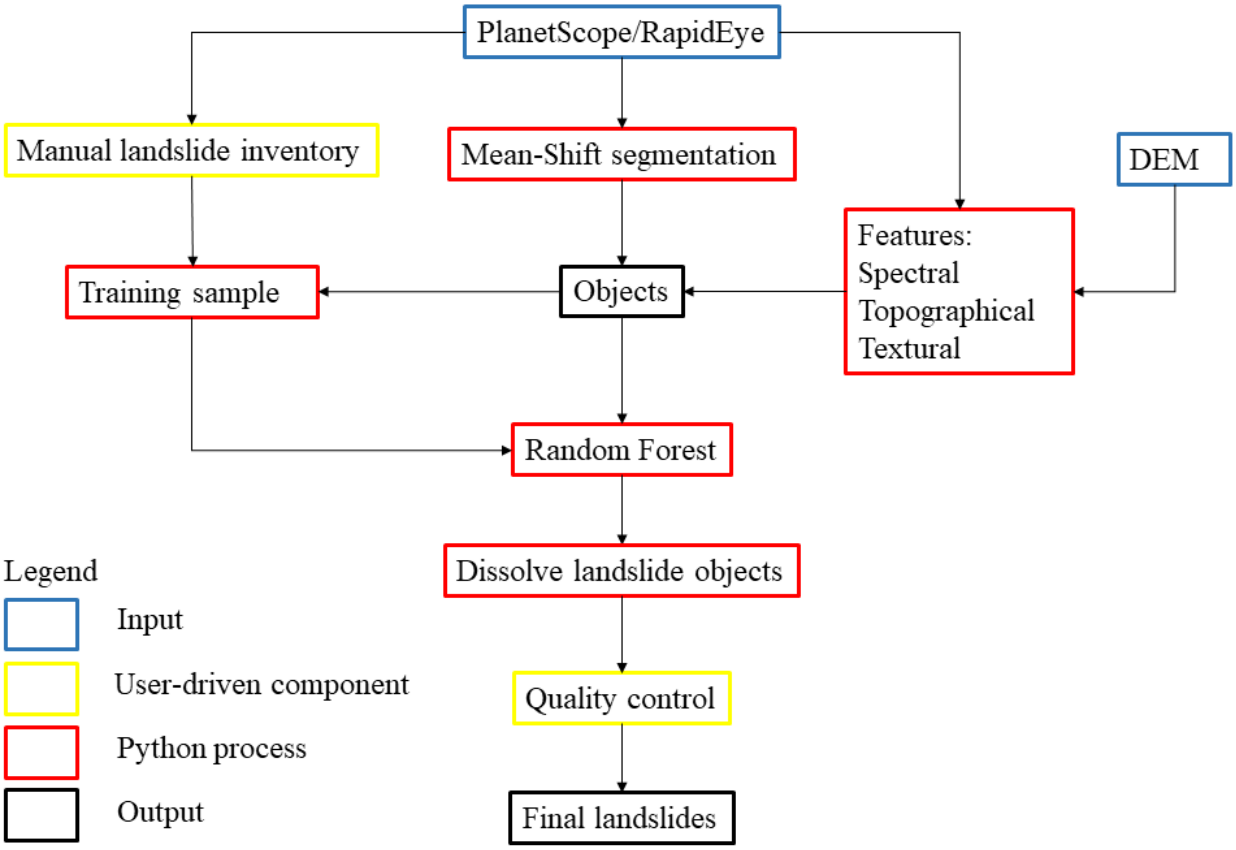

Figure 2. Flowchart showing each step of the OBIA-based SALaD system.

Segmentation is the first and most important step in OBIA. In segmentation, homogenous pixels are grouped to form objects using image segmentation algorithms. SALaD uses Large-Scale Mean-Shift segmentation (Michel *et al.*, 2014). It requires three parameters: the spatial radius (h_s) is the spatial distance between classes; the range radius (h_r) is the spectral difference between classes; and minimum object size is the minimum acceptable spatial size of the class. The SALaD system was developed and tested in the Pasang Lhamu highway area of Nepal using Rapideye imagery. Hence, we used the same segmentation parameters determined in Amatya et al. (2021). The Mean-Shift segmentation was conducted with h_s as 10 and h_r as 6 minimum object size as 10. For delineating landslides, five features: NDVI, brightness, Grey-level Co-occurrence Matrix (GLCM) mean, GLCM homogeneity and slope were used. The mean of each feature was calculated for all objects and used for landslide detection.

A random forest model (Breiman, 2001) with 500 trees was used for classifying objects to landslide or non-landslide class. The model was trained in the training area (red polygon in Figure 3). For RapidEye imagery, training data was created using a RapidEye image of December 2015. Manually mapped landslide polygons in the training area were used to create the training dataset. The training area contained 84 landslides. The segmented objects that had an overlap of 75% or more with the manual landslides were chosen as landslide objects, and all other objects within the

training area were set as non-landslide objects. The training data consisted of 809 landslide and 7331 non-landslide objects.

The same training data and segmentation parameters were used to map landslides for all RapidEye tiles used to generate landslide inventories for each year (2010-2018). For the years 2019 and 2020, we created a new training data using an image of 2019 from PlanetScope imagery for the same training area and used it to generate inventories for these two years.

Classified landslide objects were dissolved to obtain final landslide areas. The mapped landslides are shallow landslides as automated methods using optical imagery are not capable of detecting deep-seated landslides (Scaioni *et al.*, 2014). To improve our inventory, we manually removed some obvious false positives such as roads, riverbanks, houses, and barren areas, corrected erroneous landslide areas and resolved amalgamation. Since we did not use a change detection approach, we detected all landslides present in each year's tile. Landslides in a particular year that intersected with landslides in previous years were removed as a last step to differentiate landslide inventories of each year.

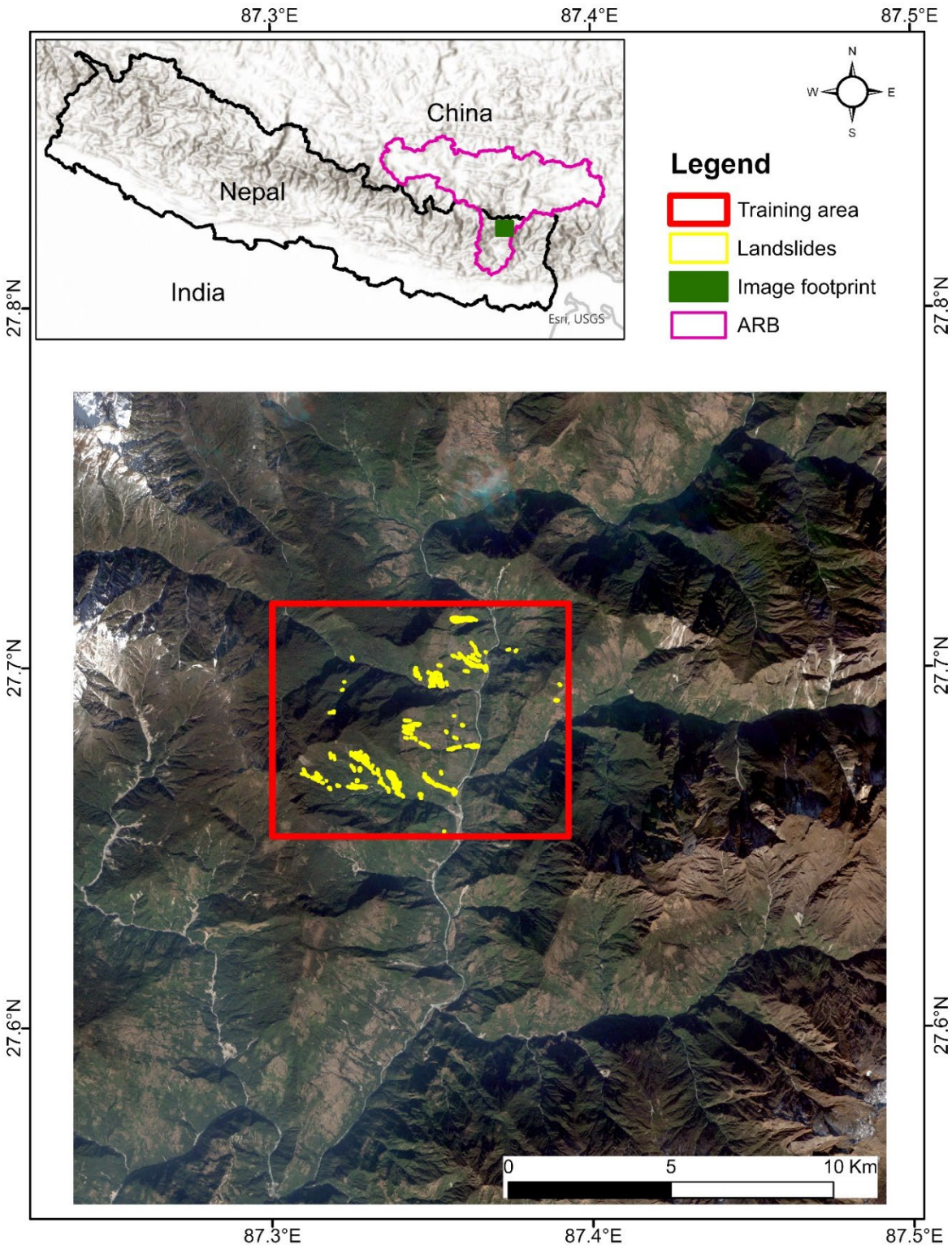

Figure 3. A 675 km^2 RapidEye image tile used for creation of training data for SALaD-based landslide mapping. The red tile highlights the 65 km^2 subset area used for training. Yellow polygons inside the training area are manually mapped landslides.

2.2. Susceptibility Mapping

Susceptibility represents the likelihood of a landslide initiating in a given location in the area of interest. The first step in generating the landslide susceptibility map is to gather the relevant datasets that describe environmental variables that may lead to increased susceptibility to landsliding. These include topographic variables slope, surface roughness, and elevation, as well as land cover classifications and lithological types. The input variables used to generate the susceptibility map are listed in Table 2. To combine the input data to generate a susceptibility model, we used the machine-learning approach based on random forest described by Emberson et al. (2021), which requires a high-quality local inventory of landslides to calibrate and validate the model outputs. The landslide inventory produced in this effort fulfills this requirement.

Table 2. Landslide susceptibility input variables.

Variable	Source	Resolution
Slope	NASADEM, slope calculation	30 m
Relief (1 km radius)	Derived from NASADEM, moving window calculation of elevation variability within a 1 km radius moving window	30 m
Average upstream slope	Derived from NASADEM, this calculates the average slope upstream from each point in the raster, calculated using a D-infinity flow accumulation algorithm	30 m
Topographic Ruggedness Index (TRI)	Derived from NASADEM, this is an estimate of local surface roughness. It measures the difference in elevation of a cell in contrast to its neighbors to provide a measure of the heterogeneity of terrain (Riley *et al.*, 1999).	30 m
Compound Topographic Index (CTI)	Derived from NASADEM. A measure of both slope and the upstream contributing area. In some locations CTI is correlated with soil parameters such as thickness (Liang and Chan, 2017). Also known as topographic wetness index (Beven and Kirkby, 1979; Sörensen *et al.*, 2006).	30 m
Topographic Position Index (TPI)	Derived from NASADEM. A measure of the relative position of a cell compared to those surrounding it; we calculate this at a 200m window (Weiss, 2001).	30m
Forest Cover	Global Forest Change 2000–2019 (Hansen *et al.*, 2013). We use the 2020 version of percent forest cover.	30 m
Land Cover	ICIMOD High Mountain Asia land cover map (ICIMOD, 2021).	30 m
Lithology	Global Lithological Map (GLiM) (Hartmann and Moosdorf, 2012). Reprojected to 30m resolution	30 m

	to match DEM; split into lithological characteristics as defined in the original study.

The second step in the process is to construct the dependent variable data (i.e., landslide source locations) to train the Random Forest classifier. In this case, the dependent variable constitutes the occurrence (or not) of landslides. Since the intention of the susceptibility model is to derive the highest likelihood of landslide source areas, we require a method to train the model on landslide source areas, rather than the entire landslide footprint. The output susceptibility model produces the likely locations of landslide sources.

We utilized the method of Marc et al. (2018) to extract the scar areas, based on the retrieval on polygon width from perimeter and area and the empirically supported assumption that the scar is 1.5 longer than wide (Domej *et al.*, 2017). The scar areas defined in this way are a first-order assumption based on the average geometry of landslides globally, but this is a consistent method that can be applied to a range of locations and we suggest that determining an exact local scar-area relationship requires field observations that are beyond the scope of this study. The training dataset of landslide scar areas was converted into a raster of landslide presence and absence at the same native resolution as the DEM (30 m, in this case) which was then provided as input to the next stage of the modeling.

To train and test the machine learning model, the data must be split into a training and test dataset. We used a training/testing split proportion of 70% and 30% of total data, respectively, which has been demonstrated to provide a good balance of model robustness and availability of training data (Xu and Goodacre, 2018). While we used pixel values for landslides, we split the data before pixelating the landslides. We generated two shapefiles of landslides, a training and test shapefile, that were then each pixelated. We define non-landslide areas as those where landslides were not mapped in the mapping interval. This means the output susceptibility analysis will be a model of likely locations for landslides that have been mapped in the mapping interval; implicit within this analysis is an assumption that these landslides are representative of the longer-term landslide trends in the study area, rather than an exception. The remaining non-landslide pixel areas within the mapped area were also split by the same proportions and combined with the respective landslide pixel areas to provide the training and testing datasets. It is important to split at the polygon, rather than pixel level, since otherwise model performance estimates will be over-optimistic (Peña and Brenning, 2015).

The random forest algorithm was run on the training data; the detailed parameters for the random forest model are shown in Table 3. The parameters were selected to avoid overfitting and to reduce the computational load. Sensitivity analysis during model development was conducted to assess the model performance with a range of values for these parameters, and the values chosen for the final model reflect the sensitivity analysis demonstrating good performance while limiting the computational time. Once fitted, the model was applied to the test subset of the data to assess the false positive rate (FPR), and true positive rate (TPR) (at a pixel basis).

Table 3. Model parameters and the justification for those values.

Random Forest Model Parameter	Value	Justification
Number of estimators (number of decision trees in forest)	100	Sensitivity analysis of model suggested limited increase in accuracy and decrease in out-of-bag error above this point
Maximum number of leaf nodes in each tree	400	Smaller values reduce computation time and overfitting, and sensitivity analysis of model suggested limited increase in accuracy and decrease in out-of-bag error above this point
Maximum number of features to consider when looking for best split of data at each node	All features	Sensitivity analysis of model suggested this has best accuracy and lowest out-of-bag error, compared to the square root of total features (which would reduce computational time).

3. Results

3.1 Landslide Mapping

Yearly distribution of landslides is shown in Figure 4. In total, 2438 landslides were mapped. In Nepal, 2223 landslides are present whereas 215 landslides are in Tibet. Landslides ranging in size from 97.65 to 357,089 m^2 were mapped. There was no imagery before 2010. As such, landslides mapped using 2010 imagery contained landslides that were induced in 2010 together with landslides induced in previous years. We considered 2010 as a base year. In 2010, there were 536 landslides with a total area of 4.35 km^2. We were able to separate landslides induced each year from 2011 onwards (Figure 5). A zoomed in view of landslides around the Upper Arun Hydropower Project is shown in Figure 6.

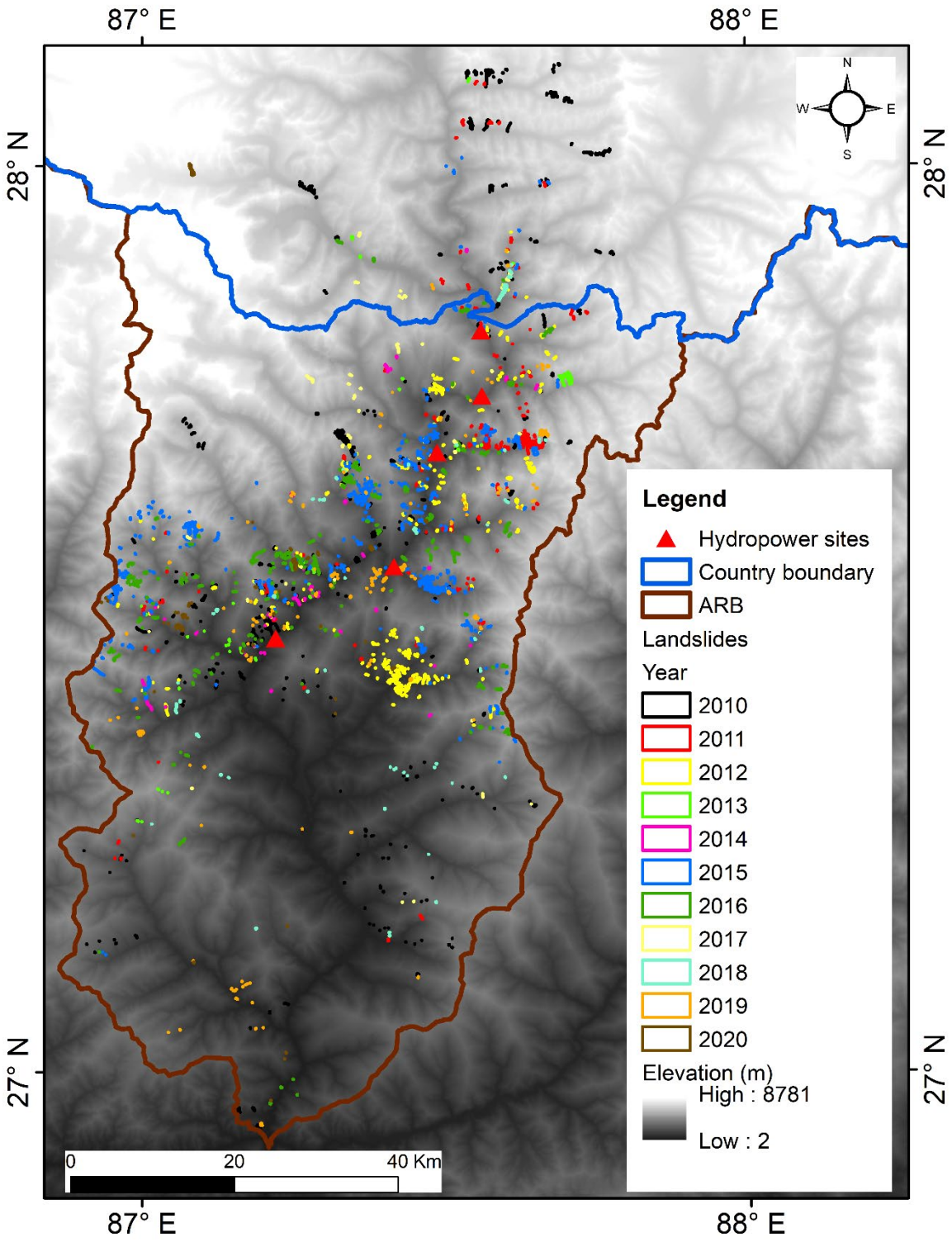

Figure 4. Lower part of Arun Basin, where majority of mapped landslides are found. Yearly distribution of landslides is highlighted by different color polygons.

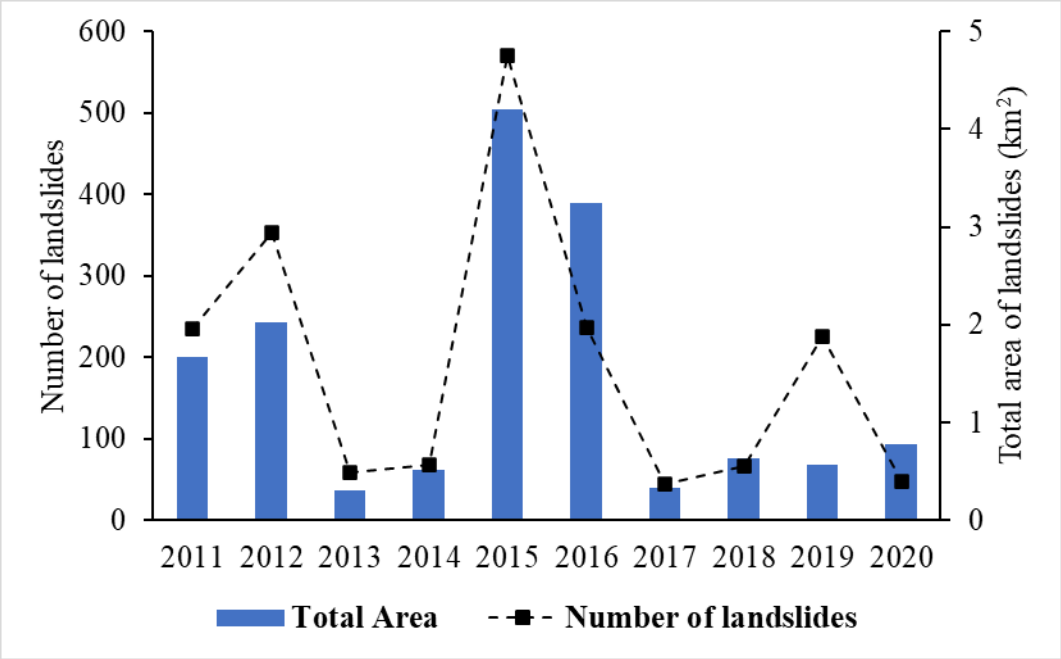

Figure 5. Yearly distribution of landslide numbers and total area of landslides.

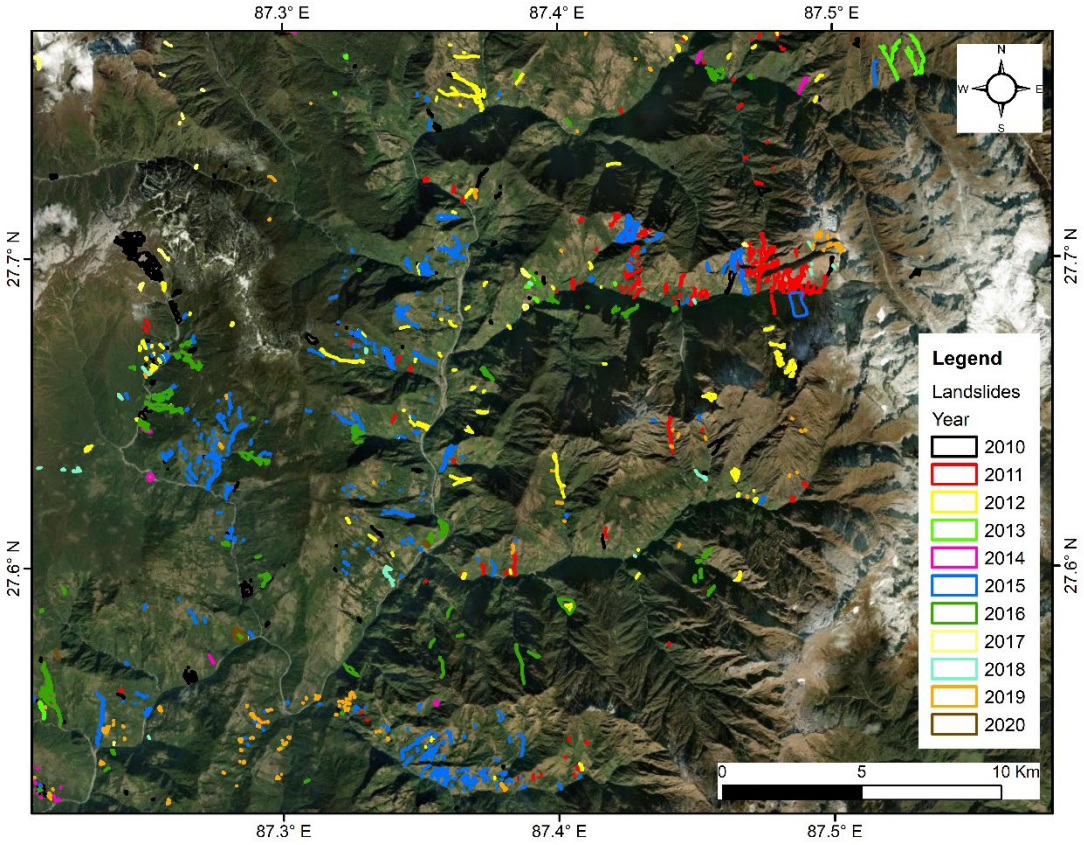

Figure 6. Zoomed in view of yearly landslides around Upper Arun Hydropower Project.

3.2 Susceptibility Mapping

A zoomed in view of the susceptibility map for areas around the hydropower sites is shown in Figure 7.

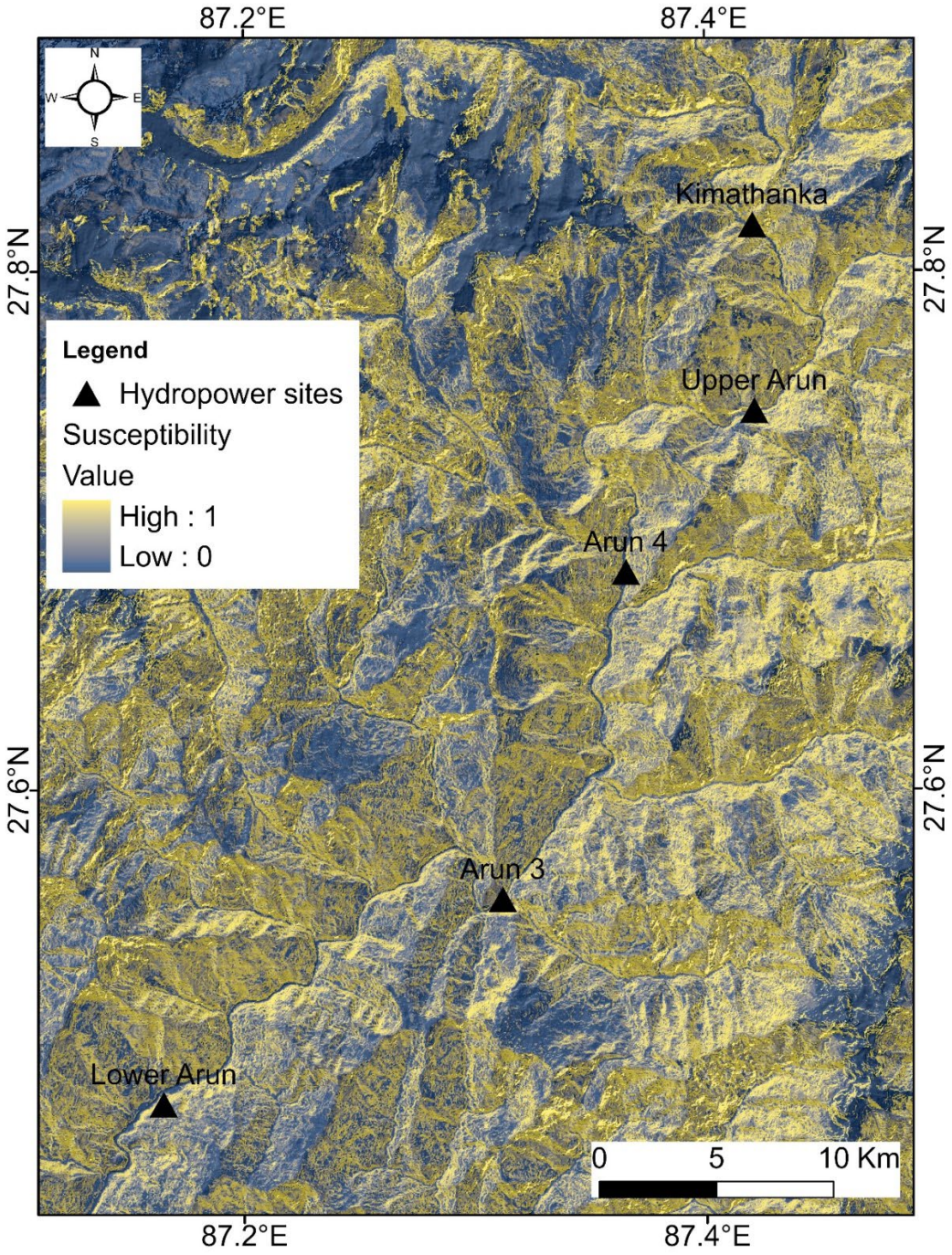

Figure 7. Susceptibility raster overlaid on a relief map in part of the study area. The hydropower sites are also shown for illustrative purposes. The susceptibility map is on a scale from 0 to 1, where 1 represents the most likely part of a given landscape to experience a landslide relative to all other parts.

The model performance was assessed using the area-under-the-curve method (AUC); a value of 1.0 indicates perfect predictive performance, with no false positives. A value of 0.5 indicates the model performance is equivalent to a random guess. The value of the model described here was 0.93 when assessed across the entire Arun basin. This indicates a good performance from the model, although it should be noted that large parts of the upper basin did not feature any significant landsliding in the mapped inventory, and the high AUC value reflects that the large landslide-sparse areas may be dominating the overall AUC value. When this value is calculated across the convex-hull of the mapped landslides in the lower basin, a value of 0.70 is obtained. If the susceptibility model were run for just this area, a potentially higher value could be obtained, but the model outputs would not be applicable to the entire basin. Although the susceptibility model relies upon prior published methods, it is important to continue to test the performance of the model. We have undertaken a 10-fold K-fold cross-validation test for the model accuracy (successful classification of landslides in the held-out cross-validation testing landslide inventory). To ensure consistency with the original method, the K-fold subsets of data are created by randomly subsetting the landslides, rather than the pixels. Based on 10-fold cross-validation, the model accuracy is 0.766 +/- 0.012 (2 standard deviations).

The importance of the input variables to the output susceptibility model can also be assessed. The decision tree algorithm within the random forest method aims to reduce the impurity of the classes at each node (i.e., split the data into landslide and non-landslide points as much as possible) based on a single parameter for each node. One measure of the importance of a parameter to the final output of the model is therefore the overall decrease in impurity of the final classification that can be attributed to that parameter (Figure 8).

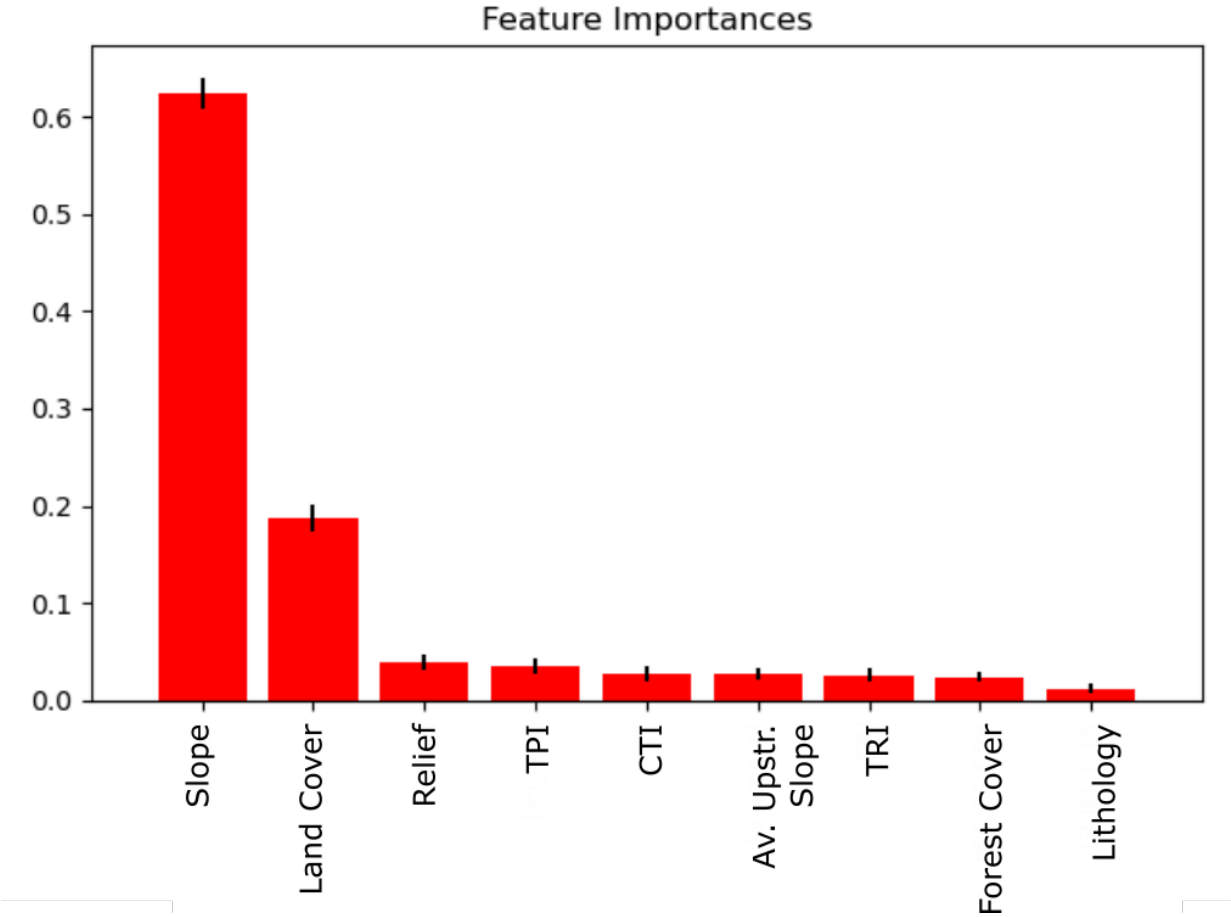

Figure 8. Importance of input parameters to output model.

The model output was, unsurprisingly, strongly influenced by slope as the most critical factor in determining landslide susceptibility. Land cover was also relevant. Other topographic factors, including relief, TPI, TRI and CTI were relevant to a much lesser extent. Lithology was only marginally relevant in the study area based on the GLiM data of Hartmann and Moosdorf (2012). The vertical black lines show the standard deviation across the different combinations of trees in the Random Forest model. Although there was a relatively small importance attached to the non-slope topographic variables, it is important to note that several topographic variables (TRI and average upstream slope in particular) are strongly correlated with slope, and so these variables may contribute more strongly than the model suggests based on the current data. The random forest method cannot attribute causal relationships between input parameters, only statistical relationships.

4. Data Availability

Multitemporal landslide inventory (2011-2020) in shapefile format is available from the NASA National Snow and Ice Data Center Distributed Active Archive Center (https://doi.org/10.5067/E4M8GRXKLCRC). Susceptibility map in GeoTIFF format is available through Figshare (https://doi.org/10.6084/m9.figshare.20986888).

5. Data Use and Reuse

The multitemporal landslide inventory (2011-2020) created for the ARB can be used for exploring the relationship between landslides and causative factors such as slope, land cover, lithology, etc (Zhang *et al.*, 2016) and climatic records (Muñoz-Torrero Manchado *et al.*, 2021). This inventory can also be used to identify areas of dense landsliding (Hölbling *et al.*, 2016) around the hydropower sites acting as a guide for further data collection and field excursions. This inventory can also enable exploration of relationship between rainfall and landsliding rate (Marc *et al.*, 2015). Landslides mapped in this effort are rainfall triggered shallow landslides visible in optical imagery. As such, the data will not support assessment of large, deep-seated landslides and earthquake induced landslides. A critical limitation is a lack of understanding on if landslides remobilized, got bigger, or otherwise in subsequent years after the initial detection due to use of polygon intersection to separate landslides induced in each year (see section 2.1.2).

Utilization of the susceptibility map can be accomplished either using the model output values as a semi-quantitative estimate of landslide likelihood, with the maximum values indicating the most likely location in the landscape, or by defining specific thresholds for certain levels of susceptibility. Ultimately, defining such levels will depend on the end-user tolerance for risk, but using an ROC-AUC curve to define thresholds (e.g. Emberson *et al.*, 2021b) can be one way to determine thresholds based on the percentage of landslides included in each zone (e.g., low risk zones could correspond to a susceptibility threshold under which only 10% of landslides were observed, while high susceptibility could correspond to a threshold under which 90% of landslides were observed). Susceptibility mapping can be paired with runout models to inform potential zones of hazard to landslide deposition and runout. Globally applicable models, such as Flow-R (Horton *et al.*, 2013) can use susceptibility maps as input. While some studies have used runout parameter values such as average velocity drawn from global observations, in this setting we lack calibration on runout parameters and as such suggest that runout validation would be important to calibrate runout models to provide qualitative estimates of potential hazard.

6. Conclusion

We used high-resolution satellite imagery from PlanetScope and RapidEye and open-source SALaD system to establish a multitemporal shallow landslide inventory for the basin. The susceptibility analysis using the inventory provides a valuable initial estimate of where landslides are likely to trigger. Since susceptibility provides guidance on the likely source of landslide materials, these outputs may be valuable in determining which hillslopes may benefit from stabilization and mitigation of landslide hazard. These datasets will become valuable for advancing landslide studies in the basin and upcoming hydropower projects.

Acknowledgements
This research was funded under NASA's High Mountain Asia Science Team (NNH19ZDA001N-HMA) and the World Bank. This work utilized data made available through the NASA Commercial Smallsat Data Acquisition (CSDA) Program. We would like to thank Pravin Karki from the World Bank, Dr. John Reynolds from Reynolds International Ltd and Professor Vishnu Prasad Pandey from Institute of Engineering, Nepal for helpful discussion during the production of data.

Conflict of interests
The authors declare that they have no conflict of interest.

References

Amatya P, Kirschbaum D, Stanley T, Tanyas H. 2021. Landslide mapping using object-based image analysis and open source tools. *Engineering Geology*, 282: 106000. https://doi.org/10.1016/j.enggeo.2021.106000.

Association IH. 2020. *No Title*. .

Beven KJ, Kirkby MJ. 1979. A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal*. Taylor & Francis, 24(1): 43–69.

Breiman L. 2001. Random forests. *Machine learning*. Springer, 45(1): 5–32.

Byers AC, Rounce DR, Shugar DH, Lala JM, Byers EA, Regmi D. 2019. A rockfall-induced glacial lake outburst flood, Upper Barun Valley, Nepal. *Landslides*. Springer, 16(3): 533–549.

Corominas J. 1996. The angle of reach as a mobility index for small and large landslides. *Canadian Geotechnical Journal*. NRC Research Press Ottawa, Canada, 33(2): 260–271.

Dahlquist MP, West AJ. 2019. Initiation and runout of post-seismic debris flows: insights from the 2015 Gorkha earthquake. *Geophysical Research Letters*, 46: 9658–9668. https://doi.org/10.1029/2019GL083548.

Domej G, Bourdeau C, Lenti L, Martino S, Pluta K. 2017. Mean landslide geometries inferred from a global database of earthquake-and non-earthquake-triggered landslides. *Italian Journal of Engineering Geology and Environment*, 2(17): pp-87.

Emberson RA, Kirschbaum DB, Stanley T. 2021. Landslide hazard and exposure modelling in data-poor regions: the example of the Rohingya refugee camps in Bangladesh. *Earth's Future*. Wiley Online Library, e2020EF001666.

Goetz J, Brenning A, Petschko H, Leopold P. 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers & geosciences*, 81: 1–11. https://doi.org/10.1016/j.cageo.2015.04.007.

Guzzetti F, Reichenbach P, Cardinali M, Galli M, Ardizzone F. 2005. Probabilistic landslide hazard assessment at the basin scale. *Geomorphology*. Elsevier, 72(1–4): 272–299.

Hansen MC, Potapov P V, Moore R, Hancher M, Turubanova SA, Tyukavina A, Thau D, Stehman S V, Goetz SJ, Loveland TR. 2013. High-resolution global maps of 21st-century forest cover change. *science*. American Association for the Advancement of Science, 342(6160): 850–853.

Hartmann J, Moosdorf N. 2012. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*. Wiley Online Library, 13(12).

Hölbling D, Betts H, Spiekermann R, Phillips C. 2016. Identifying Spatio-Temporal Landslide Hotspots on North Island, New Zealand, by Analyzing Historical and Recent Aerial Photography. *Geosciences* .

Horton P, Jaboyedoff M, Rudaz al B et, Zimmermann M. 2013. Flow-R, a model for susceptibility mapping of debris flows and other gravitational hazards at a regional scale. *Natural hazards and earth system sciences*, 13(4): 869–885. https://doi.org/10.5194/nhess-13-869-2013.

ICIMOD. 2017a. Landslide data of Koshi basin (within Nepal) of 1990 developed through remote sensing approach [Data set]. *ICIMOD*. https://doi.org/10.26066/RDS.34425.

ICIMOD. 2017b. Landslide data of Koshi basin (within Nepal) of 2010 developed through remote sensing approach (Data set]. *ICIMOD*. https://doi.org/10.26066/RDS.34426.

ICIMOD. 2021. Land cover of HKH region [Data set]. *ICIMOD*. https://doi.org/10.26066/RDS.1972511.

Jones JN, Boulton SJ, Stokes M, Bennett GL, Whitworth MRZ. 2021. 30-year record of Himalaya mass-wasting reveals landscape perturbations by extreme events. *Nature Communications*, 12(1): 6701. https://doi.org/10.1038/s41467-021-26964-8.

Liang W-L, Chan M-C. 2017. Spatial and temporal variations in the effects of soil depth and topographic wetness index of bedrock topography on subsurface saturation generation in a steep natural forested headwater catchment. *Journal of Hydrology*. Elsevier, 546: 405–418.

Marc O, Hovius N, Meunier P, Uchida T, Hayashi S. 2015. Transient changes of landslide rates after earthquakes. *Geology*. Geological Society of America, 43(10): 883–886.

Marc O, Stumpf A, Malet J-P, Gosset M, Uchida T, Chiang S-H. 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*, 6(4): 903–922. https://doi.org/10.5194/esurf-6-903-2018.

Merghadi A, Yunus AP, Dou J, Whiteley J, ThaiPham B, Bui DT, Avtar R, Abderrahmane B. 2020. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207: 103225. https://doi.org/10.1016/j.earscirev.2020.103225.

Michel J, Youssefi D, Grizonnet M. 2014. Stable mean-shift algorithm and its application to the segmentation of arbitrarily large remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*. IEEE, 53(2): 952–964.

Muñoz-Torrero Manchado A, Allen S, Ballesteros-Cánovas JA, Dhakal A, Dhital MR, Stoffel M. 2021. Three decades of landslide activity in western Nepal: new insights into trends and climate drivers. *Landslides*, 18(6): 2001–2015. https://doi.org/10.1007/s10346-021-01632-6.

NASA JPL. 2020. NASADEM Merged DEM Global 1 arc second V001 [Data set]. *NASA EOSDIS Land Processes DAAC*. https://doi.org/10.5067/MEaSUREs/NASADEM/NASADEM_HGT.001.

Paudel B, Fall M, Daneshfar B. 2021. Gis-based assessment of debris flow runout in Kulekhani Watershed, Nepal. *Geotechnical and Geological Engineering*, 39: 2755--2775. https://doi.org/10.1007/s10706-020-01655-1.

Peña MA, Brenning A. 2015. Assessing fruit-tree crop classification from Landsat-8 time series for the Maipo Valley, Chile. *Remote Sensing of Environment*. Elsevier, 171: 234–244.

Planet Team. 2017. *Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA.* .

Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F. 2018. A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*. Elsevier, 180: 60–91.

Reynolds P. 2018. Landslide hits tunnel portal at Arun-3 in Nepal. *TunnelTalk*, 5 July.

Riley SJ, DeGloria SD, Elliot R. 1999. A Terrain Ruggedness Index that quantifies topographic heterogeneity. *intermountain Journal of sciences*, 5(1–4): 23–27.

Roberts S, Jones JN, Boulton SJ. 2021. Characteristics of landslide path dependency revealed through multiple resolution landslide inventories in the Nepal Himalaya. *Geomorphology*. Elsevier, 390: 107868.

Scaioni M, Longoni L, Melillo V, Papini M. 2014. Remote Sensing for Landslide Investigations: An Overview of Recent Achievements and Perspectives. *Remote Sensing*, 6(10): 9600–9652. https://doi.org/10.3390/rs6109600.

Shakya D. 2020. Arun River in Sankhuwasabha partially dammed by landslide. *The Kathmandu Post*.

Sörensen R, Zinko U, Seibert J. 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences*. Copernicus GmbH, 10(1): 101–112.

Weiss A. 2001. Topographic position and landforms analysis. *Poster presentation, ESRI user conference, San Diego, CA*.

Xu Y, Goodacre R. 2018. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of analysis and testing*, 2: 249–262.

Zhang J, Liu R, Deng W, Khanal NR, Gurung DR, Murthy MSR, Wahid S. 2016. Characteristics of landslide in Koshi River Basin, Central Himalaya. *Journal of Mountain Science*, 13(10): 1711–1722. https://doi.org/10.1007/s11629-016-4017-0.

Zhang J, Westen CJ van, Tanyas H, Mavrouli O, Ge Y, Bajrachary S, Gurung DR, Dhital MR, Khanal NR. 2019. How size and trigger matter: analyzing rainfall-and earthquake-triggered landslide inventories and their causal relation in the Koshi River basin, central Himalaya. *Natural hazards and earth system sciences*. Copernicus GmbH, 19(8): 1789–1805.