



# Space–time landslide hazard modeling via Ensemble Neural Networks

Ashok Dahal<sup>1</sup>, Hakan Tanyas<sup>1</sup>, Cees van Westen<sup>1</sup>, Mark van der Meijde<sup>1</sup>, Paul Martin Mai<sup>2</sup>, Raphaël Huser<sup>3</sup>, and Luigi Lombardo<sup>2</sup>

<sup>1</sup>Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, Enschede, AE 7500, the Netherlands

<sup>2</sup>Physical Science and Engineering (PSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

<sup>3</sup>Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

**Correspondence:** Ashok Dahal (a.dahal@utwente.nl)

Received: 27 March 2023 – Discussion started: 6 April 2023

Revised: 16 December 2023 – Accepted: 17 January 2024 – Published: 8 March 2024

**Abstract.** Until now, a full numerical description of the spatio-temporal dynamics of a landslide could be achieved only via physically based models. The part of the geoscientific community in developing data-driven models has instead focused on predicting where landslides may occur via susceptibility models. Moreover, they have estimate when landslides may occur via models that belong to the early-warning system or to the rainfall-threshold classes. In this context, few published research works have explored a joint spatio-temporal model structure. Furthermore, the third element completing the hazard definition, i.e., the landslide size (i.e., areas or volumes), has hardly ever been modeled over space and time. However, technological advancements in data-driven models have reached a level of maturity that allows all three components to be modeled (Location, Frequency, and Size). This work takes this direction and proposes for the first time a solution to the assessment of landslide hazard in a given area by jointly modeling landslide occurrences and their associated areal density per mapping unit, in space and time. To achieve this, we used a spatio-temporal landslide database generated for the Nepalese region affected by the Gorkha earthquake. The model relies on a deep-learning architecture trained using an Ensemble Neural Network, where the landslide occurrences and densities are aggregated over a squared mapping unit of 1 km × 1 km and classified or regressed against a nested 30 m lattice. At the nested level, we have expressed predisposing

and triggering factors. As for the temporal units, we have used an approximately 6 month resolution. The results are promising as our model performs satisfactorily both in the susceptibility (AUC = 0.93) and density prediction (Pearson  $r = 0.93$ ) tasks over the entire spatio-temporal domain. This model takes a significant distance from the common landslide susceptibility modeling literature, proposing an integrated framework for hazard modeling in a data-driven context.

## 1 Introduction

The literature on physically based models for landslides shows various solutions to estimate where landslides might occur, when or how frequently they occur, and how they may evolve (e.g., Formetta et al., 2016; Bout et al., 2018). This framework allows one to describe the dynamics of a landslide from its initiation, propagation, and entrainment to the runout and deposition (e.g., Burton and Bathurst, 1998; Zhang et al., 2013). As a result, metrics such as the velocity, runout height, overall landslide area, and volume constitute standard outputs of such a modeling approach (see, van den Bout et al., 2021b, a). However, these models are often constrained to single slopes or catchments because of the spatial data requirements on geotechnical parameters. This limitation has stimulated the geoscientific community

to develop data-driven statistical models instead (Van Westen et al., 2006). These are much more versed in being extended over large regions because, rather than requiring specific geotechnical properties, they can rely on terrain attributes and remotely sensed data acting as geotechnical proxies (Van Westen et al., 2008; Frattini et al., 2010). However, in doing so, the geoscientific community has almost exclusively focused on assessing where landslides may occur, as temporal landslide inventories were scarcely available. This notion is commonly referred to as landslide susceptibility (Reichenbach et al., 2018; Titti et al., 2021). As for the low number of publications focused on estimating when or how frequently landslides may occur at a given location, the community has produced a number of near-real-time predictive landslide models for rainfall (Intrieri et al., 2012; Kirschbaum and Stanley, 2018; Ju et al., 2020) and seismic (Tanyaş et al., 2018; Nowicki Jessee et al., 2018) triggers. With regard to characteristics such as velocity, kinetic energy and runout, albeit fundamental to describing a potential landslide threat (Fell et al., 2008; Corominas et al., 2014), these cannot currently be used for data-driven modeling because no observed dataset of landslide dynamics exists to support the modeling and prediction paradigm based on Artificial Intelligence (AI) or statistical approaches. Guzzetti et al. (1999) proposed to alternatively model landslide areas that can be easily extracted from a polygonal inventory. Nevertheless, the first spatially explicit models able to estimate landslide areas have been recently proposed by Lombardo et al. (2021) and Zapata et al. (2023). In their work, the authors exclusively estimated the potential landslide size at a given location without informing whether the given location would have been susceptible in the first place. This limitation has been further addressed by Bryce et al. (2022) and Aguilera et al. (2022), implementing models that couple susceptibility with landslide area prediction. Nevertheless, even in these cases, the absence of the temporal dimension in their work implies that no current data-driven model is capable of solving the landslide hazard definition (Guzzetti et al., 1999), jointly estimating where, when (or how frequently) and how large landslides may be in a given spatio-temporal domain. Apart from spatial modeling, temporal aspects of landslides are also addressed in works of Lombardo et al. (2020); Samia et al. (2020); Ozturk et al. (2021).

The present work expands on the data-driven literature summarized above by proposing a space–time deep-learning model based on an Ensemble Neural Network (ENN) architecture. Neural Networks (NN) are not new to the landslide literature, although they have found the spotlight so far mostly for automated landslide detection (Catani, 2021; Meena et al., 2022), monitoring (Neaupane and Achet, 2004; Wang et al., 2005), and for landslide susceptibility assessment (Lee et al., 2004; Catani et al., 2005; Gomez and Kavzoglu, 2005; Grelle et al., 2014; Montrasio et al., 2014; Catani et al., 2016; Nocentini et al., 2023). Here, the main difference is that our ENN is built as an ensemble made of

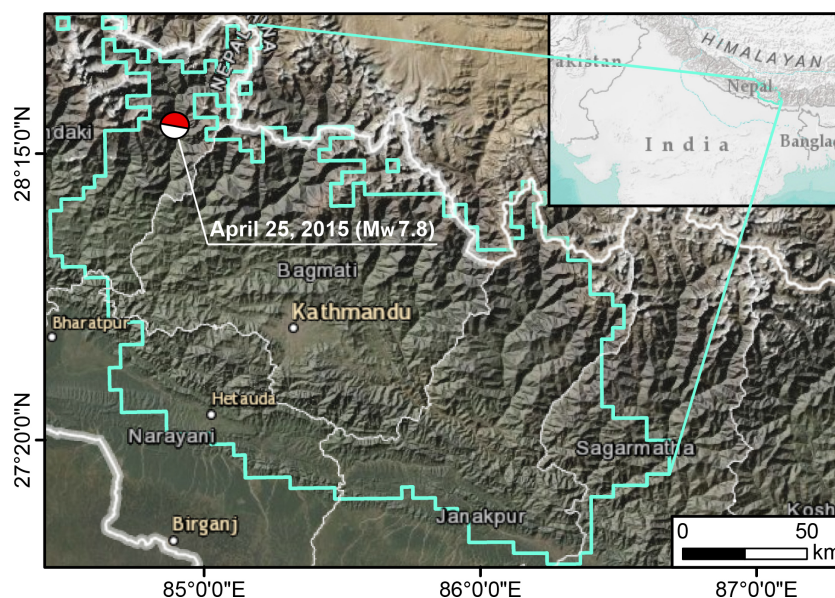
two elements, i.e., a landslide susceptibility classifier and a landslide density area regression model, both simultaneously defined over the same spatio-temporal domain. Thanks to the open data repository of Kinsey et al. (2021), we tested our space–time ENN, fully complying for the first time with the landslide hazard definition (as per Guzzetti et al., 1999).

The paper is organized as follows: Sect. 2 describes the data we used; Sect. 4 summarizes how we partitioned the study area; Sect. 5 lists the predictors we chose; Sect. 6 details our space–time ENN architecture; Sect. 7 reports our results, which are then discussed in Sect. 8, and Sect. 9 concludes our contribution with an overall summary and future plans.

## 2 Study area and landslide inventory

The 2015 Gorkha (Nepal) Earthquake is one of the strongest recent earthquakes in South Asia, and specifically along the Himalayan sector (e.g., Kargel et al., 2016). The  $M_w$  7.8 mainshock occurred on 25 April 2015, and together with a sequence of aftershocks, it was responsible for triggering more than 25 000 landslides (Roback et al., 2018). The ground motion not only affected the Nepalese terrain right after the earthquake through co-seismic landslides, but its disturbance increased landslide susceptibility in the following years, a phenomenon commonly referred to as earthquake legacy (Jones et al., 2021; Tanyaş et al., 2021a). The legacy of the Gorkha earthquake has been recently demonstrated by mapping a multi-temporal inventory, which has been publicly shared by Kinsey et al. (2021). The authors mapped landslides across the area shown in Fig. 1 from 2014 to 2018, including the co-seismic phase, as well as all pre-monsoon and post-monsoon seasons, with an approximate temporal coverage of six months. They used a time series of freely available medium-resolution satellite imagery (Landsat-8 and Sentinel-2) and aggregated the resulting landslide areas at the level of a 1 km squared lattice. Overall, they mapped three pre-seismic and seven post-seismic landslide inventories in addition to the co-seismic one. In this work, we excluded three pre-seismic inventories and selected the inventories from April 2015 onward because the effect of the ground motion and its legacy is present only after the event. As a result, from the gridded database by Kinsey et al. (2021), we extracted a total of eight landslide inventories.

As the landslide information was aggregated at a 1 km resolution, it is not possible to disentangle single landslides, one from the others. Thus, each 1 km grid reports the whole landslide area mapped by the authors each time without excluding the footprint of previous failures. For this reason, we had to include a pre-processing step where each temporal replicate was re-calculated and re-assigned with the difference in landslide area density between two original subsequent inventories. In the attempt to focus on newly activated landslides, we then considered only grid cells with an



**Figure 1.** Study area defined within the cyan polygon, where Kinney et al. (2021) mapped the multitemporal landslide inventories upon which we based the analysis in this work. The Beach Ball shows the moment tensor of the energy release from the 2015 Gorkha Earthquake.

increase in landslide area. The interpretation here is that an increase with time implies either newly formed landslides or re-activated ones. Conversely, the grids where the landslide area diminished with respect to their previous counterpart were assigned with a zero value under the assumption that no landslide took place, but rather that vegetation recovery was responsible for the estimated change. The resulting temporal inventory at different time periods over the 1 km grid is shown in Fig. 2.

### 3 Geological context

Geology largely controls the landslide initiation and flow, and thus, it is commonly used as part of landslide susceptibility studies (Fan et al., 2019). In the context of this experiment, the majority of the study area (9%) is classified as the Siwalik Formation, followed by the Himal group and a combination of many river formations such as the *Seti* and *Sarung Khola* formations (Dahal, 2012). The Siwalik Formation is mostly a Molasse deposit of the Himalayas, consisting of sandstone, mudstone, shale, and conglomerate. The river formations that predominantly outcrop in the middle Himalayas are a combination of Schist, Granite, Gneiss, Phyllite, and Quartzite. The upper Himalayan region consists instead of a combination of Schist, Gneiss, Migmatites, and Marbles (Upreti, 2001).

This general overview is summarized in the geological map by Dahal (2012). However, this map does not cover a substantial portion of the upper Himalayas, where the landslide multi-temporal inventory mapped by Kinney et al. (2021) includes many landslides. Therefore, this map is

largely unsuitable for testing a data-driven model aimed at explaining the landslide hazard associated with the above-mentioned inventory.

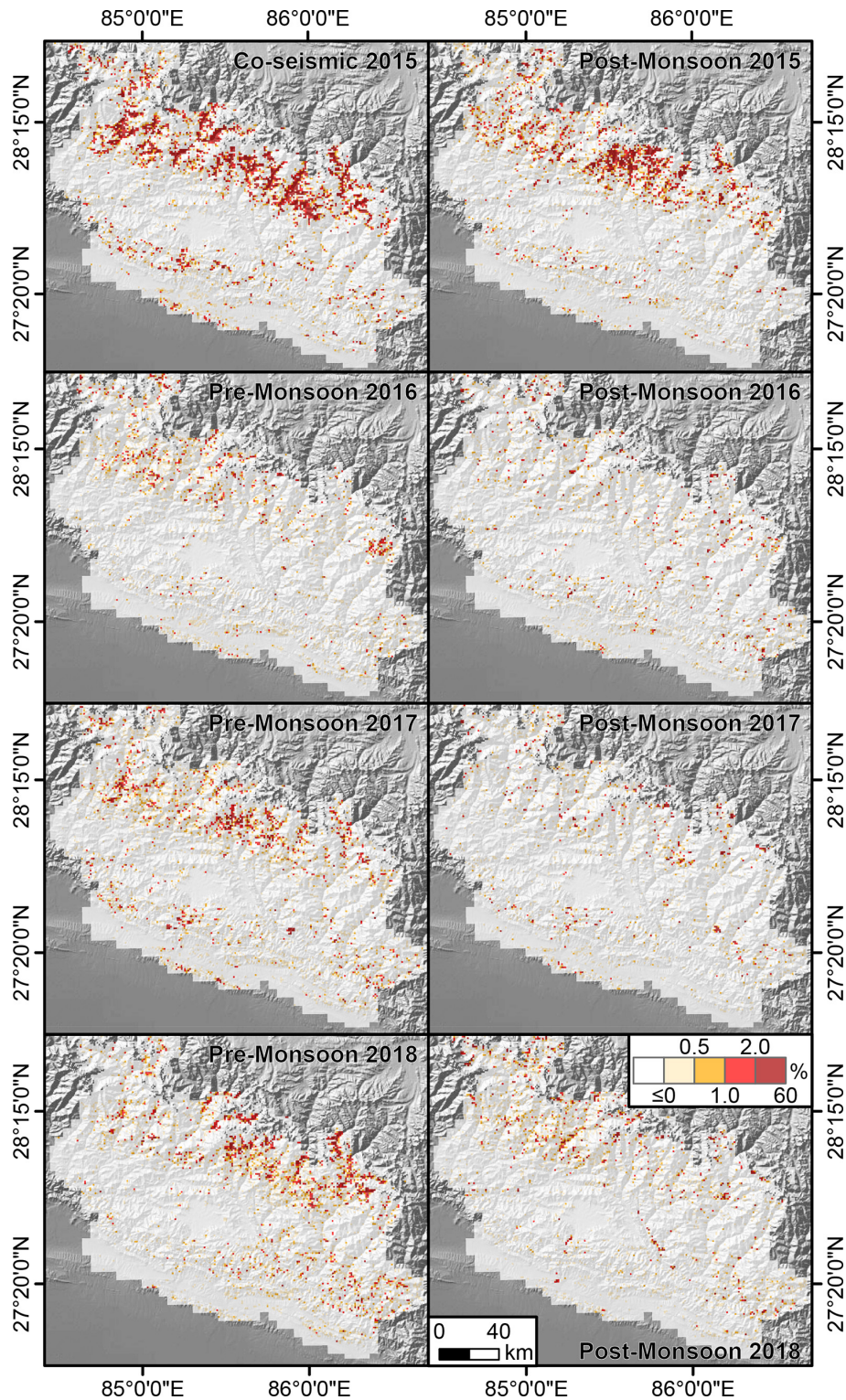
A second geological map covers the Nepalese territory, this made by Dahal (2012). However, it is limited to a very coarse resolution (1 : 1 000 000) and only reports information about the formations rather than the associated material. Even if one were to extrapolate the respective rock types from the literature, (Upreti, 2001), such formations would only summarize the complexity of the Himalayan landscape in a few classes, thus making it of limited use in the context of data-driven landslide modeling. The Department of Mines and Geology of Nepal is currently collating all the information with the intent of producing a detailed geological map at a 1 : 50 000 scale. However, this is a work in progress that still misses most of the study areas considered in this study.

This major limitation affected our ability to consider geology in this article. Therefore, although we are aware of its relevance in any landslide study, geological information is not explicitly included as a predictor in the modeling protocol presented below.

### 4 Selection of mapping units

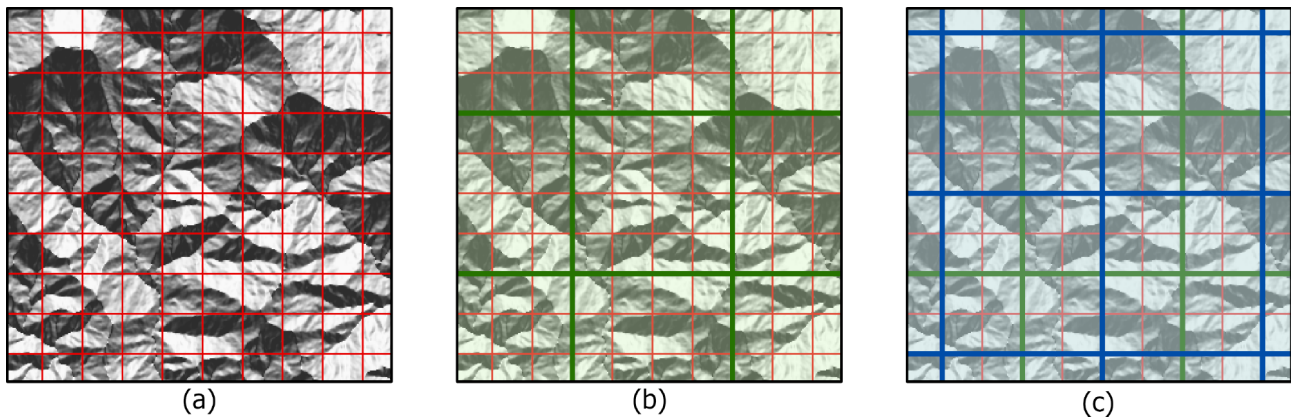
To partition our study area, we use the same mapping unit defined by Kinney et al. (2021). Because the authors aggregated the landslide information on a 1 km × 1 km square grid, our model targets are defined within the same lattice structure. As for the definition of the predictor set, unlike current data-driven practices where medium resolution mapping units are assigned with the mean and standard devia-





**Figure 2.** Landslide Area Density (% in a  $1 \text{ km}^2$  grid) calculated as the difference between two consecutive inventories mapped with different time ranges provided by Kinney et al. (2021). The colorbar is saturated between 2 % and 60 % because there are very few grid cells with such landslide area density in the data.





**Figure 3.** Panels showing the various mapping unit structures: **(a)** the covariate and existing inventory grid structure, with a  $1\text{ km} \times 1\text{ km}$  grid with  $32 \times 32$  pixels of terrain image in the background, **(b)** the patching of a  $4 \times 4$  inventory grid with a  $4\text{ km} \times 4\text{ km}$  grid, and **(c)** the shifted patch structure with a similar grid structure to **(b)**.

tion of the predictors under consideration (Ardizzone et al., 2002; Schlögel et al., 2018), here we exploit an NN structure to treat each predictor as an image. In other words, each  $1\text{ km} \times 1\text{ km}$  square grid was not summarized with its mean and standard deviation values, but rather, we provided the entire spatial distribution of predictors as an image patch to our convolutional NN (CNN) model, which is capable of extracting information from image data.

Only feeding a single grid structure to the NN would neglect any spatial dependence coming from neighboring areas (Glenn et al., 2006; Vasiliev, 2020). As landslides are dynamic phenomena, it is essential to inform the model about how the landslide distribution changes across the neighboring landscape, as well as the characteristics of the neighborhood under consideration. To do so, we extended the “spatial” vision of our ENN by creating two additional sets of lattices, each encompassing 16 grids measuring  $1\text{ km}$ , in a  $4 \times 4$  patch. Figure 3 further explains the mapping unit structures, wherein in panel (a), we can observe that the  $1\text{ km}$  red polygonal lattice created by Kinney et al. (2021) contains  $32 \times 32$  pixels of the underlying terrain characteristics.

The subplot (Fig. 3b) shows how each patch is generated through the green boxes containing 16 inventory grids. Each box will later be used as the training patches in the ENN, which in turn implies a  $128 \times 128$  pixel structure ( $32\text{ pixel} \times 4 = 128$ ) as input data. The model will then output 16 inventory grids, following the same data structure expressed at the  $4 \times 4$  patch level. Notably, if we had used the single patch arrangement shown in Fig. 3b, then the landscape characteristics along the edges of each patch would have been lost.

Therefore, we also produced a second patch arrangement, identical to the first but shifted by  $2\text{ km}$  to the east and  $2\text{ km}$  to the south. This operation returned the blue patches shown in Fig. 3c. In this way, the total data volume is also increased,

providing multiple terrain and landslide scenarios defined over the different spatial data structures.

Note that these spatial manipulation procedures are quite common for Convolutional Neural Networks (e.g., Amit and Aoki, 2017). Here, we have simply adapted them in the context of the gridded structure defined by Kinney et al. (2021).

## 5 Predictors

The predictor set we chose features a number of terrain attributes, as well as hydrological and seismic factors. These predictors are selected based on their influence on landslides, which is observed by many existing works as represented in Table 1. Our assumption is that their combined information is able to explain the distribution of landslide occurrences and area densities (the combined targets of our ENN) both in space and time. These predictors are listed in Table 1, graphically shown in Fig. 4. Below, we report a brief explanation to justify their choice.

The *Slope* carries the signal of the gravitational pull acting on potentially unstable materials hanging along the topographic profile (Taylor, 1948). *Elevation*, *Eastness* and *Northness* are common proxies for a series of processes such as moisture, vegetation, and temperature (Clinton, 2003), and their effect on slope stability (Neaupane and Piantanakulchai, 2006; Whiteley et al., 2019; Loche et al., 2022). As for the *Planar* and *Profile Curvatures*, these are known to control the convergence and divergence of overland flows (Ohlmacher, 2007). This hydrological information is also supported by the *Topographic Wetness Index* and *Distance to River* (Yesilnacar and Topal, 2005). To these finely represented predictors, we also added a number of coarser ones, representing the potential triggers behind a landslide genetic process, namely, *Rainfall* (its Maximum seasonal value (per pixel) and 95 % Confidence Interval (CI) (per pixel) within the inventory time-frame, calculated from daily CHIRPS

**Table 1.** Predictors' summary.

Type	Covariate: (Acronym   Unit)	Reference
Morphometric (30 m SRTM)	<i>Slope</i> ( <i>Slope</i>   degrees)	Zevenbergen and Thorne (1987)
Morphometric (30 m SRTM)	<i>Elevation</i> ( <i>Elevation</i>   m)	–
Morphometric (30 m SRTM)	<i>Northness</i> ( <i>Northness</i>   unitless)	Steger et al. (2016)
Morphometric (30 m SRTM)	<i>Eastness</i> ( <i>Eastness</i>   unitless)	Steger et al. (2016)
Morphometric (30 m SRTM)	<i>Profile Curvature</i> ( <i>PRC</i>   $\text{m}^{-1}$ )	Heerdegen and Beran (1982)
Morphometric (30 m SRTM)	<i>Planar Curvature</i> ( <i>PLC</i>   $\text{m}^{-1}$ )	Heerdegen and Beran (1982)
Morphometric (30 m SRTM)	<i>Topographic Wetness Index</i> ( <i>TWI</i>   unitless)	Sörensen et al. (2006)
Precipitation (~ 5 km CHRIPS)	<i>Maximum daily rainfall</i> ( <i>Max. Precip.</i>   $\text{mm d}^{-1}$ )	Funk et al. (2015)
Precipitation (~ 5 km CHRIPS)	<i>95 % CI rainfall in the inventory period</i> ( <i>95 % CI Precip.</i>   $\text{mm d}^{-1}$ )	Funk et al. (2015)
Seismic shaking (1 km USGS)	<i>Maximum Peak Ground Acceleration from main event and major aftershock</i> ( <i>Max PGA</i>   $\text{m s}^{-2}$ )	Worden and Wald (2016)
Seismic shaking (1 km USGS)	<i>St. Dev. Peak Ground Acceleration</i> ( <i>1Std. PGA</i>   $\text{m s}^{-2}$ )	Worden and Wald (2016)
Distance to River	<i>Distance to River</i> ( <i>Dist2Riv</i>   m)	–
Monsoons after Earthquake (count)	<i>Monsoons after the Earthquake</i> ( <i>Monsoons</i>   year)	–

data spanning between two subsequent landslide inventories; Funk et al., 2015) and *Peak Ground Acceleration* (its Maximum value from between main shock and aftershock and their respective standard deviation estimated using empirical ground motion prediction equations, available through the ShakeMap system of the United States Geological Survey (USGS); Worden and Wald, 2016).

The PGA is empirically estimated from local ground motion recording stations, and it has been shown to correlate with the Gorkha co-seismic landslide scenario (Dahal et al., 2023). Similar to PGA, Peak Ground Velocity (PGV) could also be used in this case to model the landslides as it is a better predictor in some cases (Maufroy et al., 2015; von Specht et al., 2019). However, very few stations actively recorded the Gorkha earthquake. This is the reason why the differences in the spatial patterns of the PGA and PGV available in the USGS ShakeMap system are negligible (Hough et al., 2016). Therefore, with the objective of explaining the co-seismic landslide scenario over such a large study area, any of the two ShakeMaps would produce similar results.

To these spatially and temporally varying predictors, we also added the number of wet seasons after the Gorkha Earthquake to inform the model about the combined effect of landscape characteristics, earthquake legacy, and meteorological stress.

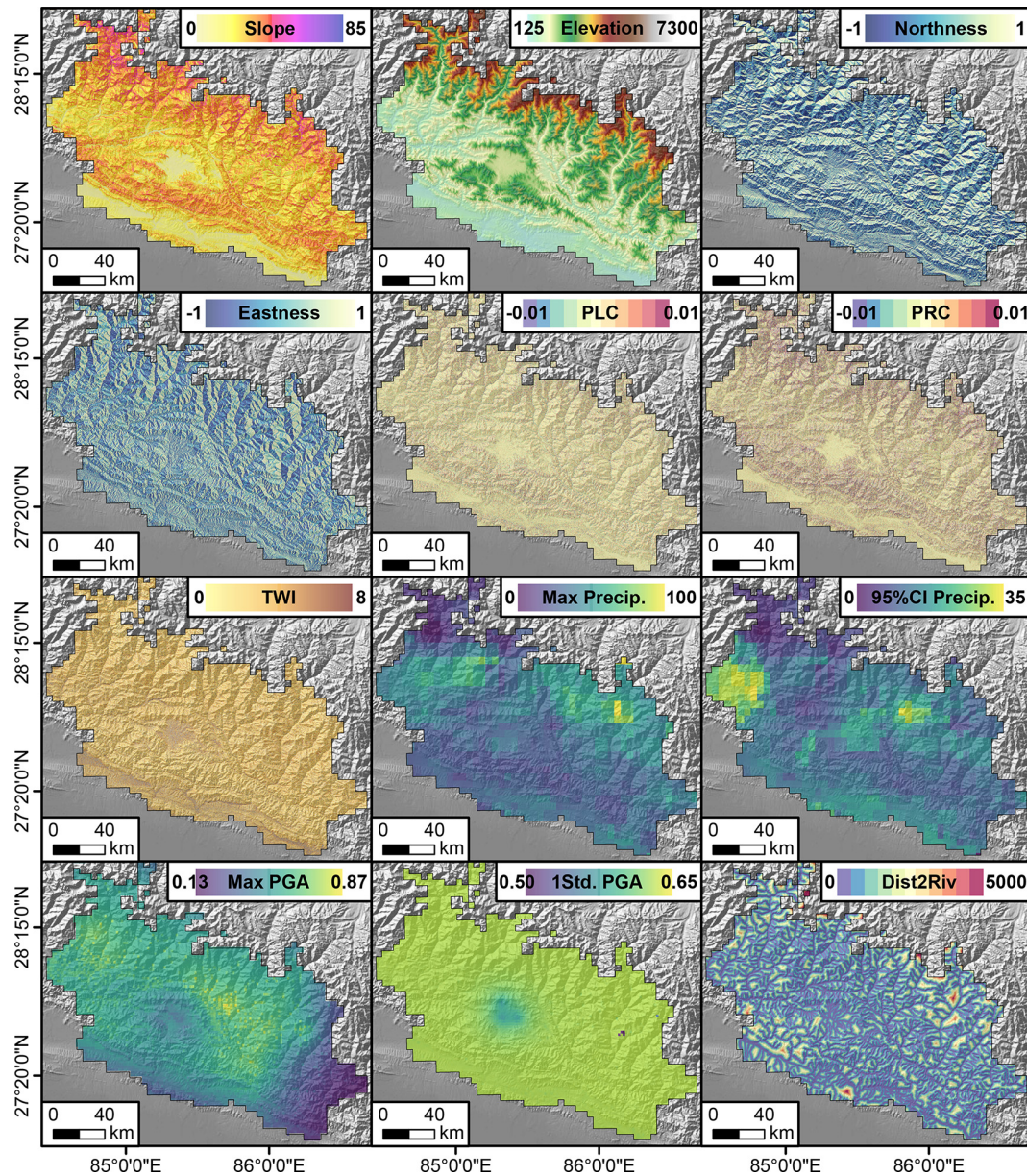
## 6 Neural networks

### 6.1 Model architecture

To contextually estimate landslide susceptibility and area density, we designed an NN with a multi-output design, relying on the same 1 km gridded data input. In short, the first model component estimates a “pseudo-probability” via a sigmoid function, whereas the second component regresses the area density information against the same set of predictors used in the previous step.

The NN design is shown in the Fig. 5. The susceptibility block is modified from the U-Net model (Ronneberger et al., 2015) with the backbone of Resnet18 (He et al., 2015), where the model processes input information through the 18 blocks of Convolution, Batch Normalization (Ioffe and Szegedy, 2015b), dropout, Rectified Linear Unit (ReLU), and Max pooling (Wu and Gu, 2015) with a total 23 556 931 trainable parameters, which are variables that need to be optimized during the training process. The convolution layer in each block convolved with the  $3 \times 3$  window, and it was initialized with the Glorot uniform initialization function (Glorot and Bengio, 2010). The convolution function was followed by a batch normalization, which works as a regularization function. This prevents the model from overfitting, and it is followed by a dropout layer. The dropout layer randomly deactivates 30 % of the neurons in the convolution layer, such that the model does not overfit. Following this, the feature space passes through the ReLU activation function, which al-





**Figure 4.** Predictors used for training the Ensemble Neural Network. The Max Precip. is one example of the maximum daily rain calculated for each inventory. The same applies to the 95 % CI Precip. calculated as the difference between the 97.5 and 2.5 percentiles of the daily rainfall distribution. Max PGA and 1Std. PGA are, respectively the maximum and one standard deviation calculated from the Peak Ground Acceleration maps of the main shock and aftershock. Dist2Riv is the Euclidean distance from each 30 m pixel to the nearest streamline. PLC, PRC, and TWI are acronyms for Planar Curvature, Profile Curvature, and Topographic Wetness Index.

lows for nonlinearity in the model, and finally, a max pooling layer is added to reduce the spatial dimension of the feature space.

The decoder part consists of the U-Net structure, but unlike the conventional U-Net model, it produces an output scaled down by a factor of 8. The schematic design of the model is shown in Fig. 6. To understand the spatial dependence between the different inventory grids (1 km  $\times$  1 km grid), we have used a 4  $\times$  4 aggregation patch as input for the

susceptibility block, which is equivalent to 128  $\times$  128 input pixels. After receiving 128  $\times$  128 pixels, the convolution operation learns the contribution of physical properties such as earthquake and rainfall intensities as well as terrain characteristics to produce the susceptibility in a 4  $\times$  4  $\times$  1 batch of 1 km  $\times$  1 km grids. We stress here that we specifically chose to use a 32  $\times$  32 pixel structure per 1 km grid to convey all the possible information to the model and provide flexibility to the neural network to learn relevant information. As a result,



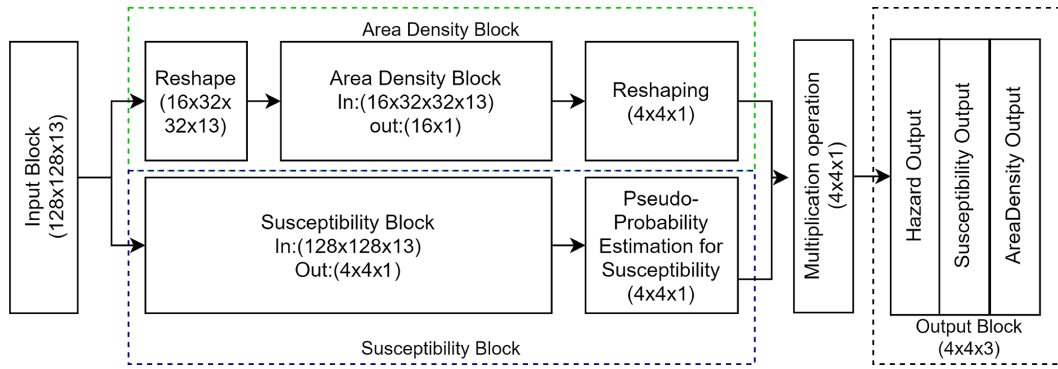


Figure 5. Designed landslide susceptibility and area density prediction model.

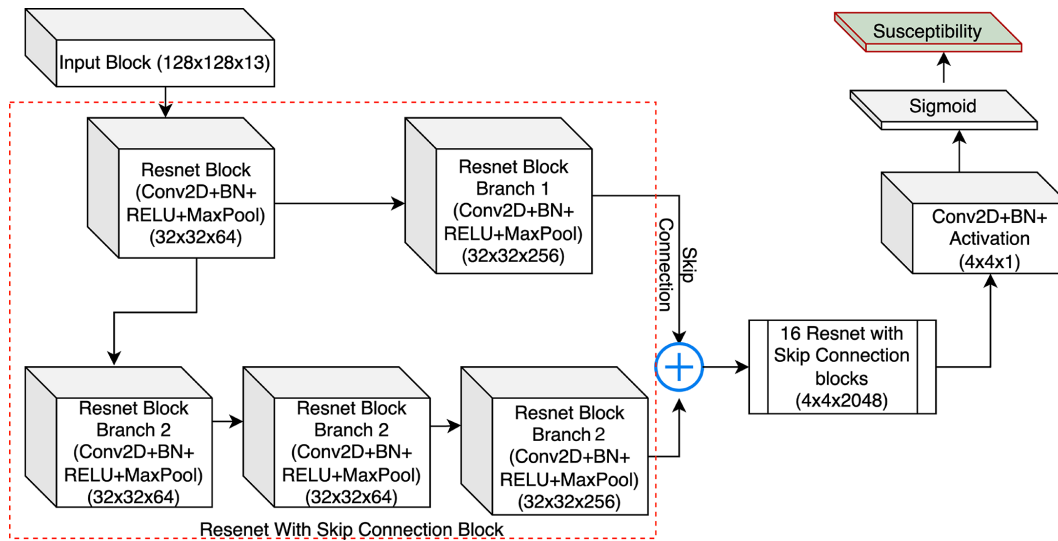


Figure 6. Susceptibility part of the model designed with a U-Net-like structure.

the model can extract the relevant information it needs from the distribution of  $32 \times 32$  pixels, rather than using arbitrary summary statistics such as the mean and standard deviation as per tradition in the geoscientific literature (e.g., Guzzetti et al., 2000; Lombardo and Tanyas, 2020). In other words, the model can learn by itself: (1) scanning  $32 \times 32$  pixel images corresponding to single 1 km grid cells and (2) matching the image characteristics to the landslide presence or absence labels.

The area density block also relies on a 1 km grid structure. However, we did not introduce the  $4 \times 4 \times 1$  neighborhood because the landslide presence or absence data present some spatial pattern beyond the extent of a 1 km grid. Conversely, the landslide area data do not present obvious spatial clusters of small or large densities.

Furthermore, it is also evident that landslides are discrete phenomena in space. This means that a large area density can be estimated at a 1 km grid, but its neighbor may not have suffered from slope failures (area density of 0). Conveying this “salt and pepper” spatial structure to the U-Net (via a

$4 \times 4$  neighboring window) tasked with regressing continuous data would negatively affect the model (unreported tests).

To address this issue, we reshaped the input data to a  $16 \times 32 \times 32 \times 13$  shape, where 16 inventory grids, each associated with 13 predictors of  $32 \times 32$  pixels, are present. The area density block is made of six dense sub-blocks, encompassing fully connected, batch normalization (Ioffe and Szegedy, 2015b) and dropout layers (Srivastava et al., 2014a). Before passing the data to the dense block, we added one Convolution block consisting of Convolution, Batch Normalization (Ioffe and Szegedy, 2015b) as well as Rectified Linear Unit and Max pooling (Wu and Gu, 2015) layers to extract the features from the input patches. Once both the area density and the susceptibility are estimated, the area density needs to be reshaped to match the data structure of the susceptibility component. To then generate landslide hazard estimates, as per the definition proposed by Guzzetti et al. (1999), we added a step where the pseudo-probability of landslide occurrence is multiplied by the landslide area density.

Notably, the developed model is spatio-temporal because it is built to explain the variability of the landslide hazard in space and time. However, the convolution layers used in this modeling approach are of a 2D nature. Therefore, the spatial structure in the data is handled via the 2D CNN, whereas the connection between subsequent spatial layers is not explicitly built in. The landslide hazard definition includes the concept of return periods to treat the temporal component, where for a given triggering situation of a given return period, the landslide hazard is estimated. The concept of return time could theoretically be included as part of the architecture presented here. However, the length of the multi-temporal inventory is relatively short. Thus, our model should not be considered suitable for long-term prediction, but rather, its validity is confined to the spatio-temporal domain under consideration or very close to it, both in space and in time (Wang et al., 2023). In other words, it should not be considered for generating long-term predictions under climate change scenarios.

## 6.2 Experimental setup

Binary classifiers are quite standard in machine or deep learning. Thus, for the susceptibility component, we opted for a focal Tversky loss function ( $\text{FTL}_c$ , see equation below for clarity), as Abraham and Khan (2018) have shown this measure to be particularly suited to imbalanced binary datasets such as ours. The major reason for choosing this loss function is the dominant absence of landslides in the dataset, complemented by many fewer cases ( $\approx 10\%$ ) where slope failures occurred. This may bias the model result and lead to a wrongly trained model if the loss function is not suitably implemented to handle this imbalance. The definition of Focal Tversky Loss can be denoted as:

$$\text{FTL}_c = \sum_c (1 - \text{TI}_c)^\gamma, \quad (1)$$

$$\text{TI}_c = \frac{\sum_{i=1}^N p_{ic} g_{ic} + \epsilon}{\sum_{i=1}^N p_{ic} g_{ic} + \alpha \sum_{i=1}^N p_{ic} g_{ic} + \beta \sum_{i=1}^N p_{i\bar{c}} g_{i\bar{c}} + \epsilon},$$

where  $\gamma$  is the focal parameter,  $p_{ic} \in [0, 1]$  is the probability that the pixel  $i$  is of the Landslide class  $c$ , and  $p_{i\bar{c}} \in [0, 1]$  is the probability that the pixel  $i$  is of the non-Landslide class  $\bar{c}$ . The  $g_{ic} \in 0, 1$  is the observed presence of landslide per pixel (i.e., the ground truth), and  $g_{i\bar{c}} \in 0, 1$  is the observed absence of landslide per pixel.  $\alpha$ , and  $\beta$  are the hyperparameters that can penalize false positives and false negatives and  $\epsilon$  value was set to  $10^{-7}$ . Furthermore,  $c$  represents the landslide presence class in the landslide classification problem; this could be represented by any positive integer in the case of multiclass classification problems. As for  $\text{FTL}_c$ , it represents the Focal Tversky Loss for binary classification, and  $\text{TI}_c$  is the Tversky Index.

To train the susceptibility component of the model, we trained a standard U-Net equipped with an early stopping functionality for a total of 500 epochs. The stopping criterion was set to detect overfit that may last for over ten epochs.

The overall data were then randomly split into training and testing sets to monitor the U-Net learning process.

As for the area density component, we opted for a loss function expressed in terms of mean absolute error (MAE, see Eq. 2 below for clarity), following the recommendations in Qi et al. (2020). The MAE is defined as:

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}, \quad (2)$$

where  $y_i$  is the observed area density and the  $\hat{y}_i$  is the predicted area density in the  $i$ th pixel, and  $n$  is the total number of samples in one batch.

To train the area density component, the imbalance in zeros and ones hindered the optimization process because the mean absolute error function did not perform well when most of the landslide densities were zeroes. This led to exploding gradients, returning all zeroes as the output. To solve this issue, we gradually increased the complexity of the task by subsampling the data and transforming the distribution of area density. The process is commonly known as curriculum learning (Wang et al., 2021), which lets the model learn a simple task at the start, and the process continues by gradually increasing the complexity of the subsequent tasks, each one linked to the previous one. To do so, we first removed all data points that contained zeros among the area density 1 km grids. We then log-transformed the target variable to convert the exponential-like distribution to a Gaussian-like distribution. Once the data were expressed according to a near-normal distribution, we trained the model for 200 epochs, including an early stopping criterion. The estimated parameters were set to initialize the subsequent steps. Specifically, with the initialization parameters available, we removed the logarithmic transformation and trained the model directly in the original landslide area density scale. This step was further run over 200 epochs, and the resulting parameters were fine-tuned to match the overall landslide area density distribution. In other words, we re-introduced the 1 km grids with zero density at this stage. Ultimately, the data were then randomly divided into 70% for calibration and 30% for validation.

The models were trained on a workstation with 160 GB Random Access Memory, a 32-core AMD Ryzen Threadripper PRO virtual CPU, and an NVIDIA RTX A4000 GPU. The computational resource used in this case made use of a shared infrastructure, with the entire training process taking 30%–40% load on the node and taking 35–40 h, depending on the available GPU memory. For the backpropagation, we used the Adam optimizer (Kingma and Ba, 2014), with an initial learning rate of  $10^{-3}$ , exponentially decreasing every 1000 steps of training. Because simultaneously training a model with two outputs based on a large and complex dataset would be extremely difficult to achieve, we opted to train the two elements separately in the beginning and combine their weights at the end of the learning process to generate a single model. This is then further trained for a few more steps to op-

timize the area density component for the nonlandslide grids. This approach is commonly known as ensemble modeling in the data-driven modeling context. The model is trained with a batch size of 16, and the training dataset is randomly divided into a 30 % validation set to check the model convergence at each epoch. The training process also featured an early stopping functionality where the model training would stop if the validation loss started to diverge. Simply put, the model weights were selected at the minimum validation loss to avoid overfitting. The number of convolution blocks, batch size, and initial learning rate were optimized through a hyperparameter tuning process, whereas other parameters were selected from the pre-existing models. For the convolution blocks, all the integer numbers of blocks were tested between 6 and 64, and we found that 18 was the best-suited number of blocks. For the batch size, four different batch sizes were tested (8, 16, 32, 64), and the fastest convergence was obtained through a batch size of 16. Moreover, learning rates from 1 to  $10^{-4}$  were tested by decreasing the learning rate by a factor of 10 each time, and we found the initial learning rate of  $10^{-3}$  to be the most stable.

### 6.3 Performance metrics

We used the following performance metrics for susceptibility and the area density components.

#### 6.3.1 Susceptibility component

To evaluate the model's performance during the training process and the inference, we used two common metrics, namely the F1 score (Meena et al., 2022; Nava et al., 2022) and the Intersection over Union (IOU) score (e.g., Huang et al., 2019; Ghorbanzadeh et al., 2020). We did not use binary accuracy because it is heavily influenced by data imbalance (Yeon et al., 2010; Li et al., 2022) and can produce high accuracy, even for poor classifications. The F1 score is calculated as:

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}},$$

$$\text{precision} = \frac{TP}{TP + FP}, \text{ recall} = \frac{TP}{TP + FN}, \quad (3)$$

where TP denotes the True Positives, FP denotes the False Positives, TN denotes the True Negatives, and FN denotes the False Negatives in the confusion matrix.

As for the IOU, this is another common metric for binary classifiers, computed as:

$$IOU = \frac{TP}{TP + FN + FP}. \quad (4)$$

We chose to use the IOU because it is a metric specifically dedicated to highlighting the accuracy in predicting the number of susceptible pixels and their location in a raster image (Monaco et al., 2020). Furthermore, to visualize how the model performs at different probability thresholds and what

the performance capacity of the model is, we also evaluated the Receiver Operating Characteristic (ROC, Fawcett, 2006) curve. This is generated at varying probability thresholds by computing pairs of True-Positive and False-Positive Rates. Moreover, we calculated the Area Under the ROC Curve (AUC) to evaluate the model's performance and to observe if the model overfits (Brenning, 2008; Brock et al., 2020).

#### 6.3.2 Area density component

To evaluate the trained model for the landslide area density, we opted to use the MAE (see Eq. 2) to monitor how the algorithm converges to its best solution, minimizing such a parameter. During the inference process, we also considered Pearson's  $R$  coefficient (Pearson, 1895), defined as:

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}, \quad (5)$$

where  $R$  denotes the correlation coefficient,  $x_i$  are the values of the  $x$ -variable in a sample,  $\bar{x}$  stands for mean of the values of the  $x$ -variable,  $y_i$  are the values of the  $y$ -variable in a sample, and  $\bar{y}$  stands for mean of the values of the  $y$ -variable.

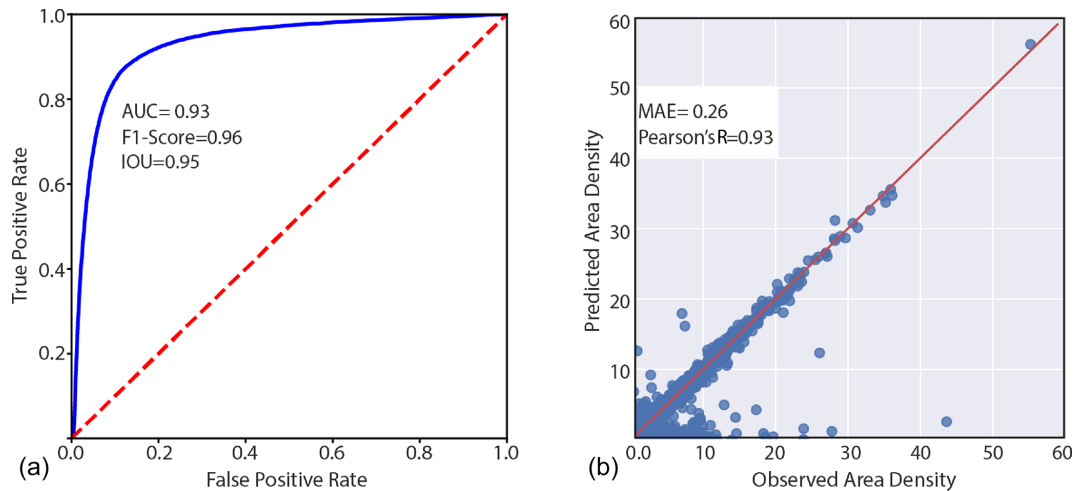
This parameter essentially provides the degree of correlation between two datasets, here taken as the observed and predicted landslide density per 1 km grid. A perfect model should have Pearson's  $R$ -value of 1, whereas two totally uncorrelated vectors would return a Pearson's  $R$ -value of 0.

## 7 Results

This section reports the model performance, initially from a purely numerical perspective. Later, we will translate this information back into maps and evaluate their temporal characteristics.

Figure 7 offers an overview of the performance that our ENN returned for its two components. The left panel reports an AUC of 0.93, associated with an F1 Score of 0.96 and IOU of 0.95. This predictive performance complies with the classification performance of outstanding data-driven models (Hosmer and Lemeshow, 2000). In the context of NNs, this is quite common because such architectures as much as other machine- or deep-learning tools and advanced statistical methods have proven to be able to reliably classify a landscape into landslide-prone or unstable slopes (e.g., Lombardo et al., 2019; Steger et al., 2021). Traditionally, the only missing element is that the vast majority of efforts so far have been spent solely in the context of pure spatial predictions, whereas the temporal dimension has been explored in a relatively smaller number of multivariate applications (Samia et al., 2017; Fang et al., 2023). Conversely, the performance of the area density component is far beyond the few analogous examples in the literature. So far, no spatially or temporally explicit model exists for landslide area





**Figure 7.** Summary of model's performance for the two components: landslide susceptibility in (a) and Area Density in (b) in the validation data.

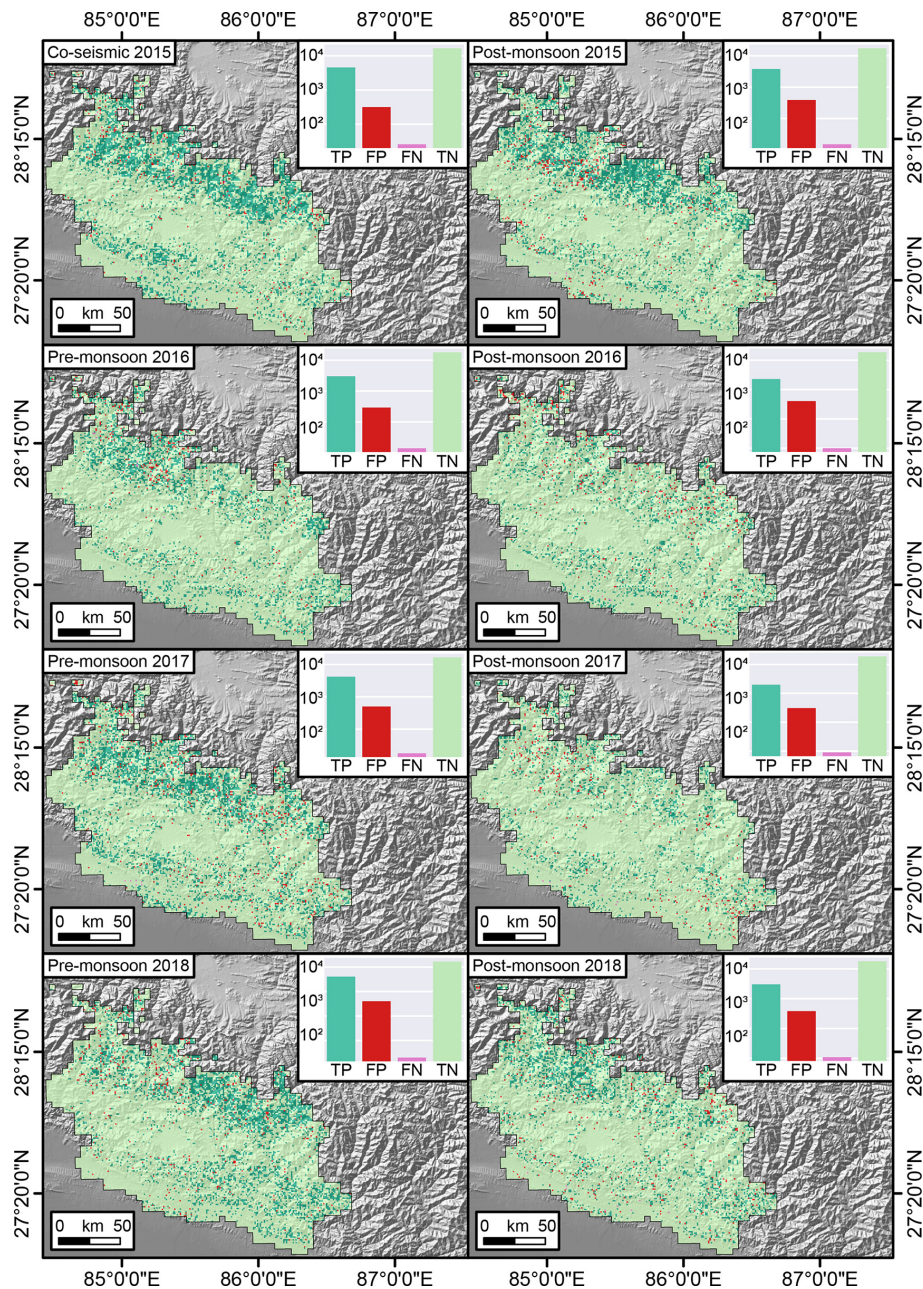
density. However, four recent articles have explored the capacity of predicting landslide areas (Lombardo et al., 2021; Aguilera et al., 2022; Bryce et al., 2022; Zapata et al., 2023). They all returned suitable predictive performance, but still far from the match seen in the second panel of Fig. 7, between observed and predicted landslide density in the out-of-sample (test) dataset. There, an outstanding alignment along the 45° line is clearly visible, together with a Pearson's  $R$  coefficient of 0.93 and a MAE of 0.26. It is important to stress that such metrics are calculated including the 1 km grids with zero landslide densities, i.e., the validation set in the study area as a whole. We also computed the same metrics exclusively at grid cells with a positive density, these resulting in a Pearson's  $R$  coefficient of 0.92 and a MAE of 0.24.

With a closer look, we can note a few exceptions, with some observations being strongly underestimated and very few cases being overestimated. This might be because we used MAE as the loss function. The MAE is a metric tailored toward the mean of a distribution, and therefore smaller values in the batch may be misrepresented without increasing the MAE. Moreover, the model is optimized by minimization of the MAE. Thus, it places more emphasis on suitably estimating large landslides and potentially underestimating the smaller ones. This problem could also have been influenced by the log-transformation step introduced at the beginning of the training process. This inevitably converted smaller values to very small ones, thus limiting their influence on the loss function even more. We are sharing this issue with the reader to provide the best description of a new modeling protocol. However, we should also mention that we consider such misrepresentation a minor problem. In fact, geoscience and risk science in general refer to the worst-case scenarios as the prediction target. Here, this is expressed by very large landslides, which appear to be correctly represented in most cases. Another worst-case scenario may be the combination

of large numbers of medium-sized landslides and their potential coalescing evolution. However, the bulk of the landslide density distribution is very well represented. This leaves most of the errors confined to the left tail (or very small values) of the landslide density distribution. These correspond to the phenomena from which one would expect the least potential threat or capacity to create damage, assuming a uniform vulnerability distribution.

These two plots offer a graphical overview of our ENN performance but they do not convey their signal in space and time. To offer a geographic and temporal overview of the same information, we opted to translate the match between observed and predicted values into maps, for both the susceptibility and the area density components. Figure 8 shows confusion maps (Titti et al., 2022; Prakash et al., 2021) where the distribution of TP, TN, FP, and FN is geographically presented for the co-seismic susceptibility as well as the following seven post-seismic scenarios. Across the whole sequence of maps, what stands out the most is that the TP and TN largely dominate the landscape, with few local exceptions. Notably, aside from the geographic translation of the confusion matrix, we reported the actual counts in logarithmic scale through the nested subpanels. There, the dominant number of TP and TN is confirmed once more and a better insight into the model misses is provided (FP and FN).

Figure 9 highlights the mismatch between observed and predicted landslide area densities. Most of the residuals are confined between  $-1$  and  $+1$  percent, with a negligible number of exceptions reaching an overestimation of  $-45\%$  and an underestimation of  $+15\%$ . Aside from these outliers, the most interesting element that stands out among these maps is the fact that the residuals do not exhibit any spatial pattern. They actually appear to be distributed randomly in both space and time.



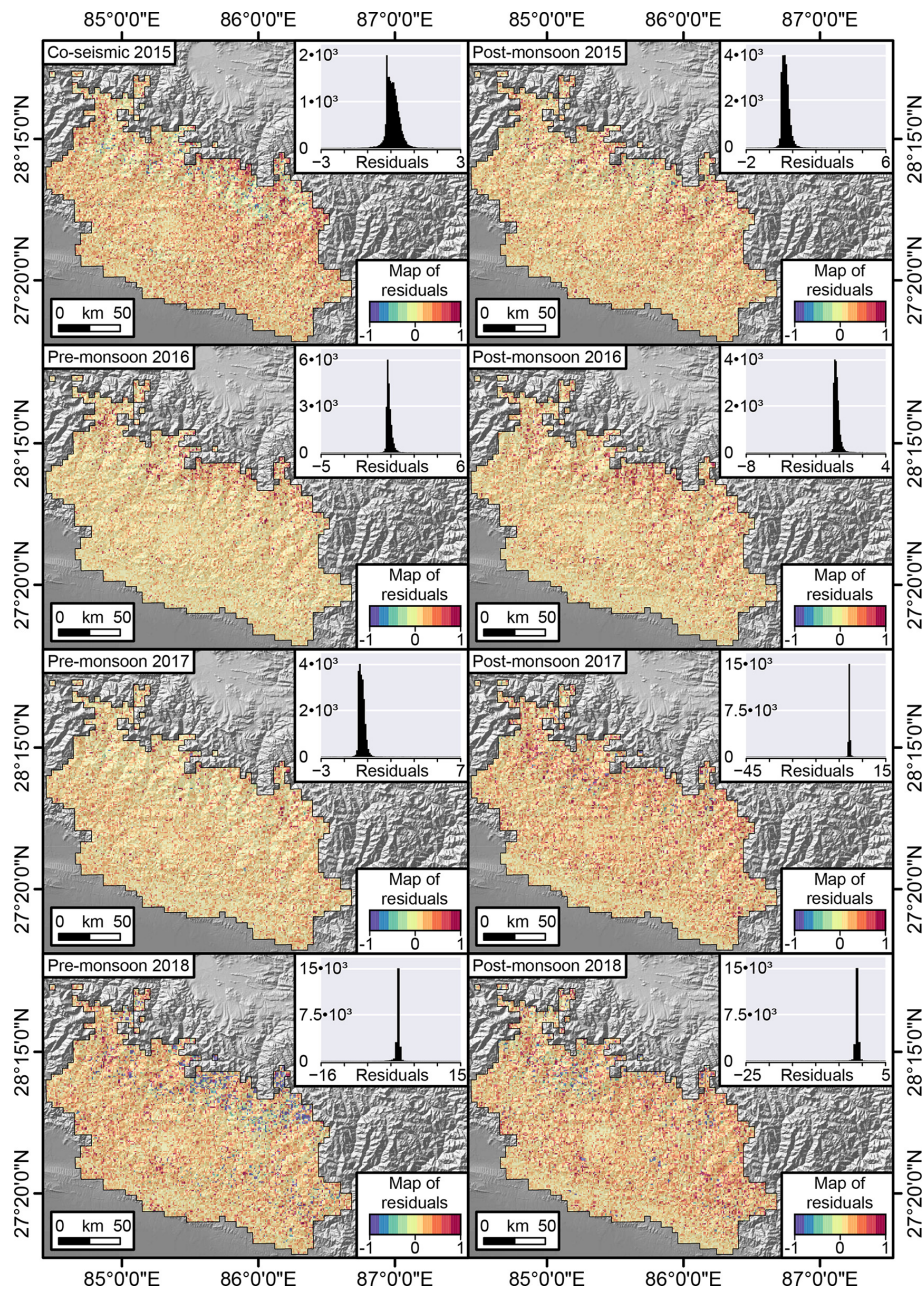
**Figure 8.** Confusion Maps offering a cartographic prediction of the performance for the susceptibility component. The TP, FP, FN, and TN are represented in the log scale.

Having stressed the predictive performance reached by our ENN, in Fig. 10, we finally offer a direct overview of the two outcomes (susceptibility and area density), as well as their product (hazard). Figure 10 reports the co-seismic case only and the post-monsoon estimates. We opted for this for reasons of practicality and visibility in a quite crowded subpanelled figure.

Reading these maps should be intuitive, but below we stress the assumptions behind the hazard one, being the first time such a map has ever been shown. The first column re-

ports the probabilities of landslide occurrence per 1 km grid. The second column shows the predicted landslide area density for the same 1 km lattice. The product of the two delivers an important element, where only coinciding high-susceptibility and high-density grids stand out. The rationale behind this is that large probability values of landslide occurrence will be inevitably canceled out whenever multiplied by low area density values. The same is valid in the opposite case. Large expected densities will be canceled out if multiplied by very low susceptibility values. Thus, the hazard





**Figure 9.** Maps displaying the pre- and post-monsoon residuals for the area density (expressed as percentages of failure in a pixel). The residuals are computed as observed landslide density minus the corresponding predicted values.

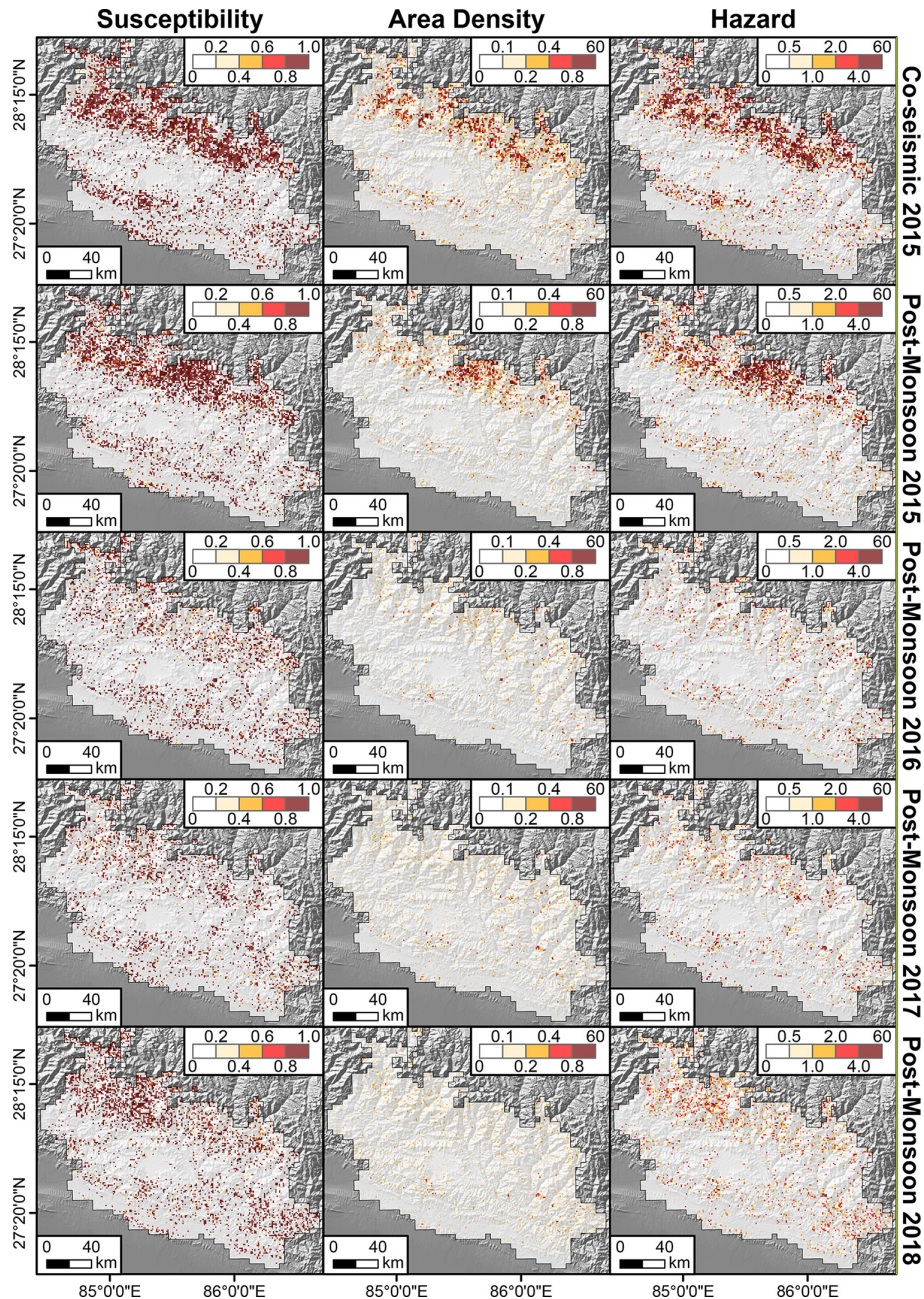
maps really do inform of the level of threat one may incur at certain 1 km grids and certain times, because a high hazard value implies that the mapping unit under consideration is not only expected to be unstable but the resulting instability is expected to lead to a large failure too.

The implications of the estimated patterns and considerations in terms of hazard will be further explored in Sect. 8. To support such discussions and highlight the link between susceptibility, hazard and their temporal evolution, we opted to

plot their signal via two-dimensional density plots, as shown in Fig. 11.

We can observe an interesting element, attributable to a concept known as earthquake legacy in the geoscientific literature. In fact, high landslide area density values associated with high susceptibility are quite well represented on the co-seismic panel and the first post-seismic one. However, as time passes, the density and proneness of the landscape appear to be estimated with lower landslide susceptibilities and area densities.





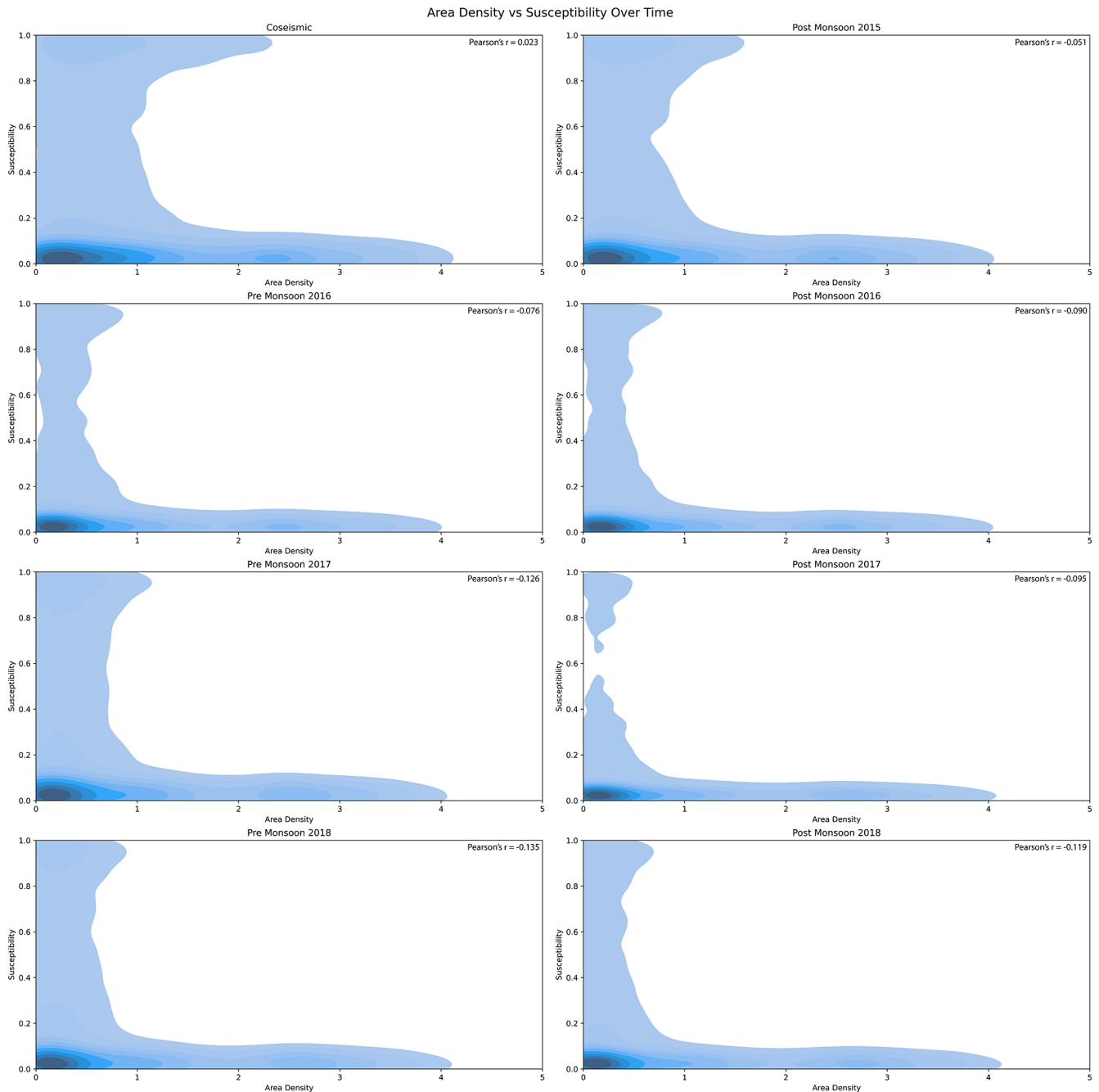
**Figure 10.** Predicted landslide susceptibility, area density, and hazard over time for co-seismic and post-monsoon seasons only, because this period had most of the landslides.

## 8 Discussion

In this section, we discuss the model’s performance, applicability, limitations, and necessary future developments in two subsections containing the supporting and opposing arguments.

### 8.1 Supporting arguments

The model results and the observations show that the deep-learning-based methods perform well in predicting landslide susceptibility and area density through a joint modeling approach. Such models can obviously provide much more information than modeling only susceptibility (Lombardo et al., 2021). Only using the susceptibility information is blind to landslide characteristics, such as how many landslides may manifest or how large they may become once



**Figure 11.** Contour plot of Area Density versus Susceptibility in different time periods showing how the area density and susceptibility are related to each other, where lighter color represents the lower density of the values and darker color represents the higher density of the values. Furthermore, it shows how in different periods after the earthquake the area density and susceptibility are distributed over space and how the range of susceptibility and area density changes.

they start moving downhill (Di Napoli et al., 2023). Thus, the combined information of which slope may be considered unstable and the expectation on the landslide can become an important source of information, not only for hazard assessment but even for risk reduction and management practitioners, once combined with the elements at risk.

Our ENN has shown the capacity to assess the two core elements, and interesting considerations can be made of its outcome. Figure 8 shows that each inventory mostly produced true positives and true negatives across the whole study site. More importantly, the number of false negatives was almost negligible. As for the false positives, their number is reason-

able and highlights locations where landslides have not manifested yet but may still occur in the future. As for the area component, Fig. 9 shows that the patterns of the residuals appear quite random in both space and time, thus fulfilling the ideal homoscedasticity requirements of a data-driven model. We can also stress that most of the residuals away from a few percentage points are confined toward negative values. This implies that our model overestimates the landslide area in a few isolated cases. However, similar to the point raised for the FP in Fig. 8, this outcome is to be expected. A negative residual indicates a location where the observed landslide area is lower than the predicted one. As most of the study site is characterized by grid cells where landslides did not occur, a negative residual points out locations that may not have exhibited landslides in the first place but whose geomorphological characteristics still indicate a likely release of a relatively larger unstable mass in the future.

Ultimately, Fig. 10 shows the constructive and destructive interference between the susceptibility and area density signals. This leads to isolating landslide-hazardous locations, which appear to be mostly located along the highest portions of the Himalayan range under consideration. There, a greater hazard is to be reasonably expected, for the higher relief is associated with higher gravitational potential and thus with a greater conversion into kinetic energy as the given landslide triggers, propagates, and finally halts.

An interesting by-product of our ENN can also be seen in Fig. 11. There, the high hazard levels estimated for the first two landslide inventories are shown to decay with time. Moreover, we can also observe that the susceptibility and area density do not necessarily correlate, meaning that the probability of landslide occurrence does not directly correlate with its size. This was also visible in the raw data shared by Kinsey et al. (2021). Such a decay supports the notion of earthquake legacy effects on landslide genetic processes (Ozturk, 2022), something still under debate in the geoscientific literature (Tanyaş et al., 2021b). Our output could bring additional information on this topic, supporting the scientific debate on landslide recovery (the time required for a given landscape to go back to pre-earthquake susceptibility conditions) by observing the predicted susceptibility change over time. Overall, multi-temporal landslide inventories and various associated parameters (e.g., number, size, area, or volume of landslides) have already been used to explore landslide recovery during post-seismic periods (e.g., Tanyaş et al., 2022). However, this has usually been done on a very generic and broad scale, usually leaving the slope scale out of the analytical process. Therefore, we see an added value in our model as it provides a comprehensive evaluation of landslide occurrences and their size. It is worth noting that examining landslide recovery is beyond the scope of this research. Yet, something worth sharing with the readers is that the decay we observe appears to have slowed down in 2017 and 2018, with a slight increase in the number of landslides, susceptibility, area density, and hazard. During those years, though, Kinsey

et al. (2021) could not regularly map landslides as they previously did. Thus, both pre-monsoon seasons in 2017 and 2018 were mapped on a longer time window than the authors did in previous years, inducing a slight temporal bias in the model.

Another element worth noting is that landslides across any given landscape are rare events. Thus, the number of presences will always be much smaller than the absences. This creates imbalanced data sets, which are often not ideally modeled in the deep-learning context (see, Johnson and Khoshgoftaar, 2019). In turn, imbalanced data sets limit the use of traditional metrics such as accuracy and the use of loss functions such as Binary Cross Entropy, because the latter will produce a high number of false negatives. We addressed this problem by adopting a Focal Tversky loss for the susceptibility component (Abraham and Khan, 2018). As for the area density component, we also faced some technical issues. Overall, around 85 % of the 1 km grids cells had a zero density value assigned to them (no landslides). In addition to this issue, the area density distribution is quite positively skewed, and regression tasks in deep learning have been mostly tested in the context of Gaussian or near-Gaussian distributions. To solve this problem, we had to split the modeling routine into a series of intermediate operations. First, we removed all zeros and used logarithmic transformation to shape the data according to a normal distribution. From this, we trained the first stage of our area density component. Once the model converged to its best solution in the log-density domain, we interrupted the training procedure, removed the log transformation, and further trained our model. This approach bypassed the need to implement even more complex NN architectures able to handle heavy-tailed distributions typical of extreme value theory (Weng et al., 2018).

An important factor to consider in a deep-learning-based modeling approach is the model overfitting and its generalization. Usually, large models can easily overfit and to avoid this, we have employed three major approaches. First, we used model regularization by adding batch normalization and dropout layers, according to standard deep-learning practices (Srivastava et al., 2014b; Ioffe and Szegedy, 2015b). Second, we included early stopping to halt the training process when divergence is observed in 30 % random validation data. Up to this point, generalization has not yet been checked. This is actually achieved by further testing over 50 % of the unseen data from which all the reported metrics are computed.

## 8.2 Opposing arguments

A major limitation we would like to recall relates to geological data availability. As discussed in the section 3, no detailed geological map covers the area where (Kinsey et al., 2021) mapped landslides, nor does the only available alternative provide enough geological classes to be meaningful in a data-driven context. In fact, the few available geological classes in Dahal (2012) present analogous proportions of landslide presences or absences and density data. This is



mainly because the few geological units necessarily cover very large extents owing to the coarse mapping resolution. In turn, this makes the available geology maps non-informative for the presented landslide modeling protocol. However, it is important to mention that a geological, or even better, a lithotechnical map, constitutes a precious layer of information for any engineering geological applications as they control landslide initiation and size. This is the reason why we consider our space–time ENN as it is currently valid only for the area we tested it for. And we would recommend re-adapting it in the case of new targets. For instance, another source of potentially relevant information could be the distance to active faults, which may be responsible for fracturing, fissuring, and rock strength degradation in general. This being said, our model still produced outstanding performance (Hosmer and Lemeshow, 2000) within the context of the Nepalese landscape under consideration, although its validity elsewhere still needs to be verified.

Another limiting factor we faced was the selected multi-temporal landslide inventory. First, a sequence of three years after the earthquake may not be enough to display the potential of space–time landslide hazard models fully. Second, owing to the resolution of satellite imagery used for mapping, some small landslides could have been omitted. To address the first issue, we have limited our predictor covariates to the temporal extent for which the mapping was carried out. This largely limits our model predictions to the time when ground truth data are available and verifiable. This model can capture the spatial and temporal variations in the landslide occurrence and size and, therefore, can be used to understand the landslide hazard for different return periods. However, we have not calculated the landslide hazard for different return periods specifically because we did not have a sufficiently long landslide occurrence sequence. This could be further improved in subsequent studies as this work mainly focuses on introducing our model as a novel methodological tool. The second limitation related to the potential omission of smaller landslides is much more difficult to address, mainly because, as universal functional approximators, the deep-learning models can only learn based on ground truth data. Therefore, this limitation cannot be removed. However, we recall that our model looks into a  $1\text{ km} \times 1\text{ km}$  grid for the area density and presence and absence of landslides. Therefore, the presence or absence of a landslide is not affected if a smaller landslide occurs together with larger and visible landslides, and it becomes a problem only when a small landslide occurs without a visible landslide within a  $1\text{ km} \times 1\text{ km}$  grid. In other words, the mapping units of the susceptibility component should be assigned with a presence value, even if in a  $1\text{ km}^2$  grid, a few small landslides may be missed. As for the area density, the effect on landslide area density may be more pronounced. However, as the landslide one may miss is small to very small in size, the effect may still be expressed within the uncertainty range of our model, potentially leading to minor omissions. This being said, outside the scope

of this paper, where this ENN architecture is introduced, we recommend future users to select an even more complete and long-term multitemporal inventory.

From the pure methodological perspective, even though the model produced outstanding results, there is still much room for improvement. As mentioned before, we addressed the heavy-tailed density distribution by using a log-transformation and L1 losses to measure the model convergence. In other words, we used a negative log-likelihood of a normal distribution to build our model, which in turn inherently assumes a normal distribution of the error. However, because the area density follows an extreme value distribution in its right tail, instead of a model built on a log-transformation and then re-trained on the original density scale, a more straightforward procedure would directly use the original data distribution and make use of performance metrics or losses that are suitable for the considered data. However, owing to a lack of mature research on existing methods of using extreme value theory with deep learning, we could not use such an approach. For further research, one of the possible approaches could be the integration of extreme value distributions (Davison and Huser, 2015) within our regression model. A similar procedure has been recently proposed to model wildfires (Richards et al., 2023; Cisneros et al., 2023).

Moreover, our model relies on a gridded partition of the geographic space under consideration. This lattice has two main elements that call for further improvements. The first is related to the size of the lattice itself. A  $1\text{ km}$  grid cell is quite far from the spatial partition required to support landslide-risk-reduction actions. Thus, the current model output can offer far richer information than the sole occurrence probabilities. However, to be actually useful for territorial management practices, the scale at which we trained should be probably downscaled at a finer resolution. The second element where our ENN can be further improved in terms of spatial structure has to do with the geomorphological significance of a lattice when used to model landslides. Such geomorphological processes, in fact, do not follow a regular gridded structure. In other words, when geoscientists go to the field, they do not see grids, whether they are a few centimeters or the  $1\text{ km}$  scale of our model. What a geomorphologist sees is a landscape partitioned into slopes. Slopes are also the same unit that geotechnical solutions aim to address. Thus, an improvement to our ENN could involve moving away from a gridded spatial partition and toward more geomorphology-oriented mapping units such as slope units (Alvioli et al., 2016; Tanyaş et al., 2022b), sub-catchments, or catchments (Shou and Lin, 2020; Wang et al., 2022). Moreover, the modeling approaches could also be further improved by the addition of landslide trigger classification (see. Rana et al., 2021), which could inform the model about which parameter (either PGA or rainfall) is responsible for causing landslides in this particular predictive context.

It is important to stress here that the structure of a Convolutional Neural Network mostly requires gridded input data. Thus, the extension toward irregular polygonal partitions such as the ones mentioned above would also require an adaptation of our ENN toward graph-based architectures (Scarselli et al., 2008).

Aside from the technical improvements that we already envision, a key problem we could not address is the lack of detailed spatio-temporal information on roadworks. Landscapes where roads are built may relapse through pronounced mass wasting (Tanyaş et al., 2022a). Nepal is known for building small roads without accounting for the required engineering solutions to maintain slope stability (McAdoo et al., 2018). For instance, Rosser et al. (2021) highlights that the elevated landslide susceptibility captured during post-seismic periods of the Gorkha earthquake could be partly associated with road construction projects. Thus, landslides are triggered on steep slopes owing to human interference, which we could not include in our model. During the very first phase of our model design, we actually tried to map those roads using freely available satellite images such as Sentinel 2 and PlanetScope. However, because the spatial resolution of those satellites is relatively coarse and the typical “self-made” roads are quite small (2–3 m in width), we could not automatize the road-mapping procedure to match our ENN spatio-temporal requirements. Therefore, rather than conveying wrong information to the model, we opted not to introduce road-network data to begin with. This is certainly a point to be improved in the future, not only for the Nepalese landscape but for any mountainous terrain where anthropogenic influence may bias the spatio-temporal distribution of landslides. Another parameter that could be added to inform our model about anthropogenic disturbances could be informal human modifications that have an influence on landslide triggering (Ozturk et al., 2022). Particularly, in this case, mountainous regions of Nepal do not have significant urban influences, and we opted not to include them. However, we recommend readers to include such features whenever necessary, depending on the study sites.

We stress that the vast majority of Neural Networks are tailored toward solving prediction tasks, and our ENN essentially offered the same outstanding performances reported in many other deep-learning applications. However, this architecture makes it very difficult to understand the causality behind the examined physical process. As our goal is to move toward a unified spatio-temporal hazard model, causality may not be a fundamental requirement at this stage. However, we envision future efforts to be directed toward more interpretable and causal machine or deep learning.

Ultimately, more can be done to clarify how our ENN should and should not be used, at least in its present form. For instance, with the current or even higher temporal frequency, our output could be used as part of parametric insurances (Horton, 2018) or large-scale risk reduction planning (Prabhakar et al., 2009). However, it is surely unsuitable for

infrastructure planning (Dhital, 2000) for the 1 km<sup>2</sup> resolution is far too coarse to be useful for detailed scale design.

## 9 Conclusions

We present a data-driven model capable of estimating where and when landslides may occur, as well as the expected landslide area density per mapping unit, a proxy for intensity. We achieved such a modeling task thanks to an Ensemble Neural Network architecture, a structure that has not yet found its expression within the geoscientific literature, making this model the first of its kind in landslide science. The implications of such a model can be groundbreaking because no data-driven model has provided an analogous level of information so far. The predictive ability of the model we propose still needs to be explored, isolating certain types of landslides, tectonic, climatic, and geomorphological settings. If similar performance is confirmed, then this can even open up a completely different toolbox for decision-makers to work with. So far, territorial management institutions have relied almost exclusively on susceptibility maps in the case of large regions and for long-term planning. The dependency on the concept of landslide susceptibility is also valid for regional and global organizations providing near-real-time or early-warning alerts for seismically or climatically triggered landslides. The model we propose can potentially link these two elements and provide a piece of even richer information, exploiting its predictive power away from the 6 month time resolution we tested here and more toward near-real-time or daily responses for various-scale applications.

*Code and data availability.* The data and code for the research can be accessed via <https://doi.org/10.5281/zenodo.10765925> (Dahal, 2024).

*Author contributions.* AD has implemented the model and run the whole analytical protocol. AD and LL have devised the research, produced scientific illustrations, and written the manuscript. HT has helped through the pre-processing phase and contributed to writing the manuscript. CvW, MvdM, PMM, and RH have commented on the manuscript to improve it before submission.

*Competing interests.* The contact author has declared that none of the authors has any competing interests.

*Disclaimer.* Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

*Acknowledgements.* We would like to thank two anonymous reviewers for their feedback and discussion. We would also like to thank the Center of Expertise in Big Geodata Science of the University of Twente for the support on computation environment.

*Financial support.* This research has been supported by the King Abdullah University of Science and Technology (grant no. URF/1/4338-01-01).

*Review statement.* This paper was edited by Filippo Catani and reviewed by two anonymous referees.

## References

- Abraham, N. and Khan, N. M.: A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation, *CoRR*, abs/1810.07842, arXiv [preprint], <https://doi.org/10.48550/arXiv.1810.07842>, 2018.
- Aguilera, Q., Lombardo, L., Tanyas, H., and Lipani, A.: On The Prediction of Landslide Occurrences and Sizes via Hierarchical Neural Networks, *Stoch. Env. Res. Risk A.*, 36, 2031–2048, 2022.
- Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., and Guzzetti, F.: Automatic delineation of geomorphological slope units with `r.slopeunits v1.0` and their optimization for landslide susceptibility modeling, *Geosci. Model Dev.*, 9, 3975–3991, <https://doi.org/10.5194/gmd-9-3975-2016>, 2016.
- Amit, S. N. K. B. and Aoki, Y.: Disaster detection from aerial imagery with convolutional neural network, in: 2017 international electronics symposium on knowledge creation and intelligent computing (IES-KCIC), Surabaya, Indonesia, 26–27 September, IEEE, 239–245, <https://doi.org/10.1109/KCIC.2017.8228593>, 2017.
- Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., and Reichenbach, P.: Impact of mapping errors on the reliability of landslide hazard maps, *Nat. Hazards Earth Syst. Sci.*, 2, 3–14, <https://doi.org/10.5194/nhess-2-3-2002>, 2002.
- Bout, B., Lombardo, L., van Westen, C., and Jetten, V.: Integration of two-phase solid fluid equations in a catchment model for flash-floods, debris flows and shallow slope failures, *Environ. Modell. Softw.*, 105, 1–16, <https://doi.org/10.1016/j.envsoft.2018.03.017>, 2018.
- Brenning, A.: Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models, *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie*, 19, 23–32, 2008.
- Brock, J., Schratz, P., Petschko, H., Muenchow, J., Micu, M., and Brenning, A.: The performance of landslide susceptibility models critically depends on the quality of digital elevation models, *Geomat. Nat. Haz. Risk*, 11, 1075–1092, 2020.
- Bryce, E., Lombardo, L., van Westen, C., Tanyas, H., and Castro-Camilo, D.: Unified landslide hazard assessment using hurdle models: a case study in the Island of Dominica, *Stoch. Env. Res. Risk A.*, 36, 2071–2084, 2022.
- Burton, A. and Bathurst, J.: Physically based modelling of shallow landslide sediment yield at a catchment scale, *Environ. Geol.*, 35, 89–99, 1998.
- Catani, F.: Landslide detection by deep learning of non-nadir and crowdsourced optical images, *Landslides*, 18, 1025–1044, 2021.
- Catani, F., Casagli, N., Ermini, L., Righini, G., and Menduni, G.: Landslide hazard and risk mapping at catchment scale in the Arno River basin, *Landslides*, 2, 329–342, 2005.
- Catani, F., Tofani, V., and Lagomarsino, D.: Spatial patterns of landslide dimension: a tool for magnitude mapping, *Geomorphology*, 273, 361–373, 2016.
- Cisneros, D., Richards, J., Dahal, A., Lombardo, L., and Huser, R.: Deep graphical regression for jointly moderate and extreme Australian wildfires, arXiv [preprint], arXiv:2308.14547, 2023.
- Clinton, B. D.: Light, temperature, and soil moisture responses to elevation, evergreen understory, and small canopy gaps in the southern Appalachians, *Forest Ecol. Manag.*, 186, 243–255, 2003.
- Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J.-P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J., and Smith, J. T.: Recommendations for the quantitative analysis of landslide risk, *B. Eng. Geol. Environ.*, 73, 209–263, 2014.
- Dahal, A., Castro-Cruz, D. A., Tanyaş, H., Fadel, I., Mai, P. M., van der Meijde, M., van Westen, C., Huser, R., and Lombardo, L.: From ground motion simulations to landslide occurrence prediction, *Geomorphology*, 441, 108898, 2023.
- Dahal, A.: ashokdahal/LandslideHazard: v1.0.0, Zenodo [data set and code], <https://doi.org/10.5281/zenodo.10765925>, 2024.
- Dahal, R. K.: Rainfall-induced landslides in Nepal, *International Journal of Erosion Control Engineering*, 5, 1–8, 2012.
- Davison, A. and Huser, R.: Statistics of Extremes, *Annu. Rev. Stat. Appl.*, 2, 203–235, <https://doi.org/10.1146/annurev-statistics-010814-020133>, 2015.
- Dhital, M. R.: An overview of landslide hazard mapping and rating systems in Nepal, *Journal of Nepal Geological Society*, 22, 533–538, 2000.
- Di Napoli, M., Tanyas, H., Castro-Camilo, D., Calcaterra, D., Cevasco, A., Di Martire, D., Pepe, G., Brandolini, P., and Lombardo, L.: On the estimation of landslide intensity, hazard and density via data-driven models, *Nat. Hazards*, 119, 1513–1530, 2023.
- Fan, X., Scaringi, G., Korup, O., West, A. J., van Westen, C. J., Tanyas, H., Hovius, N., Hales, T. C., Jibson, R. W., Allstadt, K. E., and Zhang, L.: Earthquake-Induced Chains of Geologic Hazards: Patterns, Mechanisms, and Impacts, *Rev. Geophys.*, 57, 421–503, 2019.
- Fang, Z., Wang, Y., van Westen, C., and Lombardo, L.: Space–Time Landslide Susceptibility Modeling Based on Data-Driven Methods, *Math. Geosci.*, 55, 1–20, 2023.
- Fawcett, T.: An introduction to ROC analysis, *Pattern Recogn. Lett.*, 27, 861–874, <https://doi.org/10.1016/j.patrec.2005.10.010>, 2006.
- Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., and Savage, W. Z.: Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning, *Eng. Geol.*, 102, 99–111, 2008.
- Formetta, G., Capparelli, G., and Versace, P.: Evaluating performance of simplified physically based models for shallow land-

- slide susceptibility, *Hydrol. Earth Syst. Sci.*, 20, 4585–4603, <https://doi.org/10.5194/hess-20-4585-2016>, 2016.
- Frattoni, P., Crosta, G., and Carrara, A.: Techniques for evaluating the performance of landslide susceptibility models, *Eng. Geol.*, 111, 62–72, <https://doi.org/10.1016/j.enggeo.2009.12.004>, 2010.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, *Sci. Data*, 2, 1–21, 2015.
- Ghorbanzadeh, O., Meena, S. R., Abadi, H. S. S., Piralilou, S. T., Zhiyong, L., and Blaschke, T.: Landslide Mapping Using Two Main Deep-Learning Convolution Neural Network Streams Combined by the Dempster–Shafer Model, *IEEE J. Sel. Top. Appl.*, 14, 452–463, 2020.
- Glenn, N. F., Streutker, D. R., Chadwick, D. J., Thackray, G. D., and Dorsch, S. J.: Analysis of LiDAR-derived topographic information for characterizing and differentiating landslide morphology and activity, *Geomorphology*, 73, 131–148, 2006.
- Glorot, X. and Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings, Sardinia, Italy, 13–15 May 2010, 249–256, <https://proceedings.mlr.press/v9/glorot10a.html> (last access: 2 March 2024), 2010.
- Gomez, H. and Kavzoglu, T.: Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela, *Eng. Geol.*, 78, 11–27, 2005.
- Grelle, G., Soriano, M., Revellino, P., Guerriero, L., Anderson, M., Diambra, A., Fiorillo, F., Esposito, L., Diodato, N., and Guadagno, F.: Space–time prediction of rainfall-induced shallow landslides through a combined probabilistic/deterministic approach, optimized for initial water table conditions, *Bull. Eng. Geol. Environ.*, 73, 877–890, 2014.
- Guzzetti, F., Carrara, A., Cardinali, M., and Reichenbach, P.: Landslide Hazard Evaluation: A Review of Current Techniques and Their Application in a Multi-Scale Study, Central Italy, *Geomorphology*, 31, 181–216, [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1), 1999.
- Guzzetti, F., Cardinali, M., Reichenbach, P., and Carrara, A.: Comparing Landslide Maps: A Case Study in the Upper Tiber River Basin, Central Italy, *Environ. Manage.*, 25, 247–263, <https://doi.org/10.1007/s002679910020>, 00000, 2000.
- He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, *CoRR*, abs/1512.03385, arXiv [preprint], <http://arxiv.org/abs/1512.03385> (last access: 2 March 2024), 2015.
- Heerdegen, R. G. and Beran, M. A.: Quantifying source areas through land surface curvature and shape, *J. Hydrol.*, 57, 359–373, 1982.
- Horton, J. B.: Parametric insurance as an alternative to liability for compensating climate harms, *Carbon & Climate Law Review*, 12, 285–296, 2018.
- Hosmer, D. W. and Lemeshow, S.: Applied Logistic Regression, 2nd ed. edn., Wiley, New York, <https://doi.org/10.1002/9781118548387>, 2000.
- Hough, S. E., Martin, S. S., Gahalaut, V., Joshi, A., Landes, M., and Bossu, R.: A comparison of observed and predicted ground motions from the 2015 Mw 7.8 Gorkha, Nepal, earthquake, *Nat. Hazards*, 84, 1661–1684, 2016.
- Huang, Y., Tang, Z., Chen, D., Su, K., and Chen, C.: Batching soft IoU for training semantic segmentation networks, *IEEE Signal Proc. Lett.*, 27, 66–70, 2019.
- Intrieri, E., Gigli, G., Mugnai, F., Fanti, R., and Casagli, N.: Design and implementation of a landslide early warning system, *Eng. Geol.*, 147, 124–136, 2012.
- Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International conference on machine learning, Lille, France, 7–9 July 2015, pmlr, 448–456, <https://doi.org/10.48550/arXiv.1502.03167>, 2015b.
- Johnson, J. M. and Khoshgoftaar, T. M.: Survey on deep learning with class imbalance, *Journal of Big Data*, 6, 1–54, 2019.
- Jones, J. N., Boulton, S. J., Bennett, G. L., Stokes, M., and Whitworth, M. R.: Temporal variations in landslide distributions following extreme events: Implications for landslide susceptibility modeling, *J. Geophys. Res.-Earth*, 126, e2021JF006067, <https://doi.org/10.1029/2021JF006067>, 2021.
- Ju, N., Huang, J., He, C., Van Asch, T., Huang, R., Fan, X., Xu, Q., Xiao, Y., and Wang, J.: Landslide early warning, case studies from Southwest China, *Eng. Geol.*, 279, 105917, <https://doi.org/10.1016/j.enggeo.2020.105917>, 2020.
- Kargel, J. S., Leonard, G. J., Shugar, D. H., Haritashya, U. K., Bevington, A., Fielding, E., Fujita, K., Geertsema, M., Miles, E., Steiner, J., and Anderson, E.: Geomorphic and geologic controls of geohazards induced by Nepal’s 2015 Gorkha earthquake, *Science*, 351, aac8353, <https://doi.org/10.1126/science.aac8353>, 2016.
- Kincey, M. E., Rosser, N. J., Robinson, T. R., Densmore, A. L., Shrestha, R., Pujara, D. S., Oven, K. J., Williams, J. G., and Swirad, Z. M.: Evolution of coseismic and post-seismic landsliding after the 2015 Mw 7.8 Gorkha earthquake, Nepal, *J. Geophys. Res.-Earth*, 126, e2020JF005803, <https://doi.org/10.1029/2020JF005803>, 2021.
- Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, arXiv [preprint], <https://doi.org/10.48550/ARXIV.1412.6980> (last access: 2 March 2024), 2014.
- Kirschbaum, D. and Stanley, T.: Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness, *Earths Future*, 6, 505–523, 2018.
- Lee, S., Ryu, J.-H., Won, J.-S., and Park, H.-J.: Determination and application of the weights for landslide susceptibility mapping using an artificial neural network, *Eng. Geol.*, 71, 289–302, 2004.
- Li, M., Zhang, X., Thrampoulidis, C., Chen, J., and Oymak, S.: AutoBalance: Optimized Loss Functions for Imbalanced Data, *CoRR*, abs/2201.01212, arXiv [preprint], <https://arxiv.org/abs/2201.01212> (last access: 2 March 2024), 2022.
- Loche, M., Scaringi, G., Yunus, A. P., Catani, F., Tanyaş, H., Frodella, W., Fan, X., and Lombardo, L.: Surface temperature controls the pattern of post-earthquake landslide activity, *Sci. Rep.-UK*, 12, 1–11, 2022.
- Lombardo, L. and Tanyaş, H.: Chrono-validation of near-real-time landslide susceptibility models via plug-in statistical simulations, *Eng. Geol.*, 278, 105818, <https://doi.org/10.1016/j.enggeo.2020.105818>, 2020.



- Lombardo, L., Bakka, H., Tanyas, H., van Westen, C., Mai, P. M., and Huser, R.: Geostatistical modeling to capture seismic-shaking patterns from earthquake-induced landslides, *J. Geophys. Res.-Earth*, 124, 1958–1980, <https://doi.org/10.1029/2019JF005056>, 2019.
- Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F., and Huser, R.: Space–time landslide predictive modelling, *Earth-Sci. Rev.*, 209, 103318, <https://doi.org/10.1016/j.earscirev.2020.103318>, 2020.
- Lombardo, L., Tanyas, H., Huser, R., Guzzetti, F., and Castro-Camilo, D.: Landslide size matters: A new data-driven, spatial prototype, *Eng. Geol.*, 293, 106288, <https://doi.org/10.1016/j.enggeo.2021.106288>, 2021.
- Maufroy, E., Cruz-Atienza, V. M., Cotton, F., and Gaffet, S.: Frequency-scaled curvature as a proxy for topographic site-effect amplification and ground-motion variability, *B. Seismol. Soc. Am.*, 105, 354–367, 2015.
- McAdoo, B. G., Quak, M., Gnyawali, K. R., Adhikari, B. R., Devkota, S., Rajbhandari, P. L., and Sudmeier-Rieux, K.: Roads and landslides in Nepal: how development affects environmental risk, *Nat. Hazards Earth Syst. Sci.*, 18, 3203–3210, <https://doi.org/10.5194/nhess-18-3203-2018>, 2018.
- Meena, S. R., Soares, L. P., Grohmann, C. H., van Westen, C., Bhuyan, K., Singh, R. P., Floris, M., and Catani, F.: Landslide detection in the Himalayas using machine learning algorithms and U-Net, *Landslides*, 19, 1209–1229, 2022.
- Monaco, S., Pasini, A., Apiletti, D., Colomba, L., Garza, P., and Baralis, E.: Improving wildfire severity classification of deep learning U-nets from satellite images, in: 2020 IEEE International Conference on Big Data (Big Data), Atlanta, Georgia, USA and Virtual Conference, 10–13 December 2020, IEEE, 5786–5788, <https://doi.org/10.1109/BigData50022.2020.9377867>, 2020.
- Montrasio, L., Valentino, R., Corina, A., Rossi, L., and Rudari, R.: A prototype system for space–time assessment of rainfall-induced shallow landslides in Italy, *Nat. Hazards*, 74, 1263–1290, 2014.
- Nava, L., Bhuyan, K., Meena, S. R., Monserrat, O., and Catani, F.: Rapid Mapping of Landslides on SAR Data by Attention U-Net, *Remote Sens.-Basel*, 14, 1449, <https://doi.org/10.3390/rs14061449>, 2022.
- Neaupane, K. M. and Achet, S. H.: Use of backpropagation neural network for landslide monitoring: a case study in the higher Himalaya, *Eng. Geol.*, 74, 213–226, 2004.
- Neaupane, K. M. and Piantanakulchai, M.: Analytic network process model for landslide hazard zonation, *Eng. Geol.*, 85, 281–294, 2006.
- Nocentini, N., Rosi, A., Segoni, S., and Fanti, R.: Towards landslide space–time forecasting through machine learning: the influence of rainfall parameters and model setting, *Front. Earth Sci.*, 11, 1152130, <https://doi.org/10.3389/feart.2023.1152130>, 2023.
- Nowicki Jessee, M., Hamburger, M., Allstadt, K., Wald, D., Roberson, S., Tanyas, H., Hearne, M., and Thompson, E.: A Global Empirical Model for Near-Real-Time Assessment of Seismically Induced Landslides, *J. Geophys. Res.-Earth*, 123, 1835–1859, 2018.
- Ohlmacher, G. C.: Plan curvature and landslide probability in regions dominated by earth flows and earth slides, *Eng. Geol.*, 91, 117–134, <https://doi.org/10.1016/j.enggeo.2007.01.005>, 2007.
- Ozturk, U.: Geohazards explained 10: Time-dependent landslide susceptibility, *Geology Today*, 38, 117–120, 2022.
- Ozturk, U., Pittore, M., Behling, R., Roessner, S., Andreani, L., and Korup, O.: How robust are landslide susceptibility estimates?, *Landslides*, 18, 681–695, 2021.
- Ozturk, U., Bozzolan, E., Holcombe, E. A., Shukla, R., Pianosi, F., and Wagener, T.: How climate change and unplanned urban sprawl bring more landslides, *Nature*, 608, 262–265, <https://doi.org/10.1038/d41586-022-02141-9>, 2022.
- Pearson, K.: Note on Regression and Inheritance in the Case of Two Parents, *P. R. Soc. London*, 58, 240–242, <http://www.jstor.org/stable/115794> (last access: 2 March 2024), 1895.
- Prabhakar, S., Srinivasan, A., and Shaw, R.: Climate change and local level disaster risk reduction planning: need, opportunities and challenges, *Mitig. Adapt. Strat. Gl.*, 14, 7–33, 2009.
- Prakash, N., Manconi, A., and Loew, S.: A new strategy to map landslides with a generalized convolutional neural network, *Sci. Rep.-UK*, 11, 9722, <https://doi.org/10.1038/s41598-021-89015-8>, 2021.
- Qi, J., Du, J., Siniscalchi, S. M., Ma, X., and Lee, C.-H.: On mean absolute error for deep neural network based vector-to-vector regression, *IEEE Signal Proc. Lett.*, 27, 1485–1489, 2020.
- Rana, K., Ozturk, U., and Malik, N.: Landslide geometry reveals its trigger, *Geophys. Res. Lett.*, 48, e2020GL090848, <https://doi.org/10.1029/2020GL090848>, 2021.
- Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F.: A Review of Statistically-Based Landslide Susceptibility Models, *Earth-Sci. Rev.*, 180, 60–91, <https://doi.org/10.1016/j.earscirev.2018.03.001>, 2018.
- Richards, J., Huser, R., Bevacqua, E., and Zscheischler, J.: Insights into the drivers and spatiotemporal trends of extreme mediterranean wildfires with statistical deep learning, *Artificial Intelligence for the Earth Systems*, 2, e220095, <https://doi.org/10.1175/AIES-D-22-0095.1>, 2023.
- Roback, K., Clark, M. K., West, A. J., Zekkos, D., Li, G., Gallen, S. F., Chamlagain, D., and Godt, J. W.: The size, distribution, and mobility of landslides caused by the 2015 Mw7.8 Gorkha earthquake, Nepal, *Geomorphology*, 301, 121–138, 2018.
- Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, *CoRR*, abs/1505.04597, arXiv [preprint], <http://arxiv.org/abs/1505.04597> (last access: 2 March 2024), 2015.
- Rosser, N., Kincey, M., Oven, K., Densmore, A., Robinson, T., Pujara, D. S., Shrestha, R., Smutny, J., Gurung, K., Lama, S., and Dhital, M. R.: Changing significance of landslide Hazard and risk after the 2015 Mw 7.8 Gorkha, Nepal Earthquake, *Progress in Disaster Science*, 10, 100159, <https://doi.org/10.1016/j.pdisas.2021.100159>, 2021.
- Samia, J., Temme, A. J., Bregt, A., Wallinga, J., Fausto Guzzetti, Ardizzone, F., and Rossi, M.: Characterization and Quantification of Path Dependency in Landslide Susceptibility, *Geomorphology*, 292, 16–24, <https://doi.org/10.1016/j.geomorph.2017.04.039>, 2017.
- Samia, J., Temme, A., Bregt, A., Wallinga, J., Guzzetti, F., and Ardizzone, F.: Dynamic path-dependent landslide susceptibility modelling, *Nat. Hazards Earth Syst. Sci.*, 20, 271–285, <https://doi.org/10.5194/nhess-20-271-2020>, 2020.

- Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G.: The graph neural network model, *IEEE T. Neural Networ.*, 20, 61–80, 2008.
- Schlögel, R., Marchesini, I., Alvioli, M., Reichenbach, P., Rossi, M., and Malet, J.-P.: Optimizing landslide susceptibility zonation: Effects of DEM spatial resolution and slope unit delineation on logistic regression models, *Geomorphology*, 301, 10–20, <https://doi.org/10.1016/j.geomorph.2017.10.018>, 2018.
- Shou, K.-J. and Lin, J.-F.: Evaluation of the extreme rainfall predictions and their impact on landslide susceptibility in a sub-catchment scale, *Eng. Geol.*, 265, 105434, <https://doi.org/10.1016/j.enggeo.2019.105434>, 2020.
- Sørensen, R., Zinko, U., and Seibert, J.: On the calculation of the topographic wetness index: evaluation of different methods based on field observations, *Hydrol. Earth Syst. Sci.*, 10, 101–112, <https://doi.org/10.5194/hess-10-101-2006>, 2006.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *J. Mach. Learn. Res.*, 15, 1929–1958, <http://jmlr.org/papers/v15/srivastava14a.html> (last access: 2 March 2024), 2014a.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.*, 15, 1929–1958, 2014b.
- Steger, S., Brenning, A., Bell, R., and Glade, T.: The propagation of inventory-based positional errors into statistical landslide susceptibility models, *Nat. Hazards Earth Syst. Sci.*, 16, 2729–2745, <https://doi.org/10.5194/nhess-16-2729-2016>, 2016.
- Steger, S., Mair, V., Kofler, C., Pittore, M., Zebisch, M., and Schneiderbauer, S.: Correlation does not imply geomorphic causation in data-driven landslide susceptibility modelling—Benefits of exploring landslide data collection effects, *Sci. Total Environ.*, 776, 145935, <https://doi.org/10.1016/j.scitotenv.2021.145935>, 2021.
- Tanyaş, H., Allstadt, K. E., and van Westen, C. J.: An updated method for estimating landslide-event magnitude, *Earth Surf. Proc. Land.*, 43, 1836–1847, <https://doi.org/10.1002/esp.4359>, 2018.
- Tanyaş, H., Hill, K., Mahoney, L., Fadel, I., and Lombardo, L.: The world's second-largest, recorded landslide event: Lessons learnt from the landslides triggered during and after the 2018 Mw 7.5 Papua New Guinea earthquake, *Eng. Geol.*, 297, 106504, <https://doi.org/10.1016/j.enggeo.2021.106504>, 2022.
- Tanyaş, H., Kirschbaum, D., Görüm, T., van Westen, C. J., Tang, C., and Lombardo, L.: A closer look at factors governing landslide recovery time in post-seismic periods, *Geomorphology*, 391, 107912, <https://doi.org/10.1016/j.geomorph.2021.107912>, 2021a.
- Tanyaş, H., Kirschbaum, D., and Lombardo, L.: Capturing the footprints of ground motion in the spatial distribution of rainfall-induced landslides, *B. Eng. Geol. Environ.*, 80, 4323–4345, 2021b.
- Tanyaş, H., Görüm, T., Kirschbaum, D., and Lombardo, L.: Could road constructions be more hazardous than an earthquake in terms of mass movement?, *Nat. Hazards*, 112, 639–663, <https://doi.org/10.1007/s11069-021-05199-2>, 2022a.
- Tanyaş, H., Hill, K., Mahoney, L., Fadel, I., and Lombardo, L.: The world's second-largest, recorded landslide event: Lessons learnt from the landslides triggered during and after the 2018 Mw 7.5 Papua New Guinea earthquake, *Eng. Geol.*, 297, 106504, 2022b.
- Taylor, D. W.: *Fundamentals of Soil Mechanics*, John Wiley & Sons, Wisconsin, USA, ISBN 978-1258768928, 1948.
- Titti, G., van Westen, C., Borgatti, L., Pasuto, A., and Lombardo, L.: When Enough Is Really Enough? On the Minimum Number of Landslides to Build Reliable Susceptibility Models, *Geosciences*, 11, 469, <https://doi.org/10.3390/geosciences11110469>, 2021.
- Titti, G., Sarretta, A., Lombardo, L., Crema, S., Pasuto, A., and Borgatti, L.: Mapping susceptibility with open-source tools: a new plugin for QGIS, *Front. Earth Sci.*, 229, 842425, <https://doi.org/10.3389/feart.2022.842425>, 2022.
- Upreti, B. N.: The Physiographic and Geology of Nepal and Their Bearing on the Landslide Problem, in: *Landslide Hazard Mitigation in the Hindu Kush-Himalaya*, edited by: Upreti, B. N., Tianchi, L., and Chalise, S. R. Kathmandu, International Centre for Integrated Mountain Development, 31–49, <https://doi.org/10.53055/ICIMOD.374>, 2001.
- van den Bout, B., Lombardo, L., Chiyang, M., van Westen, C., and Jetten, V.: Physically-based catchment-scale prediction of slope failure volume and geometry, *Eng. Geol.*, 284, 105942, <https://doi.org/10.1016/j.enggeo.2020.105942>, 2021a.
- van den Bout, B., van Asch, T., Hu, W., Tang, C. X., Mavrouli, O., Jetten, V. G., and van Westen, C. J.: Towards a model for structured mass movements: the OpenLISEM hazard model 2.0a, *Geosci. Model Dev.*, 14, 1841–1864, <https://doi.org/10.5194/gmd-14-1841-2021>, 2021b.
- Van Westen, C., Van Asch, T. W., and Soeters, R.: Landslide hazard and risk zonation—why is it still so difficult?, *B. Eng. Geol. Environ.*, 65, 167–184, 2006.
- Van Westen, C. J., Castellanos, E., and Kuriakose, S. L.: Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview, *Eng. Geol.*, 102, 112–131, 2008.
- Vasiliev, I. R.: Visualization of spatial dependence: an elementary view of spatial autocorrelation, in: *Practical handbook of spatial statistics*, edited by: Arlinghaus, S., CRC Press, Boca Raton, 17–30, <https://doi.org/10.1201/9781003067689>, 2020.
- von Specht, S., Ozturk, U., Veh, G., Cotton, F., and Korup, O.: Effects of finite source rupture on landslide triggering: the 2016  $M_w$  7.1 Kumamoto earthquake, *Solid Earth*, 10, 463–486, <https://doi.org/10.5194/se-10-463-2019>, 2019.
- Wang, H., Xu, W., and Xu, R.: Slope stability evaluation using back propagation neural networks, *Eng. Geol.*, 80, 302–315, 2005.
- Wang, N., Cheng, W., Marconcini, M., Bachofer, F., Liu, C., Xiong, J., and Lombardo, L.: Space–time susceptibility modeling of hydro-morphological processes at the Chinese national scale, *Eng. Geol.*, 301, 106586, <https://doi.org/10.1016/j.enggeo.2022.106586>, 2022.
- Wang, T., Dahal, A., Fang, Z., van Westen, C., Yin, K., and Lombardo, L.: From spatio-temporal landslide susceptibility to landslide risk forecast, *Geosci. Front.*, 15, 101765, <https://doi.org/10.1016/j.gsf.2023.101765>, 2023.
- Wang, X., Chen, Y., and Zhu, W.: A survey on curriculum learning, *IEEE T. Pattern Anal.*, 44, 4555–4576, 2021.
- Weng, T.-W., Zhang, H., Chen, P.-Y., Yi, J., Su, D., Gao, Y., Hsieh, C.-J., and Daniel, L.: Evaluating the robustness of neural networks: An extreme value theory approach, *arXiv [preprint]*, arXiv:1801.10578, 2018.

- Whiteley, J., Chambers, J., Uhlemann, S., Wilkinson, P. B., and Kendall, J.: Geophysical monitoring of moisture-induced landslides: a review, *Rev. Geophys.*, 57, 106–145, 2019.
- Worden, C. and Wald, D.: ShakeMap manual online: Technical manual, user's guide, and software guide, US Geol. Surv., Reston, Virginia, USA, 1–156, <https://doi.org/10.3133/tm12A1>, 2016.
- Wu, H. and Gu, X.: Max-Pooling Dropout for Regularization of Convolutional Neural Networks, CoRR, abs/1512.01400, arXiv [preprint], <http://arxiv.org/abs/1512.01400> (last access: 2 March 2024), 2015.
- Yeon, Y.-K., Han, J.-G., and Ryu, K. H.: Landslide susceptibility mapping in Injae, Korea, using a decision tree, *Eng. Geol.*, 116, 274–283, 2010.
- Yesilnacar, E. and Topal, T.: Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey), *Eng. Geol.*, 79, 251–266, 2005.
- Zapata, M. M., Steger, S., Tanyas, H., and Lombardo, L.: Modeling the area of co-seismic landslides via data-driven models: The Kaikōura example, *Eng. Geol.*, 320, 107121, <https://doi.org/10.1016/j.enggeo.2023.107121>, 2023.
- Zevenbergen, L. W. and Thorne, C. R.: Quantitative analysis of land surface topography, *Earth Surf. Proc. Land.*, 12, 47–56, 1987.
- Zhang, Y., Chen, G., Zheng, L., Li, Y., and Wu, J.: Effects of near-fault seismic loadings on run-out of large-scale landslide: a case study, *Eng. Geol.*, 166, 216–236, 2013.