Explainable artificial intelligence in geoscience: a glimpse into the future of landslide susceptibility modeling

Ashok Dahal[1] and Luigi Lombardo[2]

[1]University of Twente, Faculty of Geoinformation Science and Earth Observation
[2]University of Twente

November 26, 2022

Abstract

For decades, the distinction between statistical models and machine learning ones has been clear. The former are optimized to produce interpretable results, whereas the latter seeks to maximize the predictive performance of the task at hand. This is valid for any scientific field and for any method belonging to the two categories mentioned above. When attempting to predict natural hazards, this difference has lead researchers to make drastic decisions on which aspect to prioritize, a difficult choice to make. In fact, one would always seek the highest performance because at higher performances correspond better decisions for disaster risk reduction. However, scientists also wish to understand the results, as a way to rely on the tool they developed. Today, very recent development in deep learning have brought forward a new generation of interpretable artificial intelligence, where the prediction power typical of machine learning tools is equipped with a level of explanatory power typical of statistical approaches. In this work, we attempt to demonstrate the capabilities of this new generation of explainable artificial intelligence (ExAI). To do so, we take the landslide susceptibility context as reference. Specifically, we build an ExAI trained to model landslides occurred in response to the Gorkha earthquake (25 April 2015), providing an educational overview of the model design and its querying opportunities. The results are surprising, the performance are extremely high, while the interpretability can be extended to the probabilistic result assigned to single mapping units. This is also showcased in a web-GIS (\textcolor{blue}{https://arcg.is/0unziD}) platform we built.

Explainable artificial intelligence in geoscience: a glimpse into the future of landslide susceptibility modeling

Ashok Dahal[1], **Luigi Lombardo**[1]

[1]University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), PO Box 217, Enschede, AE 7500, Netherlands

Key Points:

- A new generation of interpretable machine learning models is tested and presented to predict landslide occurrences.
- The traditional definition of black box is left in favor of tools that can be queried to understand the artificially intelligent decision.
- A web-GIS platform has also been developed to showcase the potential of explainable artificial intelligence for geoscientific applications.

Corresponding author: Ashok Dahal, a.dahal@utwente.nl

Abstract

For decades, the distinction between statistical models and machine learning ones has been clear. The former are optimized to produce interpretable results, whereas the latter seeks to maximize the predictive performance of the task at hand. This is valid for any scientific field and for any method belonging to the two categories mentioned above. When attempting to predict natural hazards, this difference has lead researchers to make drastic decisions on which aspect to prioritize, a difficult choice to make. In fact, one would always seek the highest performance because at higher performances correspond better decisions for disaster risk reduction. However, scientists also wish to understand the results, as a way to rely on the tool they developed. Today, very recent development in deep learning have brought forward a new generation of interpretable artificial intelligence, where the prediction power typical of machine learning tools is equipped with a level of explanatory power typical of statistical approaches.

In this work, we attempt to demonstrate the capabilities of this new generation of explainable artificial intelligence (ExAI). To do so, we take the landslide susceptibility context as reference. Specifically, we build an ExAI trained to model landslides occurred in response to the Gorkha earthquake (25 April 2015), providing an educational overview of the model design and its querying opportunities. The results are surprising, the performance are extremely high, while the interpretability can be extended to the probabilistic result assigned to single mapping units. This is also showcased in a web-GIS (https://arcg.is/0unziD) platform we built.

1 Introduction

The evolution of science is marked by historical moments where discoveries or technological advancements opened up opportunities that were not there before. The history of geoscience and specifically the part of it linked to natural hazards is no different. Specifically, if we take the landslide example, before 1970's no available study attempted to estimate locations where landslides were likely to occur over a large landscape. This notion was later defined as landslide susceptibility (Reichenbach et al., 2018) and its first example dates back to Brabb et al. (1972), with a digital scan of his susceptibility map still being accessible at this link (https://pubs.usgs.gov/mf/0360/plate-1.pdf). The introduction of that document had effect that rippled even to present days. Specifically, it set the stage for a successful branch of geomorphology that has received wide attention and efforts since then. One of the main issues that document had was the fact that it relied on expert-based opinions. In other words, the definition of susceptibility classes was the result of a subjective decision. Few years later though, the introduction of Geographic Information Systems (GIS; Gates & Heil, 1980) laid the foundations to collect digital cartographic data and implement numerical operations. As a result, the geomorphological community was able to test data-driven approaches suitable to move past the subjectivity issue. This later led to the first introduction of bivariate statistical models (Naranjo et al., 1994; Soeters & Van Westen, 1994) and their multivariate extension (P. Atkinson et al., 1998; P. M. Atkinson & Massari, 1998). The latter still constitute the most common method to estimate landslide susceptibility (Reichenbach et al., 2018). Their success is due to the satisfying performance they demonstrated through the years and their high level of interpretability. The way they work is to assume a vector of landslide presence/absence data to behave across the geographic space according to a Bernoulli probability distribution, whose relation to the landslide is linearly related to a set of covariates. The latter are usually referred to as predisposing or triggering factors (Das et al., 2012; Tanyaş et al., 2022). However, the linearity assumption these models are based on, limited the performance one could obtain. Therefore, another moment of particular importance was the introduction of machine learning tools (e.g., Yesilnacar & Topal, 2005). Even the simplest of them allowed for linear combinations of nonlinear relations, providing good flexibility and performance. This is the main reason why a multitude of scientific contributions got published since then, testing each one of these new approaches, from neural network (Melchiorre et al., 2008), to decision trees (Li & Claramunt, 2006) and their subsequent stochastic versions (Vorpahl et al., 2012; Catani et al., 2013), and from support vector machines (Ballabio & Sterlacchini, 2012)'to multivariate adaptive regression splines (Marmion et al., 2009). All these newly introduced methods though, lacked in interpretability, which is why conventional statistical models still kept on being the most common modeling choice. Moreover, even statistical models received a boost in their allowed complexity, as contributions based on generalized additive models began to flourish (Brenning, 2008; J. N. Goetz et al., 2011). Since, then the two applications reached a sort of stalemate, where machine learning tool where sought for performance and statistical ones for interpretation. This is reflected even nowadays, after a decade, through the number of comparative studies, where the results of one or the other option are constantly tested to discern advantages and disadvantages (Pourghasemi & Rossi, 2017; J. Goetz et al., 2015).

The very same period has also witnessed improvements in the choice of the geographic object to partition an area under study, with unique condition units (Bednarik & Pauditš, 2010; Titti et al., 2021), slope units (Carrara, 1983; Alvioli et al., 2016) and grid cells (Dhakal et al., 2000; P. M. Atkinson & Massari, 1998) becoming the most common choices, in ascending order (Reichenbach et al., 2018).

Notably, the recent introduction of deep learning architectures has further set apart the statistical and data mining applications for landslide prediction. The classification performance of tools such as Convolutional Neural Network (Yi et al., 2020) have been shown to be even higher than their traditional machine learning counterparts (Bui et al., 2020; Fang et al., 2020) attracting the attention of a large part of the community although this is still achieved at the expense of interpretation capacity. And, their use further supported the grid cell partition because convolutional operations are commonly based on a lattice structure (Van Dao et al., 2020), with the exception of few deep learning studies adopting slope units (Hua et al., 2021).

In this complex system though, a new moment will soon mark the evolution that landslide susceptibility models have undergone since Brabb et al. (1972). Information science has put forward a huge effort to give machine learning tools the same interpretation capacity of statistically based models (Štrumbelj & Kononenko, 2014; Ribeiro et al., 2016). This has recently resulted in the seminal work of Lundberg and Lee (2017). Specifically, the authors have built the first artificial intelligence that can be queried on an element by element basis as well as a predictor by predictor basis. In other words, their model can be dissected to the level of each components it has been built upon and the results can be examined to the point of understanding why the algorithm converged to assign a specific label to a specific unit. This in an unprecedented achievement, for it opens up an entirely new field of applications in any other scientific field. In the context of landslide prediction, this can finally unify a modeling framework from which derive standard practices for susceptibility modeling. Being a complete new breakthrough, the present manuscript attempts to showcase the potential of Explainable Artificial Intelligence (ExAI, hereafter) for landslide susceptibility modeling. The remainder of the manuscript does so by providing context on the basis of the landslides triggered by the Gorkha earthquake. Moreover, a web application is also shared with the readers allowing them to explore and get accustomed to the potential of this new generation models.

2 Materials and Methods

Below, Section 2.1 will provide an overview of the data used in this work to demonstrate the potential of ExAI, whose design and web app graphical interface are presented in Section 2.2 and 2.4, respectively.

2.1 Data

2.1.1 Landslide inventory

We tested our ExAI in the area struck by the Gorkha earthquake (7.8 M_w) on the 25^{th} of April 2015. Roback et al. (2017) mapped 24,903 coseismic landslides for this event, and presented their characteristics in Roback et al. (2018), with a total landslide area of 86.5 km^2 (Nowicki Jessee et al., 2018). Figure 1 The polygonal inventory is freely accessible at the global database of earthquake-induced landslide inventories (Tanyaş et al., 2017).

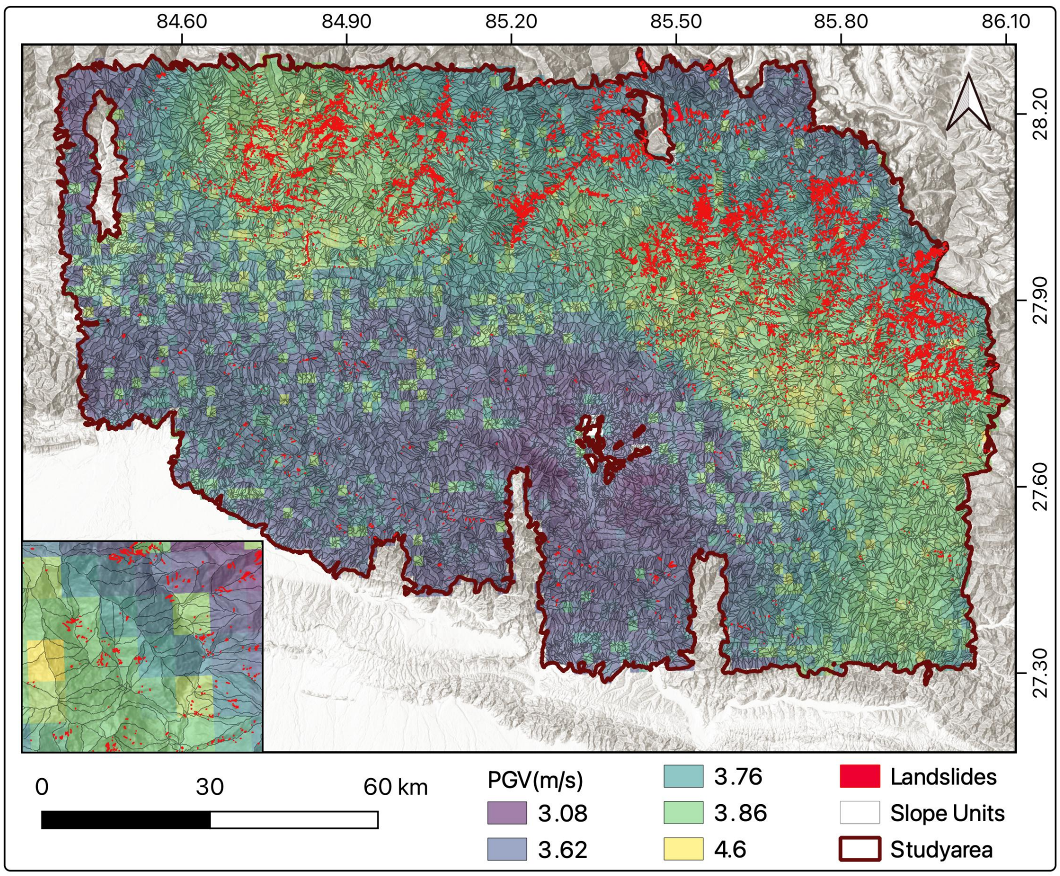


Figure 1. Overview of the study area, coseismic ground motion and associated landslides. The small panel to the bottom left shows a detail of the spatial partition we used, later explained in Section 2.1.2

Notably, this inventory is among the best coseismic ones for its quality and completeness (Tanyaş & Lombardo, 2020) for the authors characterize the polygon into source and deposition areas.

In this work, we use the spatial signal carried by this inventory as the target variable of our susceptibility model, aggregated at the level of slope units (more details below).

2.1.2 Slope unit partition

Slope units (SU) are irregular polygonal objects bound between ridges and streamlines (Carrara et al., 1995). Their use is an alternative to grid cells, which is particularly suited for regional scale susceptibility models. The recent introduction of the *r.slopeunits* software by Alvioli et al. (2016) is able to quickly generate SU under the constraint of slope exposition homogeneity, thus requiring only a digital elevation model as input data, and a few parameters to control the subdivision process. In our case, we opted for the latest version of *r.slopeunits*, capable of returning a reliable partition removing flat or near-flat areas (e.g., Alvioli et al., 2020; Lombardo & Tanyas, 2021).

Here we opted to run *r.slopeunits* with the following parameterization (after running a number of unreported tests): area_min=40000, circular variance=0.4, cleansize=20000, thresh=800000. These parameters control certain aspects of the calculations at the core of *r.slopeunits*. Specifically: *a*) area_min indicates the minimum SU area to coverge to; *b*) the circular variance controls how flexible or rigid the aspect criterion should be, with 0 being extremely rigid and 1 allowing for a large within-SU variability; *c*) cleansize refers to the dimension of spurious SU to be merged to the neighboring polygons; *d*) thresh is the SU extent *r.slopeunits* should start from.

This routine returned 16533 SU, with a mean planimetric area of 8.6×10^5 km^2 and a standard deviation of 7.8×10^5 km^2. These summary statistics attest for a slightly coarse resolution of the SU, which we opted for simply for computational reasons. In fact, as we planned to share a web-GIS platform where our model can be interactively queried, a finer SU partition would have implied a much slower interface.

2.1.3 Predictors

Our model relies on a set of predictors we chose to explain the predisposing and triggering factors that have led to the coseismic inventory mapped by Roback et al. (2017). Specifically, we opted for eight predictors, the morphometric ones originating from the 30 m SRTM digital elevation model (Van Zyl, 2001). These encompass: *i*) Slope steepness (Slp; Zevenbergen & Thorne, 1987); *ii*) horizontal (Hc Heerdegen & Beran, 1982) and *iii*) vertical (Vc Heerdegen & Beran, 1982) curvatures; *iv*) Eastness (Est) and *Nrt*) Northness (Lombardo et al., 2018). As for the expression of the *vi*) ground motion in terms of Peak Ground Velocity (PGV_Usgs), this came from the ShakeMap system of the United States Geological Survey (Worden & Wald, 2016). Vegetation density was brought in via Normalized Difference Vegetation Index (NDVI) (Pettorelli et al., 2005), computed from Landsat Imagery (Survey, 2015), whereas the antecedent precipitation (Prc) was calculated as accumulated rain over a three months period prior to the earthquake occurrence, from CHIRPS data (Funk et al., 2015).

The spatial pattern of these covariates was aggregated at the SU level by taking the mean value within the given SU, something highlighted in the remainder of the text with the suffix "_m". Notably, it is customary to express the variability of a given predictor within a SU by taking its mean behaviour as well as its variance for near Gaussian distributions (Guzzetti et al., 2006; Lombardo & Tanyas, 2020), or to use a quantile representation in situation far from the normality assumption (Castro Camilo et al., 2017; Amato et al., 2019). However, here to keep the model simple and easy to be explained, we opted to avoid adding the variability of each predictor per SU. Our explanation is that we are not trying to reach high performance through deep learning, this is something already shown in a number of contributions (e.g., Meena et al., 2022). Conversely, we seek to demonstrate the power of ExAI in susceptibility modeling.

As the last preprocessing step, we normalized all predictors between zero and one using the following transformation for each predictor:

$$X_{norm} = (X_{original} - min(X_{original}))/(max(X_{original}) - min(X_{original})) \qquad (1)$$

2.2 Explainable AI design

The deep learning model we used to test our Explainable AI was kept simple to easily diagnose the model output and to prevent it from overfitting. Its basic structure is shown in Figure 2, where the model relies on 8 input features in the input layer, followed by 12 hidden layers made out of fully connected layers of size 64 and a output layer with a sigmoid activation function. Each hidden layer is accompanied by a Rectified Linear Unit (ReLU) non linear activation (Yarotsky, 2017), followed by batch normalization (Ioffe & Szegedy, 2015) and a dropout layer (Baldi & Sadowski, 2013) with 0.3% dropout. These three elements have nowadays become standard in most deep learning architectures and we refer to the work of Schmidhuber (2015) for further details. For conciseness, here we will briefly mention that the ReLU activation allows for the model to be flexible and incorporate non-linear behaviors. Moreover, the dropout layer is used to prevent overfitting, whereas the batch normalization layer prevents weights and biases to grow unrealistically.

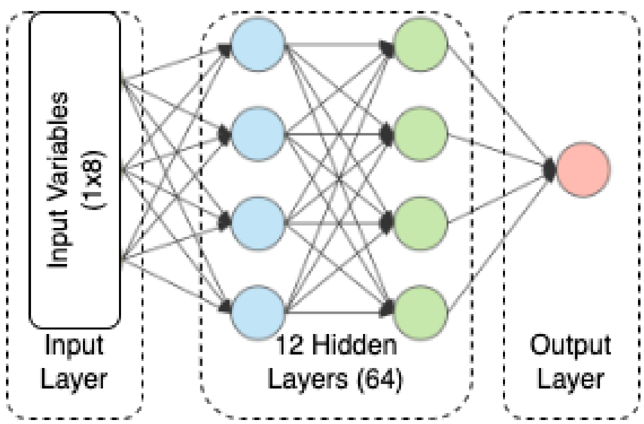


Figure 2. Simple architecture of the landslide susceptibility model which is used in our ExAI.

To train our model, we used the dataset as described in the section 2.1. The dataset was randomly divided into two disjoint training (70%) and test (30%) set. The training set was further randomly divided into 20% validation set which was randomly selected with replacement in each training epoch. All of the training performance were evaluated in the validation set (for e.g. training performance) and the model's performance itself was evaluated with the test set.

The developed model is trained using weighted binary crossentropy loss function, due to imbalance in the landslide data we provided 4× higher weights to the positive outcomes (landslides). The weight set to 4 because amount of slope-units with landslide was around 20% and without landslide was around 80%. To train the model we used Adam optimizer (Kingma & Ba, 2014) with initial learning rate of $1\times10^{-}3$ and decayed exponentially at 10000 training steps with decay rate of 0.9. The training was done with the batch size of 32 for 500 epochs with an early stopping option. This implies that the training process automatically stops once the model tends to overfit.

Once the model was fully trained and shown good performance, we calculated SHapley Additive exPlanations (SHAP) values to diagnose the model and its decisions (Lundberg & Lee, 2017). The SHAP values are calculated using the DeepSHAP method developed by Lundberg and Lee (2017). To provide a explanation about SHAP values below we present a simple practical example. Let us assume we are in the context of a simple linear regression where the target variable is regressed against only three covariates. The relative equation could be denoted as:

$$Y = \sum \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \tag{2}$$

Interpreting such simple model would be an easy task as it boils down to a linear combination of linear relations. However, machine/deep learning architectures offer the ability to extent the modeling framework even towards nonlinear combination of non-linear relations, which is something that makes the interpretation a very difficult task. For this reason, what SHAP does is to solve the predictive equation for each mapping unit of interest and storing the relative results. This provides a unique perspective on each predictor's role with respect to the others, for each slope unit in our case. In other words, to compute SHAP, one has to take the weights estimated for each predictors, multiply them for the actual predictors value and then combine them for each element in the matrix. These can then be stored and queried later on to understand how a specific probability value has been assigned to a slope unit. In the linear example mentioned above and for a single mapping unit, this would allow starting from the initial intercept value β_0 then adding the term that contributes the least to the final estimate (say $\beta_2 X_2$), then adding the second (say $\beta_3 X_3$) and the third (say $\beta_1 X_1$). As a result, SHAP allows to see changes in probability estimates as a function of each predictor offering a unique assessment tool on the final estimates and how the model has reached them.

2.3 Performance assessment

Aside from the added interpretability value of our ExAI, understanding how well it labels slope units into stable or unstable is a fundamental requirement of any binary classifier. Here, we monitored the ExAI performance via Receiver Operating Characteristic (ROC) curves and their Area Under the CUrve (AUC) as per standard (Hosmer & Lemeshow, 2000; Rahmati et al., 2019). We used this cutoff-independent metric while testing our model over different data realizations. In addition, we also produced cutoff-dependent metrics by taking the median of the probability distribution. This operation ensures the conversion of the continuous probability spectrum into two classes (stable/unstable slopes) which can be further matched to the original data to estimate True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). To complement the non-spatial information provided by the ROC curves, we opted to project these four values over the geographic space, producing in turn a confusion map (Titti et al., 2022).

We used these metrics in a number of performance tests. Specifically, we initially tested our best model, built according to the description provided in Section 2.2, and then considered it as our reference to compared against two additional cross-validation schemes. One corresponds to a purely random 10-fold cross validation (RCV hereafter), where 10% of the slope units are randomly extracted for testing, constraining the selection just once per mapping units, over ten subsequent replicates. The idea behind this validation routine is for us to assess performance while the data is perturbed the least. In fact, the random selection essentially keeps the residual spatial dependence, if any, almost intact. For this reason, the performance is expected to remain close to the reference model. Instead, to really grasp how well a model is capable of performing a susceptibility prediction task, one should always include a spatially-constrained cross-validation (SCV). A rich description on why and how to implement this technique can be found in Brenning (2012) and Pohjankukka et al. (2017). Here we briefly mention that a spatial cross-validation boils down to testing the model capabilities in an unknown region, thus in a context where the model is blind to any potential landslide clustering effect or residual spatial dependence. In turn, this usually leads to lower performance compared

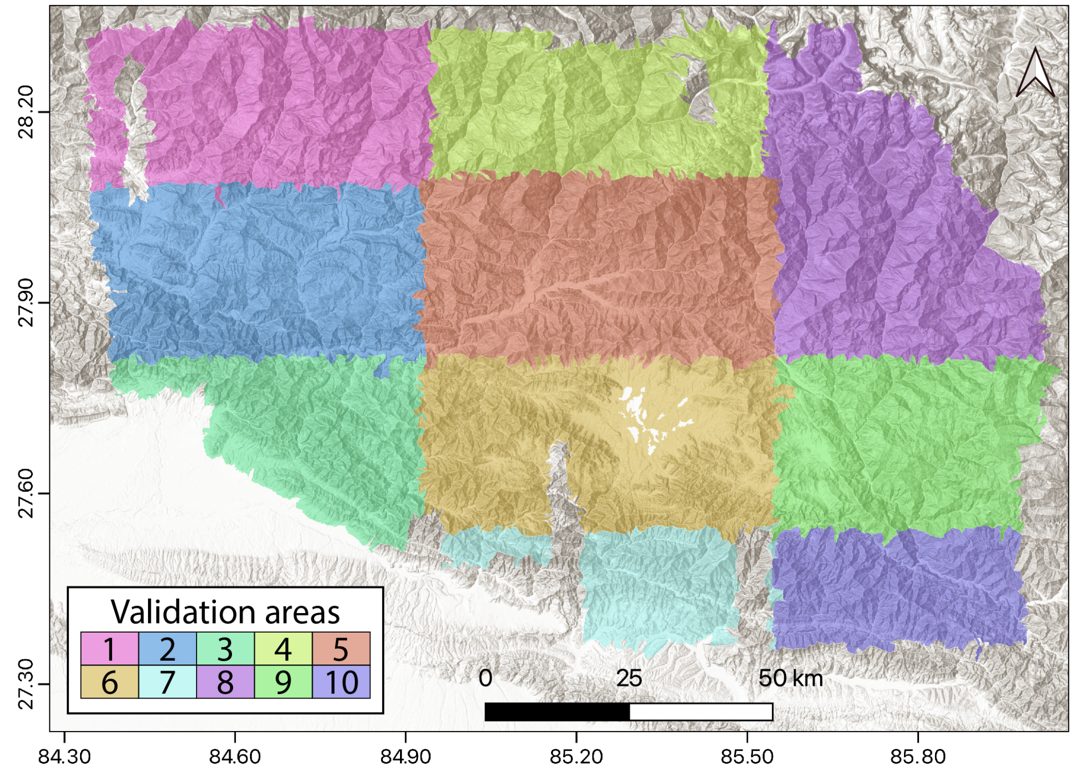


Figure 3. Aggregation of the slope unit partition into ten subregions, used for spatial cross-validation purposes.

to the reference model but also offers an overview of what to expect in one of the worst cases a classifier can face. In this work we implemented our SCV by dividing the area into ten sub-regions according to a squared lattice. Then, all the slope units falling within one area were used for testing while the remaining nine were used for calibration. This routine has been repeated ten times, until covering the whole study area and testing all the slope units partitioning it (see Figure 3).

2.4 Interactive demonstration through a web application

Explaining the potential of our ExAI simply through scientific illustrations may have not offered the same understanding as an interactive tool. For this reason, we have opted for a web application where our model results can be interactively queried to offer a more immersive experience to the readers and to anybody interested in it. The web-GIS is meant to provide the same level of query as it will be shown in the other figures in this manuscript. In addition to that, the same operation could be repeated for any slope unit in our study area, letting any user grasp why our ExAI assigned a given probability value as a linear combination of weights estimated for each predictor multiplied by their predictor value at specific locations.

Our web-GIS relies on a ArcGIS online platform, using a standard ESRI template for web applications. The ExAI output was computed outside the platform, a figure for every SU created in python and then stored in a repository where our web-GIS goes to pick any element queried by the user. When mentioning our choice of a relatively coarse slope unit partition in the previous section, we should also stress that a finer partition would have also required generating a much larger number of images, one for each slope

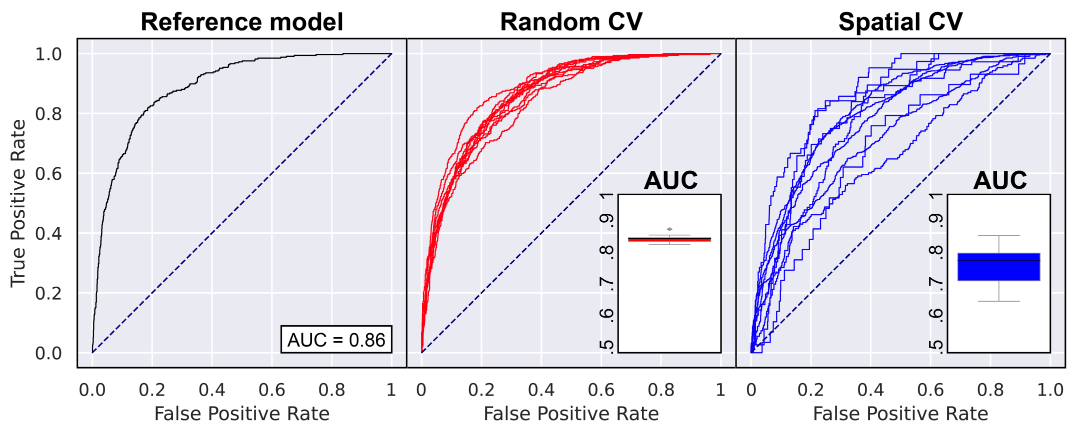


Figure 4. Panels from left to right: ROC curve and associated AUC of our reference ExAI model; ten ROC curves generated through a purely random cross-validation, with associated AUC values boxplotted at the bottom; ten ROC curves generated through spatial cross-validation, with associated AUC values boxplotted at the bottom.

unit, increasing the data storage requirements beyond the scope of a demonstrational online platform.

3 Results

Below we will initially report our ExAI performance, after which we will provide an extensive description of how the ExAI can be queried to understand why a specific probability value has been assigned to a slope unit. Ultimately, we conclude this section by illustrating our web application.

4 Perfomance overview

The most common characteristic of a machine/deep learning tools is their prediction capacity. Figure 4, offers an overview of our modeling performance. Specifically, our reference model falls in the excellent performance class according to Hosmer and Lemeshow (2000). This is also the case for the RCV, with a mean AUC of 0.86 and a very limited spread measured in a single standard deviation of 0.01. As mentioned in Section 2.3, the SCV procedure is where one would expect a significant drop in performance. This is the case also here, with a mean AUC of 0.77 and a standard deviation of 0.06. This still means that on average our model still is very close to the excellent performance class according to (Hosmer & Lemeshow, 2000). However, it points out at local performance deficiencies with a minimum AUC of 0.66.

Interestingly, this low performance is achieved for the tenth sub-region shown in Figure 3. The south-easter sector of the study area is also the one that was shaken the least by the Gorkha earthquake and this is likely the reason why our coseismic ExAI susceptibility model struggled there.

Similar considerations emerge also when looking at the three confusion maps shown in Figure 5. There, the spatial pattern of TP and FN essentially stay the same for the reference model as well as the two cross-validation schemes. This is an indicator of the consistent capacity of our ExAI to recognize unstable slope units. The main difference among the three maps becomes evident when looking at how FP substitute the TN. This is something that may be considered an issue at a first glance. However, we should re-

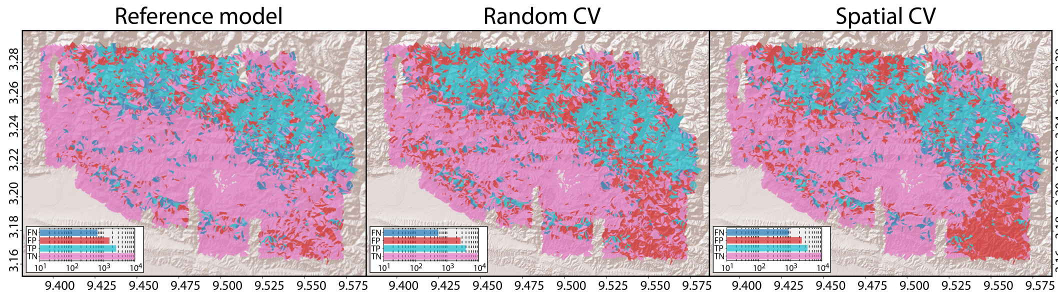


Figure 5. Confusion maps for the reference model and the two cross-validations we tested. The barplots correspond to the relative confusion matrices.

call that FP correspond to slope units that did not have a landslide in the original dataset but that the model deemed to be unstable. In other words, this is not an issue that should raise questions on the quality of our classifier. Conversely, it should be considered an indication of locations that may have not generated landslides in the occasion of the Gorkha earthquake but could still fail in the future.

5 Looking into the ExAI

Recent advancements in Artificial Intelligence have significantly pushed the boundaries of what can be queried and visualized out of an explainable AI. In this work, we tried to provide several options for the readers and selected the one we considered to be the best for our web application.

The simplest way of understanding why a given AI has assigned a specific label to a mapping unit can be done by examining the variable importance (Gunning et al., 2019; Aguilera et al., 2022). This measure expresses the influence of each predictor used in the model with respect to the others and has already found a few applications in data-driven natural hazard models (Stumpf & Kerle, 2011; J. Goetz et al., 2015; Steger et al., 2016). Here we re-created a variable importance plot in Figure 6, by using the computed SHAP values. The mean slope per slope unit and the peak ground velocity are shown to dominate the probability estimation. Then, the remaining six predictors appear to exert a similar influence onto the final susceptibility.

Another already available tool to visualize overall predictor's influence consists of response plots (Merow et al., 2013). This tool has also been featured in a number of natural hazard (Vorpahl et al., 2012; Lombardo, Fubelli, et al., 2016; Lombardo, Bachofer, et al., 2016) applications albeit to a lesser extent compared to the variable importance presented above. In this work, we reproposed a response plot graphical summary by plotting SHAP values against each predictor's domain. This is shown in Figure 7 where the two dominant predictors in the model appear to be again the mean slope steepness (Slp_m) and the mean peak ground velocity (PGV_{Usgs}) per slope unit.

The last two illustrations have been routinely included in a number of articles already for over a decade. However, this has not been sufficient to label any standard AI as explainable. The reason is due to the static vision these tools provide with respect to the modeling result. In fact, they essentially tell the same story, this being two predictors influencing more than others the final output. But, no other relevant information can be retrieved on how this happens. In other words, these plot lack the capacity to provide insight into how each predictor interacts with the others for each mapping units, leading to the final probability value. This is where our ExAI enters an uncharted territory in geosciences, providing a full description of these predictors' interactions. Be-

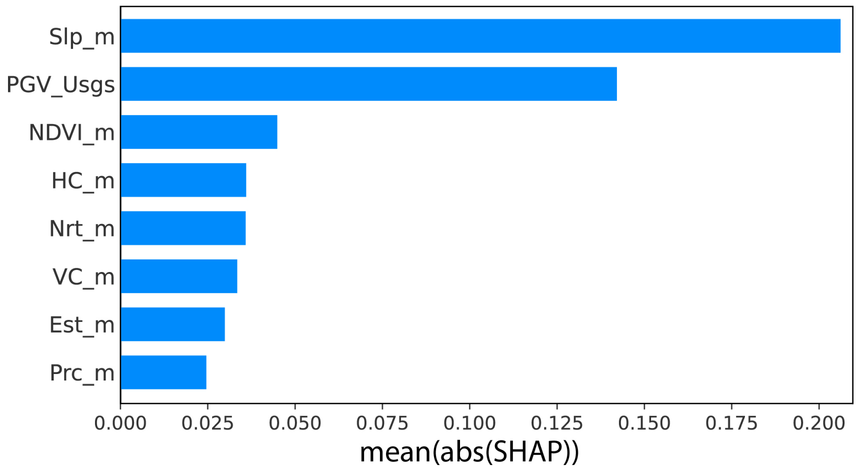


Figure 6. Variable importance plot obtained by taking the mean absolute value of SHAP, then ranked from the highest to the lowest contributor.

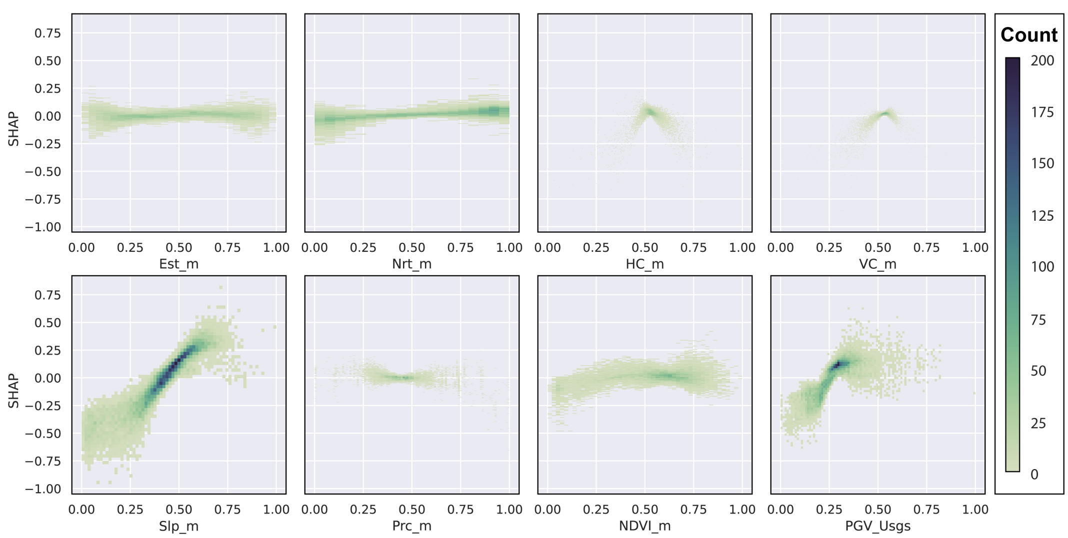


Figure 7. Response plots for each of the predictors used in the model. The x-axis reports the rescaled domain of each predictor while the y-axis corresponds to the influence each predictor exerted onto the susceptibility estimates.

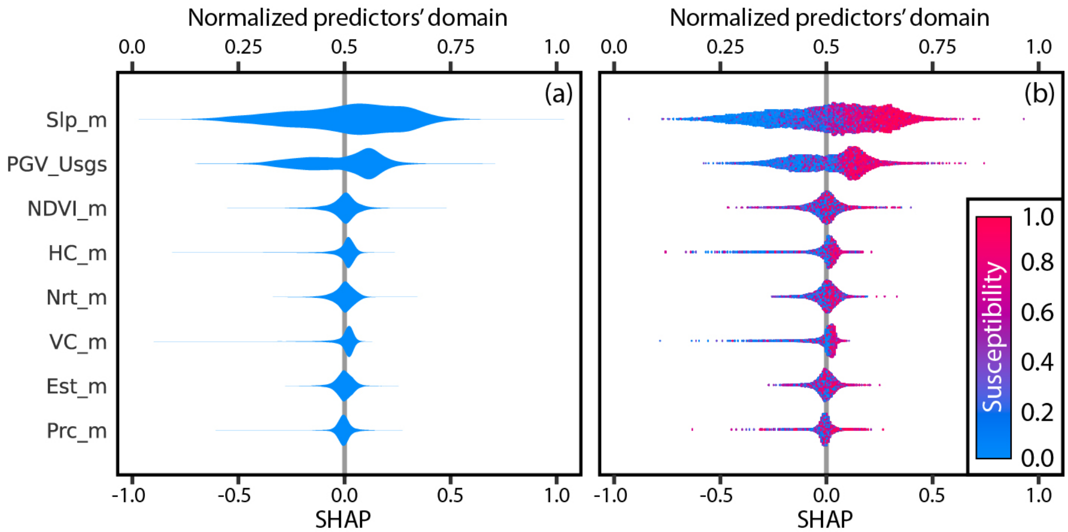


Figure 8. Panel (a) shows the SHAP distribution for each predictor expressed with a violin plot obtained considering all slope units. Panel (b) does the same but each dot corresponding to a specific slopes unit has been further colored with the susceptibility it was ultimately assigned with.

low we will provide tools to do so, presented in order of the level of information they provide.

The simplest way to additionally explore our ExAI is shown in Figure 8, at a information level which is not far off from the one provided by the two illustrations above. Specifically, panel (a) shows the overall SHAP distribution per predictor computed for the whole study area. This is something very similar to what was shown in Figure 7, to which we start adding information on specific locations. Panel (b) does accomplish exactly this task by showing the actually probability assigned to each slope unit (or dot in the figure). As a result, one can start seeing that SHAP values computed for single predictors assume essentially assume an alternating coloration per slope units until the mean NDVI ($NVDI_m$), after which an increase PGV_{Usgs} and Slp_m and associated SHAP values, also corresponds in an increase in susceptibility. This indicates a dominant effect of the last two predictors, which is a similar conclusion to what showed in previous illustrations. However, it already provides an indication that our ExAI will delve much deeper that usual tools, away from a single perspective over the whole study area and much closer to what happens at the level of the single slope unit.

The level of the single mapping unit is actually where our ExAI aims to provide information to the end user. This can be shown in Figure 9, where two slope units have been extracted as an example for demonstration. Panel (a) shows how the final susceptibility of 0.23 was reached adding the contributions of all predictors to the base probability of 0.42. We briefly point out here that 0.42 is the starting value as a result of the balanced presence/absence data we opted for. Any further imbalance in the proportion of stable and unstable slope units would lead to a lower starting value (see Frattini et al., 2010; Lombardo & Mai, 2018). Going back to Figure 9, this graphical summary is the perfect example to deliver how powerful is an ExAI, to the point where one can assess whether the susceptibility makes sense for single mapping units. However, generating these plots for each mapping unit may be too complex. For this reason, it is possible to simplify the graph while reporting the same information.

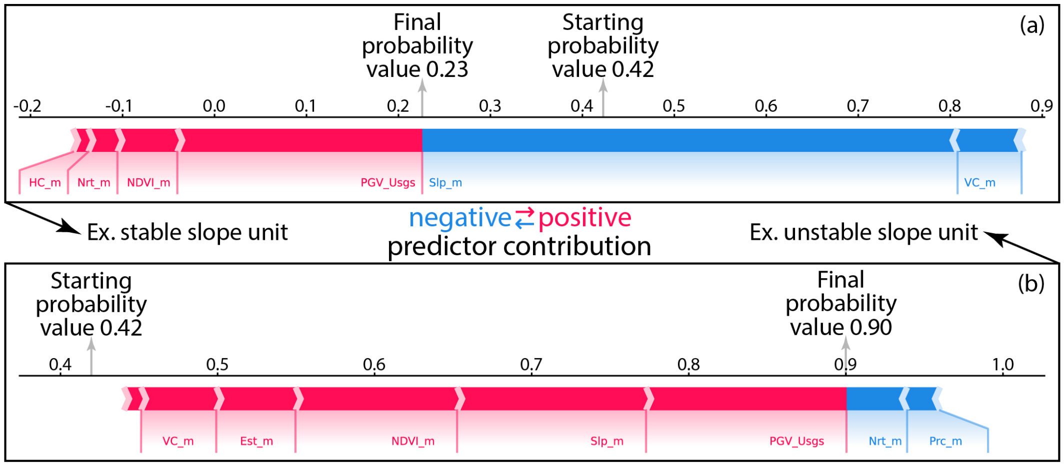


Figure 9. Panel (a) shows an example of a slope unit that started with a 0.42 probability value and whose final susceptibility reached 0.23. This value was reached due to the contribution of the other predictors, whose sign is graphically summarized through the horizontal arrows' direction and the magnitude is depicted through the horizontal arrows' length. The same is shown in panel (b) for a slope unit that started with a base probability of 0.42 and reached a final susceptibility estimate of 0.9.

This simpler yet effective overview is provided in Figure 10. There, in panel (a) we propose once more the same information provided in Figure 9 for two slope units. The way this plot can be read is to start from the bottom, where again the base susceptibility is 0.42 and then monitor the variations brought by each predictor listed on the y-axis. As for panel (b), we plotted it to demonstrate that this type of plotting makes it possible to compare as many slope unit as one desires.

The ExAI proposed by Lundberg and Lee (2017) suggests even more tools to visualize the model output. However, we consider the last illustration to be the most effective among all the available ones. For this reason, we have equipped our web application precisely with this type of visualization. The app can be accessed at the following link: https://arcg.is/0unziD. There, we have placed the final susceptibility map produced by our ExAI (see Figure 11).

Each mapping unit that constitutes the map can then be interactively queried. Specifically, by clinking on any slope unit, the system plots the ExAI according to the style explained above.

Below we present to examples captured from two adjacent SUs. Figures 12 and 13 provide two examples of how to visualize the ExAI decision within our web application, for two SUs estimated to be unstable and stable, respectively.

6 Discussion

The model we present relied on a relative small number of predictors. We opted for this structure to offer a simple and efficient visualization of the ExAI decisions. This characteristics has led our ExAI to highlight minimal contributions of terrain characteristics other that the slope steepness, in addition to which the ground motion determines most of the final probability. Nevertheless, this already provides a good idea of what ExAI can do and how its decisions can be queried in depth to understand the extent to which

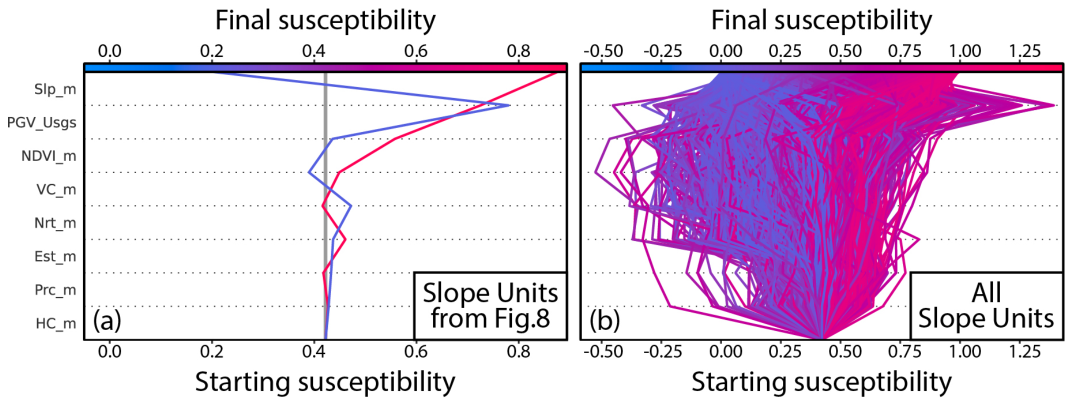


Figure 10. Panel (a) summarizes all the information presented in Figure 9 in a much more straightforward way. The variation of the probability estimates for the two slope units is compressed in a single line plot. Panel (b) makes it possible to present the whole information for all slope units partitioning the study area.

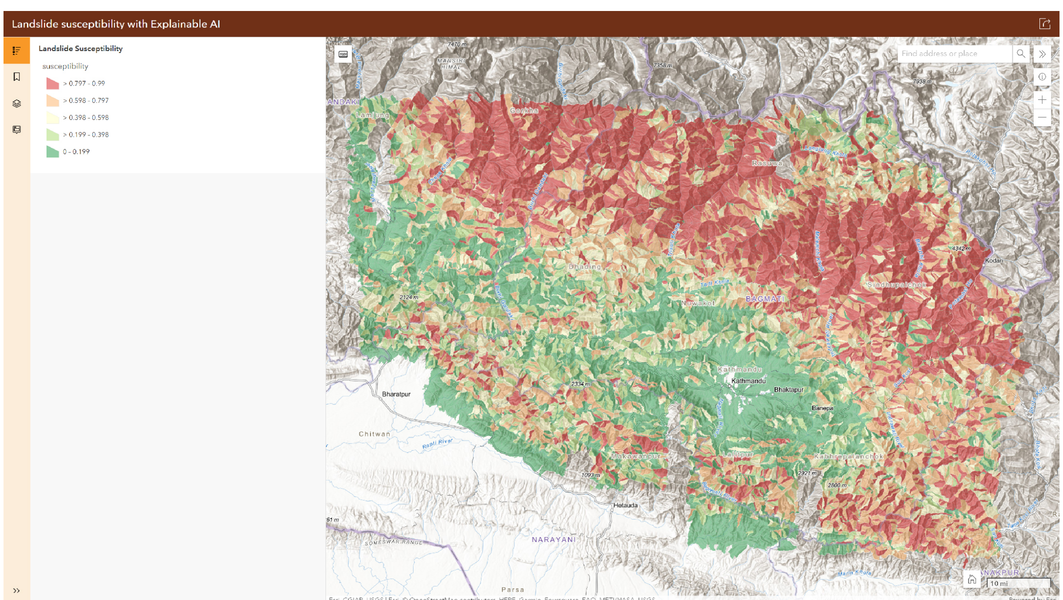


Figure 11. General overview of the web application. The susceptibility map we obtained by using our ExAI is depicted here into five equal spaced classes.

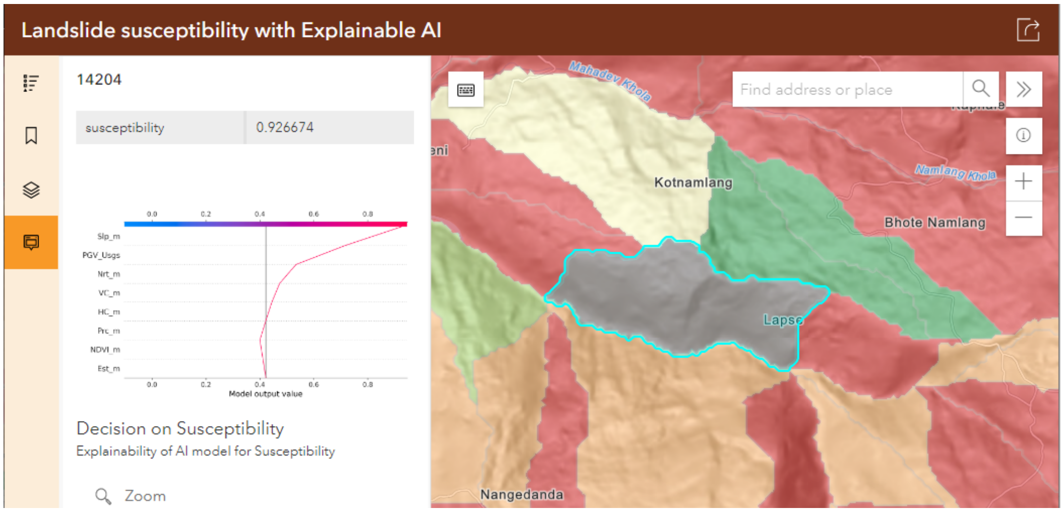


Figure 12. Example of an ExAI query for an unstable SU.

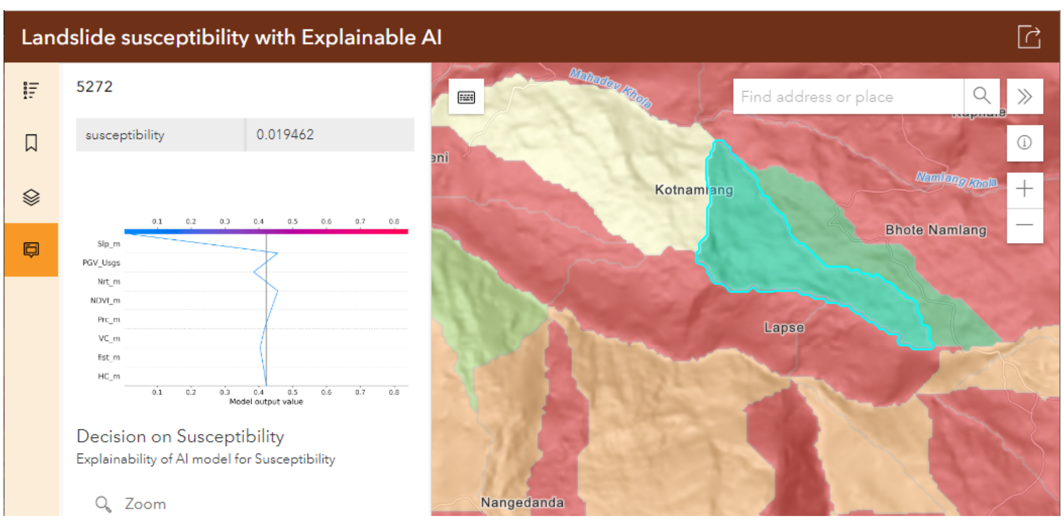


Figure 13. Example of an ExAI query for a stable SU.

one can trust the maps it produces. The traditional variable importance plot in fact, is unsuitable to provide the whole picture. Figure 6 simply illustrates the extent to which each variable dominates the outcome. However, it does not tell the user whether this makes sense from an interpretative standpoint. For instance, it indeed makes sense that Slp_m and PGV_{Usgs} does control most of the probability of landslide occurrence for the co-seismic example we considered in this work. However, the artificial intelligence behind could increase the probability at decreasing values of slope steepness and/or ground motion; something that would violate our basic understanding of the physics behind the genesis of a the failure mechanism. For this reason, response plots like Figure 7 add another level of understanding for they allow to monitor variations in SHAP with respect to each predictors' domain. This is a capability which is typical of statistical models (Lima et al., 2021; Tanyaş et al., 2022) and has found very few applications in machine learning (Park, 2015; Vorpahl et al., 2012). However, even in this case, the level of information provided is very generic and corresponds to the overall behaviour of each predictor with respect to the entire map it contributes to define. An analogous graphical representation of the model output is shown in Figure 8(a). And even if panel (b) adds some additional information through the embedded colorcoding, the model could still be locally mistaking the effect of certain predictors. In fact, at the local level, no traditional statistical models nor machine/deep learning ones have so far provided a transparent understanding of predictors contributions and how the specifically interact with each other. This aspect is now achievable through ExAI and Figures such as 9 and 10 provides a clear representation of how this can be translated into meaningful scientific illustrations. These types of graphical summaries have they have been created with the idea in mind of making black boxes into white ones. At this local level, here expressed through SUs, one can examine how reliable the probabilistic estimates are. For instance, to continue the Slp_m and PGV_{Usgs} example mentioned above, one can query a given SU, check the SHAP value and then easily cross-reference it with respect to the actual steepness and ground motion values. As a result, one can interactively realize whether steeper slopes have been assigned with a higher susceptibility or not. And, whether slopes that have undergone a greater shaking have also been estimated with a higher likelihood to host a landslide. The same is valid in the opposite situation an in any other level in between. In short, ExAI provides a window into the core calculations that the given model has gone through, helping the user to understand the extent to which the AI can be relied on. All this is essentially possible in near-real-time and our web-application is meant to highlight this specific characteristic. There, any user can query our model in a transparent manner that has not yet been reached so far within the geoscientific community. And, which we hope can become a standard as the use of ExAI becomes more common in the future.

We conclude by stressing that artificially intelligent models are usually acclaimed due to their predictive capacity, which here we tested via a suite of validation routines. The results shown in Figure 4 highlight predictive performance in line with other machine/deep learning studies, especially considering the limited number of predictors we opted for in this work. An important element the same figure highlights is the fact that despite Brenning (2012) clearly advocated for spatial cross-validations to become a standard in susceptibility modeling, this is something which is rarely done. And yet, a spatial cross-validation constitutes an important element to really assess the extent to which a given data-driven model can be used to predict natural hazard occurrences in areas outside the training set. This is an important characteristic that goes beyond the explainability or not of a given model, but it allows to estimate the minimum (worst-case scenario) one could expect when transferring the prediction elsewhere.

7 Conclusions

Explainable artificially intelligence represents the future of data-driven models in any scientific area. The prediction capacity of complex modeling architectures can be dissected into its simpler elements, allowing one to understand the reason behind a model result, leaving behind the negative connotation of the black box label and finally opening up towards white box characteristics even in the context of machine/deep learning.

Our work here introduces ExAI for landslide prediction and it is meant to offer an overview of the potential that this new generation of models can offer and will certainly offer in the future. We see ExAI as a milestone in the history of data-driven models and the extent to which these models may change the way we perceive artificially intelligent decisions is yet to be unraveled. However, we also see an opening for improvements. Currently, and this is also valid in this manuscript, ExAI is mainly integrated as part of binary classifiers. However, the information of where landslides may occur is not the only important element in the chain of hazard assessment. Another important notion would be estimating how large landslides may be once they trigger on a slope labeled as unstable. Few data-driven models have already been proposed to address this issue and we see the next step to do the same in the context of ExAI, where the expected dimension of a landslide can be precisely predicted while contextually providing information on why it may reach that extent.

Similar considerations can be extended to estimate potential losses and open up this framework towards societal risk modeling. And again, similar considerations can be extended to beyond the pure spatial context and towards spatio-temporal modeling.

In summary, ExAI applications are at an infancy stage and much is to be explored on what can be improved and how their use can be directed to address other research questions. In this work, we hoped to highlight its strength and stimulate the spread of ExAI even further. For this specific reason, we have build an interactive demonstration accessible at https://arcg.is/0unziD. Moreover, to promote reproducibility and repeatability, data and codes have also been shared in a FAIR complying repository (`https://doi.org/10.5281/zenodo.6976122`) (Dahal & Lombardo, 2022).

Acknowledgments

Ashok Dahal has implemented the ExAI, run all the analyses and created the web-GIS platform. Luigi Lombardo has worked on drafting the manuscript. Landslides are available at: https://www.sciencebase.gov/catalog/item/583f4114e4b04fc80e3c4a1a, whereas data and codes to reproduce our ExAI can be found at: `https://doi.org/10.5281/zenodo.6976122` (Dahal & Lombardo, 2022).

References

Aguilera, Q., Lombardo, L., Tanyas, H., & Lipani, A. (2022). On the prediction of landslide occurrences and sizes via Hierarchical Neural Networks. *Stochastic Environmental Research and Risk Assessment*, 1–18.

Alvioli, M., Guzzetti, F., & Marchesini, I. (2020, June). Parameter-free delineation of slope units and terrain subdivision of Italy. *Geomorphology*, *358*, 107124. Retrieved 2021-08-30, from `https://www.sciencedirect.com/science/article/pii/S0169555X20300969` doi: 10.1016/j.geomorph.2020.107124

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F. (2016). Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geoscientific Model Development*, *9*(11), 3975–3991.

Amato, G., Eisank, C., Castro-Camilo, D., & Lombardo, L. (2019). Accounting for covariate distributions in slope–unit–based landslide susceptibility models. a case study in the alpine environment. *Engineering geology*, *260*, 105237.

Atkinson, P., Jiskoot, H., Massari, R., & Murray, T. (1998). Generalized linear modelling in geomorphology. *Earth Surface Processes and Landforms: The Journal* *of the British Geomorphological Group*, *23*(13), 1185–1195.

Atkinson, P. M., & Massari, R. (1998). Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. *Computers & Geosciences*, *24*(4), 373–385.

Baldi, P., & Sadowski, P. J. (2013). Understanding dropout. *Advances in neural information processing systems*, *26*.

Ballabio, C., & Sterlacchini, S. (2012). Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study, Italy. *Mathematical geosciences*, *44*(1), 47–70.

Bednarik, M., & Pauditš, P. (2010). Different ways of landslide geometry interpretation in a process of statistical landslide susceptibility and hazard assessment: Horná súča (western slovakia) case study. *Environmental Earth Sciences*, *61*(4), 733–739.

Brabb, E., Pampeyan, H., & Bonilla, M. (1972). MG 1972. landslide susceptibility in San Mateo County, California. *US Geological Survey Miscellaneous Field Studies Map MF–360, scale*, *1*(62,500).

Brenning, A. (2008). 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), 410.

Brenning, A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. In *2012 ieee international geoscience and remote sensing symposium* (pp. 5372–5375).

Bui, D. T., Tsangaratos, P., Nguyen, V.-T., Van Liem, N., & Trinh, P. T. (2020). Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. *Catena*, *188*, 104426.

Carrara, A. (1983). Multivariate models for landslide hazard evaluation. *Journal of the International Association for Mathematical Geology*, *15*(3), 403–426.

Carrara, A., Cardinali, M., Guzzetti, F., & Reichenbach, P. (1995). GIS technology in mapping landslide hazard. In *Geographical information systems in assessing natural hazards* (pp. 135–175). Springer.

Castro Camilo, D., Lombardo, L., Mai, P., Dou, J., & Huser, R. (2017). Handling high predictor dimensionality in slope–unit–based landslide susceptibility models through LASSO–penalized Generalized Linear Model. *Environmental Modelling and Software*, *97*, 145–156.

Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, *13*(11), 2815–2831.

Dahal, A., & Lombardo, L. (2022, August). *Data: Explainable artificial intelligence in geoscience: a glimpse into the future of landslide susceptibility modeling.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.6976122` doi: 10.5281/zenodo.6976122

Das, I., Stein, A., Kerle, N., & Dadhwal, V. K. (2012). Landslide susceptibility mapping along road corridors in the Indian Himalayas using Bayesian logistic regression models. *Geomorphology*, *179*, 116–125. doi: https://doi.org/10.1016/j.geomorph.2012.08.004

Dhakal, A. S., Amada, T., Aniya, M., et al. (2000). Landslide hazard mapping and its evaluation using GIS: an investigation of sampling schemes for a grid-cell based quantitative method. *Photogrammetric engineering and remote sensing*, *66*(8), 981–989.

Fang, Z., Wang, Y., Peng, L., & Hong, H. (2020). Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. *Computers & Geosciences*, *139*, 104470.

Frattini, P., Crosta, G., & Carrara, A. (2010). Techniques for evaluating the performance of landslide susceptibility models. *Engineering Geology*, *111*(1), 62–72. doi: https://doi.org/10.1016/j.enggeo.2009.12.004

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... others (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data*, *2*(1), 1–21.

Gates, W. E., & Heil, R. J. (1980). Geographic information systems. *Journal of the Surveying and Mapping Division*, *106*(1), 105–117.

Goetz, J., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers & geosciences*, *81*, 1–11.

Goetz, J. N., Guthrie, R. H., & Brenning, A. (2011). Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology*, *129*(3–4), 376–386.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—Explainable artificial intelligence. *Science robotics*, *4*(37), eaay7120.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., & Galli, M. (2006). Estimating the quality of landslide susceptibility models. *Geomorphology*, *81*(1–2), 166–184. doi: 10.1016/j.geomorph.2006.04.007

Heerdegen, R. G., & Beran, M. A. (1982). Quantifying source areas through land surface curvature and shape. *Journal of Hydrology*, *57*(3–4), 359–373.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed ed.). New York: Wiley.

Hua, Y., Wang, X., Li, Y., Xu, P., & Xia, W. (2021). Dynamic development of landslide susceptibility based on slope unit and deep neural networks. *Landslides*, *18*(1), 281–302.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448–456).

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Li, X., & Claramunt, C. (2006). A spatial entropy-based decision tree for classification of geographical information. *Transactions in GIS*, *10*(3), 451–467.

Lima, P., Steger, S., & Glade, T. (2021). Counteracting flawed landslide data in statistically based landslide susceptibility modelling for very large areas: a national–scale assessment for Austria. *Landslides*, 1–16.

Lombardo, L., Bachofer, F., Cama, M., Märker, M., & Rotigliano, E. (2016). Exploiting Maximum Entropy method and ASTER data for assessing debris flow and debris slide susceptibility for the Giampilieri catchment (north-eastern Sicily, Italy). *Earth Surface Processes and Landforms*, *41*(12), 1776–1789.

Lombardo, L., Fubelli, G., Amato, G., & Bonasera, M. (2016). Presence–only approach to assess landslide triggering–thickness susceptibility: a test for the Mili catchment (north–eastern Sicily, Italy). *Natural Hazards*, *84*(1), 565–588.

Lombardo, L., & Mai, P. M. (2018). Presenting logistic regression–based landslide susceptibility results. *Engineering geology*, *244*, 14–24.

Lombardo, L., Saia, S., Schillaci, C., Mai, P. M., & Huser, R. (2018). Modeling soil organic carbon with Quantile Regression: Dissecting predictors' effects on carbon stocks. *Geoderma*, *318*, 148–159. doi: https://doi.org/10.1016/j.geoderma.2017.12.011

Lombardo, L., & Tanyas, H. (2020). Chrono–validation of near–real–time landslide susceptibility models via plug–in statistical simulations. *Engineering Geology*, *278*, 105818.

Lombardo, L., & Tanyas, H. (2021). From scenario–based seismic hazard to scenario–based landslide hazard: fast–forwarding to the future via statistical simulations. *Stochastic Environmental Research and Risk Assessment*, 1–14.

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In I. Guyon et al. (Eds.), *Advances in neural information processing systems 30* (pp. 4765–4774). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Marmion, M., Hjort, J., Thuiller, W., & Luoto, M. (2009). Statistical consensus methods for improving predictive geomorphology maps. *Computers & Geosciences*, *35*(3), 615–625.

Meena, S. R., Puliero, S., Bhuyan, K., Floris, M., & Catani, F. (2022). Assessing the importance of conditioning factor selection in landslide susceptibility for the province of Belluno (region of Veneto, northeastern Italy). *Natural hazards and earth system sciences*, *22*(4), 1395–1417.

Melchiorre, C., Matteucci, M., Azzoni, A., & Zanchi, A. (2008). Artificial neural networks and cluster analysis in landslide susceptibility zonation. *Geomorphology*, *94*(3-4), 379–400.

Merow, C., Smith, M. J., & Silander Jr, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069.

Naranjo, J. L., Van Westen, C., & Soeters, R. (1994). Evaluating the use of training areas in bivariate statistical landslide hazard analysis-a case study in Colombia. *ITC journal*(3), 292–300.

Nowicki Jessee, M., Hamburger, M., Allstadt, K., Wald, D., Robeson, S., Tanyas, H., … Thompson, E. (2018). A Global Empirical Model for Near–Real–Time Assessment of Seismically Induced Landslides. *Journal of Geophysical Research: Earth Surface*, *123*(8), 1835–1859.

Park, N.-W. (2015). Using maximum entropy modeling for landslide susceptibility mapping with multiple geoenvironmental data sets. *Environmental Earth Sciences*, *73*(3), 937–949.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in ecology & evolution*, *20*(9), 503–510.

Pohjankukka, J., Pahikkala, T., Nevalainen, P., & Heikkonen, J. (2017). Estimating the prediction performance of spatial models via spatial k-fold cross validation. *International Journal of Geographical Information Science*, *31*(10), 2001–2019.

Pourghasemi, H. R., & Rossi, M. (2017). Landslide susceptibility modeling in a landslide prone area in Mazandarn province, north of Iran: a comparison between GLM, GAM, MARS, and M-AHP methods. *Theoretical and Applied Climatology*, *130*(1), 609–633.

Rahmati, O., Kornejady, A., Samadi, M., Deo, R. C., Conoscenti, C., Lombardo, L., … others (2019). Pmt: New analytical framework for automated evaluation of geo–environmental modelling approaches. *Science of the total environment*, *664*, 296–311.

Reichenbach, P., Rossi, M., Malamud, B., Mihir, M., & Guzzetti, F. (2018). A review of statistically–based landslide susceptibility models. *Earth–Science Reviews*, *180*, 60–91. doi: 10.1016/j.earscirev.2018.03.001

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1135–1144).

Roback, K., Clark, M., West, A., Zekkos, D., Li, G., Gallen, S., … Godt, J. (2017). Map data of landslides triggered by the 25 April 2015 Mw 7.8 Gorkha, Nepal earthquake. *US Geological Survey data release [data set], https://doi.org/10.5066/F7DZ06F9*.

Roback, K., Clark, M. K., West, A. J., Zekkos, D., Li, G., Gallen, S. F., … Godt,

J. W. (2018). The size, distribution, and mobility of landslides caused by the 2015 Mw7.8 Gorkha earthquake, Nepal. *Geomorphology*, *301*, 121–138.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85–117.

Soeters, R., & Van Westen, C. (1994). Slope instability: the role of remote sensing and GIS in recognition, analysis and zonation. In *Natural hazards and remote sensing* (pp. 44–50). Royal Society and Royal Academy of Engineering London.

Steger, S., Brenning, A., Bell, R., & Glade, T. (2016). The propagation of inventory–based positional errors into statistical landslide susceptibility models. *Natural Hazards and Earth System Sciences*, *16*(12), 2729–2745. doi: 10.5194/nhess-16-2729-2016

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, *41*(3), 647–665.

Stumpf, A., & Kerle, N. (2011). Object-oriented mapping of landslides using Random Forests. *Remote sensing of environment*, *115*(10), 2564–2577.

Survey, U. G. (Ed.). (2015). *Landsat surface reflectance data* (Version 1: Originally posted April 20, 2015; Version 1.1: June 16, 2015; Version 1.1 updated: March 27, 2019 ed.; Tech. Rep.). Reston, VA. Retrieved from `http://pubs.er.usgs.gov/publication/fs20153034` (Report) doi: 10.3133/fs20153034

Tanyaş, H., van Westen, C., Allstadt, K., Nowicki, A. J. M., Görüm, T., Jibson, R., ... Hovius, N. (2017). Presentation and Analysis of a Worldwide Database of Earthquake–Induced Landslide Inventories. *Journal of Geophysical Research: Earth Surface*, *122*(10), 1991–2015. doi: 10.1002/2017JF004236

Tanyaş, H., Hill, K., Mahoney, L., Fadel, I., & Lombardo, L. (2022). 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. *Engineering geology*, *297*, 106504.

Tanyaş, H., & Lombardo, L. (2020). Completeness Index for Earthquake–Induced Landslide Inventories. *Engineering geology*, *264*, 105331.

Titti, G., Sarretta, A., Lombardo, L., Crema, S., Pasuto, A., & Borgatti, L. (2022). Mapping susceptibility with open-source tools: a new plugin for QGIS. *Front Earth Sci*, *229*.

Titti, G., van Westen, C., Borgatti, L., Pasuto, A., & Lombardo, L. (2021). When Enough Is Really Enough? On the Minimum Number of Landslides to Build Reliable Susceptibility Models. *Geosciences*, *11*(11), 469.

Van Dao, D., Jaafari, A., Bayat, M., Mafi-Gholami, D., Qi, C., Moayedi, H., ... others (2020). A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena*, *188*, 104451.

Van Zyl, J. J. (2001). The Shuttle Radar Topography Mission (SRTM): a breakthrough in remote sensing of topography. *Acta Astronautica*, *48*(5-12), 559–565.

Vorpahl, P., Elsenbeer, H., Märker, M., & Schröder, B. (2012). How can statistical models help to determine driving factors of landslides? *Ecological Modelling*, *239*, 27–39.

Worden, C., & Wald, D. J. (2016). ShakeMap manual online: Technical manual, user's guide, and software guide. *US Geol. Surv.*, 1–156.

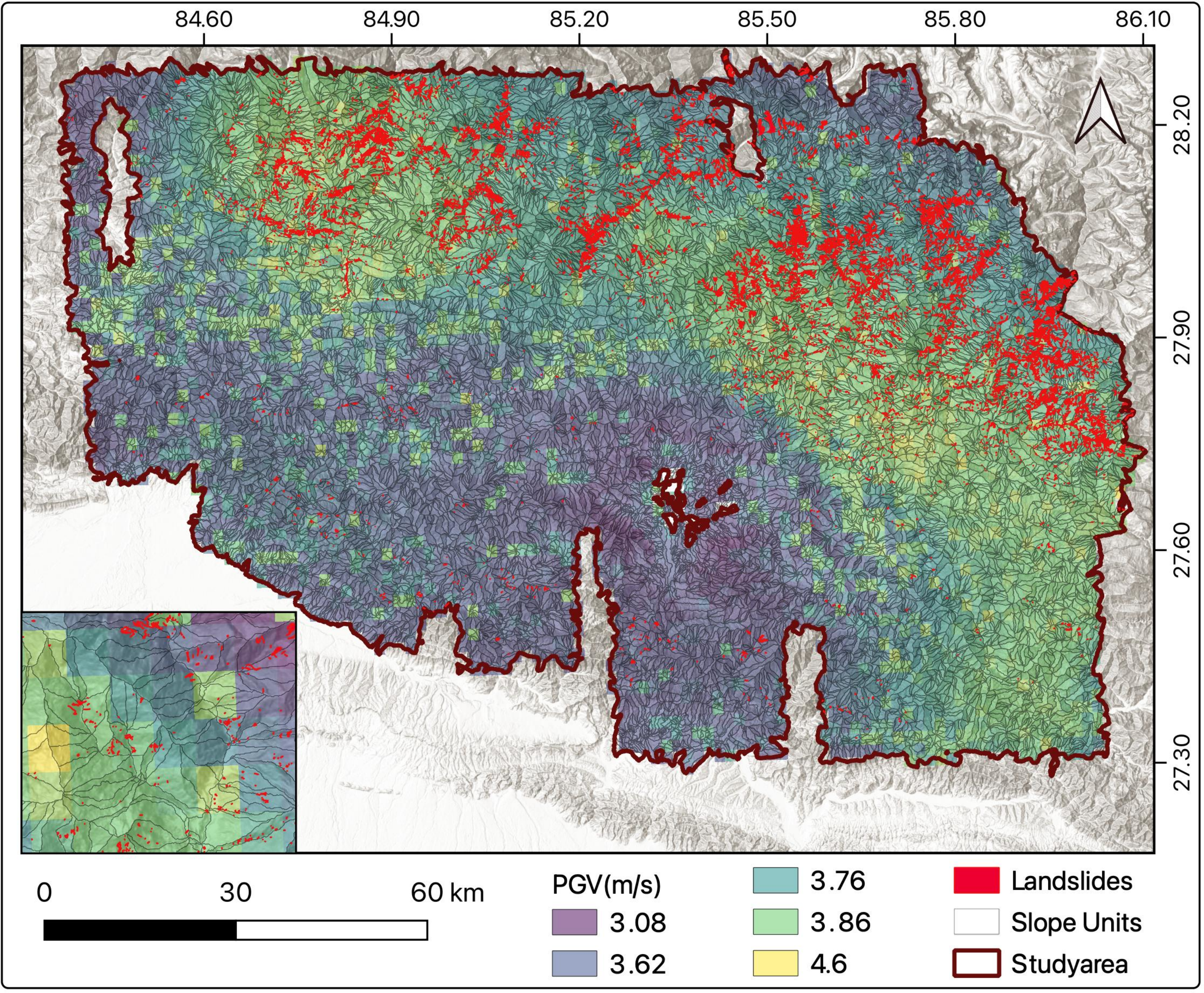
Yarotsky, D. (2017). Error bounds for approximations with deep ReLU networks. *Neural Networks*, *94*, 103–114.

Yesilnacar, E., & Topal, T. (2005). Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology*, *79*(3–4), 251–266.

Yi, Y., Zhang, Z., Zhang, W., Jia, H., & Zhang, J. (2020). Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: A case study in Jiuzhaigou region. *Catena*, *195*, 104851.

Zevenbergen, L. W., & Thorne, C. R. (1987). Quantitative analysis of land surface topography. *Earth surface processes and landforms*, *12*(1), 47–56.

Figure 01.



84.60
84.90
85.20
85.50
85.80
86.10
28.20
27.90
27.60
27.30
0
30
60 km
PGV(m/s)
3.08
3.62
3.76
3.86
4.6
Landslides
Slope Units
Studyarea

Figure 02.

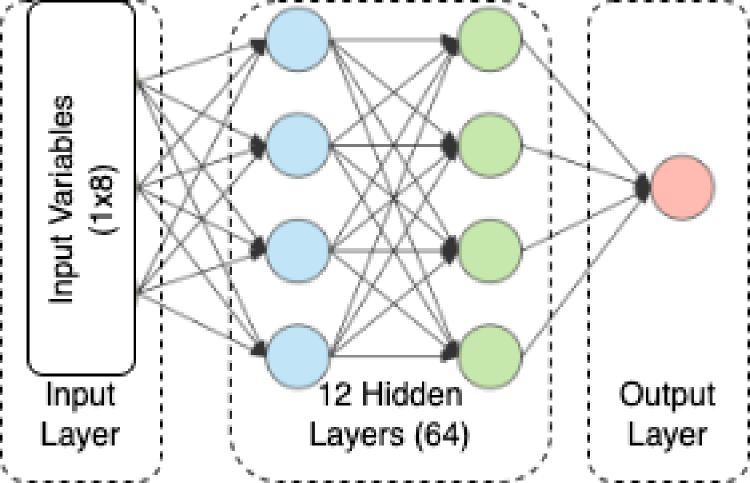
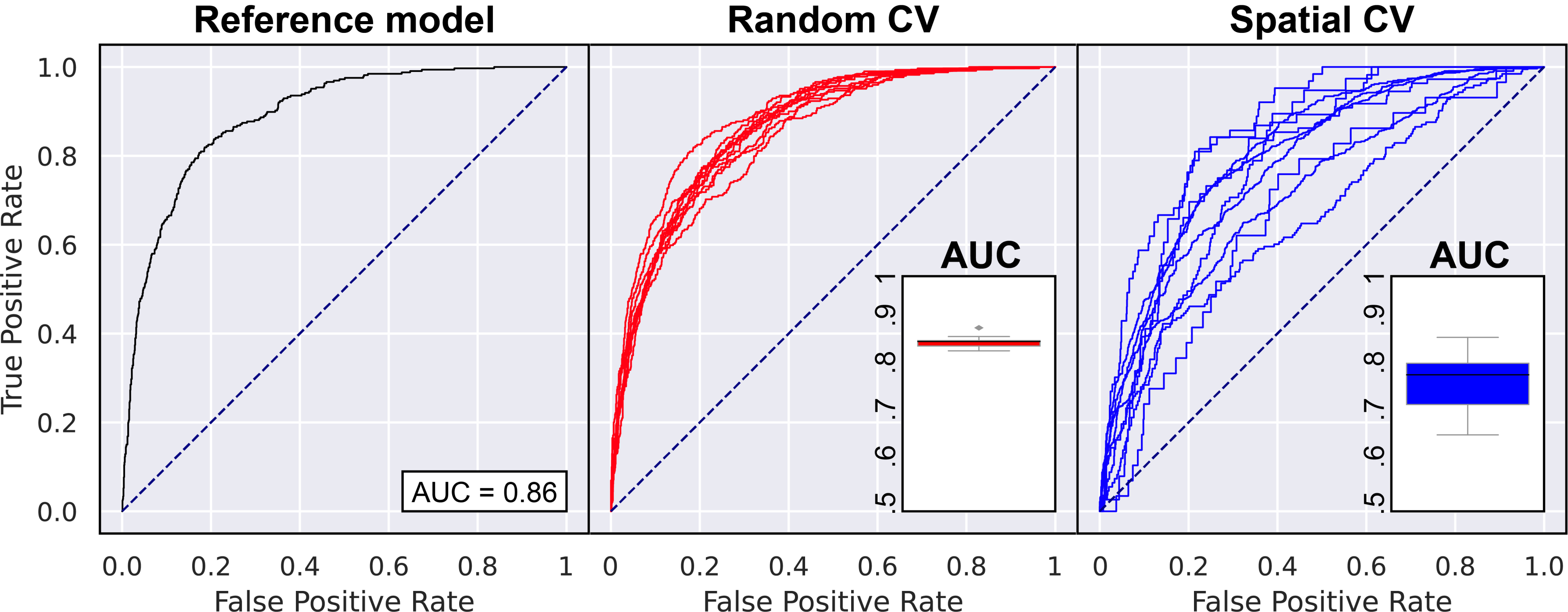

Input Variables (1x8)
Input Layer
12 Hidden Layers (64)
Output Layer

Figure 03.


Validation areas
1
2
3
4
5
6
7
8
9
10
0
25
50 km
28.20
27.90
27.60
27.30
84.30
84.60
84.90
85.20
85.50
85.80

Figure 04.



Reference model
Random CV
Spatial CV
True Positive Rate
False Positive Rate
False Positive Rate
False Positive Rate
AUC = 0.86
AUC
AUC
1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.2 0.4 0.6 0.8 1
0 0.2 0.4 0.6 0.8 1
0 0.2 0.4 0.6 0.8 1.0
.5 .6 .7 .8 .9 1
.5 .6 .7 .8 .9 1

Figure 05.


Reference model
Random CV
Spatial CV
3.16
3.18
3.20
3.22
3.24
3.26
3.28
9.400
9.425
9.450
9.475
9.500
9.525
9.550
9.575
FN
FP
TP
TN
10^1
10^2
10^3
10^4

Figure 06.


Slp_m
PGV_Usgs
NDVI_m
HC_m
Nrt_m
VC_m
Est_m
Prc_m
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
mean(abs(SHAP))

Figure 07.

SHAP
Est_m
Nrt_m
HC_m
VC_m
Slp_m
Prc_m
NDVI_m
PGV_Usgs
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.00 0.25 0.50 0.75 1.00
Count
200
175
150
125
100
75
50
25
0

Figure 08.

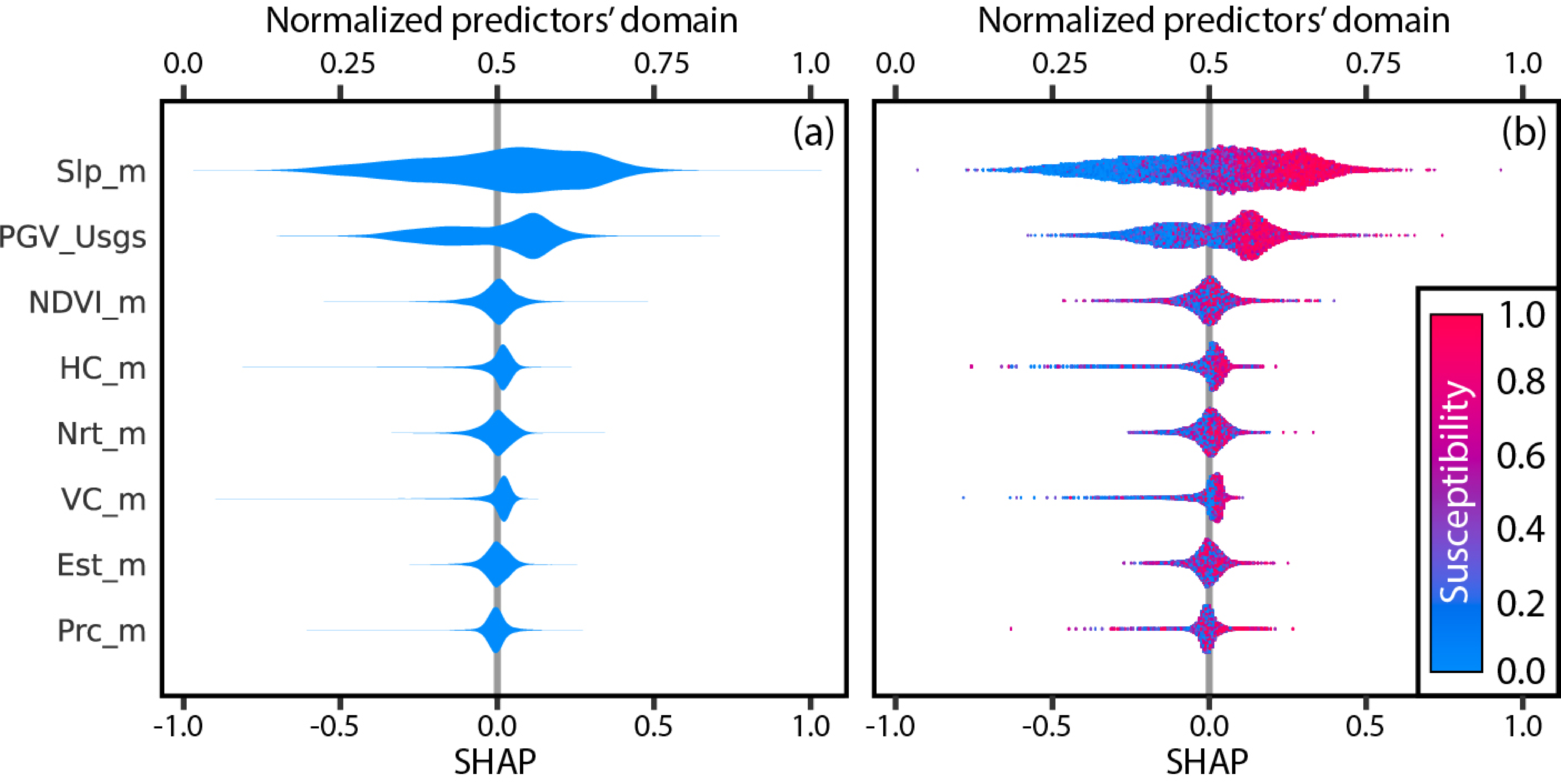
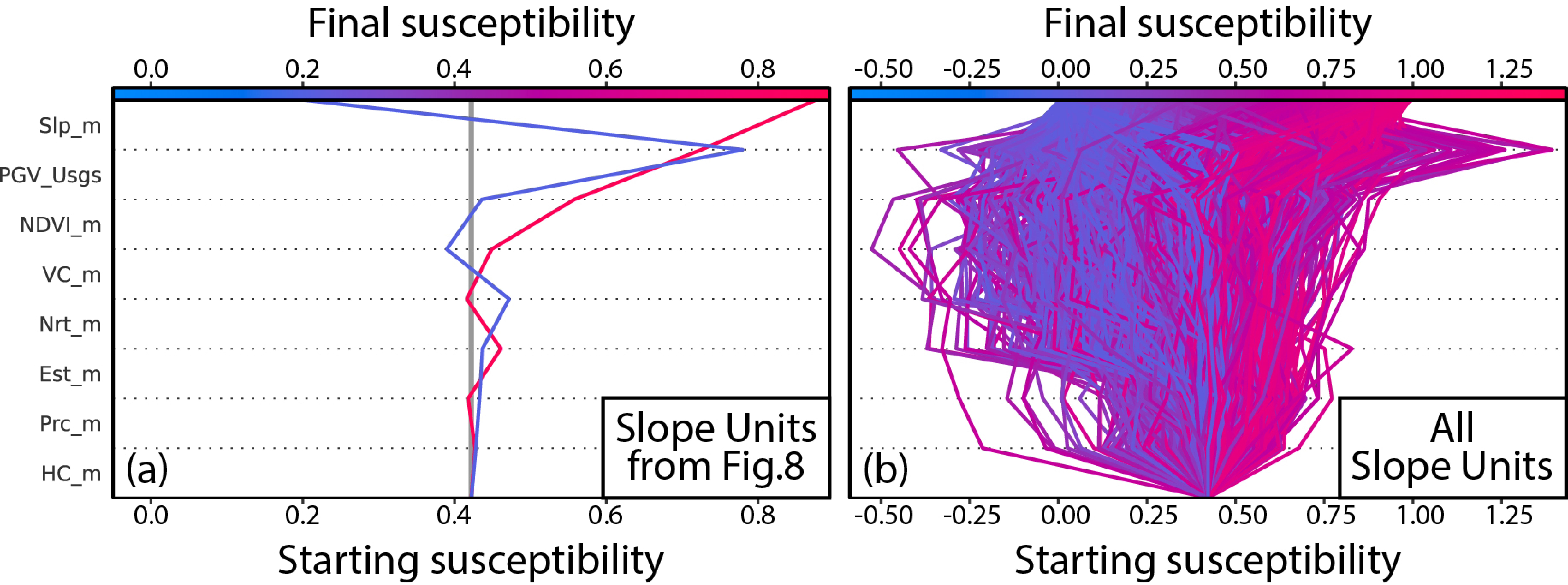

Normalized predictors' domain
0.0
0.25
0.5
0.75
1.0
(a)
(b)
Slp_m
PGV_Usgs
NDVI_m
HC_m
Nrt_m
VC_m
Est_m
Prc_m
-1.0
-0.5
0.0
0.5
1.0
SHAP
Susceptibility
1.0
0.8
0.6
0.4
0.2
0.0

Figure 09.

(a)
Final probability value 0.23
Starting probability value 0.42
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
HC_m
Nrt_m
NDVI_m
PGV_Usgs
Slp_m
VC_m
Ex. stable slope unit
negative ⇄ positive
predictor contribution
Ex. unstable slope unit
(b)
Starting probability value 0.42
Final probability value 0.90
0.4
0.5
0.6
0.7
0.8
0.9
1.0
VC_m
Est_m
NDVI_m
Slp_m
PGV_Usgs
Nrt_m
Prc_m

Figure 10.



Final susceptibility
0.0
0.2
0.4
0.6
0.8
Slp_m
PGV_Usgs
NDVI_m
VC_m
Nrt_m
Est_m
Prc_m
HC_m
(a)
Slope Units from Fig.8
Starting susceptibility
Final susceptibility
-0.50
-0.25
0.00
0.25
0.50
0.75
1.00
1.25
(b)
All Slope Units
Starting susceptibility

Figure 11.

Landslide susceptibility with Explainable AI

Landslide Susceptibility

susceptibility

- > 0.797 - 0.99
- > 0.598 - 0.797
- > 0.398 - 0.598
- > 0.199 - 0.398
- 0 - 0.199

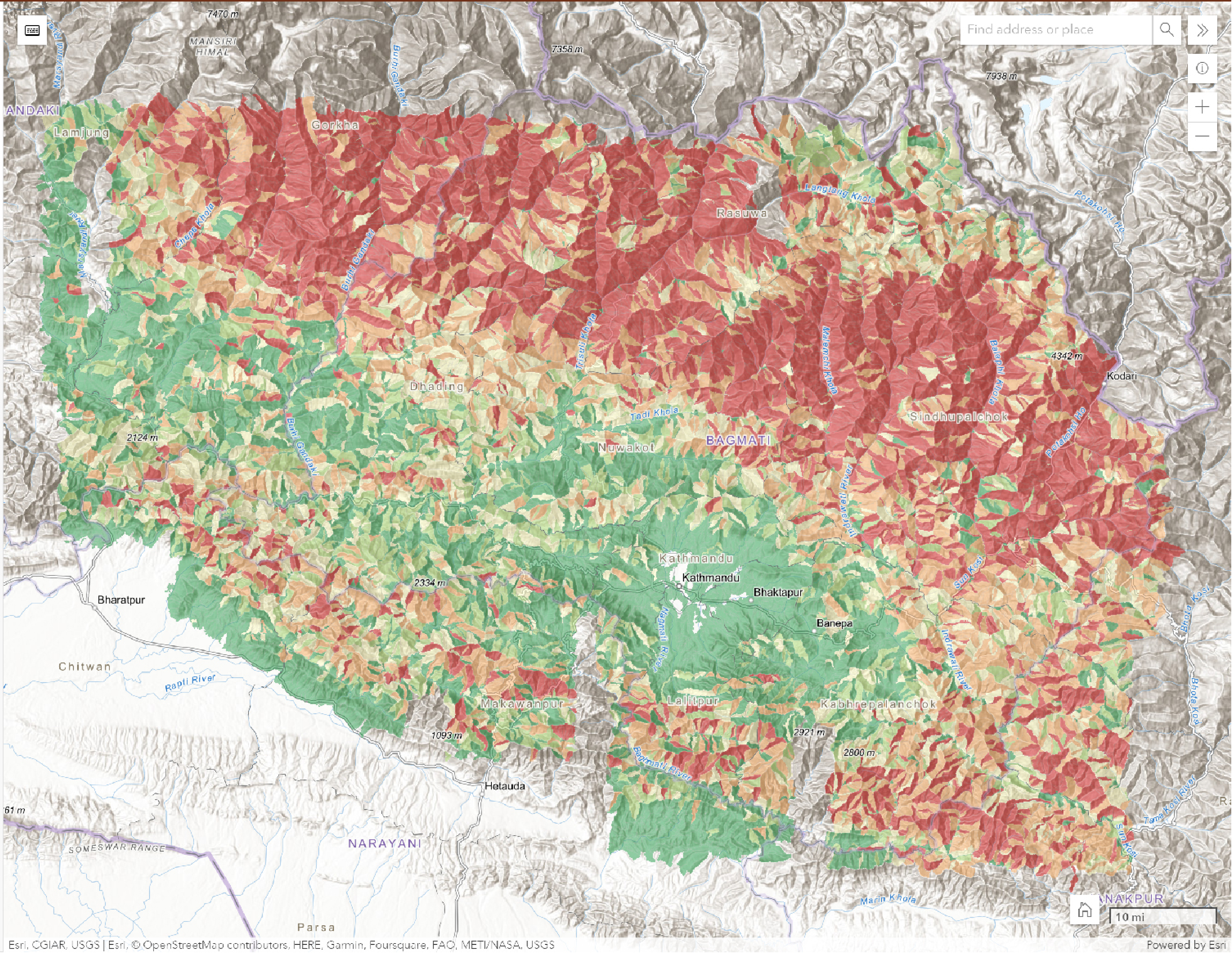
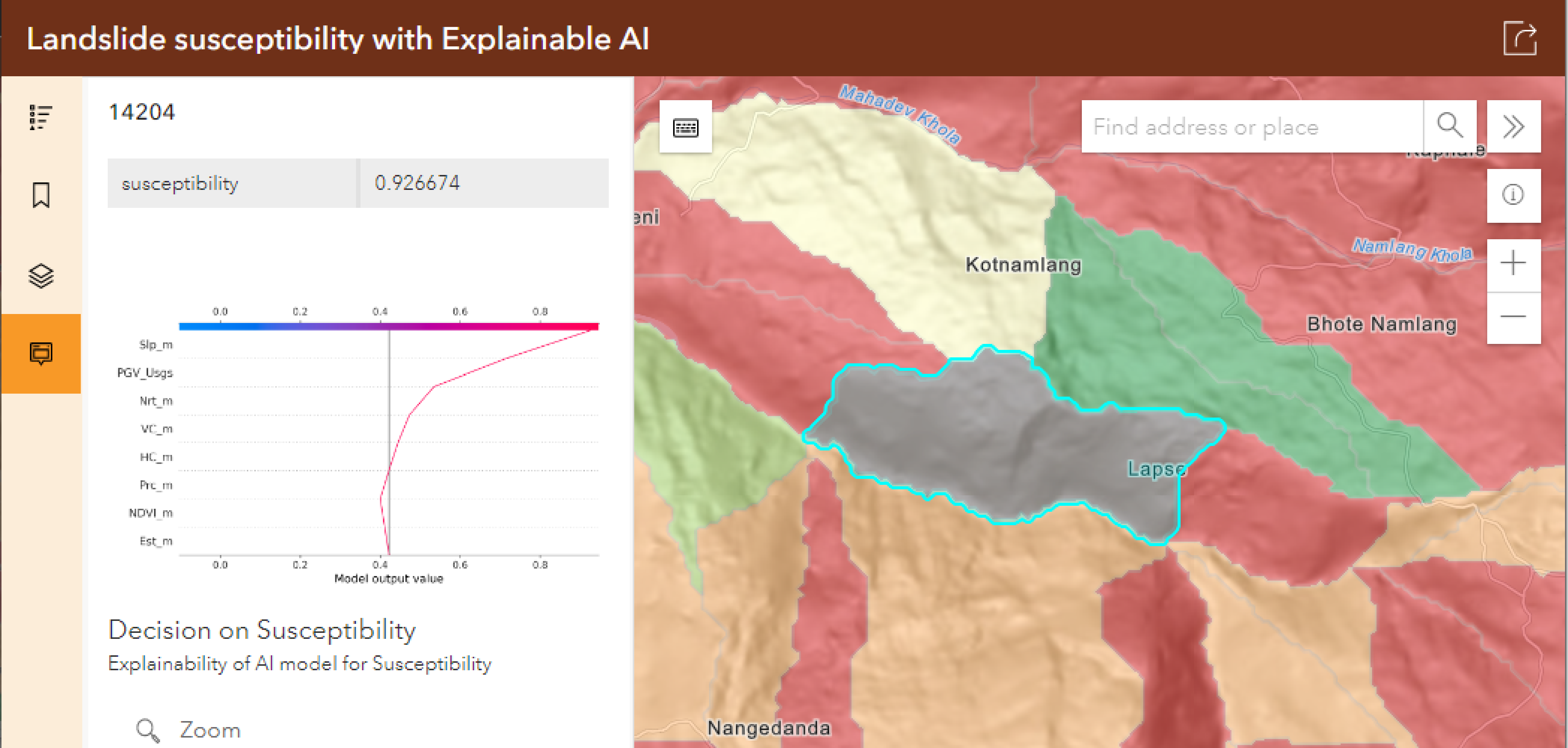


Figure 12.



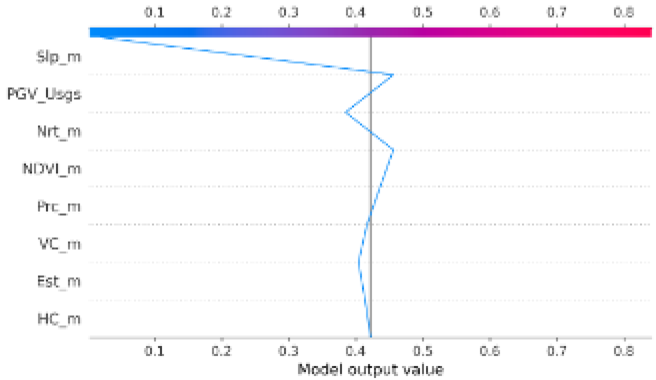
Landslide susceptibility with Explainable AI
14204
susceptibility
0.926674
0.0
0.2
0.4
0.6
0.8
Slp_m
PGV_Usgs
Nrt_m
VC_m
HC_m
Prc_m
NDVI_m
Est_m
Model output value
Decision on Susceptibility
Explainability of AI model for Susceptibility
Zoom
Find address or place
Mahadev Khola
Kotnamlang
Namlang Khola
Bhote Namlang
Lapse
Nangedanda

Figure 13.

Landslide susceptibility with Explainable AI

5272

susceptibility	0.019462



Decision on Susceptibility

Explainability of AI model for Susceptibility

Zoom