Geoscience Frontiers 15 (2024) 101800

Contents lists available at ScienceDirect

Geoscience Frontiers

journal homepage: www.elsevier.com/locate/gsf

Research Paper

On the use of explainable AI for susceptibility modeling: Examining the spatial pattern of SHAP values

Nan Wang [a,b], Hongyan Zhang [a,*], Ashok Dahal [c], Weiming Cheng [b,d,e,f], Min Zhao [g,h,i], Luigi Lombardo [c]

[a] Key Laboratory of Geographical Processes and Ecological Security in Changbai Mountains, Ministry of Education, School of Geographical Sciences, Northeast Normal University, Changchun 130024, China
[b] State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[c] University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), PO Box 217, Enschede, AE 7500, Netherlands
[d] University of Chinese Academy of Sciences, Beijing 100049, China
[e] Jiangsu Center for Collaborative Innovation in Geographic Information Resource Development and Application, Nanjing 210023, China
[f] Collaborative Innovation Center of South China Sea Studies, Nanjing 210093, China
[g] State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China
[h] Key Laboratory of Environmental Change and Natural Disaster, Beijing Normal University, Beijing 100875, China
[i] Center for Geodata and Analysis, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Article history:
Received 18 September 2023
Revised 27 December 2023
Accepted 24 January 2024
Available online 1 February 2024
Handling Editor: Biswajeet Pradhan

Keywords:
Hydro-morphological processes
SHAP maps
Explainable AI
China

ABSTRACT

Hydro-morphological processes (HMP, any natural phenomenon contained within the spectrum defined between debris flows and flash floods) are globally occurring natural hazards which pose great threats to our society, leading to fatalities and economical losses. For this reason, understanding the dynamics behind HMPs is needed to aid in hazard and risk assessment. In this work, we take advantage of an explainable deep learning model to extract global and local interpretations of the HMP occurrences across the whole Chinese territory. We use a deep neural network architecture and interpret the model results through the spatial pattern of SHAP values. In doing so, we can understand the model prediction on a hierarchical basis, looking at how the predictor set controls the overall susceptibility as well as doing the same at the level of the single mapping unit. Our model accurately predicts HMP occurrences with AUC values measured in a ten-fold cross-validation ranging between 0.83 and 0.86. This level of predictive performance attests for an excellent prediction skill. The main difference with respect to traditional statistical tools is that the latter usually lead to a clear interpretation at the expense of high performance, which is otherwise reached via machine/deep learning solutions, though at the expense of interpretation. The recent development of explainable AI is the key to combine both strengths. In this work, we explore this combination in the context of HMP susceptibility modeling. Specifically, we demonstrate the extent to which one can enter a new level of data-driven interpretation, supporting the decision-making process behind disaster risk mitigation and prevention actions.

1. Introduction

Hydro-morphological processes (HMP) define a large spectrum of phenomena that include debris flows, debris floods, flash floods, etc., essentially reflecting the dynamics of a mixture of water and debris moving under the effect of gravity (van den Bout et al., 2023). Because of their impulsive and stochastic nature, they can pose a significant threat to most global communities (Kobiyama and Goerl, 2007). As a result, HMP prediction is one the most emergent topics among researchers working on natural hazards (Gariano and Guzzetti, 2016). Historically, this has been attempted and achieved with satisfying results through statistical methods, in the case of debris flows (Carrara et al., 2008), mud flows (Ozdemir, 2009), earth flows (Can et al., 2005), debris floods (Santangelo et al., 2011), flash floods (Marchi et al., 2010) and even riverine floods (Merz et al., 2009). These approaches share some degree of dissimilarity, but they also have something consistently in common: the need to understand the given HMP under consideration and predict its occurrence probability. The term "understand" here refers to the inference that statistical solutions offer when explaining the distribution of HMP presences and absences in space (or

* Corresponding author.
E-mail address: zhy@nenu.edu.cn (H. Zhang).

https://doi.org/10.1016/j.gsf.2024.101800

more rarely in space and time, Fang et al., 2023a,b) according to a set of predictors (Amato et al., 2023). However, statistical models are not performance-oriented tools, which is the reason why recent advancements in artificial intelligence have produced valid alternatives (e.g., Merghadi et al., 2020). In such cases, machine and deep learning models are employed to maximize the HMP prediction capacity (Kern et al., 2017). However, this happens at the expense of interpretation. In fact, most of the standard machine learning models become so complex that it is impossible to understand why a given probability has been assigned to a given mapping unit (Korup and Stolle, 2014; Goetz et al., 2015). Only in recent years, the computer science community has worked out potential solutions to combine the performance of machine/deep learning and the interpretation of statistical modelling, giving birth to the concept of explainable AI (XAI; Gunning, 2017; Samek et al., 2017). As a result, XAI has started to attract the attention of researchers even in the field of natural hazards, in the hope of performing predictive tasks with high precision but also understanding the processes underlying the observed data (Li, 2022; Tehrani et al., 2022).

The probabilistic estimation of locations prone to experience HMPs is a notion commonly referred to as susceptibility mapping (Guzzetti et al., 2006) and constitutes an integral part of the hazard and risk standard definitions (e.g., Fell et al., 2008; Domeneghetti et al., 2013). In a data-driven context, the susceptibility is usually quantified using statistical models that either linearly or nonlinearly relate the effect of a set of covariates to the distribution of presence/absence hazard data in the study area. The simpler case belongs to the family of Generalized Linear Models (GLMs), which still constitute the most common method in the literature (Reichenbach et al., 2018; Lima et al., 2022). As for more flexible approaches, these are usually built in the framework of Generalized Additive Models (GAMs; Brenning, 2008; Di Napoli et al., 2023). The regression coefficients estimated for each covariate lead to the model interpretation in both cases. For GLMs, this is done by examining the sign and magnitude of a single regression coefficient (Brenning, 2005; Lombardo and Mai, 2018). In contrast, for GAMs, this is done over a number of regression coefficients that together define a function associated with each covariate (Loche et al., 2022b; Steger et al., 2022). The role of each model component is then interpreted by reading the sign of the coefficients, with positive values indicating a marginal (assuming all other covariates contributions are fixed) increase of the final susceptibility and negative values indicating the opposite (Shirzadi et al., 2017; Loche et al., 2022a). Another appealing advantage of statistical-based models is their capability to capture and display spatial effects (Song et al., 2020), such as spatially varying coefficients models (e.g., Geographically Weighted Regression, Fotheringham et al., 2003) or (e.g., Spatially Varying Regression, Opitz et al., 2022). However, restricted by the data size and the relationships' complexity, statistical models are usually computationally challenging when dealing with big spatial data (Lombardo et al., 2019).

This level of understanding is generally lost in the case of machine learning tools, where the prediction rule becomes so complex that even visualizing it does not really help understand why the stable or unstable label was assigned to a given catchment (e.g., Yeon et al., 2010). In this context, local interpretation methods such as LIME (Local Interpretable Model-agnostic Explanation; Ribeiro et al., 2016), and SHAP (SHapley Additive exPlanations, Lundberg and Lee, 2017), offer the opportunity to flexibly model, visualize and interpret complex geographical phenomena. Rather than providing the feature importance for the whole model, local interpretation methods allow giving detailed feature contributions at the level of each mapping unit. As a result, the integration of machine/deep learning tools with locally interpretable techniques (Collini et al., 2022; Chang et al., 2023; Dahal and Lombardo, 2023) has been explored in a number of geographical studies (Lubo-Robles et al., 2020; Li, 2022; Ullah et al., 2023). These achievements open up a new explainable modeling avenue built by computing and visualizing the SHAP patterns in space, and ultimately by interpreting individual predictions.

China has suffered severe destructive HMPs in recent years (He et al., 2018; Liu et al., 2018a; Wang et al., 2020). Therefore, it is important to use this unfortunate information and understand which areas may undergo analogous disasters in the years to come. The Chinese geoscientific community has worked together for this objective, producing a number of documents where the susceptibility to HMP has been assessed at various scales (Lin et al., 2022; Wang et al., 2022b). Following the international trends where machine learning solutions are the preferred architectures to solve prediction tasks, most of the national efforts have prioritized performance (e.g., Zhao et al., 2022a). However, seeking model performance only highlights susceptible locations, thus neglecting the required knowledge necessary to understand why HMP may hit specific areas rather than others. In turn, this implies that decision-makers may not be sufficiently supported in planning suitable mitigation actions. For this reason, we test the extent to which deep learning solutions can be explained by examining the SHAP results and their spatial pattern across the whole Chinese territory. Specifically, due to the continental scale of the study area, we opted for a catchment partition, assigning the presence label if at least one HMP has been locally recorded in the Chinese HMP catalogue (more details in Wang et al., 2021). To offer an interactive experience for the reader, we also created a web-GIS platform where our model results can be queried and used to understand the potential of explainable AI tools.

The paper is organized as follows: Section 2 presents the HMP data, the mapping unit and the variables used in this study; Section 3 describes the adopted methodology for the susceptibility model and how to produce interpretable deep learning results. The analytical protocol we implemented is outlined in Section 4, from calibration to performance assessment and model explanation. In Section 5, we explore the implications of local interpretation and the possible improvements to this work. Ultimately, the conclusions are drawn in Section 6.

2. Materials

2.1. HMP inventory

In this work, we accessed the digital collection of HMP records put together thanks to the China National Flash Flood Disasters Prevention and Control Project (Liu et al., 2018b, 2021; Xiong et al., 2019, 2020). This project is a large-scale national initiative that has involved many administrations and research centers across China, to collect, standardize and digitize HMP occurrence data in the last fifty years. Here, we selected HMP locations mapped between 1985 and 2015, and only kept the records with a complete metadata description (x, y, and time in year-month-date format). We adopted this filter to remove noisy and imprecise information, leading to 24,956 selected HMPs (Fig. 1).

2.2. Mapping units

The choice of a suitable mapping unit boils down to three criteria. The first links the mapping unit to the process one wants to model. For instance, landslides are often modeled at the slope unit scale because half-basins can reflect the morphodynamic response to slope failures (Carrara et al., 1995; Alvioli et al., 2022). Conversely, HMPs can manifest, travel and develop involving whole

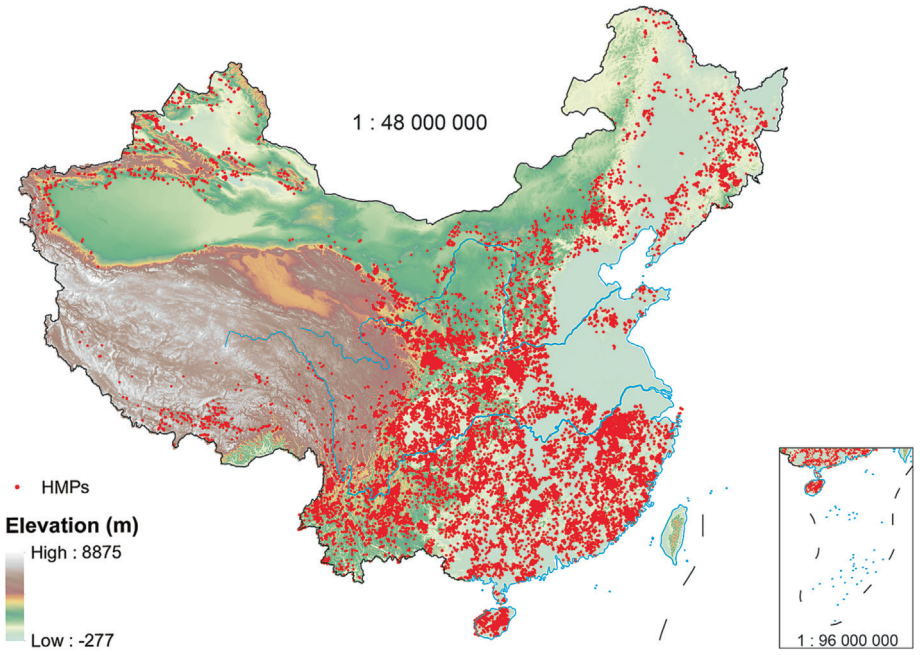

Fig. 1. Geomorphological settings of HMPs in China.

catchments, thus making these units the most appropriate choice for flow-type hazards (Lin et al., 2021; Wang et al., 2022b).

The second criterion relates to the computational burden a given mapping unit choice inevitably leads to. For instance, choosing an extremely small mapping unit compared to the extent of the study area may lead to data matrices made of several million rows (or billions of elements overall). Such dimensions are computationally challenging and either may end up limiting the complexity of the model one may choose or impose the need for dedicated computational facilities (Lombardo et al., 2020). The third criterion consists of the data aggregation step required for medium to coarse mapping units. Remote sensing technologies lead to characterize the earth's surface on a very fine scale. For instance, global digital elevation models are now expressed at the scale of a few meters (Moreira et al., 2004). As a result, from thousand to million pixels may be contained in a single catchment. Therefore, one usually needs to summarize the distribution of values expressed at the pixel scale to a much coarser hierarchical level (e.g., Jacobs et al., 2020). This is usually done by computing mean and standard deviation values, but one can also opt for a much more detailed quantile description at times (e.g., Camilo et al., 2017).

In this study, we selected a catchment partition, by using the Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales (HydroSHEDS database, https://hydrosheds.org/). This data contains several levels of details, from which we selected the 12th level. This resulted in a partition made of 73,587 catchments for the whole Chinese territory. The catchment size spans from 0.1 km^2 to 667 km^2, with an average area of 130 km^2 and a 95% confidence interval of 231 km^2.

2.3. Environmental variables

We chose our predictor set to reflect the environmental conditions responsible for the HMP hazard occurrences, listing terrain, climatic and anthropic influences. As also introduced before, the native covariate resolution differed among covariate groups, and was also inconsistent with respect to the catchment partition. We then adopted the strategy of calculating the mean values per catchment for the following numerical predictors: elevation, slope, planar and profile curvatures. Stream/catchment features (including form factor (Horton, 1932), relief ratio (Schumm, 1956), elongation ratio (Schumm, 1956), and drainage density (Strahler, 1952) are morphometric characteristics representative of the catchment hydrology, thus they did not require any aggregation step. As for NDVI (Normalized Difference Vegetation Index), settlement area and rainfall, these required a dual aggregation step, calculating the respective mean values over 30 years and then per single catchment. Notably, we could have also calculated standard deviation values but the interpretation of such summary statistics becomes very difficult. Because in this work we seek a clear explanation of the predictors' role, we opted to leave out these measures, the additional information they would introduce to the model, and the possible performance increase this would imply. Therefore, we selected a total of 12 variables, whose acronyms and sources are reported in Table 1.

Table 1
Overview of environmental variables used in this study.

Variable	Description	Source
Elv	mean of elevation	SRTM, https://earthexplorer.usgs.gov/
Slp	mean of slope	
Prc	mean of profile curvature	
Plc	mean of plan curvature	
Rr	relief ratio	HydroSHEDS, https://hydrosheds.org/
Ff	form factor	
Er	elongation ratio	
Dd	drainage density	
Wr	wandering ratio	
NDVI	mean of NDVI	GIMMS NDVI, https://data.tpdc.ac.cn/
MaxRain	maximum daily rainfall	Meteorological Data, https://data.cma.cn/
Sa	settlement area	WSF2015, https://developers.google.com/earthengine/datasets/

3. Methodology

The modeling protocol we followed includes two steps, one where a "black box" neural network is built to produce HMP susceptibility estimates and a second one where the box gets opened for interpretation calculating SHAP values and assessing their spatial patterns per predictor. These are illustrated in Fig. 2, through a general flowchart.

3.1. Susceptibility model

Deep learning models have been proven to be effective in susceptibility modeling in recent studies (Bui et al., 2020; Panahi et al., 2021; Zhao et al., 2022a). To demonstrate the explainability of our model, we opted for an Artificial Neural Networks (ANN; Yilmaz, 2009), although we stress here that SHAP values (the building blocks of explainable AI; Baptista et al., 2022) can be computed even for other data-driven approaches such as random forest (e.g., Titti et al., 2022) or support vector machine (e.g., Yu et al., 2012) to mention a few.

The basic structure of our ANN model consists of nodes and connections that are organized into three layers, i.e., the input layer, the hidden layer, and the output layer. Among them, the hidden layer is used herein to prevent the ANN from falling into bad local minima (De Villiers and Barnard, 1993). In this work, we kept the structure and parameters of the ANN model to be simple, with 12 variables in the input layer, together with 12 hidden layers made out of fully connected layers of size 64 and an output layer with a sigmoid activation function (Albawi et al., 2017). We implemented a ReLU non-linear activation and adopted 30% dropout in a dropout layer, which could be used to prevent overfitting (Li and Yuan, 2017).

As for the explainable component, we used DeepLIFT, and more details are provided in Section 3.2.

3.1.1. Model calibration

We randomly divided the dataset into the training (70%) and testing (30%) parts. In each training epoch, 20% of the training dataset was further randomly selected with replacement to evaluate the training performance. The model was trained via a weighted binary cross-entropy loss function, and some of the important parameters were set as follows:

- optimizer: Adam optimizer
- learning rate: 0.001
- decay steps: 10,000
- decay rate: 0.9
- early stopping option: 500

3.1.2. Model validation

The model performance was evaluated on the testing dataset to monitor the generalization ability stemming from the calibration. We recall that the input of a susceptibility model is a vector of presence/absence data, i.e., an array of zeroes and ones. However, the output is not discrete but rather continuously expressed in probabilities. Therefore, to assess the performances of any binary classifier, the first requirement is always the classification of the probability spectrum into a sequence of binary information to be matched against the initial presence/absence observation. This procedure entails the selection of a probability cutoff and for this reason, performance metrics of binary classifiers either fall in the cutoff-dependent or cutoff-independent categories. Here we use both criteria, using a single confusion matrix for the cutoff-dependent analyses. A confusion matrix is made of four elements, reflecting all possible combinations between observed and predicted presence/absence data (Townsend, 1971). As a result, one can define True Positives (TP) and Negatives (TN) for presences and absences that are respectively matched. As for False Positives (FP) and Negatives (FN), these two correspond to model errors, for misclassified absences and presences, respectively. Therefore, it is of utmost importance to select an appropriate probability

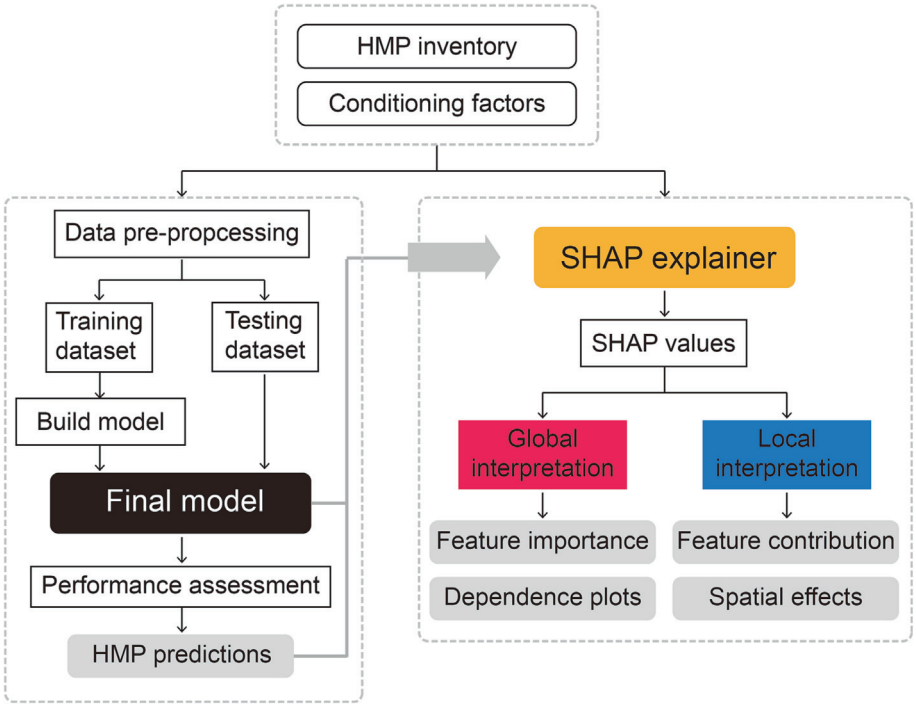

Fig. 2. Flowchart of the methodology in this study.

cutoff, as a wrong choice can drastically change the confusion matrix. For balanced datasets (equal number of presence absences) a straightforward choice is to set the cutoff at 0.5 because the resulting probability distributions are typically bell-shaped. However, in case of unbalanced data, the resulting probability distributions become heavily skewed, with the predominant class pulling the probability spectrum (Ramyachitra and Manikandan, 2014). The latter case is the typical situation one may find in HMP datasets (and luckily for most natural hazards) because the number of occurrences is much lower than the number of absences (Frattini et al., 2010). To address this issue, we opted for a two-stepped approach. The first step is actually part of the model architecture where we used a class-weight binary cross-entropy criterion (Aljohani et al., 2021). This criterion allows one to add a penalty to the model's error measured on the class of interest. In our dataset, the number of absences is approximately seven times the number of presences. Therefore, the model would naturally learn to recognize zeroes (absences) better than ones (presences). However, this issue can be addressed by increasing the weight of the error in the classification of the unstable catchments (by a factor of seven in our case), effectively minimizing the unbalance in the data proportion. In the second step, we a posteriori used a standard procedure based on the Youden Index to select the best probability cutoff (Fluss et al., 2005). We recall here that the Youden Index can be calculated as Eq. (1):

$$J = \frac{TP}{TP + FN} + \frac{TN}{TN + FP} - 1 \quad (1)$$

However, even if the retrieved cutoff is the best numerical solution, it still remains only one of the possible solutions. For this reason, we complemented this cutoff-dependent assessment together with Receiver Operating Characteristic (ROC) curves and their integral (AUC, Area Under the Curve ROC) for the cutoff-independent analyses. These curves are generated by plotting pairs of FP/(FP + TN) and TP/(TP + FN) computed for a large number of possible probability cutoffs. As a result, the function linking all pairs sorted by cutoff can be used to calculate its integral, whose resulting value (AUC) indicates how the model performed irrespectively of any specific cutoff. These metrics have then also been assessed over a bootstrapping procedure that randomly selected a 10% subset from the total for further testing.

3.2. Explainable model

The most important goal of explainable deep learning models is to demonstrate how the predictions are reached, highlighting the role (Li, 2022). SHapley Additive exPlanations (SHAP), which originated from the game theory, can be used to quantify the contribution of each predictor to the model (Strumbelj and Kononenko, 2014) (Fig. 3). Therefore, we computed SHAP values (Lundberg and Lee, 2017b) for each catchment partitioning the Chinese landscape, allowing to summarize predictors' contributions to the global model and also their relevance at the scale of a single mapping unit. We recall here that SHAP values can be estimated using a number of approaches ranging from Kernel SHAP (e.g., Roshan and Zafar, 2022), Tree SHAP (e.g., Wang et al., 2022a), and Deep SHAP (e.g., Singh et al., 2020). Among these, the latter consists of a high-speed approximation algorithm for SHAP values, whose estimates are reached through a DeepLIFT (Deep Learning Important FeaTures) approach (Panati et al., 2022). Specifically, DeepLIFT is a method used for decomposing the output of a neural network on a specific input by back-propagating the contributions of all neurons in the network to each feature of the input Shrikumar et al., 2017). SHAP values' main strength is to generate locally additive feature attribution via the Eqs. (2) and (3).

$$\widehat{y_i} = shap_0 + shap(X_{1i}) + shap(X_{2i}) + \cdots + shap(X_{ji}) \quad (2)$$

$shap_0 = E(\widehat{y_i})$ (3) where $\widehat{y_i}$ is the model prediction for the catchment i, $shap_0$ is the mean value of predictions cross all catchments, and $shap(X_{ji})$ is the SHAP values of the j^{th} variable for the catchment. In this way, the SHAP values start from the initial intercept value $shap_0$, which is the mean value of all predictions, and then add the least contributed term $shap(X_{1i})$, followed by the second least $shap(X_{2i})$, and so on. Finally, the absolute SHAP value reflects each variable's importance for the final prediction (Molnar, 2020).

In this work, we implemented SHAP in open source python package ("shap").

4. Results

In this section, we will initially look into an overall assessment of model performance, and later dive into global and local interpretations of the established model. As part of the last procedure, we will also present a step that even other recent explainable AI contributions in natural hazard research have not yet explored (Dikshit and Pradhan, 2021; Zhou et al., 2022). This corresponds to the ability to generate maps of SHAP values for each predictor under consideration and plot them according to the spatial variation each SHAP map presents. Notably, this is done for each catchment partitioning the study area, assigning a colorbar to each SHAP map and allowing for further interpretation. The resulting geographic overview offers a unique perspective on variable contributions and we believe this to be an important element that future explainable AI solutions should be equipped with. This section will be concluded with the estimated susceptibility map.

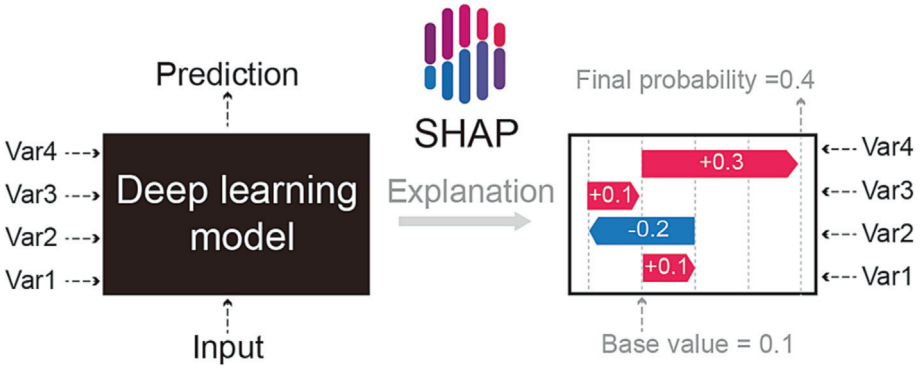

Fig. 3. An illustration demonstrating the SHAP-explained deep learning models (modified from Lundberg and Lee, 2017a).

4.1. Model performance

Our neural network architecture produced performance in the range of excellent results according to the classification system proposed by (Hosmer et al., 2013). This is shown in Fig. 4a, where this panel contains both the ROC curves generated from the random cross validation procedure as well as the AUC values estimated at each bootstrap replicate. The latter is summarized with a boxplot where the median AUC is 0.85, and the two extremes of the AUC distribution are confined between 0.83 and 0.86. As described in Section 3.1.2, this is a perspective independent of the probability cutoff one may opt for to translate susceptibility values back into presence/absence classes. To complement this assessment, we also report the probability density function of the susceptibility spectrum, together with the estimated Youden Index (Y = 0.52) in Fig. 4b. This cutoff leads to the confusion matrix and confusion maps (Nicu et al., 2023) shown in Fig. 4c. There, we summarize the frequency distribution for each class of the confusion matrix and plot the corresponding geographic distribution expression across China. We recall here that this confusion matrix relates to the predictive performance assessment. What we observe is that the classification generally reflects the original distribution of presence/absence HMP data, with the dominant class represented by TN. However, the high number of TP (7347 out of 8821 = 83% accuracy) and low number of FN (the complementary 17%) indicate the model's ability to recognize susceptible catchments. In turn, this implies that the FP catchments (15,862 out of 64,768 = 24%) highlighted in the confusion map may surely be the result of a model error. But, they may also represent locations that the model actually recognizes to likely host HMPs in the future. Answering the question as to whether these FP may be due to misclassification or if they may actually be susceptible but have not yet experienced HMP occurrence is not straightforward. However, examining FP actually constitutes the reason behind susceptibility modeling, and the accuracy we observed in recognizing presence data warrants trusting the model prediction. Notably, these are mostly located in the central and southeast sectors of China.

4.2. Model interpretation

4.2.1. Global interpretation

The most traditional way to understand how a machine-learning models work is to list the variable importance ranking (e.g., Band et al., 2020; Hosseini et al., 2020; Zhao et al., 2022b). Here, we also produce the same graphics in Fig. 5 but use SHAP values to sort each predictor according to the impact it may have over the final susceptibility. Among all the variables we considered, NDVI, settlement area, maximum daily rainfall, elevation, and slope steepness appear to be the dominating ones. One of the interesting aspects of using SHAP values instead of traditional variable importance is that SHAP is not bound to positive values, but it ranges from negative to positive ones. The way to read SHAP values essentially matches the interpretation of regression coefficients in statistical models. The magnitude of the SHAP value indicates the influence on the final susceptibility whereas the sign indicates whether the given predictor contributes to increasing or decreasing the probability estimates. For instance, most of the predictors have a positive contribution to the pattern of relative probabilities in space. This is not the case for the elongation rate (Er) of the catchment as well as the planar (Plc) and profile (Prc) curvatures.

An additional solution to assess variable contribution in traditional machine learning consists of response plots (e.g., Park, 2015). Here we also produce an analogous illustration but again as a function of SHAP values. Specifically, we plot the SHAP estimates against the normalized variables' domain for each catchment and for each predictor under consideration. This is shown in Fig. 6, where the resulting scatterplots present the marginal effects (assuming all other covariate effects to be fixed) adding another dimension to the static view offered by the variable importance. Here we can distinguish portions of each variable domain and how they individually contribute to increasing or decreasing the susceptibility. For instance, NDVI, maximum daily rainfall (MaxRain), and form factor (Ff) revealed a weak positive effect on the HMP occurrences, whereas the elongation ratio showed a slightly negative association with HMPs.

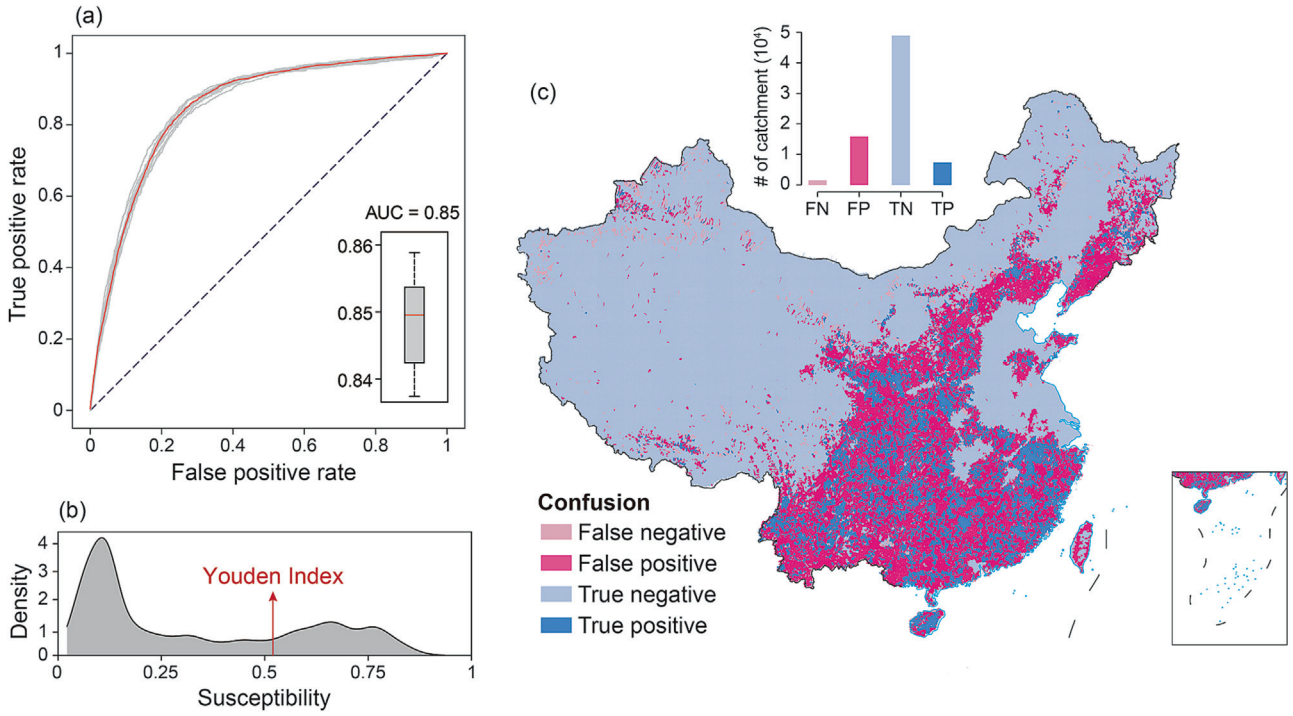

Fig. 4. The ROC curves (a) and confusion map (b) for the validation model.

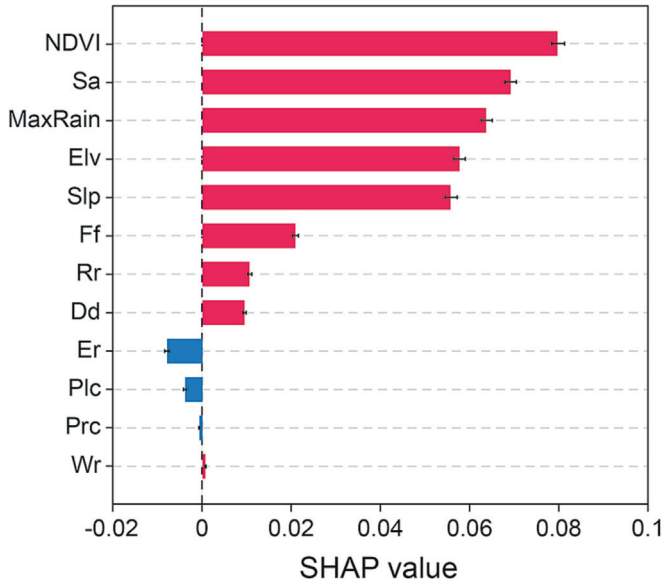

Fig. 5. Variable importance expressed in terms of SHAP values.

This plot essentially corresponds to the limit of model explainability of traditional machine learning studies. The next session is dedicated to further exploring predictors' effects and understanding their contribution to the HMP susceptibility model.

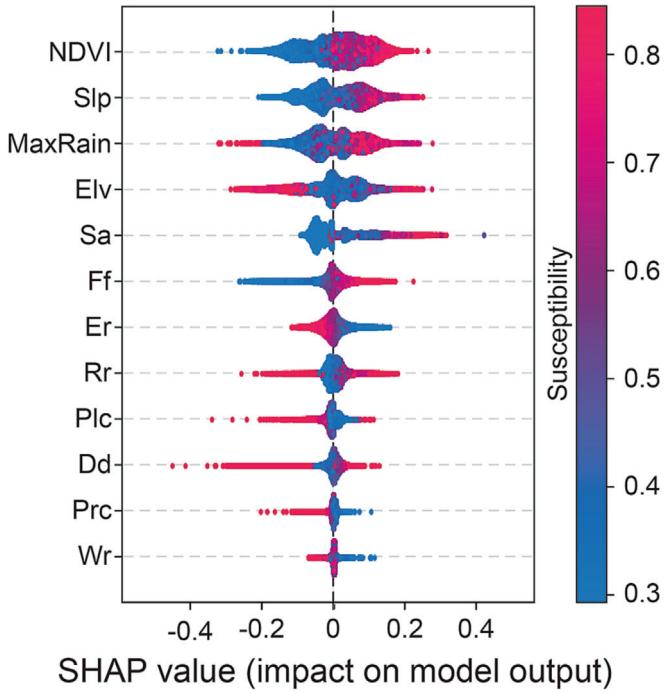

Fig. 7. The SHAP value distribution for each variable against the susceptibility. Each dot corresponding to a specific catchment, the color map showed the final susceptibility.

4.2.2. Local interpretation

The first step to deepen our understanding of the model results focuses on moving from global to individual catchment predictions. Fig. 7 illustrates an intermediate level between the two options by plotting SHAP values for each normalized predictor domain. This further adds another exploratory dimension by plotting the actual susceptibility estimate for each catchment in a violin plot. In such a way, one can quickly visualize whether a given predictor behaves linearly or not. For instance, the elongation ratio shows high susceptibility values on the left side of the violin plot, transitioning to low probabilities at greater elongation ratio values. Conversely, elevation is initially associated with high susceptibil-

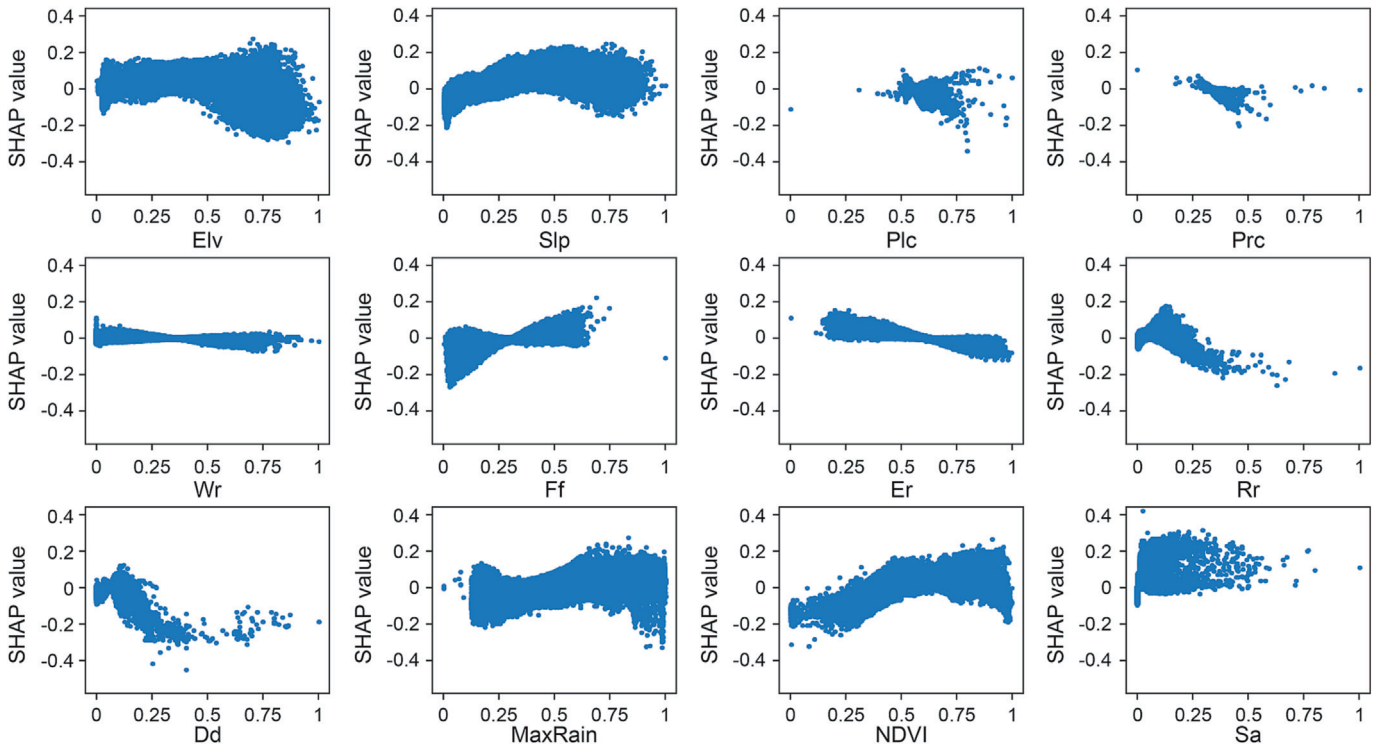

Fig. 6. Scatter plots for each variable used in the model.

ity, then moves to non-susceptible catchments and transitions to the right side of the violin to high susceptibility once more.

Fig. 8 is the first level of localized interpretation of the model results. This plot is built by showing the base and final probabilities for two random catchments, highlighting how each predictor has contributed to the final susceptibility estimate. We recall here that the base probability value is analogous to a model intercept for a statistical model and its definition depends on the proportion of presence/absence data across the whole study area (Frattini et al., 2010; Petschko et al., 2014). For instance, Fig. 8a and 8b both start from the same probability value of 0.32 and respectively reach a final susceptibility of 0.21 and 0.52. The magnitude and sign of each predictor contributing to this value change are color coded in the figure, with the actual numerical variation written to further improve readability. It is important to stress that the same variable does not bring the same level of change to the two catchments. For instance, elongation ratio (Er) has a much larger contribution in Fig. 8a than it has in Fig. 8b. This is a characteristic of SHAP values, as they essentially visualize the combinations of predictor weight and relative predictor value for each individual mapping unit. As explanatory as this illustration may be, it is difficult to use this level of detail for each catchment.

For this reason, another level of model exploration is offered by computing the combination of each predictor contribution and plotting the ranked probability from the base value to the final one, for each catchment. This provides an alternative option for end users to look into how the susceptibility varies, and for the whole Chinese HMP susceptibility, this can be visualized in Fig. 9. Implications of the information conveyed will be presented in Section 5.

So far, this level of model explainability was already presented in few recent articles, featuring mostly susceptibility studies (e.g., Pradhan et al., 2023a, 2023b; Sun et al., 2023; Zhang et al., 2023) and even some interesting InSAR deformation experiments (Al-Najjar et al., 2023). However, what they all missed is translating the information offered by the SHAP values across the geographic space, which is what we will present in the next section.

4.3. Geographic view of predictors' effects

As mentioned above, the strength of using SHAP seen so far for model explainability can be taken a step further. Here we propose to do so by looking into the spatial patterns of SHAP values for each predictor. Such a procedure can offer the added value of hierarchically understanding not only the variable at the global and individual catchment level but also exploring relative contributions and how they vary across the Chinese landscape. This is shown in Fig. 10. There, with the exception of the wandering ratio, form factor, elongation ratio, and relief ratio, all other variables' impacts on susceptibility showed distinctive spatial patterns. For instance, this is evident in the positive influence of elevation across the Yungui Plateau and Hexi Corridor (Fig. 10a). In the most mountainous areas, the slope exhibited a positive impact on HMPs, and in the plain areas, it showed a negative impact (Fig. 10b). As for the maximum daily rainfall, a positive contribution can be observed in eastern China (Fig. 10j), and a similar pattern can also be detected in the NDVI (Fig. 10k).

The combination of all the exploratory tools we present here is what we believe can become a new standard for the future generation of landslide susceptibility studies.

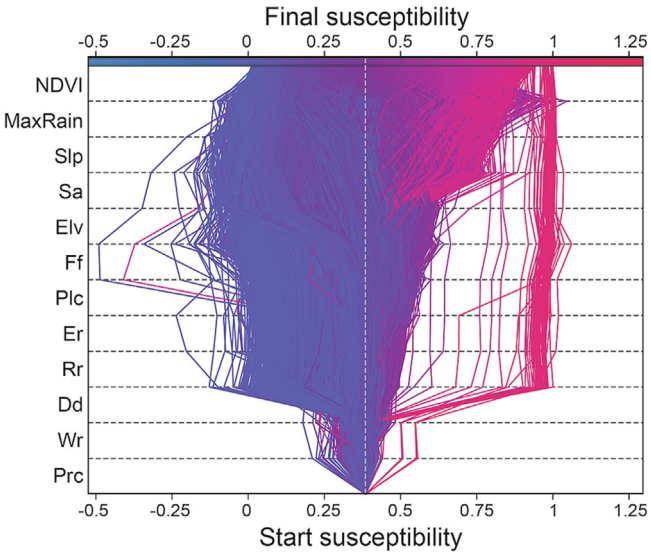

Fig. 9. The variation of the probability estimates for all catchments partitioning the study area.

4.4. Susceptibility mapping

Ultimately, we summarized the resulting susceptibility map for HMPs across the entire Chinese territory in Fig. 11. There, we reclassified the susceptibility spectrum, binning the probability

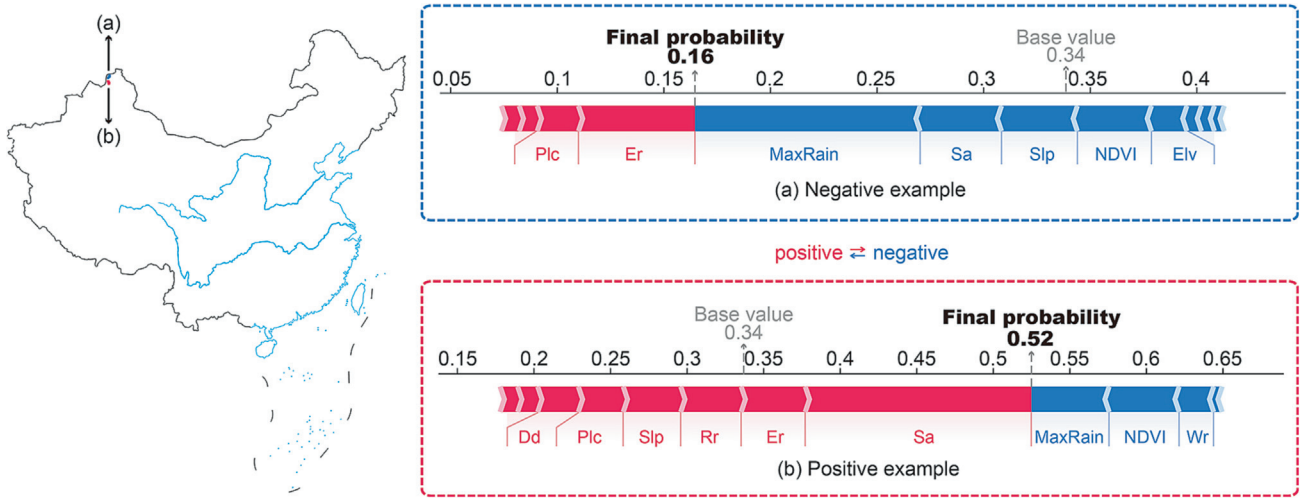

Fig. 8. Examples of catchments that were detected as the negative (a) and positive (b) ones.

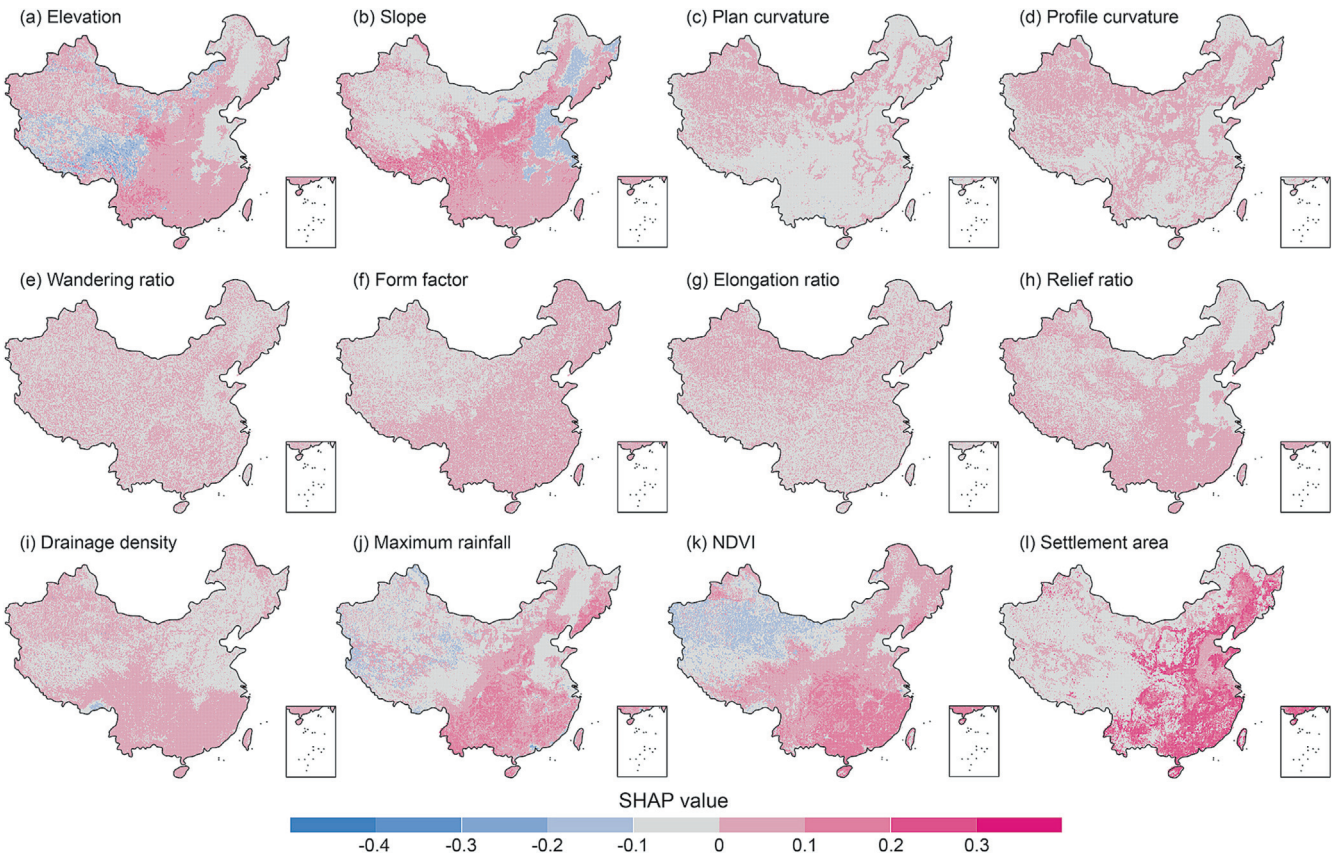

Fig. 10. Spatial effects of variables to HMPs detected via the SHAP values, where the pink colors indicate positive contributions and the blue colors represent negative contributions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

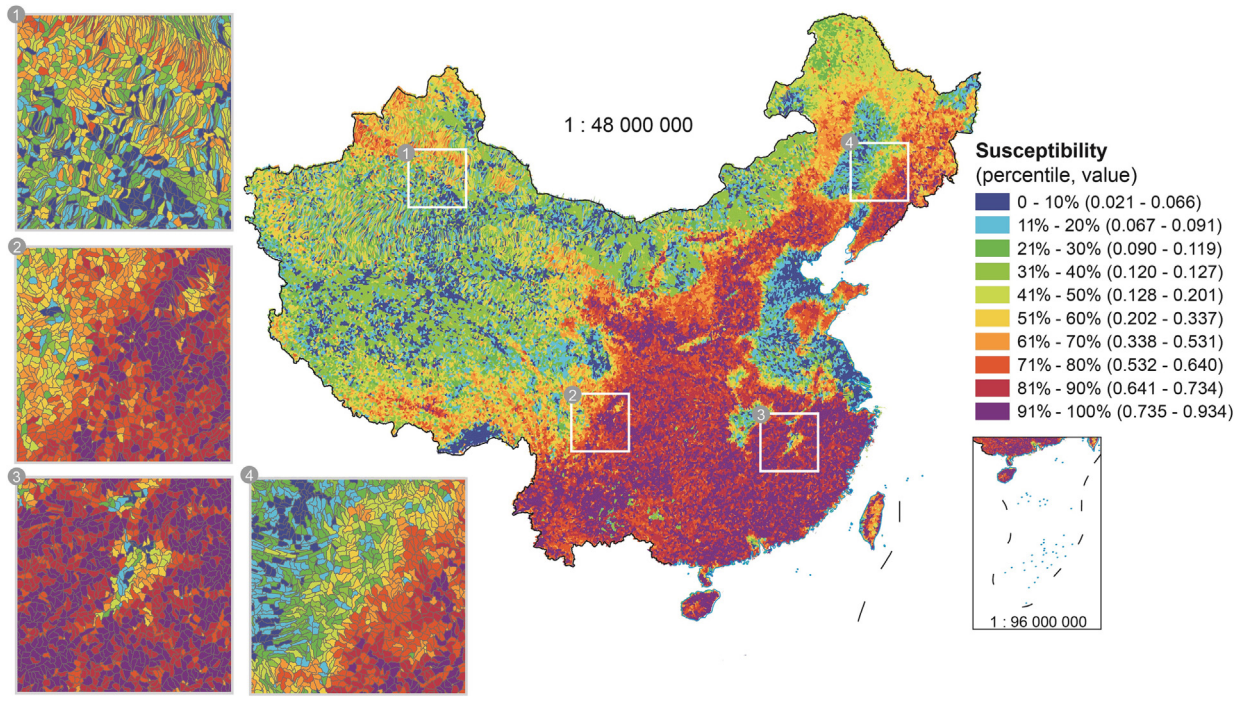

Fig. 11. The final mean susceptibility map of HMPs in China.

values at a decile interval. In general, the areas that present a higher susceptibility are prone to be in southeast China, whereas the low values tend to show in the northwest. However, it is difficult to recognize details in such a vast landscape. For this reason, we also plotted four static zooms, offering a closer view of the susceptibility patterns and the catchment sizes/shapes. Nevertheless, even zooming into the map does not offer a clear view and explainability of the susceptibility estimates. Therefore, we built a webGIS application where each catchment can be queried and the relative SHAP values interactively queried (https://arcg.is/0eGGT8).

5. Discussions

5.1. From global to local model interpretations

Standard approaches to understanding why machine learning models return certain outputs are generally based on variable importance ranks. In this contribution, we stress how important it is to extend this traditional view to welcome the SHAP-oriented model explanation instead. The main reason behind this has to do with the static view that variable importance plots offer. Conversely, SHAP-based graphics expand toward variable interaction processes, adding another dimension to the explainability potential of machine learning solutions. An intuitive explanation of how SHAP values work can be thought of comparing these parameters to the regression coefficients typical of statistical models. Traditionally, one interprets the outcome of a statistical model according to the sign and magnitude of the regression coefficients estimated for each corresponding covariate. However, this is usually represented as a single linear values or nonlinear functions of values for the whole model under consideration. A machine/deep learning model does not estimate regression coefficients but rather multiple weights associated to the range of each predictor. However, SHAP values can be thought of as parameters conveying the role of each predictor onto the final susceptibility. Main difference with respect to a statistical model is that SHAP values are obtained by solving the predictive equation for each mapping unit under consideration. Therefore, rather than obtaining a general overview of each predictors' contribution, SHAP values present how each predictor have modified the HMP occurrence probability for each catchment. Therefore, the level of explanation one can be obtained from using SHAPs is arguably even more than what regression coefficients or weight can offer, in traditional statistical or machine learning applications, respectively.

This becomes clear in Fig. 9, where a closer inspection highlights a cluster of catchments with final susceptibility close to 1. These catchments all start from the same starting point as all others (susceptibility = 0.32), but their predicted value stays essentially the same because of the Dd influence. We recall here that Dd stands for drainage density, whose dominant effect can be geomorphologically justified. As for how this parameter specifically contributes on an individual catchment basis, one can then dive into graphics such as Fig. 8, where the second example purposely reports a catchment where the Dd is responsible for a marginal increase in the final susceptibility. Analogous considerations arise for the other dominant factors, including NDVI, maximum daily rainfall, slope, and settlement area. These results well align with other HMP studies (Ragettli et al., 2017; Zhao et al., 2018). However, as informative as these explainable components may be, they still only offer a non-spatial view of the model output. Therefore, to further enrich the model interpretation, here we demonstrate an additional use of SHAP values. In fact, being SHAPs calculated for individual predictors and for individual mapping units, one can easily translate their combination in map form (see Fig. 10). As a result, one can visualize and query a unique spatial pattern for each predictor and assess their effect and consistency/heterogeneity across the geographic space. For instance, the influence of the NDVI was previously shown to be among the most important HMP predictors. In Fig. 10k though, the spatial dimension is added to this consideration, showing how its model contribution varies across the landscape, with the largest positive contribution depicted across South China, transitioning to smaller SHAP values in Central and Northeast China. Even such a view though is nothing but a static image of the predictors' contribution. With this idea in mind, we decided to prompt the reviewers in thinking about the potential of spatially querying SHAP values, especially, if this can be done through webGIS applications. At this link https://arcg.is/0eGGT8, we provided an interactive cloud-based platform where we stored all our modeling results. There, each catchment susceptibility can be visualized together with each predictor SHAP value responsible for the local HMP probability. We believe that such a way of summarizing model results can become a future standard for susceptibility modeling, in view of maximizing the opportunities of the digital era for risk assessment. In fact, the webGIS application not only reports model results but also offers the ability to bring information on exposure together. To do so, we used data accessed at this link, https://risk.preventionweb.net/ where information on population density and land economical value is reported at the global scale with a 1 km resolution along the coastlines and 5 km resolution inland (Koks et al., 2019). The exposure information complements the susceptibility, allowing for risk-oriented considerations. This is shown in Fig. 12, where we plot the summary of the population and land value as a function of the predicted HMP probability. The catchments labeled as susceptible that contextually report high population and/or high financial value would represent those that may need further attention for tailored risk mitigation strategies.

5.2. Supporting and opposing arguments

For a long time, despite the higher performance offered by machine learning solutions, statistical models have still represented the preferred alternative for researchers equally interested in comprehending and interpreting why a given data-driven model has produced a certain prediction. Recent advancements in SHAP-explained deep learning modeling have the potential to unify these two fundamental aspects within the very same tool. This is particularly relevant in the context of spatial big data, where machine learning ensures performance, efficiency, and computational speed. For this reason, here we tried to push the boundaries of current explainable AI applications on HMP prediction, testing it over a very large dataset reflecting a continental scale.

The performance and level of interpretation provided, support the choice of this approach. As part of the explainability characteristics, we particularly stress the relevance of converting SHAP values into map form, here called SHAP maps. What is also important to stress is that the calculation of SHAP values is not limited to Neural Network architectures but rather SHAP metrics can be computed for any machine/deep learning model. Therefore, the interpretability of the whole spectrum of such models has recently received quite a substantial boost, especially if one would explore SHAP values both non-spatially, as per previous literature, or in a spatial manner, as per the current contribution. The latter resulting geographic view allows for considerations of variable contributions and potential interactions in a straightforward way. This could support decision-making processes, especially if beyond the static map perspective, SHAP values are interactively queried in webGIS applications. The webGIS app we built is meant to showcase these aspects together with considerations of potential HMP risk. In fact, one can dynamically overlay susceptibility estimates, their SHAP corresponding contributors, and exposure information to lay down

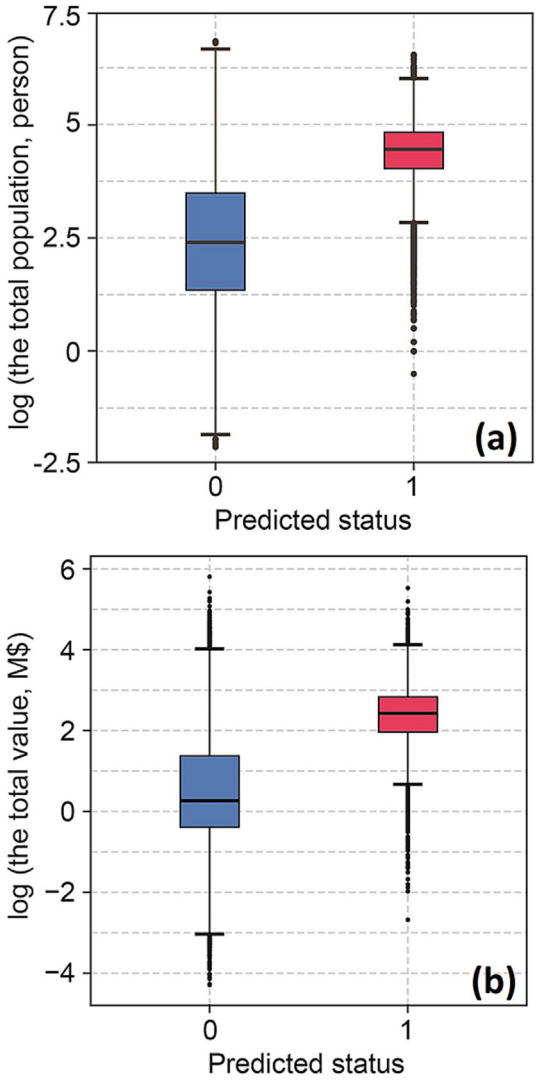

Fig. 12. Summary plot of log values of (a) population and (b) land value vs. the HMP probability.

a comprehensive platform for territorial management and civil protection agencies.

As for the weaknesses behind the experiment we present here, it should be stressed that the temporal component is still missing. Therefore, even putting together HMP susceptibility and exposure data, these are not enough to fully characterize the expected risk but rather constitute an approximation of it. Conversely, the susceptibility model should be extended toward its space–time counterpart (Lombardo et al., 2020; Steger et al., 2022). Such a shift would ensure two possible applications of explainable AI, one where the prediction is performed as a nowcasting/forecasting service for specific events (Collini et al., 2022), and one where its potential can be tapped in for long term scenario-building based on the return time of the HMP trigger.

Another important element to be discussed is the generalization of the current model. Machine and deep learning models tend to become overly specific, limiting the ability to generalize their results in areas other than the ones they have been trained for. Here the scale of our model covers the whole Chinese mainland, demonstrating already the flexibility of our model over such a large and diverse territory. Ideally, we would have tested our model for generalization purposes even in neighboring countries. However, the response variable and the predictor set we collected here are China-specific and therefore testing model transferability was not possible at this stage. In the future, we plan to use more generic predictors, potentially computing them from global data and cloud platforms such as Google Earth Engine (Ahmed et al., 2023), thus ensuring that generalization test could be included.

Ultimately, here we would like to stress once more the elements of innovation presented in this manuscript with respect to the available literature. The first novelty resides in the exploration of SHAP values in their spatial form. So far, geoscientists have presented SHAP values as means to explain machine/deep learning architectures only by presenting them into static plots such as scatter plots or bar plots and so on. In this work, we highlight the added value of translating SHAP values into maps. The second element of novelty corresponds to the scale at which we applied our explainable AI. Machine and deep learning models often become overly specific of the area they have been applied to, something that has limited their applicability mostly to catchment or regional susceptibility models. Here we take the opportunity to test a similar modeling architecture but extending the spatial application to the continental scale corresponding to the Chinese territory. The last main innovative element we present has to do with the interactive accessibility of our results to the audience. In fact, most of the illustrations nowadays are still confined to static figures presented as part of scientific article. Here, we created a Web Application where our modeling results are presented, allowing for anyone interested to perform spatial queries even on individual catchments, thus obtaining information on susceptibility estimates, SHAP plots, population at risk and total land economic worth under potential threat. We conclude here by mentioning that risk oriented considerations are also a very minor component of the geoscientific literature and a unified web platform where all the results are presented together can be hardly found in the relevant literature.

The future challenges we foresee mainly correspond to the extension of the present protocol beyond susceptibility studies. In fact, susceptibility only corresponds to one of the three requirements of the hazard definition (Guzzetti et al., 1999). The other requirements correspond to the temporal prediction of the HMP under consideration as well as its intensity. For floods, the latter corresponds to water height or peak flow (Schumann et al., 2007), whereas for landslides this may corresponds to velocity (He et al., 2023) or planimetric area (Bryce et al., 2022). Therefore, future developments may involve using a regression framework rather than a classification one, using one of the above intensity parameters as the response variable of the deep learning routine. Interestingly, such parameters usually have a very heavy-tailed structure, therefore requiring for specific solutions in the direction of extreme value theory applications (Yadav et al., 2022) and their implementation as part of deep learning routines (Cisneros et al., 2023). Aside from the regression framework, a natural extension would lead to a space–time deep-learning model rather than a pure spatial application.

6. Conclusion

We tested a SHAP-explained deep learning architecture across the whole Chinese territory. Our work showcases a hierarchical overview of predictors' contributions to the final susceptibility, offering both global to local perspectives. The combination of a suite of non-geographic SHAP summaries already represents a step forward compared to traditional alternatives, not only in machine learning but also for the more explainable statistical solutions. This takes another explainable dimension when predictors' effects are examined over the geographic space for individual catchments, something we exemplified in a dedicated webGIS application (accessible at https://arcg.is/0eGGT8) to allow for user interactions.

There, we report not only the model results (final HMP susceptibility and SHAP values) but also relevant information on exposure. We believe this modeling approach will constitute the future standard for data-driven solutions not only for HMP but for any natural hazard predictive model. As pointed out in the discussions, we believe risk assessments will be possible once the temporal dimension will be added to the model, something we are already working on.

To promote analogous analyses, we refer the readers to Dahal and Lombardo (2023), where the core code upon which this contribution is based is shared with the public.

CRediT authorship contribution statement

Nan Wang: Conceptualization, Methodology, Software, Writing – original draft. **Hongyan Zhang:** Supervision. **Ashok Dahal:** Supervision. **Weiming Cheng:** Funding acquisition. **Min Zhao:** Data curation, Validation. **Luigi Lombardo:** Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (grant no. 42201452), and the Fundamental Research Funds for the Central Universities (grant no. 2412022QD003). We especially thank for the support from the China Institute of Water Resources and Hydropower Research (IWHR).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gsf.2024.101800.

References

Ahmed, M., Tanyas, H., Huser, R., Dahal, A., Titti, G., Borgatti, L., Francioni, M., Lombardo, L., 2023. Dynamic rainfall-induced landslide susceptibility: a step towards a unified forecasting system. Int. J. Appl. Earth Obs. Geoinf. 125, 103593.

Albawi, S., Mohammed, T. A. and Al-Zawi, S. (2017) Understanding of a convolutional neural network. In 2017 International Conference on Engineering and Technology (ICET), pp. 1–6.

Aljohani, N.R., Fayoumi, A., Hassan, S.-U., 2021. A novel focal-loss and class-weight-aware convolutional neural network for the classification of in-text citations. J. Inf. Sci. 0165551521991022.

Al-Najjar, H.A., Pradhan, B., Beydoun, G., Sarkar, R., Park, H.J., Alamri, A., 2023. A novel method using explainable artificial intelligence (XAI)-based Shapley Additive Explanations for spatial landslide prediction using Time-Series SAR dataset. Gondw. Res. 123, 107–124.

Alvioli, M., Marchesini, I., Pokharel, B., Gnyawali, K., Lim, S., 2022. Geomorphological slope units of the Himalayas. J. Maps, 1–14.

Amato, G., Fiorucci, M., Martino, S., Lombardo, L., Palombi, L., 2023. Earthquake-triggered landslide susceptibility in Italy by means of Artificial Neural Network. Bull. Eng. Geol. Environ. 82 (5), 160.

Band, S.S., Janizadeh, S., Chandra Pal, S., Saha, A., Chakrabortty, R., Melesse, A.M., Mosavi, A., 2020. Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. Remote Sens. (Basel) 12 (21), 3568.

Baptista, M.L., Goebel, K., Henriques, E.M., 2022. Relation between prognostics predictor evaluation metrics and local interpretability SHAP values. Artif. Intell. 306, 103667.

Brenning, A., 2005. Spatial prediction models for landslide hazards: review, comparison and evaluation. Natural Hazards and Earth System Science 5 (6), 853–862.

Brenning, A., 2008. Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. Hamburger Beitr ̈agezur Physischen Geographie Und Landschafts ̈okologie 19 (23–32), 410.

Bryce, E., Lombardo, L., van Westen, C., Tanyas, H., Castro-Camilo, D., 2022. Unified landslide hazard assessment using hurdle models: a case study in the Island of Dominica. Stoch. Env. Res. Risk A 36 (8), 2071–2084.

Bui, D. T., Hoang, N.-D., Mart ́ınez-́Alvarez, F., Ngo, P.-T. T., Hoa, P. V., Pham, T. D., Samui, P. and Costache, R. (2020) A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. Sci. Total Environ. 701, 134413.

Camilo, D.C., Lombardo, L., Mai, P.M., Dou, J., Huser, R., 2017. Handling high predictor dimensionality in slope-unit-based landslide susceptibility models through LASSO-penalized Generalized Linear Model. Environ. Model. Softw. 97, 145–156.

Can, T., Nefeslioglu, H.A., Gokceoglu, C., Sonmez, H., Duman, T.Y., 2005. Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. Geomorphology 72 (1–4), 250–271.

Carrara, A., Crosta, G., Frattini, P., 2008. Comparing models of debris-flow susceptibility in the alpine environment. Geomorphology 94 (3–4), 353–378.

Carrara, A., Cardinali, M., Guzzetti, F. and Reichenbach, P. (1995) GIS technology in mapping landslide hazard. In Geographical Information Systems in Assessing Natural Hazards, Advances in Natural and Technological Hazards Research, pp. 135–175. Dordrecht: Kluwer, Springer. ISBN 978-90-481-4561-4 978-94-015-8404-3.

Chang, L., Xing, G., Yin, H., Fan, L., Zhang, R., Zhao, N., Huang, F., Ma, J., 2023. Landslide susceptibility evaluation and interpretability analysis of typical loess areas based on deep learning. Natural Hazards Research 3 (2), 155–169.

Cisneros, D., Richards, J., Dahal, A., Lombardo, L. and Huser, R., 2023. Deep graphical regression for jointly moderate and extreme Australian wildfires. arXiv preprint arXiv:2308.14547.

Collini, E., Palesi, L.I., Nesi, P., Pantaleo, G., Nocentini, N., Rosi, A., 2022. Predicting and understanding landslide events with explainable AI. IEEE Access 10, 31175–31189.

Dahal, A., Lombardo, L., 2023. Explainable artificial intelligence in geoscience: A glimpse into the future of landslide susceptibility modeling. Comput. Geosci. 176, 105364.

De Villiers, J., Barnard, E., 1993. Backpropagation neural nets with one and two hidden layers. IEEE Trans. Neural Netw. 4 (1), 136–141.

Di Napoli, M., Tanyas, H., Castro-Camilo, D., Calcaterra, D., Cevasco, A., Di Martire, D., Pepe, G., Brandolini, P. and Lombardo, L., 2023. On the estimation of landslide intensity, hazard and density via data-driven models. Natural Hazards, In press.

Dikshit, A., Pradhan, B., 2021. Interpretable and explainable AI (XAI) model for spatial drought prediction. Sci. Total Environ. 801, 149797.

Domeneghetti, A., Vorogushyn, S., Castellarin, A., Merz, B., Brath, A., 2013. Probabilistic flood hazard mapping: effects of uncertain boundary conditions. Hydrol. Earth Syst. Sci. 17 (8), 3127–3140.

Fang, Z., Wang, Y., van Westen, C., Lombardo, L., 2023a. Space-Time Landslide Susceptibility Modeling Based on Data-Driven Methods. Math. Geosci., 1–20

Fang, Z., Tanyas, H., Gorum, T., Dahal, A., Wang, Y., Lombardo, L., 2023b. Speech-recognition in landslide predictive modelling: A case for a next generation early warning system. Environ. Model. Softw. 170, 105833.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W.Z., et al., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. Eng. Geol. 102 (3–4), 99–111.

Fluss, R., Faraggi, D., Reiser, B., 2005. Estimation of the Youden Index and its associated cutoff point. Biometrical Journal: Journal of Mathematical Methods in Biosciences 47 (4), 458–472.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons.

Frattini, P., Crosta, G., Carrara, A., 2010. Techniques for evaluating the performance of landslide susceptibility models. Eng. Geol. 111 (1), 62–72.

Gariano, S.L., Guzzetti, F., 2016. Landslides in a changing climate. Earth Sci. Rev. 162, 227–252.

Goetz, J., Brenning, A., Petschko, H., Leopold, P., 2015. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput. Geosci. 81, 1–11.

Gunning, D. (2017) Explainable Artificial Intelligence (XAI). Defense Advanced Research Projects Agency (DARPA), nd Web 2(2), 1.

Guzzetti, F. et al., 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31 (1-4), 181–216.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., Galli, M., 2006. Estimating the quality of landslide susceptibility models. Geomorphology 81 (1–2), 166–184.

He, B., Huang, X., Ma, M., Chang, Q., Tu, Y., Li, Q., Zhang, K., Hong, Y., 2018. Analysis of flash flood disaster characteristics in China from 2011 to 2015. Nat. Hazards 90, 407–420.

He, K., Tanyas, H., Chang, L., Hu, X., Luo, G., Lombardo, L., 2023. Modelling InSAR-derived hillslope velocities with multivariate statistics: A first attempt to generate interpretable predictions. Remote Sens. Environ. 289, 113518.

Horton, R.E., 1932. Drainage-basin characteristics. Trans. Am. Geophys. Union 13 (1), 350–361.

Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied Logistic Regression, Volume 398. John Wiley & Sons.

Hosseini, F.S., Choubin, B., Mosavi, A., Nabipour, N., Shamshirband, S., Darabi, H., Haghighi, A.T., 2020. Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method. Sci. Total Environ. 711, 135161.
Jacobs, L., Kervyn, M., Reichenbach, P., Rossi, M., Marchesini, I., Alvioli, M., Dewitte, O., 2020. Regional susceptibility assessments with heterogeneous landslide information: Slope unit-vs. pixel-based approach. Geomorphology 356, 107084.
Kern, A.N., Addison, P., Oommen, T., Salazar, S.E., Coffman, R.A., 2017. Machine learning based predictive modeling of debris flow probability following wildfire in the intermountain Western United States. Math. Geosci. 49, 717–735.
Kobiyama, M., Goerl, R.F., 2007. Quantitative method to distinguish flood and flash flood as disasters. SUISUI Hydrological Research Letters 1, 11–14.
Koks, E.E., Rozenberg, J., Zorn, C., Tariverdi, M., Vousdoukas, M., Fraser, S.A., Hall, J., Hallegatte, S., 2019. A global multi-hazard risk analysis of road and railway infrastructure assets. Nat. Commun. 10 (1), 2677.
Korup, O., Stolle, A., 2014. Landslide prediction from machine learning. Geol. Today 30 (1), 26–33.
Li, Z., 2022. Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. Comput. Environ. Urban Syst. 96, 101845.
Li, Y. and Yuan, Y. (2017) Convergence analysis of two-layer neural networks with relu activation. Advances in Neural Information Processing Systems 30.
Lima, P., Steger, S., Glade, T., Murillo-Garc´ıa, F.G., 2022. Literature review and bibliometric analysis on data-driven assessment of landslide susceptibility. J. Mt. Sci. 19 (6), 1670–1698.
Lin, Q., Lima, P., Steger, S., Glade, T., Jiang, T., Zhang, J., Liu, T., Wang, Y., 2021. National-scale data-driven rainfall induced landslide susceptibility mapping for China by accounting for incomplete landslide data. Geosci. Front. 12, (6) 101248.
Lin, Q., Steger, S., Pittore, M., Zhang, J., Wang, L., Jiang, T., Wang, Y., 2022. Evaluation of potential changes in landslide susceptibility and landslide occurrence frequency in China under climate change. Sci. Total Environ. 850, 158049.
Liu, C., Guo, L., Ye, L., Zhang, S., Zhao, Y., Song, T., 2018a. A review of advances in China's flash flood early-warning system. Nat. Hazards 92 (2), 619–634.
Liu, Y., Yang, Z., Huang, Y., Liu, C., 2018b. Spatiotemporal evolution and driving factors of China's flash flood disasters since 1949. Sci. China Earth Sci. 61 (12), 1804–1817.
Liu, Y., Huang, Y., Wan, J., Yang, Z., Zhang, X., 2021. Analysis of human activity impact on flash floods in China from 1950 to 2015. Sustainability 13 (1), 217.
Loche, M., Alvioli, M., Marchesini, I., Bakka, H., Lombardo, L., 2022a. Landslide susceptibility maps of Italy: Lesson learnt from dealing with multiple landslide types and the uneven spatial distribution of the national inventory. Earth Sci. Rev. 104125
Loche, M., Scaringi, G., Yunus, A. P., Catani, F., Tanya s, H., Frodella, W., Fan, X. and Lombardo, L. (2022b) Surface temperature controls the pattern of post-earthquake landslide activity. Scientific Reports 12(1), 988.
Lombardo, L., Bakka, H., Tanyas, H., van Westen, C., Mai, P.M., Huser, R., 2019. Geostatistical modeling to capture seismic-shaking patterns from earthquake-induced landslides. J. Geophys. Res. Earth 124 (7), 1958–1980.
Lombardo, L., Mai, P.M., 2018. Presenting logistic regression-based landslide susceptibility results. Eng. Geol. 244, 14–24.
Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F., Huser, R., 2020. Space-time landslide predictive modelling. Earth Sci. Rev. 103318
Lubo-Robles, D., Devegowda, D., Jayaram, V., Bedle, H., Marfurt, K. J. and Pranter, M. J. (2020) Machine learning model interpretability using SHAP values: Application to a seismic facies classification task. In SEG International Exposition and Annual Meeting.
Lundberg, S. M. and Lee, S.-I. (2017a) A Unified Approach to Interpreting Model Predictions. In Advances in Neural Information Processing Systems 30, eds I. Guyon, U. V.
Lundberg, S. M. and Lee, S.-I. (2017b) A unified approach to interpreting model predictions. Advances in neural information processing systems 30.
Marchi, L., Borga, M., Preciso, E., Gaume, E., 2010. Characterisation of selected extreme flash floods in Europe and implications for flood risk management. J. Hydrol. 394 (1–2), 118–133.
Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R., Abderrahmane, B., 2020. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth Sci. Rev. 103225
Merz, B., Elmer, F., Thieken, A., 2009. Significance of "high probability/low damage" versus "low probability/high damage" flood events. Nat. Hazards Earth Syst. Sci. 9 (3), 1033–1046.
Molnar, C. (2020) Interpretable machine learning. Lulu. com.
Moreira, A., Krieger, G., Hajnsek, I., Hounam, D., Werner, M., Riegger, S. and Settelmeyer, E. (2004) TanDEM-X: a TerraSAR-X add-on satellite for single-pass SAR interferometry.
Nicu, I. C., Elia, L., Rubensdotter, L., Tanya s, H. and Lombardo, L. (2023) Multi-hazard susceptibility mapping of cryospheric hazards in a high-Arctic environment: Svalbard Archipelago. Earth System Science Data 15(1), 447–464.
Opitz, T., Bakka, H., Huser, R., Lombardo, L., 2022. High-resolution bayesian mapping of landslide hazard with unobserved trigger event. Ann. Appl. Stat. 16 (3), 1653–1675.
Ozdemir, A., 2009. Landslide susceptibility mapping of vicinity of Yaka Landslide (Gelendost, Turkey) using conditional probability approach in GIS. Environ. Geol. 57, 1675–1686.
Panahi, M., Jaafari, A., Shirzadi, A., Shahabi, H., Rahmati, O., Omidvar, E., Lee, S., Bui, D.T., 2021. Deep learning neural networks for spatially explicit prediction of flash flood probability. Geosci. Front. 12, (3) 101076.
Panati, C., Wagner, S., Br¨uggenwirth, S., 2022. Feature Relevance Evaluation using Grad-CAM, LIME and SHAP for Deep Learning SAR Data Classification. In: 2022 23rd International Radar Symposium (IRS), pp. 457–462.
Park, N.-W., 2015. Using maximum entropy modeling for landslide susceptibility mapping with multiple geoenvironmental data sets. Environ. Earth Sci. 73, 937–949.
Petschko, H., Brenning, A., Bell, R., Goetz, J., Glade, T., 2014. Assessing the quality of landslide susceptibility maps—case study lower austria. Nat. Hazards Earth Syst. Sci. 14 (1), 95–118.
Pradhan, B., Dikshit, A., Lee, S., Kim, H., 2023a. An explainable AI (XAI) model for landslide susceptibility modeling. Appl. Soft Comput. 142, 110324.
Pradhan, B., Lee, S., Dikshit, A., Kim, H., 2023b. Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model. Geosci. Front. 14, (6) 101625.
Ragettli, S., Zhou, J., Wang, H., Liu, C., Guo, L., 2017. Modeling flash floods in ungauged mountain catchments of China: A decision tree learning approach for parameter regionalization. J. Hydrol. 555, 330–346.
Ramyachitra, D., Manikandan, P., 2014. Imbalanced dataset classification and solutions: a review. International Journal of Computing and Business Research (IJCBR) 5 (4), 1–29.
Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. Earth Sci. Rev. 180, 60–91.
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.
Roshan, K., Zafar, A., 2022. Using Kernel SHAP XAI Method to Optimize the Network Anomaly Detection Model. In: 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 74–80.
Samek, W., Wiegand, T. and M¨uller, K.-R. (2017) Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296 .
Santangelo, N., Santo, A., Di Crescenzo, G., Foscari, G., Liuzza, V., Sciarrotta, S., Scorpio, V., 2011. Flood susceptibility assessment in a highly urbanized alluvial fan: the case study of Sala Consilina (southern Italy). Nat. Hazards Earth Syst. Sci. 11 (10), 2765–2780.
Schumann, G., Hostache, R., Puech, C., Hoffmann, L., Matgen, P., Pappenberger, F., Pfister, L., 2007. High-resolution 3-D flood information from radar imagery for flood hazard management. IEEE Trans. Geosci. Remote Sens. 45 (6), 1715–1725.
Schumm, S.A., 1956. Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. Geol. Soc. Am. Bull. 67 (5), 597–646.
Shirzadi, A., Shahabi, H., Chapi, K., Bui, D.T., Pham, B.T., Shahedi, K., Ahmad, B.B., 2017. A comparative study between popular statistical and machine learning methods for simulating volume of landslides. Catena 157, 213–226.
Shrikumar, A., Greenside, P., Kundaje, A., 2017. Learning important features through propagating activation differences. In International Conference on Machine Learning, 3145–3153.
Singh, A., Sengupta, S., Lakshminarayanan, V., 2020. Explainable deep learning models in medical image analysis. Journal of Imaging 6 (6), 52.
Song, C., Shi, X., Wang, J., 2020. Spatiotemporally varying coefficients (stvc) model: A bayesian local regression to detect spatial and temporal nonstationarity in variables relationships. Ann. GIS 26 (3), 277–291.
Steger, S., Moreno, M., Crespi, A., Zellner, P.J., Gariano, S.L., Brunetti, M.T., Melillo, M., Peruccacci, S., Marra, F., Kohrs, R., et al., 2022. Deciphering seasonal effects of triggering and preparatory precipitation for improved shallow landslide prediction using generalized additive mixed models. In: Natural Hazards and Earth System Sciences Discussions, pp. 1–38.
Strahler, A.N., 1952. Dynamic basis of geomorphology. Geol. Soc. Am. Bull. 63 (9), 923–938.
Strumbelj, E., Kononenko, I., 2014. Explaining prediction models and individual predictions with feature contributions. Knowledge and Information Systems 41, 647–665.
Sun, D., Chen, D., Zhang, J., Mi, C., Gu, Q., Wen, H., 2023. Landslide Susceptibility Mapping Based on Interpretable Machine Learning from the Perspective of Geomorphological Differentiation. Land 12 (5), 1018.
Tehrani, F.S., Calvello, M., Liu, Z., Zhang, L., Lacasse, S., 2022. Machine learning and landslide studies: recent advances and applications. Nat. Hazards 114 (2), 1197–1245.
Titti, G., Napoli, G.N., Conoscenti, C., Lombardo, L., 2022. Cloud-based interactive susceptibility modeling of gully erosion in Google Earth Engine. Int. J. Appl. Earth Obs. Geoinf. 115, 103089.
Townsend, J.T., 1971. Theoretical analysis of an alphabetic confusion matrix. Percept. Psychophys. 9, 40–50.
Ullah, I., Liu, K., Yamamoto, T., Zahid, M., Jamal, A., 2023. Modeling of machine learning with shap approach for electric vehicle charging station choice behavior prediction. Travel Behav. Soc. 31, 78–92.
van den Bout, B., Jetten, V.G., van Westen, C.J., Lombardo, L., 2023. A breakthrough in fast flood simulation. Environ. Model. Softw. 168, 105787.
Wang, N., Cheng, W., Wang, B., Liu, Q., Zhou, C., 2020. Geomorphological regionalization theory system and division methodology of China. J. Geog. Sci. 30 (2), 212–232.

Wang, N., Cheng, W., Marconcini, M., Bachofer, F., Liu, C., Xiong, J., Lombardo, L., 2022. Space-time susceptibility modeling of hydro-morphological processes at the Chinese national scale. Eng. Geol. 301, 106586.

Wang, N., Cheng, W., Lombardo, L., Xiong, J., Guo, L. (2021) Statistical spatiotemporal analysis of hydro-morphological processes in China during 1950–2015. Stochastic Environmental Research and Risk Assessment pp. 1–21.

Wang, D., Thun´ell, S., Lindberg, U., Jiang, L., Trygg, J. and Tysklind, M. (2022a) Towards better process management in wastewater treatment plants: Process analytics based on SHAP values for tree-based machine learning methods. Journal of Environmental Management 301, 113941.

Xiong, J., Li, J., Cheng, W., Wang, N., Guo, L., 2019. A GIS-based support vector machine model for flash flood vulnerability assessment and mapping in China. ISPRS Int. J. Geo Inf. 8 (7), 297.

Xiong, J., Pang, Q., Cheng, W., Wang, N., Yong, Z., 2020. Reservoir risk modelling using a hybrid approach based on the feature selection technique and ensemble methods. Geocarto Int., 1–22

Yadav, R., Huser, R., Opitz, T. and Lombardo, L., 2022. Joint modeling of landslide counts and sizes using spatial marked point processes with sub-asymptotic mark distributions. arXiv preprint arXiv:2205.09908.

Yeon, Y.-K., Han, J.-G., Ryu, K.H., 2010. Landslide susceptibility mapping in Injae, Korea, using a decision tree. Eng. Geol. 116 (3–4), 274–283.

Yilmaz, I., 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat—Turkey). Comput. Geosci. 35 (6), 1125–1138.

Yu, L., Porwal, A., Holden, E.-J., Dentith, M.C., 2012. Towards automatic lithological classification from remote sensing data using support vector machines. Comput. Geosci. 45, 229–239.

Zhang, J., Ma, X., Zhang, J., Sun, D., Zhou, X., Mi, C., Wen, H., 2023. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. J. Environ. Manage. 332, 117357.

Zhao, G., Pang, B., Xu, Z., Yue, J., Tu, T., 2018. Mapping flood susceptibility in mountainous areas on a national scale in China. Sci. Total Environ. 615, 1133–1142.

Zhao, G., Liu, R., Yang, M., Tu, T., Ma, M., Hong, Y., Wang, X., 2022a. Large-scale flash flood warning in China using deep learning. J. Hydrol. 604, 127222.

Zhao, P., Masoumi, Z., Kalantari, M., Aflaki, M., Mansourian, A., 2022b. A GIS-based landslide susceptibility mapping and variable importance analysis using artificial intelligent training-based methods. Remote Sens. (Basel) 14 (1), 211.

Zhou, X., Wen, H., Li, Z., Zhang, H., Zhang, W., 2022. An interpretable model for the susceptibility of rainfall-induced shallow landslides based on SHAP and XGBoost. Geocarto Int. 37 (26), 13419–13450.