https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

Landslide susceptibility mapping with explainable AI techniques: Evidence from Bavaria, Germany

Veronika Buchauer[1,2,3], Marta Sapena[1,4], Christian Geiß[1,5], Patrick Aravena Pelizari[1], and Hannes Taubenböck[1,6]

[1]German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Münchner Straße 20, 82234 Weßling, Germany
[2]Department of Aerospace and Geodesy, Technical University of Munich, Arcisstraße 21, 80333 Munich, Germany
[3]Munich Center for Machine Learning, Oettingenstraße 67, 80538 Munich, Germany
[4]Image Processing Laboratory, Universitat de València, 46980 Valencia, Spain
[5]Department of Geography, University of Bonn, Meckenheimer Allee 166, 53115 Bonn, Germany
[6]Department of Global Urbanization and Remote Sensing, University of Würzburg, John-Skilton-Str. 4a, 97074 Würzburg, Germany

Correspondence: Patrick Aravena Pelizari (patrick.aravenapelizari@dlr.de)

Abstract. Landslides threaten infrastructure, ecosystems, and human safety, particularly in mountainous regions. Climate change with increasingly intense rainfall, together with growing populations and assets in hazard-prone areas, increases the need for accurate and interpretable landslide susceptibility assessments. This study presents a region-wide landslide susceptibility map modeled for entire Bavaria, Germany, based on more than 11,000 recorded landslide events. Using slope units, which are terrain-based spatial mapping entities following natural drainage lines and ridges, the model captures landslide-prone areas in a more terrain-consistant manner than traditional grid-based approaches. To generate the landslide susceptibility map, we employ a dense neural network architecture. The model is trained on the landslide inventory and a wide range of landslide-influencing factors derived from high-resolution topographic, geological, and land cover data and achieves strong predictive performance (ROC AUC = 0.953, PR AUC = 0.844). Model interpretability is approached using the SHapley Additive exPlanations (SHAP) framework, which provides both global and local insights into the factors influencing landslide susceptibility, revealing a strong predictive influence of geology, soil properties and terrain heterogeneity. The resulting susceptibility map is compared with an existing map, which is based on manual assessments, and shows good performance, particularly for deep-seated landslides. However, evaluation using newly recorded landslides reveals limitations in the model's generalizability. Many newly recorded events occur in regions that were underrepresented in the original inventory and are therefore wrongly assigned low susceptibility values. This demonstrates how spatial incompleteness and selection bias in landslide inventories directly propagate into susceptibility maps, leading to systematic underestimation of hazard. Overall, this study highlights that while explainable machine learning enables robust and more interpretable regional susceptibility mapping, the quality and spatial completeness of landslide inventories are critical for reliable hazard assessment and mitigation.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

1 Introduction

Landslides are a widespread and often underestimated natural hazard that causes severe damage to infrastructure, ecosystems, and human lives (Munich Re, 2025). They can destroy roads, buildings, and utilities, disrupt forests and waterways, and often cause fatalities and long-term socioeconomic consequences (Kennedy et al., 2015). Their frequency and magnitude are expected to grow under changing climate conditions, with more intense rainfall, flooding, and changes in groundwater levels which destabilize slopes (IPCC, 2022). By the end of the century, moderate climate scenarios project an expansion of landslide-prone areas by 5 to 12 % and a rise in the population exposed to landslide risk by approximately 15 % (Jemec Auflič et al., 2023; Jaedicke et al., 2011). These trends highlight the urgent need for reliable landslide susceptibility and hazard maps as essential tools for planning, disaster risk reduction, and sustainable land management.

Given that landslides are common in mountainous regions, the issue of increased landslide risk is particularly relevant in Bavaria, Germany, which has hilly regions in the north and the Alps in the south. The Bavarian State Office for the Environment (*LfU*) has developed landslide and rockfall susceptibility maps to identify unstable areas. However, these maps are currently incomplete and cover only 51 % of Bavaria. Furthermore, the areas most prone to landslides, such as the Alps and the Franconian Jura, are not up-to-date, with landslide inventories and resulting susceptibility maps dating back to 2013 and 2014, and the methodology being unclear. Currently, no other landslide susceptibility map covers the whole of Bavaria or Germany, making it challenging to assess landslide risks thoroughly across the region (Bayerisches Landesamt für Umwelt, 2020a, b).

Landslide susceptibility mapping involves identifying areas *where* these mass movements are likely to occur based on terrain conditions and predisposing factors, without considering their temporal probability or potential harm (Pradhan et al., 2023; Guzzetti et al., 2005). Susceptibility maps therefore express the spatial likelihood of landslide occurrence, in contrast to hazard maps, which additionally take the expected time frame or magnitude of events into account, and risk maps, which further include exposure and vulnerability of elements at risk (Guzzetti et al., 2005; Varnes and IAEG, 1984). Susceptibility maps are a key part of understanding and managing such hazards. The reliability of the these maps depends on multiple factors, including the quality of the landslide inventory, the choice of the mapping units used for analysis, the availability of data and proxies for landslide influencing factors and their adequate selection (i.e., explanatory features), as well as the modeling approach (Reichenbach et al., 2018; Merghadi et al., 2020; Lima et al., 2022; Ba et al., 2018; Jacobs et al., 2020; Steger et al., 2017). Among these, the completeness and spatial accuracy of the inventory are especially important, as they provide the fundamental training data that guides the model's learning process. A model is only as good as the data it learns from. Incomplete or inconsistent inventories can introduce bias in training, distort model validation, and reduce the transferability of results to unseen areas (Huang et al., 2024; Steger et al., 2017).

Beyond the quality of the training data, methodological advances have also helped to increase the accuracy of the results. An important design question for susceptibility mapping is the segmentation approach for the area of interest. Often, susceptibility models are based on grid-cell data, which are computationally simple but do not have any physical relation to the landslide process (Jacobs et al., 2020; Reichenbach et al., 2018). Mass movements can occur at different scales and affect slopes of various sizes. This is better represented by a more flexible and locally-adapted segmentation of the study area. Topography-

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

dependent polygons can vary in size and account for this (Woodard et al., 2024). These polygons are called slope units. The slope units are delineated according to drainage and divide lines, which provides a more accurate representation of slopes than uniform and rigid grids (Carrara, 1988; Reichenbach et al., 2018). Slope units aim to maximize within-unit homogeneity and between-unit heterogeneity based on geomorphological properties like slope, aspect, or circular variance (Alvioli et al., 2016). From a computational modeling perspective, estimating susceptibility using grid units can be more resource-intensive than using slope units, primarily due to the larger total number of terrain segmentations used for prediction (Chang et al., 2023). This is especially relevant for studies covering large areas, as in this work. However, pre-processing to retrieve the slope units can also be computationally expensive, depending on the spatial resolution of the DEM used for segmentation. Overall, between grids and slope units the computational costs are mostly just shifted from the modeling stage to pre-processing, making it difficult to determine which approach is more resource-intensive. Regarding predictive performance, models based on slope units have shown better accuracy and Area under the Curve (AUC) metrics compared to grid-based estimations (Ba et al., 2018; Jacobs et al., 2020). The development of automated delineation algorithms, such as `r.slopeunits`, have made slope-unit-based mapping more accessible and reproducible at regional scales (Alvioli et al., 2016).

Similarly, the shift from traditional statistical models, such as logistic regression, to data-driven machine learning (ML) approaches like random forests, support vector machines, and artificial neural networks (ANN), has shown significant improvements in predictive accuracy (Merghadi et al., 2020; Chang et al., 2023; Lima et al., 2022). Among these methods, ANNs are especially well suited for problems where input features have strong nonlinear relationships and complex interactions, as they can flexibly learn patterns directly from the data without model assumptions about underlying relationships (Fischer, 2006). In Sect. 3.2, we demonstrate that such nonlinear effects are present in our model, supporting the choice of an ANN over other ML approaches. However, in favor of more flexible modeling, ANNs sacrifice interpretability, a problem that is being addressed by explainable artificial intelligence (XAI) techniques such as SHapley Additive exPlanations (SHAP). XAI allows the quantification and assessment of how individual features contribute to the estimated susceptibility of single observations (i.e., local interpretability) or the entire study area (i.e., global interpretability) (Lundberg and Lee, 2017; Dahal and Lombardo, 2023).

Embedded in this research context, this study develops a data-driven landslide susceptibility map for the Bavarian region in Germany, based on a uniquely extensive landslide inventory. In summary, the main contributions of this study are as follows: (1) the development of a region-wide susceptibility model trained on more than 11,000 mapped landslides and a large feature set, allowing robust learning of non-linear relationships using an ANN; (2) the interpretation of model behavior using SHAP to identify the most relevant spatial predictors and to improve transparency in susceptibility modeling; and (3) an assessment of inventory-related uncertainties and a comparison with existing susceptibility maps, contributing to the discussion on inventory quality and the comparability of landslide susceptibility products.

In the remaining part of this paper, Sect. 2 describes the study area, input data and methods, including slope unit delineation, feature extraction, the ANN architecture, and SHAP analysis. Section 3 presents the susceptibility mapping results and the interpretation of model behavior. The results are compared with the existing *LfU* susceptibility maps currently used by public authorities and validated against an updated landslide inventory. In the discussion (Sect. 4), issues related to model general-

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

ization, potential biases in the landslide inventory, and possible improvements to slope unit delineation are addressed. Finally, Sect. 5 summarizes the main findings and concludes the study.

2 Materials and methods

In this section, we present the data and methods used to produce a landslide susceptibility map for the state of Bavaria using a large landslide inventory compiled from *LfU* and NASA. At the time of model training, the inventory comprised approximately $11,000$ geolocated landslide events, though ongoing fieldwork suggests that spatial coverage is incomplete across the region. This potential source of spatial bias is accounted for and examined within the modeling workflow. An ANN model is trained using nine thematic input maps, resulting in more than 80 explanatory features describing topographic, geological, and land-cover conditions. Modeling is performed at the slope-unit level to improve geomorphological relevance and spatial coherence. The ANN allows to capture complex nonlinear relationships among the extensive feature set and the observed landslide occurrences, while SHAP analysis is employed to quantify feature contributions and improve the model's interpretability. To assess the influence of inventory incompleteness, the trained model is evaluated against an updated landslide inventory containing events recorded after the training phase. This additional validation highlights limitations in model generalization and illustrates the sensitivity of data-driven susceptibility models to spatial biases and inconsistencies in the training data. Figure 1 provides a full overview of the study workflow.

2.1 Study area

Bavaria is the largest federal state in southeastern Germany, bordering Austria to the south and the Czech Republic to the east. Its landscape is diverse: the Alps form a high-mountain environment in the south, followed by the Alpine foreland, while further north the terrain transitions into the Franconian Jura and other upland regions with lower relief. These differences in topography, geology, precipitation, and land use lead to varying conditions for slope instability across the state. Although landslides are comparatively well documented in the Bavarian Alps and other intensively studied regions, a consistent, area-wide landslide susceptibility map for the whole of Bavaria is still missing (Bayerisches Landesamt für Umwelt, 2023c, 2020a). Providing such a map supports planning and risk-informed decision-making throughout the state.

2.2 Data

2.2.1 Landslide inventories

For the analysis, multiple data sources are integrated. One crucial dataset is the landslide inventory, which contains the geographic coordinates of past landslides and additional information about their characteristics. In order to obtain as comprehensive a database as possible, we combine two sources: the *LfU* GEORISK-Objects (Bayerisches Landesamt für Umwelt, 2023c) and NASA Cooperative Open Online Landslide Repository (COOLR) (NASA, 2024). The *LfU* GEORISK dataset, was downloaded on 30 January 2024 and includes $11,188$ geolocated landslides. NASA COOLR included 5 events that were not

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

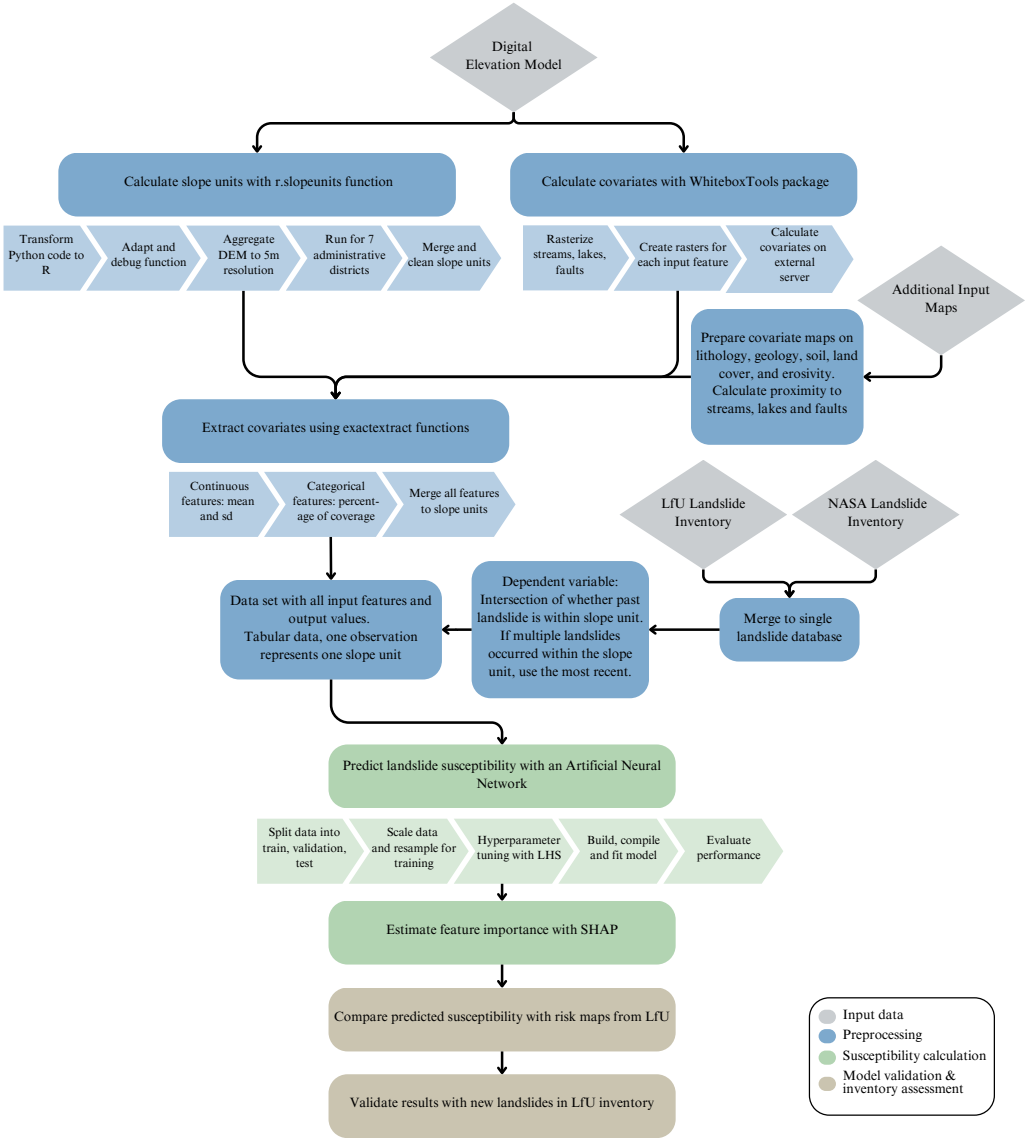

Figure 1. Workflow overview: step-by-step process visualization of data preparation, analysis and validation. The gray rectangles represent input data sources, blue steps are regarding data preparation. The lower part illustrates the susceptibility estimation using an ANN and the post-analysis step of calculating SHAP values in green. The final comparison and validation of the model results is shown in brown.

already recorded in the *LfU* inventory, bringing the total number of 11,193 observations. This data provides the target labels used to train and evaluate the ANN-based landslide susceptibility model.

The two inventories differ in their compilation strategies and data sources. The *LfU* GEORISK inventory is maintained by the Bavarian Environment Agency and combines historical records, verified civilian reports, and results from systematic site

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

inspections conducted by experts. These inspections apply various geodetic and geotechnical measurement techniques, such as extensometers, inclinometers, repeated geodetic surveys, and terrain change detection using laser scanning or photogrammetry. Historically, mapping activities focused on areas with higher susceptibility, particularly the Alps and the Franconian Jura, resulting in spatial differences in inspection intensity across Bavaria (Bayerisches Landesamt für Umwelt, 2020b, 2024a). NASA COOLR aggregates data from the Global Landslide Catalog (GLC), the Landslide Reporter Catalog, and selected external inventories. The Global Landslide Catalog has been compiled since 2007 at NASA Goddard Space Flight Center and is largely based on media reports, susceptibility databases, scientific publications, and other publicly available sources, with a strong emphasis on rainfall-triggered events (NASA, 2024).

In early 2024, the *LfU* inventory focused on regions with higher landslide activity, particularly the Alps and the Franconian Jura. Later, in mid-2024, *LfU* updated the inventory by adding 802 additional landslide events. For more clarity, in the remainder of this work, the version from early 2024 is called the "original" *LfU* landslide inventory, while the newly added 802 events are referred to as the "updated" *LfU* inventory. Most of these newly included records correspond to older events rather than recent occurrences and are primarily located in previously sparsely recorded parts in the districts of Günzburg, Erding, Landshut, Mühldorf am Inn, and Rottal-Inn (Bayerisches Landesamt für Umwelt, 2023c). Since the *LfU* inventory used to train the model was downloaded in early 2024, these regions were underrepresented at that time. This introduces sample selection bias and inventory incompleteness, whose effects on landslide susceptibility prediction are analyzed in this study. Nevertheless, the used original inventory contains over 11,000 records, providing a valuable dataset. The data preparation, model training and testing are conducted using this incomplete original inventory. The updated database is then used to assess the effect of the incompleteness on our ANN model's results (Sect. 3.4). Figure 2 shows the spatial distribution of recorded landslides from the two inventories.

2.2.2 Environmental and topographic input maps

Another key dataset for susceptibility modeling is the Digital Elevation Model (DEM), which captures the surface height excluding vegetation and buildings. We used a high-resolution dataset with elevation data at 1 m resolution available for Bavaria (Landesamt für Digitalisierung Breitband und Vermessung Bayern, 2024a). While this high resolution DEM captures fine structures, it requires significant storage space and computing power. As the aim of this study is to generate a continuous susceptibility map of Bavaria, the high resolution of the DEM caused computational problems in several pre-processing steps. Therefore, following a sensitivity analysis, the DEM was resampled to a resolution of 5 m, which had a minimal effect on the delineation of slope units, while enabling more efficient processing. Consequently, the resampled DEM was used for both the delineation of slope units and the derivation of topography-related features, which were then used as input variables in the model. A full list of features is available in Table B1 in the appendix.

Beyond DEM-derived features, other relevant landslide-influencing factors are obtained from geological and environmental datasets. These include, the *BÜK200 ("Bodenübersichtskarte 1:200.000")* soil map (Bundesanstalt für Geowissenschaften und Rohstoffe, 2022), which classifies soil types into 36 categories (aggregated into 13 categories to reduce dimensionality in the ANN model), and the *dIGK25 ("Digitale Ingenieurgeologische Karte 1:25.000")* subsoil map, representing 33 geological

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

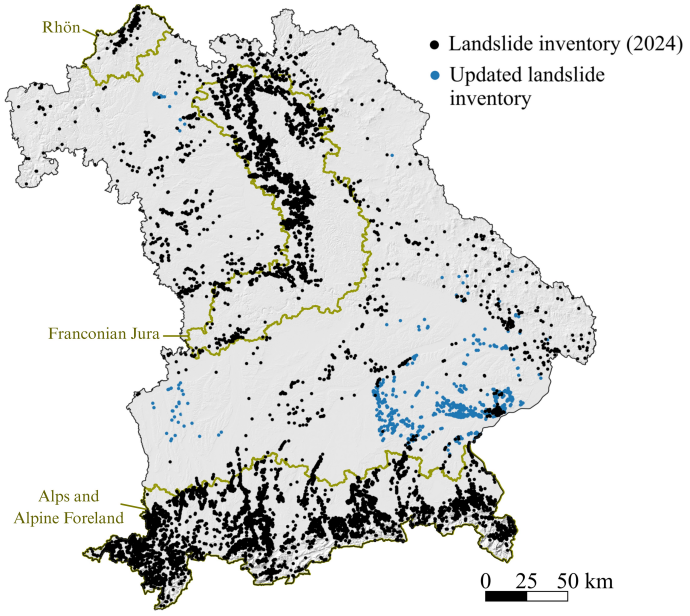

Figure 2. Locations of the landslide events from the two inventories. The events shown in black are used for landside susceptibility modeling. These are mainly in the Alps, Franconian Jura and the Rhön in northern Bavaria. The blue locations show updated events from mid-2024, which are used to further discuss the susceptibility results. These are mostly located in the south-eastern part of Bavaria.

types (aggregated into six broader classes, again for dimensionality reduction) (Bayerisches Landesamt für Umwelt, 2023b). Additionally, the *HÜK250 ("Hydrogeologische Übersichtskarte 1:250.000")* hydrogeological map provides information on surface permeability and geochemical rock type (Bundesanstalt für Geowissenschaften und Rohstoffe, 2019). Information about land cover, derived from Sentinel-2 imagery at a resolution of 10 m and including seven categories, was used to provide information about the type of coverage for the slope unit (German Aerospace Center (DLR), 2020; Weigand et al., 2020). Information on rainfall erosivity, was included to represent the spatial variability of rainfall-induced erosion potential across Bavaria (Uber et al., 2024, 2023). In addition, four structural and hydrological factors were derived from the stream network, such as distances to streams, lakes, and faults (Bundesanstalt für Gewässerkunde, 2024; Bayerisches Landesamt für Umwelt, 2018a, b, 2023a).

2.2.3 Landslide susceptibility maps

As discussed later in Sect. 3.3, the model results are also compared with existing but incomplete susceptibility maps produced by *LfU*. These maps, derived from the GEORISK inventory, are being developed progressively across Bavaria, starting with the most landslide-prone areas. However, early susceptibility maps have not been updated, leaving some of the most affected regions represented by outdated maps, in some cases from 2013 or 2014. Currently, only 47 of 96 regional districts in Bavaria

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

have complete susceptibility maps and seven districts have partial coverage, leaving only $51\,\%$ of Bavaria with susceptibility mapping (Bayerisches Landesamt für Umwelt, 2020a, 2024b).

LfU produces four types of susceptibility maps at 1:25,000 scale covering different landslide processes and scenarios: areas with actively sliding deep-seated landslides, areas at risk of new or reactivating deep-seated landslides, slopes susceptible to shallow failures under current conditions, and slopes susceptible to shallow failures under extreme deforestation. The two shallow slope failure maps are derived from physically-based simulations computing initiation zones and runout areas from slope stability analysis, geology, and land cover information. The deep-seated landslide maps follow a more empirical approach, where each documented landslide is assessed through field surveys and classified based on observed morphological indicators and geological conditions. Susceptible zones are then manually delineated as polygons, representing the estimated future process space and a safety buffer of $20 - 30\,\mathrm{m}$ (Bayerisches Landesamt für Umwelt, 2020b). While this approach reflects detailed expert knowledge, publicly available documentation provides limited information on the criteria used to identify potential new deep-seated landslides or to classify them into active, potentially reactivating, and inactive categories. As these zones are derived from case-by-case expert interpretation rather than a spatial predictive model, the maps inherently reflect the spatial coverage and completeness of the underlying field surveys, and the manual procedure makes it difficult to extend coverage to new areas or to incorporate newly recorded events in a timely manner (Bayerisches Landesamt für Umwelt, 2020b). These limitations of the *LfU* susceptibility maps legitimize to recalculate landslide susceptibility using a fully data-driven, spatially continuous modeling framework and a harmonized regional dataset. In this study, the *LfU* susceptibility maps will therefore only serve as a comparative framework for evaluating model outcomes.

2.3 Slope unit delineation, feature extraction, and data preparation

The delineation of slope units is an integral part of the data preparation. Each slope unit represents a single observation in our model, with all input features aggregated across the unit. The trained model then estimates the landslide susceptibility for each slope unit. This means from a conceptual point of view that susceptibility is estimated for the entire slope that could potentially be affected by a landslide, from the scar to the accumulation zone.

2.3.1 Segmentation of slope units

Despite their growing relevance, automated and reproducible tools for slope unit delineation remain limited. In this work, slope units are generated using the `r.slopeunits` algorithm (Alvioli et al., 2016). The original implementation requires manual parameter entry, cannot process multiple DEM tiles or run in parallel, and lacks metadata recording. To improve reproducibility and scalability, the workflow was therefore adapted: segmentation is performed using an `R` wrapper and the `rgrass` package (Bivand et al., 2022), which manages parameter settings, metadata documentation, and saving of temporary outputs, while keeping the core algorithm unchanged.

Slope units are delineated from the input DEM based on hydrological principles. Using the `r.watershed` module in GRASS GIS, flow accumulation is computed and drainage lines and divides are identified. These hydrological features form the basis for defining initial half basins, which are subsequently divided in a top-down strategy according to terrain-homogeneity

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

criteria. The initial flow-accumulation threshold t controls the size of the first half basins; the minimum unit size a and minimum circular variance c enforce constraints on size and aspect homogeneity. If c is set too low, the algorithm tends to over-segment connected slopes. The reduction factor r determines how quickly the threshold t decreases over the iterations, influencing both runtime and the granularity of the delineation. Optional parameters such as $maxarea$ and $cleansize$ limit maximum unit size and remove small artifacts in the segmentation (Alvioli et al., 2016).

At each iteration, cells with flow accumulation above the current threshold are treated as streams, catchments are defined accordingly, and half basins are split. The algorithm proceeds until either the maximum number of iterations is reached or all candidate units meet the criteria for circular variance, minimum size, and, if specified, maximum area (Alvioli et al., 2016). A final cleaning step removes very small or narrow artifacts by merging them with neighboring slope units. Because the removal of narrow units can itself produce new fragments below the cleaning size threshold, we adapt the original routine and re-aggregate these remaining pieces to neighboring units.

Parameter selection followed a quantitative and qualitative evaluation across three test sites representing the main terrain types found in Bavaria: flat, hilly, and mountainous. Multiple configurations of t, a, c, r, $maxarea$ and $cleansize$ were tested and compared by examining the distribution of slope unit sizes and total counts at each site. Configurations producing either very large units that failed to capture terrain variability or a large number of implausibly small fragments were discarded. The remaining candidates were then assessed visually by overlaying the resulting slope unit boundaries on the DEM and evaluating their alignment with drainage lines and ridges.

We found the best delineation is achieved using $200\,\text{ha}$ as initial threshold t, $50\,\text{ha}^2$ as minimum area size, 0.2 as minimum circular variance, a reduction factor of 5, the maximum number of iterations set to 30, a maximum area size of $350\,\text{ha}^2$ and the cleaning size at $3\,\text{ha}$. This results in a comprehensive slope unit map for Bavaria with $317,788$ slope units and a mean size of $22.7\,\text{ha}$. Figure 3 shows an example of the resulting slope unit segmentation.

Given the size of our study area and the resolution of the data, applying the slope unit delineation across all of Bavaria requires addressing computational limits. Running `r.slopeunits` on the full $1\,\text{m}$ DEM is not feasible, since the algorithm generates a complete segmentation for the entire input at every iteration. We address this limitation by resampling the DEM to a $5\,\text{m}$ resolution, reducing noise and lowering the processing time substantially, as well as by splitting the study area into seven administrative units (*Regierungsbezirke*) (Landesamt für Digitalisierung Breitband und Vermessung Bayern, 2023) with an overlap of $5\,\text{km}$ and running delineation independently in every one of these.

Since slope units are delineated independently within each buffered administrative unit, slope units located near the boundaries may be generated in both neighboring subsets. In the overlapping areas, slope units with identical geometries are therefore kept only once and slope units cut by the buffer boundary are removed to avoid incomplete units. In some cases, slope units delineated on both side of a boundary differ slightly in their geometry. These mismatched slope units are kept, as their spatial extent and aggregated feature values are very similar, leading to comparable landslide susceptibility estimations. As a result, a few slope units overlap the boundaries of at least two administrative units, which have slightly different shapes. We use these units in the model to avoid gaps and to ensure a continuous susceptibility map of Bavaria, since their effect on the final product

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

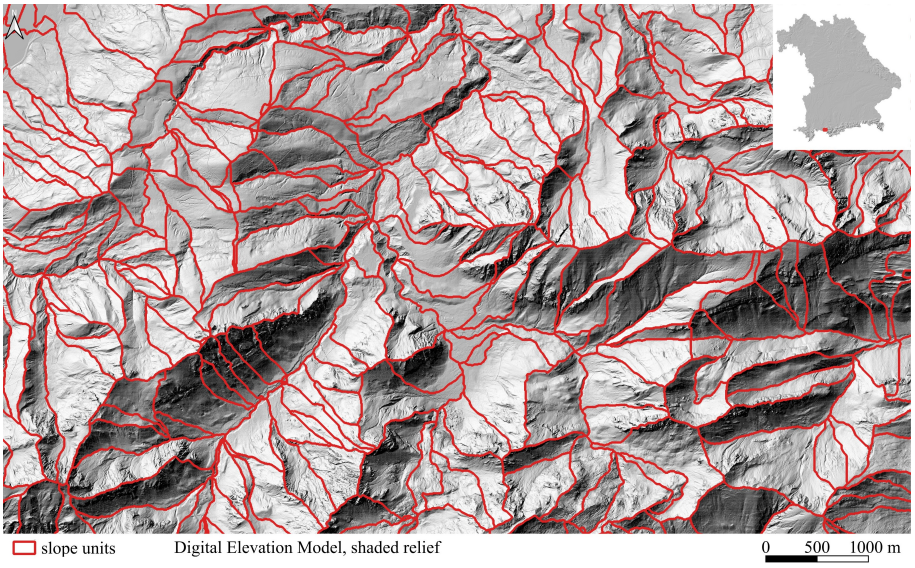

Figure 3. Example of the slope units in a part of the Ammergau Alps. The boundaries follow the ridges of the mountains in an accurate manner. The terrain in the background is illustrated using the DEM from Landesamt für Digitalisierung Breitband und Vermessung Bayern (2024b), licensed under CC BY 4.0.

can be considered negligible in the context of the overall analysis. Finally, the slope units are spatially joined with the landslide inventory, meaning that each unit carries the attributes of the most recent landslide event occurring within its boundaries.

2.3.2 Feature extraction and data preparation

After creating the slope units for the landslide susceptibility model, we extract the input features for the model separately for each of the seven administrative units of Bavaria.

Most topography-related variables are derived from the 5 m DEM (as shown in Table B1 in the appendix) using the `whiteboxtools` package (Wu and Brown, 2019). These include slope, aspect, plan and profile curvature, ruggedness measure, length-slope (LS) factor, terrain relief intensity and variance, and the convergence index. Since the convergence index is not available in `whiteboxtools`, it was derived using `starsExtra` (Dorman, 2020). Additionally, hydrology-related features such as flow accumulation, topographic wetness index, stream power index and downslope distance gradient are also computed. This results in 14 landslide-influencing factors derived from the 5 m resolution DEM. Besides the DEM, we use additional features such as rainfall erosivity, land cover, the Euclidean distances to the nearest streams, lakes, and faults, as well as geology, permeability, geochemical rock type and soil type from other input maps.

The selection of input variables follows established practice in landslide susceptibility modeling (Reichenbach et al., 2018; Lima et al., 2022). In general, landslides are often explained by the interaction of (i) topographic preconditions, (ii) hydrological

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

and erosion processes, and (iii) geological and land cover properties (Reichenbach et al., 2018). The chosen variables reflect these process domains.

Topographic variables are among the most widely used inputs (Reichenbach et al., 2018). Slope steepness is the most fundamental and theoretically grounded feature, as it directly controls slope stability and has been shown to be the single most effective variable in susceptibility modeling (Taylor, 1948; Reichenbach et al., 2018; Fabbri et al., 2003; Carrara et al., 1991). Curvature influences terrain processes such as erosion and runoff (Lima et al., 2022), while additional metrics such as the convergence index provide a practical representation of slope curvature and have been linked to small and medium-sized landslide events (Nakileza and Nedala, 2020; Regmi et al., 2014). Other terrain variables such as aspect and elevation are less universally justified and mainly reflect local terrain conditions (Reichenbach et al., 2018). Advanced terrain metrics, including ruggedness measure and terrain relief intensity, capture local topographic roughness associated with slope instability (Reichenbach et al., 2018; Lombardo et al., 2021; Carrara et al., 1991).

Hydrological indices such as the topographic wetness index and flow accumulation suggest the tendency of a location to accumulate water and are commonly used to approximate moisture conditions influencing slope stability (Reichenbach et al., 2018; Lima et al., 2022). Related metrics such as the downslope distance gradient (downslope index) describe the potential for water to flow downslope from a given position (Hjerdt et al., 2004; Lindsay, 2023). Erosion-related indices including the LS factor and the stream power index quantify the influence of topography on soil erosion and the erosive power of flowing water (Pourghasemi et al., 2013; Lindsay, 2023). Rainfall erosivity is included as a precipitation-related variable representing the spatial differences in soil erosion due to rainfall (Uber et al., 2024). Proximity to streams, lakes, and faults captures destabilizing effects related to erosion, groundwater influence, and structural weaknesses (Reichenbach et al., 2018; Lombardo et al., 2021; Pourghasemi et al., 2013; Varnes, 1978).

Geological variables, including soil type, subsoil type, geochemical rock type, and permeability, are used according to a review of over 300 studies in about $73\,\%$ of the susceptibility studies (Lima et al., 2022). They represent material properties controlling shear strength and drainage capacity (Reichenbach et al., 2018; Varnes, 1978). Land cover, included in around $63\,\%$ of reviewed publications (Lima et al., 2022), accounts for the stabilizing effect of vegetation root and changes in evapotranspiration (Reichenbach et al., 2018).

All features are aggregated for each slope unit using the `exactextract` package (Baston, 2024). For continuous variables, we calculate the mean and standard deviation of all the pixels within the slope unit (Jacobs et al., 2020). For categorical variables, we calculate the percentage of the slope unit area that is covered by each category. A list of input features, used metrics and data sources is available in Table B1 in the appendix. The final dataset therefore includes $317,788$ slope units with 84 input features representing landslide-influencing factors.

We then combine the landslide inventory with the input features and perform a descriptive statistical analysis to explore the differences between slope units with and without a recorded landslide event. The violin plots in Fig. 4 show significant differences in some selected features. Slope units with landslides tend to be higher in elevation, more rugged and steeper. They also demonstrate greater internal slope variability, indicating more heterogeneous terrain. These differences are statistically significant for all the features shown ($p < 0.01$, unpaired t-test).

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

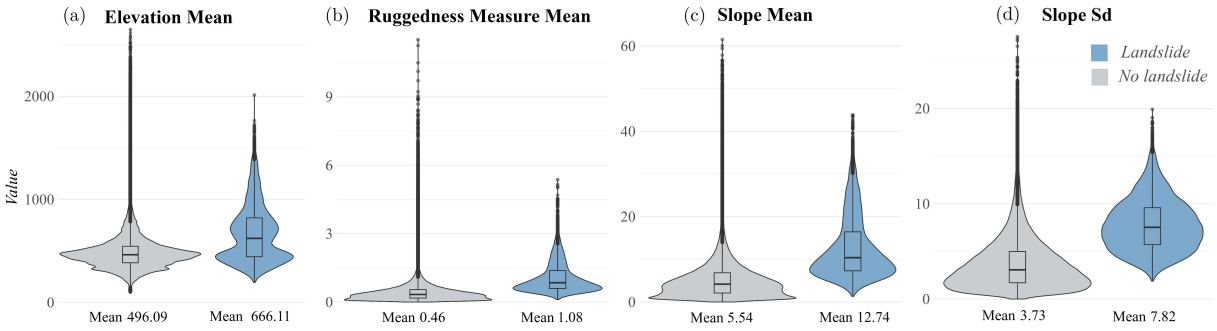

Figure 4. Violin plots showing the values of selected features for slope units with and without recorded landslide events. These plots show at a descriptive level, that non-landslide areas differ significantly from landslide areas in our inventory.

Since landslides are considered rare events, there are far fewer slope units with recorded landslides in our dataset (only $2.6\,\%$) than without, resulting in a strong class imbalance. This makes training for the susceptibility estimation challenging. To reduce the class imbalance, we first remove units with very low probability of having landslides, such as those dominated by water or artificial surfaces ($> 90\,\%$ coverage) and very flat units (slope mean $< 1°$ and standard deviation $< 0.5°$), discarding $11,720$ slope units. The data are then randomly split into $80\,\%$ for training (including a $10\,\%$ validation subset) and $20\,\%$ for testing, resulting in $220,368$ training samples, $24,486$ validation samples, and $61,213$ test samples. Then, we apply random oversampling of the minority class to the training data to reach approximately $10\,\%$ landslide cases, which performs in our model better than without resampling but does not enforce a fully balanced dataset. The validation and test sets remain untouched and reflect the original class distribution to ensure unbiased model evaluation. Before modeling, all features are z-score standardized.

2.4 Model architecture and training

To estimate landslide susceptibility, we develop a fully connected ANN. ANNs are data-driven models that learn nonlinear relationships between input features and observed outcomes directly from the data, without requiring predefined functional forms or assumptions (Prosise, 2022; Bishop, 1995). This makes ANNs well suited for susceptibility modeling, where relationships between input features and slope instability are often nonlinear, interacting, and difficult to represent with simpler statistical models, particularly in imbalanced datasets.

An ANN consists of an input layer, one or more hidden layers, and an output layer. The input layer receives the input features, which are successively transformed through one or more hidden layers before reaching the output layer. In each neuron, incoming signals are weighted, summed with a bias term, and passed through an activation function. Through this layered structure and nonlinear activation, the network is able to learn complex dependencies and hierarchical feature representations that cannot be captured by linear models (Goodfellow et al., 2016; Prosise, 2022).

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

We employ the rectified linear unit (ReLU) activation function in the hidden layers to promote efficient training and mitigate vanishing gradient effects. For the output layer, we use a sigmoid activation that transforms the model output values between 0 and 1. This output can be understood as the susceptibility for a landslide in a given observation, i.e., a slope unit (Goodfellow et al., 2016; Prosise, 2022). We use the Adam algorithm for optimization due to its stability and efficiency in high-dimensional parameter spaces (Kingma and Ba, 2014). The learning rate is reduced by a factor of 0.1 if the validation loss does not improve for 10 epochs.

To avoid overfitting and improve generalization, two regularization strategies are applied: *dropout*, with a probability of 0.45, randomly deactivating neurons during training, and *early stopping* ends training if the validation loss does not improve more than 0.001 over 25 epochs. Hyperparameters, including the number of neurons per layer, learning rate, and dropout probability, are tuned using Latin Hypercube Sampling (LHS) (McKay et al., 1979). LHS ensures that each parameter range is sampled evenly across its domain, providing a more comprehensive exploration of the hyperparameter space than simple random sampling for grid search.

The final model architecture comprises 84 input features and six layers with 200, 600, 450, 200, 75, and 20 neurons, respectively, trained for up to 100 epochs. The loss is defined by the binary cross entropy (BCE); however, weighted BCE and binary focal cross entropy are tested to address the class imbalance in the data. The latter two loss functions, also in combination with various resampling techniques like SMOTE (Chawla et al., 2002) and Borderline-SMOTE (Han et al., 2005), did not perform as good as BCE with random resampling.

2.5 Shapley values for explainable AI

Neural networks offer several advantages over simpler models, including the ability to handle high-dimensional data, capture complex nonlinear relationships, and often achieve higher predictive performance (Goodfellow et al., 2016). A major drawback, however, is their black-box nature: the magnitude and direction of each feature's effect on the final prediction are not directly reported. Several approaches address this, of which *Shapley Additive Explanations (SHAP)* (Lundberg and Lee, 2017) provide a theoretically grounded framework for post-estimation model interpretation, which use local surrogate models to explain individual predictions. Surrogate models, in general, approximate the outputs of a complex model with a simpler representation, and local surrogates approximate the original model locally only around one prediction (in this case, a single slope unit with its landslide susceptibility estimation) (Molnar, 2022; Ribeiro et al., 2016). To extract this approximation, SHAP builds on game-theoretic Shapley values, which quantify the average marginal contribution of a feature across all possible feature subsets (Shapley, 1953).

SHAP combines local surrogate modeling with Shapley values to produce locally interpretable explanations. The computation of a SHAP value for a single feature i needs the predicted model outputs from all subsets of features excluding i, but the number of possible subsets rises exponentially with the amount of features, making it slow for high-dimensional models like ours (Molnar, 2022). Therefore, we apply the Monte Carlo approximation proposed by (Štrumbelj and Kononenko, 2014), where a random sample of feature subsets is drawn for each feature to efficiently estimate SHAP values.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

SHAP provides interpretable insights for complex ML models, offering local interpretability on feature effects (single slope unit) and global interpretability (all predictions across Bavaria). In this study, SHAP enables the identification input features that have the greatest impact on the prediction of landslide susceptibility, as well as the ones that are less relevant to include in the analysis. By using an exhaustive feature set with 84 variables, we try to determine which input data might be relevant for the landslide susceptibility estimation in Bavaria and which factors can be omitted in future work.

2.6 Comparison against *LfU* susceptibility maps

To contextualise the model's performance beyond standard accuracy metrics, the predicted susceptibility map is compared with the landslide susceptibility maps published by *LfU*, which cover four categories: present deep-seated landslides, potentially new or reactivating deep-seated landslides, shallow slope failures, and shallow slope failures under deforestation scenarios. To enable this comparison, the slope units are overlaid with the *LfU* polygons, so that each slope unit can be linked to the corresponding *LfU* susceptibility category. For each of the four categories, the mean predicted susceptibility of all slope units falling within the respective *LfU* zones is calculated. Since the shallow slope failure maps consist of many small, dispersed polygons while the deep-seated landslide zones form larger, more coherent areas, a simple mean would overrepresent the numerous small shallow-failure polygons. An area-weighted mean is therefore computed additionally, giving more weight to larger polygons. To summarize model agreement across all four categories combined, the same area-weighted mean is calculated for the overall susceptibility zone. Finally, density plots are used to visualize the distribution of predicted susceptibility values within each *LfU* category, complemented by a map section that illustrates spatial agreement and discrepancies between the two products. This provides a more complete picture of the overlap between our ANN based results and the *LfU* maps.

3 Results

This section presents the results of the ANN susceptibility model. First, the landslide susceptibility map and its predictive performance are evaluated. Second, the global and local feature importance is analyzed. Finally, the susceptibility map is compared to the existing, yet incomplete, *LfU* landslide susceptibility map and validated against the updated landslide inventory.

3.1 Landslide susceptibility map and model performance

Figure 5 shows the estimated susceptibility in the map, with darker colors indicating a higher likelihood of landslides. At first glance, a visual comparison with the landslide locations from the inventory (Fig. 2) shows a good agreement between predictions and reference data, which suggests a good model performance. High susceptibility is mainly concentrated in the Alps, the Franconian Jura, and the Rhön Mountains in northern Bavaria. The distribution of susceptibility values for slope units with and without recorded landslides (Fig. 5) supports this finding. Slope units affected by past landslides show in the boxplot a median susceptibility slightly above 0.5, whereas unaffected units have median values close to 0. The density plot reveals that most stable areas are assigned low susceptibility, while the areas with past landslides show a bimodal susceptibility

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

distribution, with peaks around 0.25 and between 0.5 and 0.6. Very few landslide-affected units receive predictions below 0.1, suggesting a good overall fit.

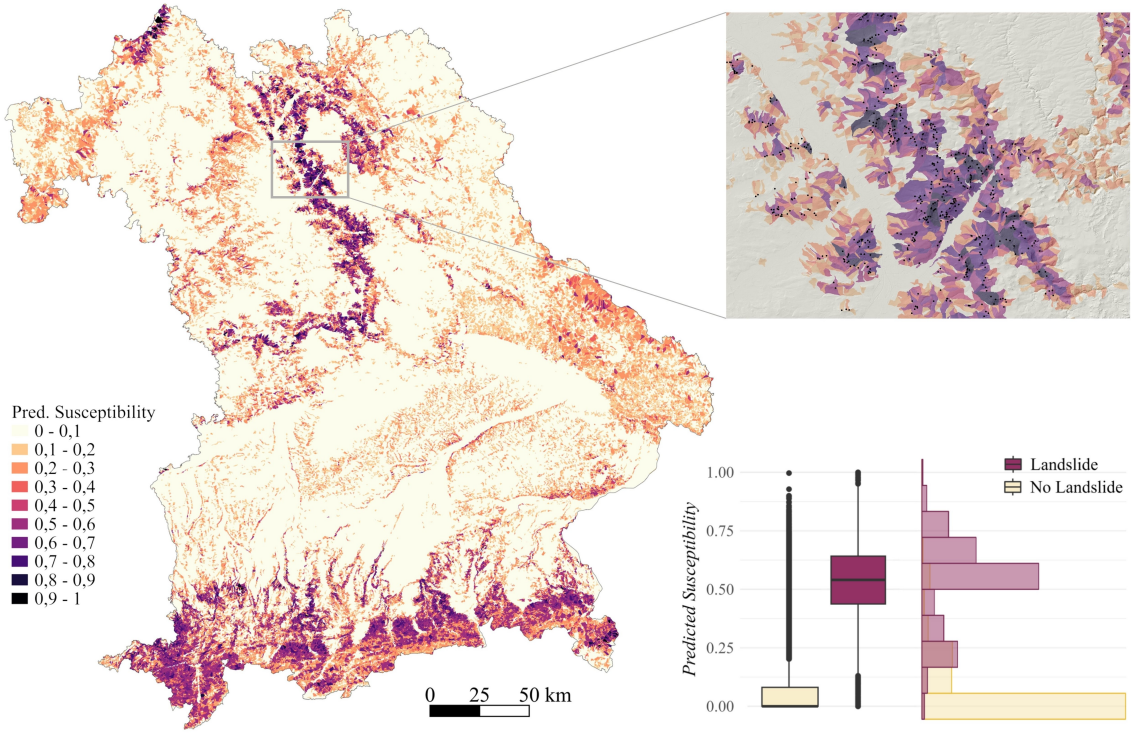

Figure 5. Estimated landslide susceptibility map of Bavaria with a boxplot of susceptibility values by class. Highest values (dark purple) occur in the Alps, the Franconian Jura, and the Rhön, with moderate susceptibility in the Bavarian Forest. A zoomed map section shows detailed model predictions and landslide locations overlaid on the DEM (Landesamt für Digitalisierung Breitband und Vermessung Bayern, 2024b), licensed under CC BY 4.0. The boxplot and accompanying histogram illustrate higher median and mean susceptibility values for slope units containing landslides, highlighting the model's discriminative performance.

Figure 5 also presents a detailed view of the susceptibility map for parts of the Franconian Jura. Predicted susceptibility values are shown for individual slope units alongside historical landslides and a hillshaded DEM. No events from the updated inventory are shown here, as these landslides are concentrated in other parts of Bavaria. Most recorded landslides occur in areas of high estimated susceptibility, illustrating the model's ability to generalize unstable terrain based on learned feature patterns.

Model performance is evaluated using metrics suitable for strongly imbalanced data. While the training set is resampled to include 10 % landslide cases, the test set remains with the original distribution, with approximately 2.6 % positive observations. Based on the distributions shown in Fig. 5, a threshold of around 0.2 appears appropriate for binary classification into "high susceptibility" and "low susceptibility". At this level, only few "no landslide" observations are misclassified, while nearly all past landslide locations fall in the "high susceptibility" class.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

Correctly identifying susceptible areas is particularly important in the field of landslide risk. High sensitivity ensures that most high-risk areas are detected, minimizing the chances of missing locations where landslides might occur. Missing a susceptible area (a false negative) can have serious consequences, so prioritizing sensitivity over specificity is generally best in this context (Frattini et al., 2010). Although lower specificity means that some stable areas are incorrectly labeled as susceptible (false positives), this is an acceptable outcome in landslide susceptibility mapping, as false positives can still be monitored at relatively low cost compared to the potential risk of false negatives (Frattini et al., 2010). Also, new landslides may occur in areas where no historic landslides were detected, highlighting the primary goal of susceptibility mapping: to identify locations at risk of landslides beyond those that have already experienced such events.

Confusion matrices for the test data are computed for four thresholds (Table 1). Although accuracy of the test data is one of the most commonly used evaluation metrics, it may not be ideal for imbalanced data. However, for the sake of completeness, it is included in Table 1. In contrast, the balanced accuracy provides a more reliable measure for performance in this context, and it is calculated as the average of sensitivity and specificity. The highest balanced accuracy occurs at a threshold of 0.2, further supporting that this is a good threshold to distinguish between high and low susceptibility slopes.

Threshold	Accuracy	95% CI	Sensitivity	Specificity	Balanced Accuracy
0.15	0.808	(0.804, 0.811)	0.944	0.804	0.874
0.20	0.859	(0.856, 0.862)	0.915	0.858	0.887
0.25	0.917	(0.915, 0.919)	0.833	0.919	0.876
0.50	0.957	(0.955, 0.958)	0.668	0.965	0.816

Table 1. Confusion matrix statistics for different thresholds. At the decision threshold of 0.20 we reach the highest balanced accuracy and a good trade off between sensitivity and specificity. Naturally, with increasing threshold the sensitivity decreases and specificity rises.

The ROC Area under the curve (ROC AUC) metric, widely used in the classification context, evaluates the ability of a model to distinguish between two classes without relying on a specific threshold. By assessing the performance for test data across all possible thresholds, the metric provides a global picture of the classifier (Alqadhi et al., 2024b; Ramsundar and Zadeh, 2018). The susceptibility model achieves a ROC AUC of 0.953, indicating strong overall discrimination ability. However, given the class imbalance, the Precision–Recall AUC (PR AUC) is more informative for evaluating performance in the minority class. The model achieves a PR AUC of 0.844, reflecting good capability in detecting the minority class, avoiding the masking effect of the majority class distribution (Geron, 2019; Saito and Rehmsmeier, 2015).

3.2 Model explainability

To interpret the ANN, Shapley values are computed for each feature and observation. These values can quantify both the magnitude and direction of a feature's contribution to the individual predictions (Molnar, 2022).

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

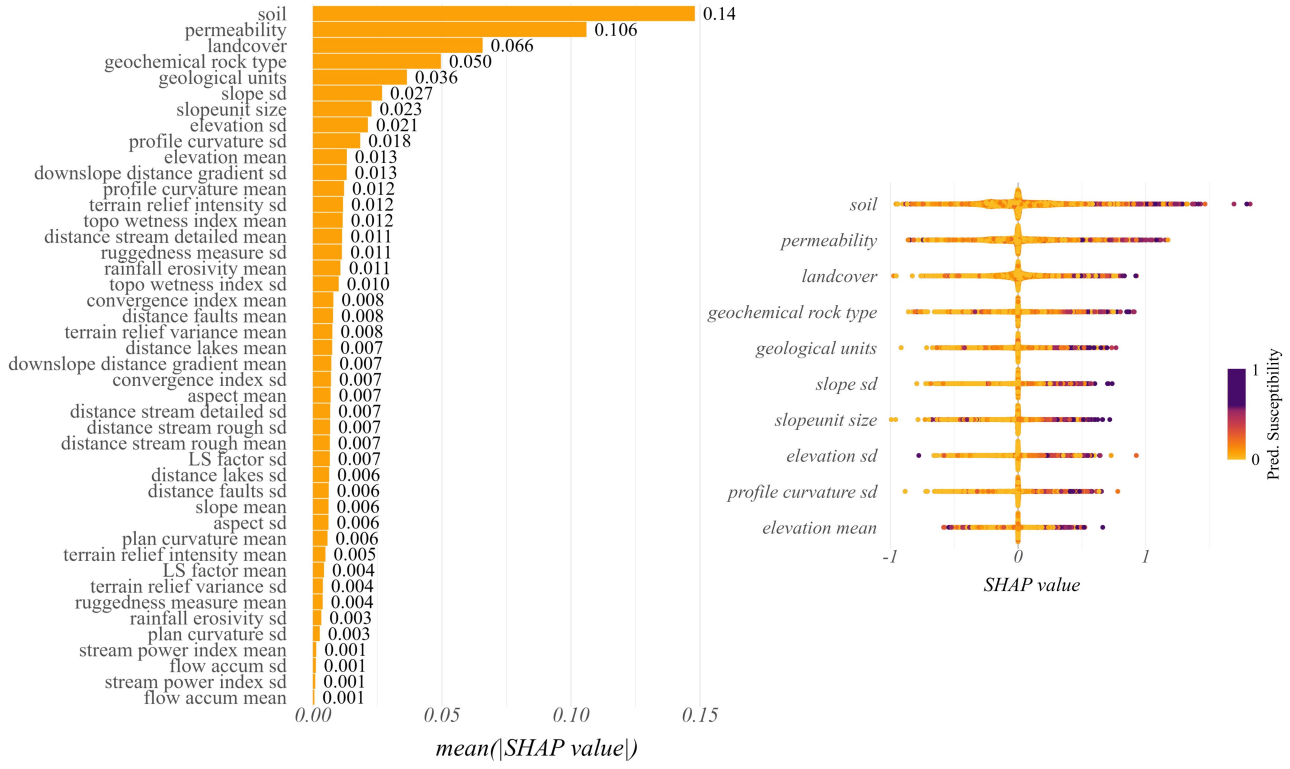

Figure 6. SHAP feature importance and beeswarm plot for collapsed features. Global importance is calculated as the mean SHAP value across all slope units and shown on the left, with categorical variables aggregated into single features. The right panel presents the beeswarm plot for the ten most important features, illustrating the distribution of SHAP values for individual observations and their influence on the estimated susceptibility.

First, we assess the global feature importance using the absolute Shapley values averaged across all observations (Fig. 6, left). The categorical features are aggregated across their individual classes. The results indicate that *soil type*, *permeability*, *land cover*, *geochemical rock type*, and *geological units* have the strongest influence on predicted landslide susceptibility. This finding is rather unexpected, as in many studies topographic features such as *slope* and *elevation* play a bigger role in the prediction (Alqadhi et al., 2024a; Inan and Rahman, 2023; Pradhan et al., 2023; Teke and Kavzoglu, 2024). Nevertheless, the *standard deviations of slope* and *elevation* within slope units, which somewhat reflect the slope's ruggedness, also rank among the most influential factors. In contrast, hydrological variables such as *flow accumulation* and *stream power index* play only a minor role in our ANN.

To provide a more detailed perspective on the SHAP values and also show the direction of the effects, Fig. 6 illustrates on the right side the 10 most influential features, with each point representing a SHAP value for a specific feature and observation. The points are colored according to the predicted susceptibility value of the corresponding slope unit. The figure demonstrates that higher SHAP values tend to positively influence the prediction, resulting in a higher estimated susceptibility. Additionally,

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

features such as *soil type* and *permeability* show a wider range of SHAP values than *elevation*, whose influence remains comparatively moderate.

Figure 7 shows the values of the single slope units, with the x-axis indicating the feature values rescaled to their original scale, and the y-axis displaying the SHAP values. The color gradient highlights observations with high predicted susceptibility in shades of red and purple. Slope units with low *elevation* are associated with negative SHAP values and low predicted susceptibility, whereas mid-range *elevation* (approximately 700–1500 m) show more dispersed SHAP values and oftentimes high susceptibility estimates. At *elevations* above about 2000 m, susceptibility decreases again, likely due to very steep Alpine terrain where rockfalls dominate over landslides (Bayerisches Landesamt für Umwelt, 2020b). The findings regarding the distribution of mean elevation within a slope unit are supported by existing literature (Inan and Rahman, 2023).

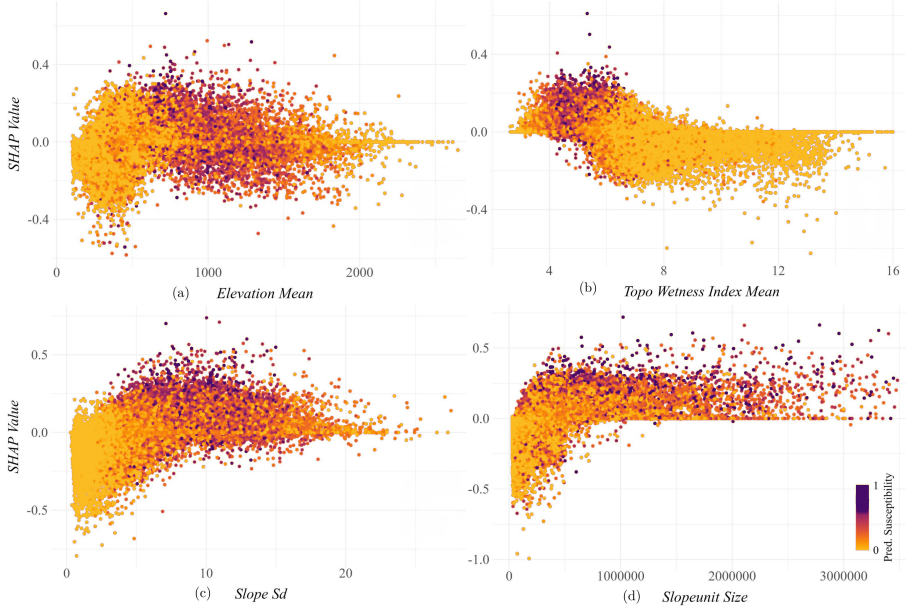

Figure 7. Scatterplots of four features and their SHAP values. The x-axis portraits the original feature values, while the y-axis gives information on the SHAP values. Every point in the scatterplot represents a single observation and is colored in its respective susceptibility estimation. For all four features in the plot, patters are visible between feature values, model importance and final predictions.

The SHAP values of the *topographic wetness index* indicates that lower values of this index tend to positively affect the predicted susceptibility, visible by the color gradient and the predominantly positive SHAP values in this range. With increasing index values the effect reverses and higher *topographic wetness* negatively impacts the predicted susceptibility in this model. This behavior is unexpected and contrasts with previous studies (Inan and Rahman, 2023; Le et al., 2024), as higher topographic wetness index values are generally associated with greater landslide susceptibility through water accumulation, increased soil saturation, and elevated pore water pressure, which collectively reduce shear strength and promote slope failures (Van Asch et al., 1999). Furthermore, The *standard deviation of slope* within a slope unit shows a positive relationship with susceptibility,

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

indicating that internally heterogeneous terrain is more prone to failure. A similar pattern is observed with the *size of the slope units*: smaller slope units negatively impact the prediction, while larger units results in a positive effect on susceptibility values, suggesting that larger cohesive slopes are more unstable.

To get more detailed insights into how the individual susceptibility predictions are created, Fig. 8 shows an example for a slope unit with low predicted susceptibility alongside one with high predicted susceptibility. These plots give insight into which features, along with their exact feature values, have a big impact on the final predictions for these two individual observations. In this representation, categorical features are aggregated to a single variable, resulting in feature values of 1, as the individual categories represent percentages that sum up to 1 in the aggregation.

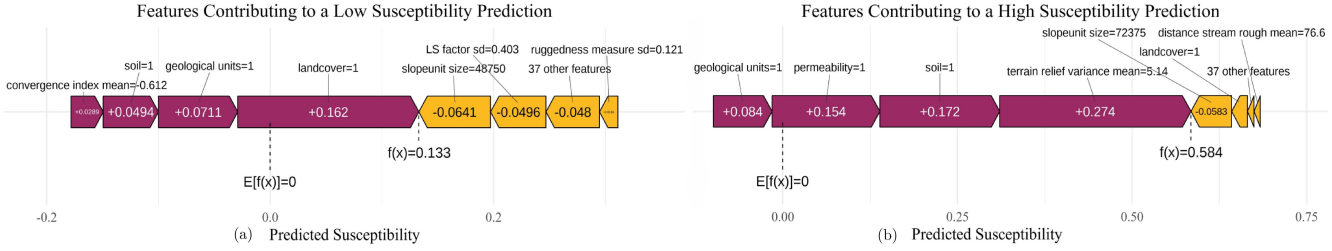

Figure 8. Force plots for individual high and low susceptibility observations. The figure shows easy local interpretability of two exemplary slope units, where the contributing feature values, magnitude of importance, directions of the effects and final susceptibility estimations are visible. Pink arrows correspond to positive effects of a feature on the predicted susceptibility, yellow arrows are negative effects.

In summary, the global SHAP-based analysis clarifies which input variables are most influential in determining estimated landslide susceptibility and identifies features that play only a minor role in the model. While these results reflect feature importance within the model framework, they should not be interpreted as evidence of causal relationships or physical landslide mechanisms (Lundberg, 2019). Nevertheless, this information is highly valuable for making practical modeling decisions. In particular, it supports informed variable selection when computational constraints require a smaller feature set or simpler model architectures. Estimating feature importance from a limited or preselected subset of variables can be misleading because it may overemphasize certain features while obscuring the relevance of others. By assessing feature importance within an intentionally large input space, this study provides a robust basis for identifying variables that genuinely contribute to model performance. At the same time, the SHAP scatterplots in Fig. 7 highlight the strongly nonlinear relationships between input features and predicted landslide susceptibility. These patterns highlight the importance of using more flexible, nonlinear ML models, as linear or logistic regressions reach their methodological limits here.

3.3 Comparison with *LfU* susceptibility map

A section of the four landslide susceptibility maps published by *LfU* is illustrated on the left side of Fig. 9. Two of them cover deep-seated landslides in large coherent polygons: actively sliding slopes in red and areas at risk of new or reactivating

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

deep landslides in orange. The other two cover shallow slope failures under current vegetation conditions (yellow) and under extreme deforestation scenarios (green), represented as numerous smaller, dispersed polygons.

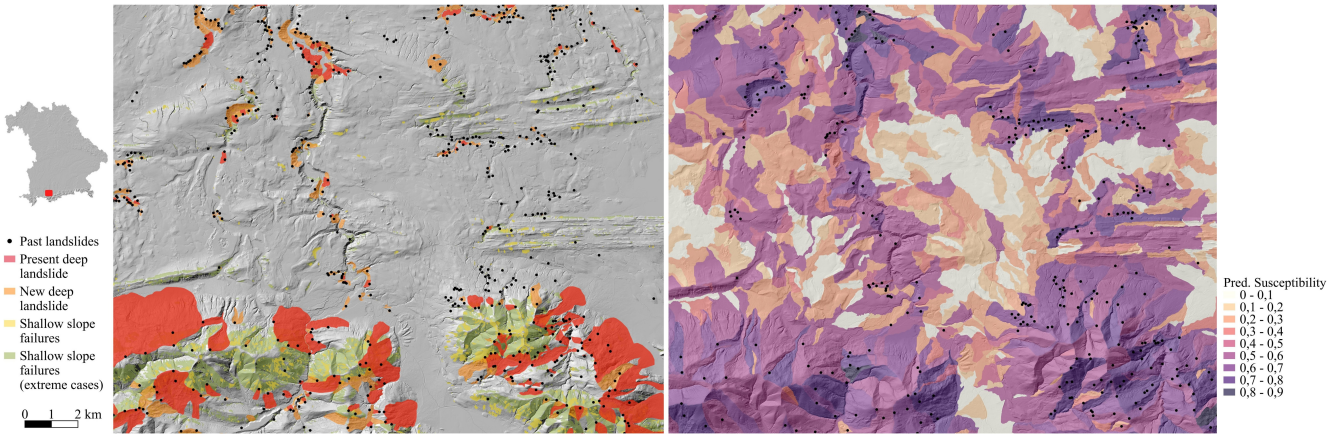

Figure 9. Landslide susceptibility zones published by Bayerisches Landesamt für Umwelt (2020a), licensed under CC BY 4.0. On the left side, the susceptibility zones of present deep landslides, new deep landslides, shallow slope failure and shallow slope failures under deforestation conditions are shown in four different colors. On the right, the map contains the same map section and illustrates the predicted susceptibility from the ANN. Additionally, the locations of landslides in the landslide inventory are depicted as black dots. For more context, the background contains the shaded relief from the DEM (Landesamt für Digitalisierung Breitband und Vermessung Bayern, 2024b), licensed under CC BY 4.0

For a visual comparison of our predicted susceptibility map and the susceptibility maps from *LfU*, Fig. 9 displays the same map section also on the right side with the map estimated by our ANN. While *LfU* identifies relatively few susceptible areas, our predicted susceptibility map is more conservative, classifying more slopes as potentially unstable. Conversely, *LfU* zoning of susceptible areas missed out most of the inventoried landslides, which may appear overly optimistic. Our ANN model predicts high susceptibility in most of the landslide locations.

The agreement between our susceptibility map and the *LfU* maps shows, that on average the predicted susceptibility in areas classified as susceptible by the *LfU* map is 0.376 across the four maps. Table 2 shows that the susceptibility maps for deep landslides have higher estimated susceptibility values compared to shallow slope failures. This is most likely due to the large number of slides in the landslide inventory compared to the shallower flows ($82.4\,\%$ vs $5.5\,\%$). The weighted mean of predicted susceptibility, adjusted by polygon size, shows higher values, which confirms that especially the large susceptible slope units are predicted well. Present deep landslides have the highest predicted susceptibility with 0.583, while shallow slope failures are the least well-predicted, with a weighted mean of 0.356. This difference in the estimation of the landslide types may be due to the different methodology used by *LfU* in the map production, or the imbalanced landslide inventory used for the ANN, that overrepresents deep slides compared to shallow slope failures. The overall weighted mean of the predicted susceptibility is 0.484, which is a satisfactory result given our defined threshold of 0.2 for high susceptibility.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

	Present deep landslides	New deep landslides	Shallow slope failures	Shallow slope failures (extreme)	All
Mean pred. susceptibility	0.520	0.516	0.345	0.399	0.376
Weighted mean pred. susceptibility (by size)	0.583	0.575	0.356	0.422	0.484

Table 2. Mean predicted susceptibility in susceptible areas detected by Bayerisches Landesamt für Umwelt (2020a). For the four types of susceptibility maps, we calculate the average susceptibility values within the susceptibility zones. Overall deep landslides are detected better in the ANN model than shallow slope failures.

On closer examination of the distribution of the predicted susceptibility within *LfU*'s susceptible areas (Fig. A1 in the appendix), we found that most deep landslides have a susceptibility above 0.5, with only few areas close to 0. In contrast, shallow slope failures have a higher portion of areas with low estimated susceptibility. A second cluster is formed around the value 0.25, while many areas are correctly estimated as high susceptibility (over the threshold 0.2) here as well. This pattern suggests that the susceptibility model performs better in identifying deep landslides than shallow slope failures.

Overall, the quantitative comparison shows a high agreement between areas of high predicted susceptibility and official *LfU* susceptibility zones, particularly for deep landslides. Together with the strong predictive performance metrics (high ROC AUC and PR AUC, as well as high sensitivity and balanced accuracy at a threshold of 0.2), this shows the potential of using slope unit based methods with an ANN as a data-driven approach in landslide risk assessments.

3.4 Generalizability and spatial bias

Despite the high performance metrics in Sect. 3.1, when we applied the model to the updated inventory, which included 802 additional landslide events overlapping 672 slope units (626 of which had no prior landslides), the predictions revealed several challenges. The histogram of predicted susceptibilities for these updated events (Fig. 10) shows that most slope units are assigned very low probabilities, with a mean susceptibility estimation of 0.22. Many estimations have zero or very low susceptibility, indicating that the model generally fails to recognize landslides in underrepresented areas within the inventory. Lowering the decision threshold between high and low susceptibility predictions from 0.2 to 0.15 yields only two-thirds true positives, with one-third of the updated events missed (223 of 672 as false negatives). Therefore, although the model achieves high accuracy with a test set of landslides from the inventory, it struggles to generalize and predict updated recorded landslides in areas not represented in the inventory. This suggests that metrics like ROC AUC or PR AUC can mask spatial or sample-selection biases, as discussed below.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

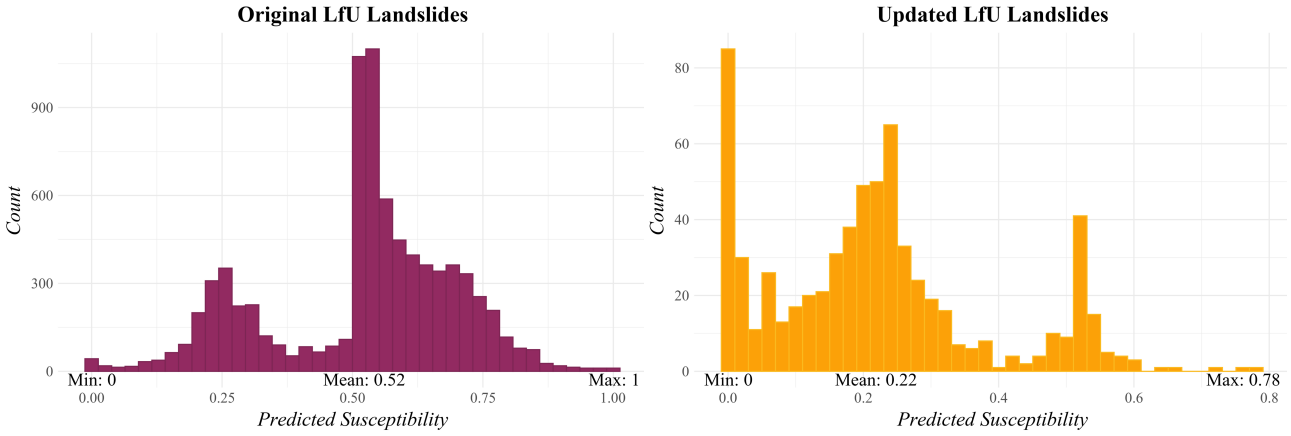

Figure 10. Histogram of estimated landslide susceptibility in areas with updated landslide events. The plot combines the distribution of the susceptibility values in areas with recently added landslides in yellow with the distribution for areas with original landslides in pink. Additionally, the minimum, maximum, and mean estimated values are reported.

4 Discussion

This study developed a region-wide landslide susceptibility map for Bavaria using a dense neural network and slope-unit-based terrain segmentation. While the model achieved strong predictive performance and identified influential features, such as geology, land cover, and terrain roughness, evaluation against updated inventories further highlights challenges related to model generalizability and inventory-related biases. The following discussion critically examines these aspects, including the effects of inventory incompleteness and the influence of spatial aggregation and the Modifiable Areal Unit Problem (MAUP) in slope unit segmentation.

4.1 Impact of inventory incompleteness on susceptibility predictions

We compared our predicted landslide susceptibility with existing susceptibility maps and updated recorded events. The model performs well in locations represented in the training data but struggles to generalize to underrepresented areas. This indicates that inventory incompleteness and spatial bias are the primary causes of the observed limitations. In other words, the training data do not represent the full range of landslide processes across Bavaria.

As shown in Fig. 2, the updated landslide locations are strongly clustered in a few areas (e.g. Erding, Landshut). This pattern reflects the historical development of the inventory: *LfU* initially focused on mapping the Alps and Franconian Jura and is only gradually expanding fieldwork to other regions. As a result, the spatial distribution of recorded landslides partly reflects differences in mapping effort rather than the true distribution of slope instability. Areas that have not yet been systematically surveyed appear artificially stable in the inventory, even though landslides may have occurred there.

This spatial bias directly affects the model. The 626 slope units with updated landslides were assigned as class 0 (no landslide) during training, teaching the network that these units represent stable conditions. When landslides were later recorded in

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

the same units but not included in training, the model continued to predict low susceptibility. This demonstrates how missing data in the inventory propagate into model bias, leading to systematic underestimation of risk in entire regions.

The problem is further compounded by a fundamental limitation of landslide inventory data. While reported landslides can be treated as positives, the absence of a record does not confirm stability. Slope units without a landslide are not verified negatives; they are unknowns. They may simply not have been surveyed or events may have gone undetected. By treating all unrecorded slope units as stable during training, the model implicitly assumes that the inventory is complete, which is rarely the case. This issue has been explicitly discussed in landslide susceptibility research, where the choice between presence-only and presence–absence approaches has been shown to significantly affect model outcomes and spatial predictions (Zhu et al., 2018; Steger et al., 2017). However, it remains difficult to fully resolve without reliable absence data or uncertainty quantification.

Our findings are consistent with recent research showing that spatially aggregated missing samples can severely bias susceptibility models. Huang et al. (2024) demonstrate that clustered gaps in an inventory lead to significant local prediction errors, whereas randomly omitted samples have milder effects. In our case, the updated landslides are concentrated in previously under-surveyed regions, resulting in a strong local underestimation of susceptibility.

From a methodological perspective, ANNs learn associations between feature patterns and the predicted class during training. They can generalize these learned associations to new slope units with similar characteristics. However, if certain landslide types or environmental conditions are absent from the training data, the model cannot learn to recognize them as unstable. The updated landslides likely reflect combinations of characteristics that were underrepresented or absent in the original inventory, such as different rock types or soil conditions.

This difference is visible in the feature distributions (Fig. A2). Updated landslides occur at lower *elevations* and are less scattered, suggesting more homogeneous altitude conditions. The permeability category identified as most important in the global SHAP analysis (*moderate to low permeability*) occupies noticeably less area in slope units with updated landslides than in those with original landslides. Moreover, the updated landslides occur on *slopes* that are less steep and show lower within-unit slope variability. Since slope variation was identified as an important predictor in the SHAP analysis, these differences likely contribute to the underestimation. In other words, the updated events are out-of-distribution examples, and introduce feature combinations that the ANN never learned to associate with instability. As a result, the learned decision boundary does not generalize well to these contexts.

Similar effects have been reported in previous studies. Steger et al. (2017) showed that a biased inventory led to an entire geological unit being falsely classified as stable. In our case, the regions with updated landslides would remain substantially underestimated unless the inventory is expanded.

Another related issue is model overconfidence. The histogram in Fig. 10 shows pronounced overconfidence in low-risk areas. Many slope units with updated landslides were assigned a susceptibility of zero. This indicates poor calibration for out-of-distribution samples and suggests high epistemic uncertainty. Such behavior limits the suitability of the current susceptibility map for early warning applications. Approaches such as explicit uncertainty quantification and uncertainty mapping could help identify unreliable regions and prevent overinterpretation of predictions (Schlögl et al., 2025). Likewise, spatial cross-

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

validation or transferability indices, as recommended by Steger et al. (2017), would likely expose the lack of generalization to newly surveyed areas.

In summary, the additional evaluation with out-of-distribution data highlights the central role of inventory quality and completeness for model validity. Although a fully unbiased inventory is rarely available, the integration of a relatively small number of updated landslides already revealed substantial weaknesses in the model. The original inventory contained spatial selection bias, which propagated into the model as systematic false negatives. This illustrates the limits of generalization when training data are spatially or thematically incomplete.

In practice, susceptibility and hazard maps derived from biased inventories should therefore be interpreted with caution. Until the inventory becomes more spatially complete, predictions in less-mapped regions remain highly uncertain. Improving inventory completeness or incorporating uncertainty quantification is essential for robust susceptibility modeling and reliable hazard assessment.

To mitigate spatial sampling bias in future work, spatially stratified sampling could be implemented during training. Instead of random sampling, training data could be selected to ensure balanced representation across regions or feature combinations. This would reduce overtraining on well-surveyed areas such as the Alps and Franconian Jura and promote a more representative learning process. Stratification could be applied geographically, for example by balancing samples across administrative regions, oversampling areas with few recorded landslides and undersampling densely surveyed regions. Alternatively, stratification could be performed in feature space along gradients of slope, elevation, or geology to avoid skewed associations caused by overrepresented conditions.

However, this strategy requires knowledge of spatial survey coverage, which is available in this case due to the documented progress of *LfU* fieldwork. In regions where survey intensity is unknown, it is not possible to distinguish between true low susceptibility and under-surveyed areas. Moreover, stratified sampling can only redistribute existing samples; it cannot compensate for entirely missing landslide records. It may reduce overtraining in densely surveyed regions, but systematic underestimation in genuinely unmapped areas would persist.

4.2 Challenges in slope unit segmentation

Another key challenge in this work is the segmentation of slope units. Since there is no consensus or objective method for defining the optimal delineation (Dahal and Lombardo, 2023; Schlögel et al., 2018; Moreno et al., 2024; Ahmed et al., 2023), the chosen segmentation is partly subjective and adapted to specific locations, which can affect the extracted input feature values. As with all analyses based on spatial aggregation, slope unit-based approaches are affected by the Modifiable Areal Unit Problem (MAUP), meaning that both scale and zoning can bias the results (Openshaw, 1984; Manley, 2014). In the context of landslide susceptibility, these effects influence the extracted input feature values and consequently, the model predictions.

Larger units might smooth the values of landslide-influencing factors. This is known as scale effects, which means the aggregation of the observed patterns and relationships (Manley, 2014; Openshaw, 1984). Larger slope units may mask local instability on individual slopes, while finer partitions can highlight small-scale variability but increase noise. As the scale changes, statistical indicators, such as averages, correlations, or the calculated feature values, can vary, potentially obscuring or

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

exaggerating predictions. Zoning effects, in contrast, occur when different delineations at the same scale produce varying results (Manley, 2014; Openshaw, 1984). Topography-based slope units generally provide a more meaningful spatial representation than arbitrary boundaries; however, even within this framework, inconsistencies can arise. In the present dataset, some slope units cover the area of a possible landslide comprehensively, from scarp to runoff zone, while others only represent steep sections, missing the runoff. This heterogeneity can affect the coherence of the predictions and impact susceptibility mapping, as ideally the map should either reflect high susceptibility across the entire area affected by a landslide or focus solely on the steep, actively sliding sections, rather than a mix of both.

To minimize the negative effects of scaling and zonation, it is essential to segment the slope units properly. This involves selecting the appropriate scale for creating the slope units and the most objective delineation based on topography, ensuring that each polygon represents a complete slope (Alvioli et al., 2020). However, achieving this in large areas, such as Bavaria, is time-consuming and computationally intensive. In this study, parameter selection was guided by an evaluation across three test sites representing the main terrain types in Bavaria, combining quantitative comparison of slope unit size distributions with visual assessment of boundary alignment. Nevertheless, the final parameter choice holds an element of subjectivity, as evaluation of boundary quality was not fully standardized, and the selected configuration may not be equally optimal across all terrain types in such a large and geomorphologically diverse study area.

A step forward should involve objective optimization of slope unit segmentation. These are e.g. automated parametrization techniques for the segmentation of terrain using remote sensing data, which are widely applied across various fields (Drăguţ et al., 2014; Espindola et al., 2006; Flanders et al., 2003). In the context of landslide susceptibility, for instance, Alvioli et al. (2016) enhanced the `r.slopeunits` algorithm by introducing an optimization approach that calculates metrics for internal and external terrain aspect variance. This optimization function is applied in other studies as well, but due to its computational costs it is only suitable for small areas of interest (Alvioli et al., 2020; Bornaetxea et al., 2018; Schlögel et al., 2018). Previous works also suggest that DEM resolution plays a critical role, with resolutions around 10 m yielding more robust slope units (Schlögel et al., 2018) than using lower resolution DEMs. Applying these methods in a small test area and comparing the results with the slope-unit delineation would help to identify potential scale or zoning biases and give insight into the robustness and predictive performance of the landslide susceptibility map.

5 Conclusions

This study presents a regional-scale landslide susceptibility assessment for Bavaria based on slope units, a dense neural network, and explainable AI methods. Using more than 11,000 recorded landslide events and a wide range of topographic, geological, soil, land-cover, and hydrological landslide-influencing factors as input features, the model achieves high predictive performance on a test set (ROC AUC = 0.95, PR AUC = 0.84) and shows a high sensitivity in identifying unstable slopes. The resulting susceptibility map provides the first consistent, data-driven overview of landslide-prone areas across Bavaria at the regional scale.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

The slope-unit-based approach proved well suited for modelling such a large area, as it allows input features to be meaningfully aggregated while keeping computational costs manageable. The SHAP analysis adds transparency and interpretability to the model by highlighting geology, soil properties, permeability, land cover, and terrain heterogeneity as key features influencing landslide susceptibility in Bavaria.

A central finding of this work is the strong dependence of our model on the quality of the landslide inventory. While the model performs well on the inventory, its transferability to unseen locations (quantified by over 800 updated landslide events in the inventory) reveals clear limitations in its ability to generalize. Many updated landslides occurred in areas that were previously underrepresented in the inventory and were therefore assigned low susceptibility values. This demonstrates how spatial incompleteness and selection bias in landslide inventories can directly translate into systematic underestimation of susceptibility in poorly mapped regions. As a result, high overall performance metrics based on random test splits alone are not sufficient to fully assess model reliability across the entire study area.

Overall, this study shows that explainable ML models can offer valuable insights for regional landslide susceptibility mapping, but their reliability strongly depends on the quality and completeness of the underlying landslide inventory. Ensuring adequate spatial representativity of the training and evaluation data remains essential for robust susceptibility and hazard assessment.

Code and data availability. The code repository can be made available upon request during the review process and will be made publicly accessible upon acceptance. The Landslide Susceptibility Map for Bavaria is planned to be published via a web service. Due to the high spatial resolution of the map, the potentially sensitive nature of the information, and model uncertainties, the full spatial dataset is not provided for direct download. Access to the data may be granted upon request for scientific purposes. For the input data sources, we refer to the references provided in Section 2.2 and Appendix B1.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

Appendix A: Figures

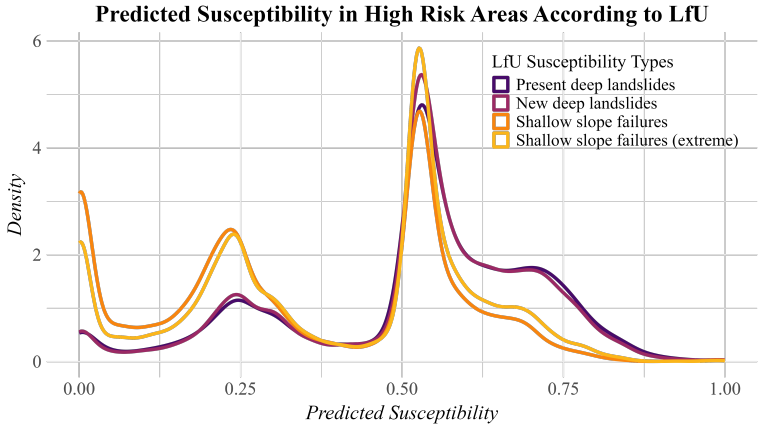

Figure A1. Density plot of predicted susceptibility compared to *LfU* susceptible areas. The plot shows the distribution of estimated susceptiblity for slope units that are within the four types of susceptible areas published by *LfU* (Bayerisches Landesamt für Umwelt, 2020a). A large portion of the estimated values lay at values over 0.5, indicating a good overlap between the here calculated susceptibility and *LfU*'s susceptibility map.

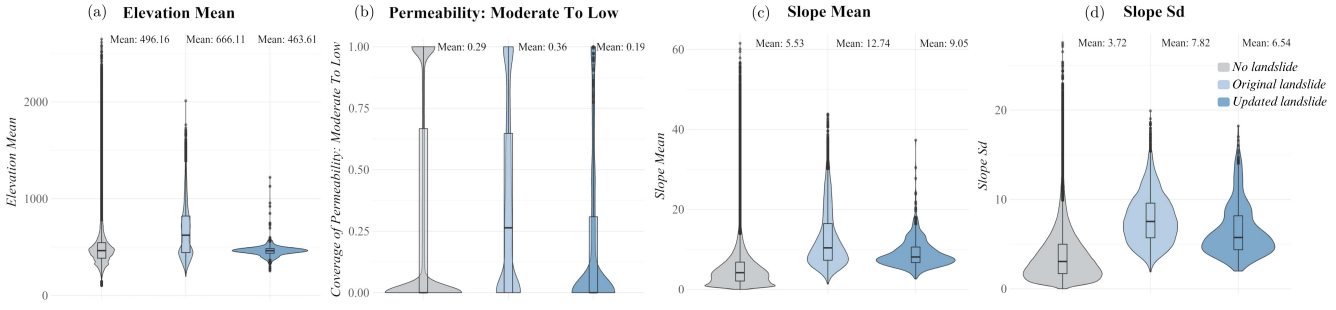

Figure A2. Violin plots detecting differences in original and updated landslide inventories. The figures show on a descriptive level how the original and updated landslide locations vary systematically. Three of the four features have a high global feature importance in the ANN model.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

Appendix B: Table of input features for the ANN model

Variable	Description	Statistics	Source
Elevation	Ground surface height in meters, excluding buildings or vegetation.	mean and sd	*DEM*, 5 m resolution (Landesamt für Digitalisierung Breitband und Vermessung Bayern, 2024a)
Slope	Steepness of the ground surface in degrees.	mean and sd	*DEM*
Flow accumulation	D8 flow accumulation: Number of upslope cells contributing flow to a given pixel.	mean and sd	*DEM*
Aspect	Slope orientation.	mean and sd	*DEM*
Plan curvature	Curvature of the terrain within a horizontal plane.	mean and sd	*DEM*
Profile curvature	Curvature of the terrain in the vertical plane, impacts the acceleration of downhill water flow.	mean and sd	*DEM*
Ruggedness measure	Elevation differences between a grid cell and its queen neighbors.	mean and sd	*DEM*
LS factor	Length-slope factor, combines slope length and steepness to explain how topography influences soil loss.	mean and sd	*DEM*
Stream power index	Erosive power of water flow by considering both terrain and flow accumulation.	mean and sd	*DEM*
Topographic wetness index	Propensity for a site to be saturated to the surface given its contributing area and local slope characteristics.	mean and sd	*DEM*
Downslope distance gradient	Potential for downhill water flow.	mean and sd	*DEM*
Terrain relief intensity	Elevation deviations within 1 km neighborhood.	mean and sd	*DEM*
Terrain relief variance	Elevation deviations within 1 km neighborhood, indicating how the relief is distributed throughout the landscape.	mean and sd	*DEM*
Convergence index	Structure of the terrain in terms of convergent and divergent areas.	mean and sd	*DEM*

Continued on next page

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

Variable	Description	Statistics	Source
Distance to stream course	Euclidean distance to the nearest streams using a generalized stream network.	mean and sd	*Waterbody-DE* (Bundesanstalt für Gewässerkunde, 2024)
Distance to stream detailed	Euclidean distance to the nearest streams based on a detailed stream network.	mean and sd	*FGN25* (Bayerisches Landesamt für Umwelt, 2018a)
Distance to lakes	Euclidean distance to the nearest lake.	mean and sd	*SEV25* (Bayerisches Landesamt für Umwelt, 2018b)
Distance to faults	Euclidean distance to the nearest geological fault.	mean and sd	*dGK25* (Bayerisches Landesamt für Umwelt, 2023a)
Slope unit size	Area of slope unit in square meters.	size	Own slope unit delineation
Rainfall erosivity	Potential for soil erosion caused by rainfall, specifically using current soil erosion values.	mean and sd	4 km resolution (Uber et al., 2023)
Land cover	Seven land cover classes.	% of land cover class	*Land Cover DE*, 10 m resolution (German Aerospace Center (DLR), 2020)
Geochemical rock type	Mineralogical-chemical composition of the ground, eight categories.	% of rock type class	*HÜK250* (Bundesanstalt für Geowissenschaften und Rohstoffe, 2019)
Permeability	Capacity of rock to allow water flow, with higher values representing greater flow capacity, 11 classes.	% of permeability class	*HÜK250*
Geological unit	Subsoil types at the ground surface, six classes.	% of geological class	*dIGK25* (Bayerisches Landesamt für Umwelt, 2023b)
Soil	Soil types across the study area, 13 classes.	% of soil class	*BÜK200* (Bundesanstalt für Geowissenschaften und Rohstoffe, 2022)

Table B1: Description of landslide-influencing factors used as input variables in the susceptibility modeling. Definitions follow Lindsay (2023); Uber et al. (2024); Lombardo et al. (2021).

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

Author contributions. Conceptualization: VB, MS, CG, HT; methodology, data curation, formal analysis, investigation, visualization: VB, MS; software: VB; data validation: VB, MS; supervision, funding acquisition: MS, HT, CG; writing – original draft: VB; writing – review and editing: MS, PAP, HT, CG, VB. All authors have read and agreed to the published version of the manuscript.

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

Financial support. The Free State of Bavaria funded this research via the Bavarian State Ministry of Economic Affairs, Regional Development and Energy as part of the “Earth Observation Innovation Laboratory for Climate Adaptation and Mitigation” project (www.EO4CAM.de).

Disclaimer. AI was used to improve the grammar, readability, and clarity of some parts of the text. However, AI was not used to generate scientific content, objectives, or citations. Parts of this paper are taken from Veronika Buchauer’s master thesis at University of Bamberg, which was submitted under the title "Explainable AI for Landslide Susceptibility Modeling in Bavaria: Integrating Neural Networks and SHAP Analysis" on 15 November 2024.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

References

Ahmed, M., Tanyas, H., Huser, R., Dahal, A., Titti, G., Borgatti, L., Francioni, M., and Lombardo, L.: Dynamic rainfall-induced landslide susceptibility: A step towards a unified forecasting system, International Journal of Applied Earth Observation and Geoinformation, 125, 103 593, https://doi.org/10.1016/j.jag.2023.103593, 2023.

Alqadhi, S., Mallick, J., and Alkahtani, M.: Integrated deep learning with explainable artificial intelligence for enhanced landslide management, Natural Hazards, 120, 1343–1365, https://doi.org/10.1007/s11069-023-06260-y, 2024a.

Alqadhi, S., Mallick, J., Alkahtani, M., Ahmad, I., Alqahtani, D., and Hang, H. T.: Developing a hybrid deep learning model with explainable artificial intelligence (XAI) for enhanced landslide susceptibility modeling and management, Natural Hazards, 120, 3719–3747, https://doi.org/10.1007/s11069-023-06357-4, 2024b.

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., and Guzzetti, F.: Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling, Geoscientific Model Development, 9, 3975–3991, https://doi.org/10.5194/gmd-9-3975-2016, 2016.

Alvioli, M., Guzzetti, F., and Marchesini, I.: Parameter-free delineation of slope units and terrain subdivision of Italy, Geomorphology, 358, https://doi.org/10.1016/j.geomorph.2020.107124, 2020.

Ba, Q., Chen, Y., Deng, S., Yang, J., and Li, H.: A comparison of slope units and grid cells as mapping units for landslide susceptibility assessment, Earth Science Informatics, 11, 373–388, https://doi.org/10.1007/s12145-018-0335-9, 2018.

Baston, D.: exactextractr: Fast Extraction from Raster Datasets using Polygons, [Software], https://isciences.gitlab.io/exactextractr/, 2024.

Bayerisches Landesamt für Umwelt: Fließgewässernetz 25, [Data set], accessed: 07.03.2024, 2018a.

Bayerisches Landesamt für Umwelt: Seenverzeichnis 25, [Data set], accessed: 07.03.2024, 2018b.

Bayerisches Landesamt für Umwelt: Gefahrenhinweiskarte Bayern, [Data set], https://www.lfu.bayern.de/umweltdaten/geodatendienste/pretty_downloaddienst.htm?dld=georisiken.xml, accessed: 30.01.2024, 2020a.

Bayerisches Landesamt für Umwelt: Methoden-Bericht zur Georisiken im Klimawandel - Vorgehen und technische Details, Tech. rep., Bayerisches Landesamt für Umwelt (LfU), Augsburg, https://www.bestellen.bayern.de/application/eshop_app000002?SID=1636211080&ACTIONxSETVAL(artdtl.htm,APGxNR:,AARTxNR:lfu_bod_00133,AKATxNAME:StMUG,APGxNR:,USERxARTIKEL:suchergebnisse.htm,USERxPORTAL:FALSE)=Z, accessed: 22.01.2024, 2020b.

Bayerisches Landesamt für Umwelt: Digitale Geologische Karte 1:25.000 (dGK25), [Data set], https://www.lfu.bayern.de/umweltdaten/geodatendienste/index_detail.htm?id=11ffe982-2913-4d67-b688-b1747fb94789&profil=WMS, accessed: 02.04.2024, 2023a.

Bayerisches Landesamt für Umwelt: Digitale Ingenieurgeologische Karte von Bayern 1:25.000 (dIGK25), [Data set], https://www.lfu.bayern.de/umweltdaten/geodatendienste/index_detail.htm?id=bbadf6af-af32-440a-ab30-b75558dfa0e6&profil=WMS, accessed: 02.04.2024, 2023b.

Bayerisches Landesamt für Umwelt: GEORISK-Objekte, [Data set], https://www.lfu.bayern.de/umweltdaten/geodatendienste/pretty_downloaddienst.htm?dld=georisiken.xml, accessed: 30.01.2024 and 13.08.2024, 2023c.

Bayerisches Landesamt für Umwelt: Beobachtung von Hangbewegungen, https://www.lfu.bayern.de/geologie/massenbewegungen/hangbewegungen/index.htm, accessed: 27.09.2024, 2024a.

Bayerisches Landesamt für Umwelt: Was sind Gefahrenhinweiskarten?, https://www.lfu.bayern.de/geologie/massenbewegungen/gefahrenhinweiskarten/index.htm, accessed: 30.09.2024, 2024b.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

Bishop, C. M.: Neural Networks for Pattern Recognition, Oxford University Press, Inc., USA, ISBN 0198538642, https://doi.org/10.1093/oso/9780198538493.001.0001, 1995.

Bivand, R., Krug, R., Lovelace, R., Neteler, M., Jeworutzki, S., and Vanderhaeghe, F.: rgrass: Interface Between 'GRASS' Geographical Information System and 'R', [Software], https://doi.org/10.32614/CRAN.package.rgrass, 2022.

Bornaetxea, T., Rossi, M., Marchesini, I., and Alvioli, M.: Effective surveyed area and its role in statistical landslide susceptibility assessments, Natural Hazards and Earth System Sciences, 18, 2455–2469, https://doi.org/10.5194/nhess-18-2455-2018, 2018.

Bundesanstalt für Geowissenschaften und Rohstoffe: Hydrogeologische Übersichtskarte 1:250.000 von Deutschland (HÜK250), [Data set], https://www.bgr.bund.de/DE/Themen/Grundwasser/Projekte/Flaechen-Rauminformationen/Huek250/huek250.html, accessed: 31.01.2024, 2019.

Bundesanstalt für Geowissenschaften und Rohstoffe: Bodenübersichtskarte 1:200.000 (BÜK200), [Data set], https://www.bgr.bund.de/DE/Themen/Boden/Flaechen-Rauminformationen/Bodenkundliche-Flaecheninformationen/BUEK200/buek200_node.html, accessed: 15.04.2024, 2022.

Bundesanstalt für Gewässerkunde: Waterbody-DE, [Data set], https://geoportal.bafg.de/smartfinderClient/?lang=de#/datasets/iso/7b86c167-cad5-4769-b32e-91a30f667e41, accessed: 07.02.2024, 2024.

Carrara, A.: Drainage and Divide Networks Derived from High-Fidelity Digital Terrain Models, in: Quantitative Analysis of Mineral and Energy Resources, pp. 581–597, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-009-4029-1_34, 1988.

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., and Reichenbach, P.: GIS techniques and statistical models in evaluating landslide hazard, Earth Surface Processes and Landforms, 16, 427–445, https://doi.org/10.1002/esp.3290160505, 1991.

Chang, Z., Catani, F., Huang, F., Liu, G., Meena, S. R., Huang, J., and Zhou, C.: Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors, Journal of Rock Mechanics and Geotechnical Engineering, 15, 1127–1143, https://doi.org/10.1016/j.jrmge.2022.07.009, 2023.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P.: SMOTE: Synthetic Minority Over-sampling Technique, Journal of Artificial Intelligence Research, 16, 321–357, https://doi.org/10.1613/jair.953, 2002.

Dahal, A. and Lombardo, L.: Explainable artificial intelligence in geoscience: A glimpse into the future of landslide susceptibility modeling, Computers & Geosciences, 176, 105 364, https://doi.org/https://doi.org/10.1016/j.cageo.2023.105364, 2023.

Dorman, M.: starsExtra: Miscellaneous Functions for Working with 'stars' Rasters, [Software], https://doi.org/10.32614/CRAN.package.starsExtra, 2020.

Drăguţ, L., Csillik, O., Eisank, C., and Tiede, D.: Automated parameterisation for multi-scale image segmentation on multiple layers, ISPRS Journal of Photogrammetry and Remote Sensing, 88, 119–127, https://doi.org/10.1016/j.isprsjprs.2013.11.018, 2014.

Espindola, G. M., Camara, G., Reis, I. A., Bins, L. S., and Monteiro, A. M.: Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation, International Journal of Remote Sensing, 27, 3035–3040, https://doi.org/10.1080/01431160600617194, 2006.

Fabbri, A. G., Chung, C.-J. F., Cendrero, A., and Remondo, J.: Is Prediction of Future Landslides Possible with a GIS?, Natural Hazards, 30, 487–503, https://doi.org/10.1023/b:nhaz.0000007282.62071.75, 2003.

Fischer, M. M.: Neural Networks: A General Framework for Non-Linear Function Approximation, Transactions in GIS, 10, 521–533, https://doi.org/https://doi.org/10.1111/j.1467-9671.2006.01010.x, 2006.

Flanders, D., Hall-Beyer, M., and Pereverzoff, J.: Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction, Canadian Journal of Remote Sensing, 29, 441–452, https://doi.org/10.5589/m03-006, 2003.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

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

German Aerospace Center (DLR): Land Cover DE - Sentinel-2, [Data set], https://doi.org/10.15489/1ccmlap3mn39, accessed: 16.04.2024, 2020.

Geron, A.: Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reilly Media, Inc., 2nd edn., ISBN 1492032646, 2019.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, http://www.deeplearningbook.org, 2016.

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., and Ardizzone, F.: Probabilistic landslide hazard assessment at the basin scale, Geomorphology, 72, 272–299, https://doi.org/https://doi.org/10.1016/j.geomorph.2005.06.002, 2005.

Han, H., Wang, W.-Y., and Mao, B.-H.: Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning, pp. 878–887, https://doi.org/10.1007/11538059_91, 2005.

Hjerdt, K. N., McDonnell, J. J., Seibert, J., and Rodhe, A.: A new topographic index to quantify downslope controls on local drainage, Water Resources Research, 40, https://doi.org/10.1029/2004WR003130, 2004.

Huang, F., Mao, D., Jiang, S.-H., Zhou, C., Fan, X., Zeng, Z., Catani, F., Yu, C., Chang, Z., Huang, J., Jiang, B., and Li, Y.: Uncertainties in landslide susceptibility prediction modeling: A review on the incompleteness of landslide inventory and its influence rules, Geoscience Frontiers, 15, 101 886, https://doi.org/https://doi.org/10.1016/j.gsf.2024.101886, 2024.

Inan, M. S. K. and Rahman, I.: Explainable AI Integrated Feature Selection for Landslide Susceptibility Mapping Using TreeSHAP, SN Computer Science, 4, 482, https://doi.org/10.1007/s42979-023-01960-5, 2023.

IPCC: Climate Change 2022: Impacts, Adaptation and Vulnerability Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/https://doi.org/10.1017/9781009325844, 2022.

Jacobs, L., Kervyn, M., Reichenbach, P., Rossi, M., Marchesini, I., Alvioli, M., and Dewitte, O.: Regional susceptibility assessments with heterogeneous landslide information: Slope unit- vs. pixel-based approach, Geomorphology, 356, 107 084, https://doi.org/10.1016/j.geomorph.2020.107084, 2020.

Jaedicke, C., Sverdrup-Thygeson, K., Syre, E., Nadim, F., Kalsnes, B., and Vangelsten, B. V.: Expected changes in climate-driven landslide activity (magnitude, frequency) in Europe in the next 100 years, SafeLand. Living with landslide risk in Europe: Assessment, effects of global change, and risk management strategies: project deliverable reports, https://www.ngi.no/globalassets/bilder/prosjekter/safeland/rapporter/d3.7.pdf, accessed: 01.11.2024, 2011.

Jemec Auflič, M., Bezak, N., Šegina, E., Frantar, P., Gariano, S. L., Medved, A., and Peternel, T.: Climate change increases the number of landslides at the juncture of the Alpine, Pannonian and Mediterranean regions, Scientific Reports, 13, 23 085, https://doi.org/10.1038/s41598-023-50314-x, 2023.

Kennedy, I. T. R., Petley, D. N., Williams, R., and Murray, V.: A Systematic Review of the Health Impacts of Mass Earth Movements (Landslides), PLoS Currents, https://doi.org/10.1371/currents.dis.1d49e84c8bbe678b0e70cf7fc35d0b77, 2015.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, 2014.

Landesamt für Digitalisierung Breitband und Vermessung Bayern: ALKIS® Verwaltungsgebiete, [Data set], https://geodaten.bayern.de/opengeodata/OpenDataDetail.html?pn=verwaltung, accessed: 26.01.2024, 2023.

Landesamt für Digitalisierung Breitband und Vermessung Bayern: Digitales Geländemodell 1m (DGM), [Data set], https://geodaten.bayern.de/opengeodata/OpenDataDetail.html?pn=dgm1, accessed: 04.03.2024, 2024a.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

Landesamt für Digitalisierung Breitband und Vermessung Bayern: Geländemodell, https://www.ldbv.bayern.de/produkte/landschaftsinformationen/gelaende.html, accessed: 25.09.2024, 2024b.

Le, X.-H., Choi, C., Eu, S., Yeon, M., and Lee, G.: Quantitative evaluation of uncertainty and interpretability in machine learning-based landslide susceptibility mapping through feature selection and explainable AI, Frontiers in Environmental Science, 12, https://doi.org/10.3389/fenvs.2024.1424988, 2024.

Lima, P., Steger, S., Glade, T., and Murillo-García, F. G.: Literature review and bibliometric analysis on data-driven assessment of landslide susceptibility, Journal of Mountain Science, 19, 1670–1698, https://doi.org/10.1007/s11629-021-7254-9, 2022.

Lindsay, J. B.: WhiteboxTools User Manual, https://www.whiteboxgeo.com/manual/wbt_book/, accessed: 10.10.2024, 2023.

Lombardo, L., Tanyas, H., Huser, R., Guzzetti, F., and Castro-Camilo, D.: Landslide size matters: A new data-driven, spatial prototype, Engineering Geology, 293, 106 288, https://doi.org/10.1016/j.enggeo.2021.106288, 2021.

Lundberg, S.: Be careful when interpreting predictive models in search of causal insights, https://shap.readthedocs.io/en/latest/example_notebooks/overviews/Be%20careful%20when%20interpreting%20predictive%20models%20in%20search%20of%20causal%20insights.html, accessed: 06.03.2026, 2019.

Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems, edited by Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., vol. 30, Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf, 2017.

Manley, D.: Scale, Aggregation, and the Modifiable Areal Unit Problem, in: Handbook of Regional Science, pp. 1157–1171, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-23430-9_69, 2014.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, Technometrics, 21, 239, https://doi.org/10.2307/1268522, 1979.

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., and Abderrahmane, B.: Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance, Earth-Science Reviews, 207, 103 225, https://doi.org/10.1016/j.earscirev.2020.103225, 2020.

Molnar, C.: Interpretable Machine Learning, 2 edn., https://christophm.github.io/interpretable-ml-book, accessed: 11.11.2024, 2022.

Moreno, M., Lombardo, L., Crespi, A., Zellner, P. J., Mair, V., Pittore, M., van Westen, C., and Steger, S.: Space-time data-driven modeling of precipitation-induced shallow landslides in South Tyrol, Italy, Science of The Total Environment, 912, 169 166, https://doi.org/10.1016/j.scitotenv.2023.169166, 2024.

Munich Re: Risiken durch Naturkatastrophen - Schäden nehmen tendenziell zu, https://www.munichre.com/de/risiken/naturkatastrophen.html, eaccessed: 22.02.2026, 2025.

Nakileza, B. R. and Nedala, S.: Topographic influence on landslides characteristics and implication for risk management in upper Manafwa catchment, Mt Elgon Uganda, Geoenvironmental Disasters, 7, 27, https://doi.org/10.1186/s40677-020-00160-0, 2020.

NASA: NASA Cooperative Open Online Landslide Repository (COOLR), [Data set], https://maps.nccs.nasa.gov/arcgis/apps/MapAndAppGallery/index.html?appid=574f26408683485799d02e857e5d9521, accessed: 30.01.2024, 2024.

Openshaw, S.: The Modifiable Areal Unit Problem, vol. 38 of *Concepts and techniques in modern geography : CATMOG, 38*, Geo Books, Norwich, ISBN 0860941345, https://www.tib.eu/de/suchen/id/TIBKAT%3A017804361, 1984.

Pourghasemi, H., Pradhan, B., Gokceoglu, C., and Moezzi, K. D.: A comparative assessment of prediction capabilities of Dempster–Shafer and Weights-of-evidence models in landslide susceptibility mapping using GIS, Geomatics, Natural Hazards and Risk, 4, 93–118, https://doi.org/10.1080/19475705.2012.662915, 2013.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

Pradhan, B., Dikshit, A., Lee, S., and Kim, H.: An explainable AI (XAI) model for landslide susceptibility modeling, Applied Soft Computing, 142, 110 324, https://doi.org/10.1016/j.asoc.2023.110324, 2023.

Prosise, J.: Applied Machine Learning and AI for Engineers, O'Reilly Media, Inc., https://learning.oreilly.com/library/view/applied-machine-learning/9781492098041/, 2022.

Ramsundar, B. and Zadeh, R. B.: TensorFlow for Deep Learning: From Linear Regression to Reinforcement Learning, O'Reilly Media, Inc., 1st edn., ISBN 1491980451, https://learning.oreilly.com/library/view/tensorflow-for-deep/9781491980446/, 2018.

Regmi, N. R., Giardino, J. R., McDonald, E. V., and Vitek, J. D.: A comparison of logistic regression-based models of susceptibility to landslides in western Colorado, USA, Landslides, 11, 247–262, https://doi.org/10.1007/s10346-012-0380-2, 2014.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F.: A review of statistically-based landslide susceptibility models, Earth-Science Reviews, 180, 60–91, https://doi.org/10.1016/j.earscirev.2018.03.001, 2018.

Ribeiro, M. T., Singh, S., and Guestrin, C.: "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, KDD '16, pp. 1135–1144, Association for Computing Machinery, New York, NY, USA, ISBN 9781450342322, https://doi.org/10.1145/2939672.2939778, 2016.

Saito, T. and Rehmsmeier, M.: The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets, PLOS ONE, 10, e0118 432, https://doi.org/10.1371/journal.pone.0118432, 2015.

Schlögel, R., Marchesini, I., Alvioli, M., Reichenbach, P., Rossi, M., and Malet, J.-P.: Optimizing landslide susceptibility zonation: Effects of DEM spatial resolution and slope unit delineation on logistic regression models, Geomorphology, 301, 10–20, https://doi.org/10.1016/j.geomorph.2017.10.018, 2018.

Schlögl, M., Graser, A., Spiekermann, R., Lampert, J., and Steger, S.: Brief communication: Visualizing uncertainties in landslide susceptibility modelling using bivariate mapping, Natural Hazards and Earth System Sciences, 25, 1425–1437, https://doi.org/10.5194/nhess-25-1425-2025, 2025.

Shapley, L. S.: A Value for n-Person Games, in: Contributions to the Theory of Games (AM-28), vol. 2, pp. 307–318, Princeton University Press, https://doi.org/10.1515/9781400881970-018, 1953.

Steger, S., Brenning, A., Bell, R., and Glade, T.: The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements, Landslides, 14, 1767–1781, https://doi.org/10.1007/s10346-017-0820-0, 2017.

Štrumbelj, E. and Kononenko, I.: Explaining prediction models and individual predictions with feature contributions, Knowledge and Information Systems, 41, 647–665, https://doi.org/10.1007/s10115-013-0679-x, 2014.

Taylor, D. W.: Fundamentals of Soil Mechanics, John Wiley & Son, New York, 1948.

Teke, A. and Kavzoglu, T.: Exploring the decision-making process of ensemble learning algorithms in landslide susceptibility mapping: Insights from local and global explainable AI analyses, Advances in Space Research, 74, 3765–3785, https://doi.org/10.1016/j.asr.2024.06.082, 2024.

Uber, M., Haller, M., Brendel, C., Hillebrand, G., and Hoffmann, T.: Past, present and future rainfall erosivity in Central Europe (v1.0), [Data set], https://doi.org/10.5281/zenodo.7628957, accessed: 07.02.2024, 2023.

Uber, M., Haller, M., Brendel, C., Hillebrand, G., and Hoffmann, T.: Past, present and future rainfall erosivity in central Europe based on convection-permitting climate simulations, Hydrology and Earth System Sciences, 28, 87–102, https://doi.org/10.5194/hess-28-87-2024, 2024.

Van Asch, T., Buma, J., and Van Beek, L.: A view on some hydrological triggering systems in landslides, Geomorphology, 30, 25–32, https://doi.org/https://doi.org/10.1016/S0169-555X(99)00042-2, 1999.

https://doi.org/10.5194/egusphere-2026-1647
Preprint. Discussion started: 8 April 2026

Varnes, D. J.: Slope Movement Types and Processes, in: Special Report 176: Landslides: Analysis and Control, edited by Schuster, R. and Krizek, R., chap. 2, pp. 11–33, Transportation research board, Washington D.C., 1978.

Varnes, D. J. and IAEG: Landslide Hazard Zonation: A Review of Principles and Practice, UNESCO, Paris, 1984.

Weigand, M., Staab, J., Wurm, M., and Taubenböck, H.: Spatial and semantic effects of LUCAS samples on fully automated land use/-land cover classification in high-resolution Sentinel-2 data, International Journal of Applied Earth Observation and Geoinformation, 88, 102 065, https://doi.org/https://doi.org/10.1016/j.jag.2020.102065, 2020.

Woodard, J. B., Mirus, B. B., Wood, N. J., Allstadt, K. E., Leshchinsky, B. A., and Crawford, M. M.: Slope Unit Maker (SUMak): an efficient and parameter-free algorithm for delineating slope units to improve landslide modeling, Natural Hazards and Earth System Sciences, 24, 1–12, https://doi.org/10.5194/nhess-24-1-2024, 2024.

Wu, Q. and Brown, A.: whitebox: 'WhiteboxTools' R Frontend, [Software], https://doi.org/10.32614/CRAN.package.whitebox, 2019.

Zhu, A.-X., Miao, Y., Yang, L., Bai, S., Liu, J., and Hong, H.: Comparison of the presence-only method and presence-absence method in landslide susceptibility mapping, CATENA, 171, 222–233, https://doi.org/https://doi.org/10.1016/j.catena.2018.07.012, 2018.