Landslide Susceptibility Assessment Tools v1.0.0b - Project Manager Suite: A new modular toolkit for landslide susceptibility assessment.

Jewgenij Torizin[1], Nick Schüßler[1], Michael Fuchs[1]

[1]Federal Institute for Geosciences and Natural Resources (BGR), Hanover, 30655, Germany

Correspondence to: Jewgenij Torizin (jewgenij.torizin@bgr.de)

Abstract. This paper introduces the Landslide Susceptibility Assessment Tools – Project Manager Suite (LSAT PM), an open-source, easy-to-use software written in Python. Primarily developed to conduct landslide susceptibility analyses (LSA), it is not limited to this issue and applies to any other research dealing with supervised spatial binary classification. ~~With its standardized project framework,~~ LSAT PM provides efficient interactive data management supported by handy tools~~.~~ in a standardized project framework. The application utilizes open standard data formats ensuring data transferability to all geographic information systems. LSAT PM has a modular structure ~~allowing to extend~~that allows extending the existing toolkit by additional ~~analyses~~tools. The LSAT PM v1.0.0b implements heuristic and data-driven methods ~~such as the~~: analytical hierarchy process, weights of evidence, logistic regression, and artificial neural networks. The software was developed and tested over the years in different projects dealing with landslide susceptibility assessment. The emphasis on model uncertainties and statistical model evaluation makes the software a practical modeling tool~~. Also, it provides the possibility~~ to explore and evaluate different native and foreign LSA models~~, even those not created with LSAT PM~~. The software distribution package includes comprehensive documentation. A dataset for testing purposes of the software is available. LSAT PM is subject to continuous further development.

1 Introduction

~~Gravitational mass movements or, more general, landslides~~Landslides occur in all mountainous parts of the world ~~mainly in association~~, significantly contributing to disastrous socio-economic consequences and claiming thousands of casualties every year (Petley, 2012; Froude and Petley, 2018). These phenomena are frequently associated with other natural hazards such as severe rainfalls, floods, and earthquakes (e.g., Polemio and Petrucci, 2000; Crozier, 2010; Keefer, 2002; Kamp et al., 2008). ~~Landslides contribute to disastrous socio-economic consequences and yearly claim thousands of casualties worldwide (Petley, 2012; Froude and Petley, 2018).~~

Per definition, a comprehensive assessment of landslide hazards addresses the spatial and temporal landslide occurrence based on three questions: "~~where?", "when~~Where?", "When?" and "~~how~~How large?" (e.g., Varnes, 1984, ~~Guzzeti~~Guzzetti et al., 1999, Tanyaş et al., 2018; Reichenbach et al., 2018). The landslide susceptibility analysis (LSA) depicts the probability of spatial landslide occurrence (e.g., Brabb, 1985; Guzzetti et al., 2005) covering the spatial domain of the hazard analysis.

Addressing the temporal domain of landslide hazard assessment is much more challenging due to the local character of the phenomenon and a common lack of multi-temporal landslide inventories covering sufficient periods (e.g., Aleotti and Chowdhury, 1999; Van Westen et al., 2006). Therefore, most case studies at regional scales focus on LSA as the most feasible part of the landslide hazard analysis.

35 Regional LSA is usually done based on qualitative (heuristic or knowledge-driven) and quantitative methods (e.g., Reichenbach et al. 2018). The quantitative techniques comprise physically-based and data-driven statistical~~,~~ and machine learning (ML) approaches. ~~Based on the~~ (e.g., Aleotti and Chowdhury, 1999). The desired analysis scale and ~~available~~ data~~,~~ ~~the application~~ availability usually govern the decision ~~of some methods might be limited~~which method to use (e.g., Van Westen et al., 2008; Balzer et al., 2020).

40 ~~With the advances in remote sensing~~In the past decades, advances in remote sensing made significant progress allowing efficient data acquisition at regional scales. Additionally, digitalization forced a general boost of data mining techniques. Together with the development of GIS software packages and open-source statistical and machine learning libraries, the data-driven methods for LSA have gained popularity. These methods belong mainly to supervised binary classification (e.g., Torizin, 2016). In the supervised classification, we use a set of recorded observations (labels) and independent explanatory

45 factors (features) such as different geomorphologic, hydrologic, and geological conditions to train a statistic function~~, the so-called~~ (classifier~~.~~). The ~~classifier's task is, in a narrower sense, to distinguish based on available features whether~~classifier estimates the likelihood of a specific countable element in a study area (e.g., raster pixel ~~or unique condition unit) is a~~) to be in the specified target class (e.g., landslide or ~~non~~no landslide~~.~~).

Because classification is one of the fundamental tasks in statistics and ML, many different classifiers exist. Consequently, in

50 numerous case studies, the academic community continuously applies and compares classification algorithms and their variations, which were initially developed for other purposes but are sufficiently general to be used for LSA. While some classifiers might outperform others, the drawn conclusions are often valid only for particular settings and study designs (e.g., Balzer et al~~.~~., 2020). Under other circumstances, such as different data quality or distribution, it is very likely that some of the other classifiers perform on par or better. Reviewing the LSA research of the past 30 years, Reichenbach et al. (2018) counted

55 about 163 different data-driven methods, emphasizing the problem of excessive experimentation with statistical classifiers rather than focusing on the LSA's reliability. Many of these methods have never been adopted or seriously considered by practitioners skeptically following the academic research at their own pace and utilizing a comparably small part of it. Thus, despite the many academic publications dealing with regional data-driven LSA, ~~there are~~only a few practical solutions have been adopted in national landslide risk assessment strategies yet (e.g., Balzer et al., 2020).~~.~~ Also, ~~in the~~ user-friendly standalone

60 software ~~development~~developed in this field~~,~~ is rare compared to the available geotechnical software applications ~~in geo-engineering, user-friendly standalone applications dealing with regional LSA and especially with its data-driven methods are rare. Most.~~ Many available tools exist as academic code~~, implying~~ generated to support specific case studies (e.g., Merghadi, 2018; Merghadi, 2019; GitHub, 2021a; GitHub, 2021b) and imply that the user has the necessary programming or scripting skills to set up and run the tools. ~~Thus, the user must program pipes or deal with several apps or packages at different assessment~~

65   ~~steps.~~ Despite considerable efforts to adopt the earth science curricula to digital transformation (e.g., Hall-Wallace, 1999; Makkawi et al., 2006, Senger et al., 2021), the required computational literacy to deal with ~~current data-driven LSA~~those applications is ~~still~~ not a broad standard in geosciences. However, there is a positive trend. Rising education possibilities on e-learning platforms with exhaustive offers in programming narrow the gap between geosciences and data sciences. ~~So~~Bouziat et al. (2020) noted that the geoscience community increasingly uses Python (Van Rossum & Drake, 2009) for data processing,

70   R statistical package (R Core Team, 2013) for statistical analysis, and custom web services for sharing results.

Since 2010, ~~most~~many LSA tools have been available as plugins or extensions ~~for ESRI ArcGIS, QGIS, and GRASS~~in different GIS (~~such as,~~ e.g., LSAT (Polat, 2021~~; Torizin, 2012;~~), LSM Tool Pack (Sahin et al., 2020), or ArcMAP Tool (Jebur et al., 2015~~;~~) in ESRI ArcGIS, SZ-plugin (Titti & Sarretta, 2020) in QGIS (QGIS Development Team, 2009), r.landslide (Bragagnolo et al., 2020a~~),~~) in GRASS GIS (GRASS Development Team, 2021), and RSAGA (Brenning, 2008) in SAGA, as

75   well as scripts in R statistical package, e.g., LAND-SE (Rossi and Reichenbach, 2016), and very few standalone applications, e.g., GeoFIS (Osna et al., 2014).

With the Landslide Susceptibility Assessment Tools – Project Manager Suite (LSAT PM), we introduce an open-source (~~under~~ GNU General Public License v3), standalone~~,~~ and easy-to-use tool that supports scientific principles of openness, knowledge integrity, and replicability. Doing so, we want to share our experience in ~~the implementation of~~implementing heuristic and

80   data-driven LSA methods. Our primary goal is not to introduce as many algorithms as possible for LSA but to provide easy access to a selection of state-of-the-art methods representing groups of different approaches. Providing a convenient framework for model building, evaluation, and uncertainty assessment, we want to highlight ~~such methods'~~the capabilities and limitations~~, making them more transparent~~ of those methods.

2 LSAT PM Software

85   LSAT PM is a desktop application designed to support decision-makers and the scientific community in generating and evaluating landslide susceptibility models based on heuristic and data-driven approaches. ~~LSAT PM's core is the weights of evidence (WoE) method (e.g., Bonham Carter et al., 1989) widely applied in LSA in the past decades (e.g., Mathew et al., 2007; Moghaddam et al., 2007; Thiery et al., 2007; Neuhäuser and Terhorst, 2012; Teerarungsigul et al., 2015). It is simple to understand and provides a transparent computation algorithm. Enhanced by assessing uncertainties (e.g., Torizin et al., 2018;~~

90   ~~Torizin et al., 2021), WoE becomes a robust tool for rapid analysis providing a good reference model to test against when exploring new methods. For example, it can be beneficial to investigate the data's dependencies or run several sensitivity analyses based on transparent WoE before applying more sophisticated multivariate statistical analysis techniques, e.g., logistic regression (LR) or ML algorithms such as artificial neural networks (ANN).~~

~~Besides WoE, LSAT PM offers two other popular data-driven LSA methods:~~ ~~LR (e.g., Lee, 2005; Budimir et al., 2015;~~

95   ~~Lombardo & Mai, 2018;)~~ ~~and ANN (e.g., Lee and Evangelista, 2006; Pradhan and Lee, 2010; Bragagnolo et al., 2020b),~~ ~~both based on the sklearn library (Pedregosa et al., 2011; Buitinck et al., 2013). For heuristic analyses, LSAT PM embeds a module~~

for the analytical hierarchy process (AHP). ~~This decision support method finds application primarily for areas with insufficient observational data (e.g., Balzer et al., 2020; Panchal & Shrivastava, 2020).~~

2.1 Development history

100 The development of Landslide Susceptibility Assessment Tools (LSAT) started in 2011 with Python scripting within ESRI ArcGIS 10.0 Toolbox to support technical cooperation (TC) projects (Torizin, 2012). TC projects are usually not ~~aimed at~~ cutting-edge research but summarise, adapt, and implement scientific outcomes by following the best-practice approach. Since then, LSAT was continuously improved and tested at different development stages in case studies in Indonesia (Torizin et al., 2013), Thailand (Teerarungsigul et al., 2015), Pakistan (Torizin et al., 2017), China (Torizin et al., 2018) and Germany (Balzer

105 et al., 2020). Working with different data of varying quality helped us ~~to~~ develop efficient methodical workflows. It also enabled us to better understand the limitations of some methods and design practical approaches to assess model uncertainties (e.g., Torizin et al., 2021).

~~In the beginning, like many other tools, LSAT relied on a host GIS (ESRI ArcGIS) (Torizin, 2012).~~ In 2017, we started to prototype LSAT as a standalone application ~~with~~bearing the extension "Project Manager Suite" ~~extension~~ in Python 2 and

110 later in Python 3. This development began within the sino-german scientific cooperation project (Tian et al., 2017) and continued in a cooperation project with german geological surveys (Balzer et al., 2020).

2.2 Software architecture and capabilities

LSAT PM v1.0.~~0a0b~~ is written in Python 3 ~~(Van Rossum & Drake, 2009).~~ Graphical framework PyQt5 provides the basis for the ~~Graphical User Interface~~graphical user interface (GUI). Functionalities of the software rely on different third-party libraries

115 and Python packages. Geospatial Data Abstraction software Library (GDAL) (GDAL/OGR Contributors, 2021) and its Python bindings provide the core functionality for dealing with spatial data. The highly efficient *NumPy* (Harris et al., 2020) provides array computations through the analyses. Implemented ML algorithms rely on the powerful *sklearn* library (Pedregosa et al., 2011; Buitinck et al., ~~are~~2013). *Matplotlib* (Hunter, 2007) provides the basis for generating analysis plots. *Openpyxl* (Gazoni and Clark, 2018) and *python-docx* (Canny, 2018) packages allow the export of ~~the core functions~~analysis results as convenient

120 MS Office files and automatized analysis reports generation.

The software consists of the main GUI with integrated independent modules (widgets), building the software's functionality. Due to the LSAT PM development history, we built the most functionalities around the weights of evidence (WoE) method, which was the initial analysis module of LSAT PM. WoE (e.g., Bonham-Carter et al., 1989) belongs to the bivariate statistical methods frequently applied in LSA in the past decades (e.g., Mathew et al., 2007; Moghaddam et al., 2007; Thiery et al., 2007;

125 Neuhäuser et al., 2012; Teerarungsigul et al., 2015). It is simple to understand and provides a transparent computation algorithm. With enhanced uncertainty assessment (e.g., Torizin et al., 2018; Torizin et al., 2021), WoE becomes a robust tool for rapid analysis providing a good reference model to test against when exploring new methods. For example, it can be beneficial to investigate the data dependencies or run several sensitivity analyses based on transparent WoE before applying

more sophisticated multivariate statistical analysis techniques, e.g., logistic regression (LR) or *black-box* ML algorithms such as artificial neural networks (ANN). LSAT PM includes both LR (e.g., Lee, 2005; Budimir et al., 2015; Lombardo and Mai, 2018) and ANN (e.g., Lee and Evangelista, 2006; Pradhan and Lee, 2010; Bragagnolo et al., 2020b), and also a module for heuristic analyses based on the analytical hierarchy process (AHP). This decision support method finds application primarily for areas with insufficient observational data (e.g., Balzer et al., 2020; Panchal and Shrivastava, 2020).

Currently, LSAT PM can utilize Tagged Image File Format (GeoTiff) raster data for model parameters and vector data formats, such as Tagged Image File Format (GeoTiff), ESRI shapefiles, Keyhole Markup Language (KML), and JavaScript Object Notation (GeoJSON).) for inventories. Further GDAL-supported formats are incorporable on demand.

Complementary to the spatial data output in the same formats as input, LSAT PM supports table exportexporting tables to Microsoft Excel-Sheets, graphs to portable network graphic files (PNG), and automated analysis reports to MicrosoftMS Word Documents documents.

For spatial analysis, LSAT PM implements basic geoprocessing functionalities for data preparation. Morphological analyses, such as slope, aspect, topographic position index (TPI), and many others, can be performed based on raster datasets in GeoTiff format. Functions such as euclidean distance and raster classification are also available. A simple data viewer visualizes raster data. Table 1 provides an overview of implemented tools.

However, although LSAT PM provides some GIS capabilities for geoprocessing, it cannot be characterized as a solid GIS application and was never supposed to become one. It has a slim structure tailored to manage and prepare the data for the targeted binary spatial classification.

Table : Overview of LSAT PM's tools.

Category	Tool domain	Tools (with GUI support)
PROJECT	Project	• Create project
		• Open project
		• Project info
	Settings	• Language
	Context Menu	• Attribute table (raster, vector)
		• Data properties
		• View data
DATA	Import	• Import raster
		• Import inventory
	Vector Tools	• Random sampling
		• Subset by attribute
		• Geoprocessing tools
	DEM Tools	• Slope
		• Aspect
		• Hillshade
		• Topographic positional index (TPI)
		• Roughness

		• Topographic ruggedness index (TRI)
		• Euclidean distance
		• Contingency
	Raster Tools	• Combine
		• Reclassify
		• Lookup
		• Sensitivity reclassification
		• Reproject
	Visualization	• Viewer
		• Weights of Evidence (WoE)
	Analysis	• Logistic Regression (ANN)
		• Artificial Neural Networks (ANN)
ANALYSIS		• Analytic Hierarchy Process (AHP)
		• Model Builder
	Model Management	• Model Info
		• Zoning

3 Test data

150   A dataset to test the functionalities of the software is available separately from https://doi.org/10.5281/zenodo.5109620. The example or test dataset is an excerpt of the data collected in the German-Pakistani technical cooperation project Georisk Assessment Northern Pakistan (GANP), carried out by the Federal Institute for Geosciences and Natural Resources and Geological Survey of Pakistan (e.g., Torizin et al., 2017). The dataset covers about 664 square kilometers, including parts of the Kaghan and Siran Valleys in Khyber Pakhtunkhwa, Northern Pakistan. Table 2 and Fig.1 provide an overview of the test

155   data.

A severe Kashmir Earthquake struck the test data area on October 8th, 2005, with a moment magnitude of 7.8, triggering thousands of landslides. The collected landslide inventory results from the visual interpretation of optical satellite images available through Google Earth (GE), which, during the data acquisition, consisted mainly of imagery from Quickbird (up to 0.60 m ground resolution), IKONOS (~4 m ground resolution), SPOT (SPOT5 about 5 m ground resolution), and Landsat (15

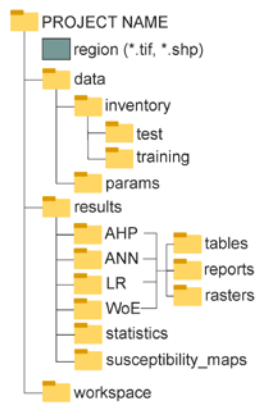
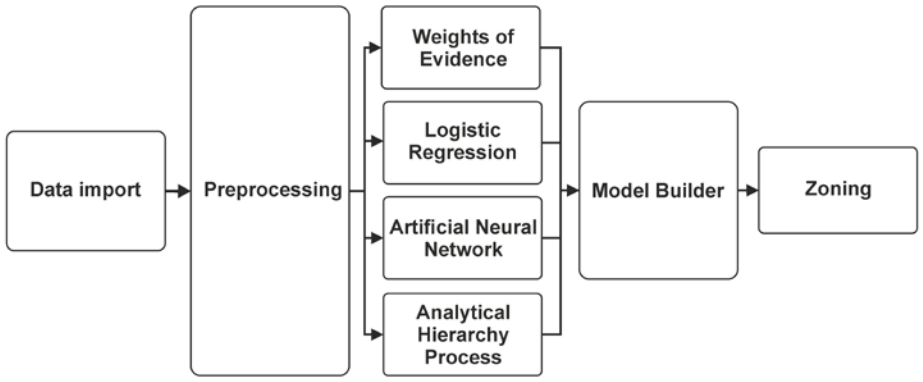
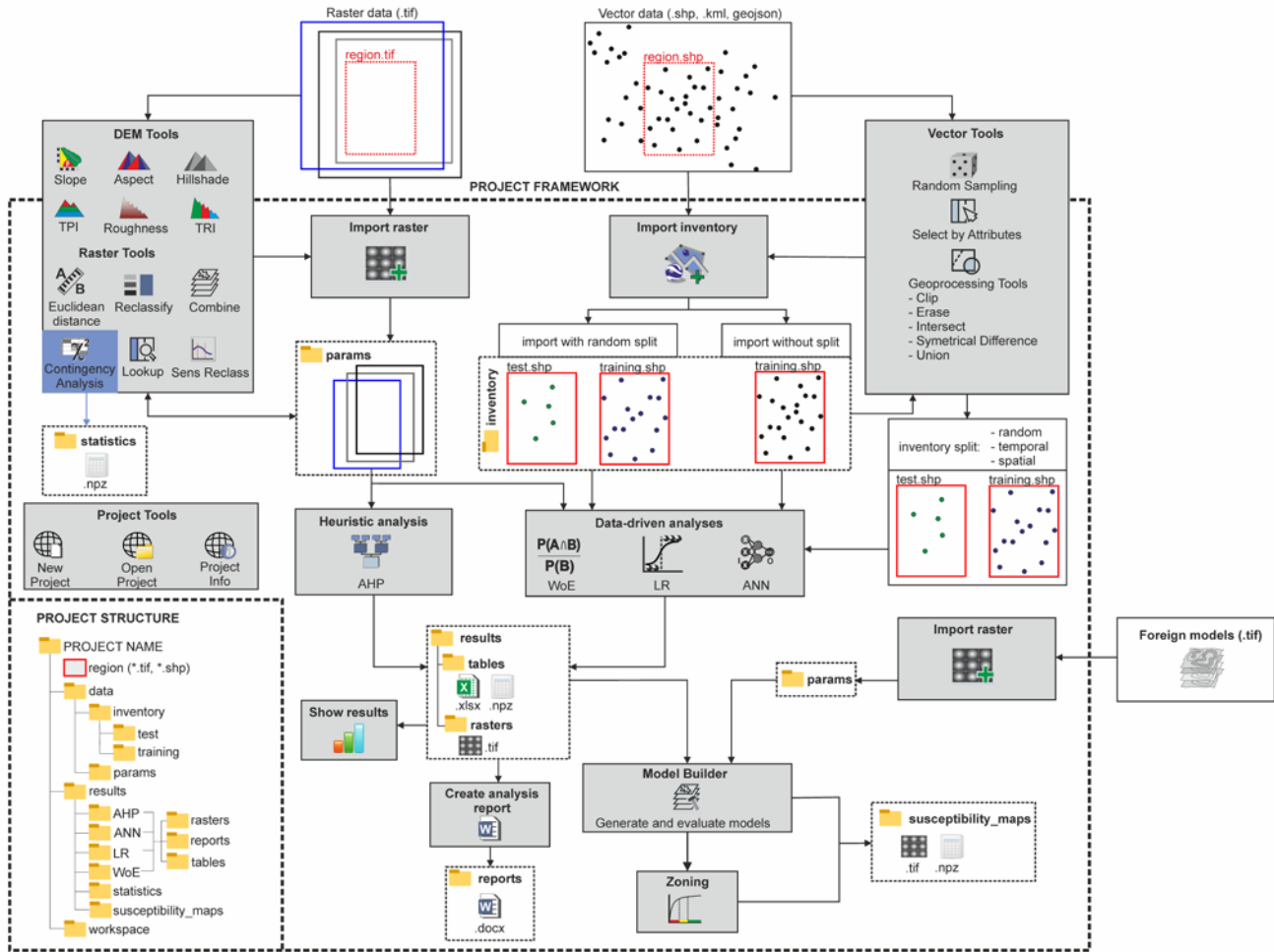
160   m ground resolution) (Torizin et al., 2017). In total, the landslide inventory includes 3819 events for the test area depicted as polygons. Landslide sizes range from about 12 square meters to about 88444 square meters representing the depletion area of the landslides (as far as it was possible to determine by visual interpretation of imagery).

The digital elevation model is the ALOS Global Digital Surface Model (AW3D30) (JAXA, 2017) with a ground resolution of approximately 30 m (Fig. 1a). The geological information and the tectonic features (faults) were derived from the geological

165   map of Calkins et al. (1975) (Fig. 1b). The landcover results from the supervised classification on Landsat imagery performed by Fuchs et al. (2015) (Fig. 1c).

4 Working with LSAT PM

LSAT PM provides handy tools to set up the model through data exploration, preprocessing, analysis, model evaluation, and postprocessing.

175   The logical workflow in Fig. 2̶1 schematically sketches the working process with LSAT PM. The following sections briefly introduce the single steps of this workflow and their corresponding modules. For moreMore technical details, we refer to are obtainable from the software documentation distributed standalone or as part of the installer package (see also Sect. 8̶7).

Line 180: **Figure 1: Standardized project structure and the logical workflow of LSAT PM.**

4~~3~~.1 Project

After launching the software, the user can create a new or open an existing project. ~~The LSAT PM project has a standardized predefined structure, as shown in Fig. 2. The project structure includes~~ A project is a structured folder system that stores data with a specified spatial extent and spatial reference ~~that needs~~(region). The spatial extent and spatial reference need to be assigned on project creation manually or by selecting an already existing raster dataset as a project reference~~.~~ file (recommended).

~~4~~The LSAT PM project has a standardized predefined structure, as shown in the bottom left of Fig. 1. The local project files overview (*Catalog* in Fig. 2) in the main GUI helps manage the project data providing a data-type dependent context menu.

3.2 Data import

190 Data import distinguishes between *Import Raster* and *Import Inventory* (Fig. ~~3~~2). The first tool imports raster datasets considered to represent independent exploratory factors ~~or~~ (features~~,~~), the second imports vector-based datasets for observational data~~, the so-called~~ representing inventory or labels.

The raster import tool ensures ~~the data's~~ consistency by validating imported datasets against the project reference ~~datasets~~dataset (region). If not consistent with the project reference, the data gets warped ~~accordingly~~. This procedure is

195 comparable to the concept of regions found in GRASS GIS ~~(GRASS Development Team, 2021)~~ and helps avoid processing errors due to specific resolution and spatial reference inconsistencies. ~~During the~~The import~~, LSAT PM~~ procedure generates data copies; thus, original files remain unchanged.

Inventory import ~~supports~~generally does the same for vector datasets as input. Additionally, it includes the random splitting of the dataset into training and test datasets ~~(optional)~~. Using this option, the user can specify the percentage ratio of training

200 and test datasets. The ~~data split~~splitting option is ~~an essential step in modern data-driven approaches (e.g., Xu and Goodacre, 2018), which is employed to evaluate the data-driven model by a test dataset~~ not ~~involved in the model's training.~~ mandatory and skippable because inventory subsetting is possible later using one of the tools described in Sect. 3.3 (see also Fig. 1).

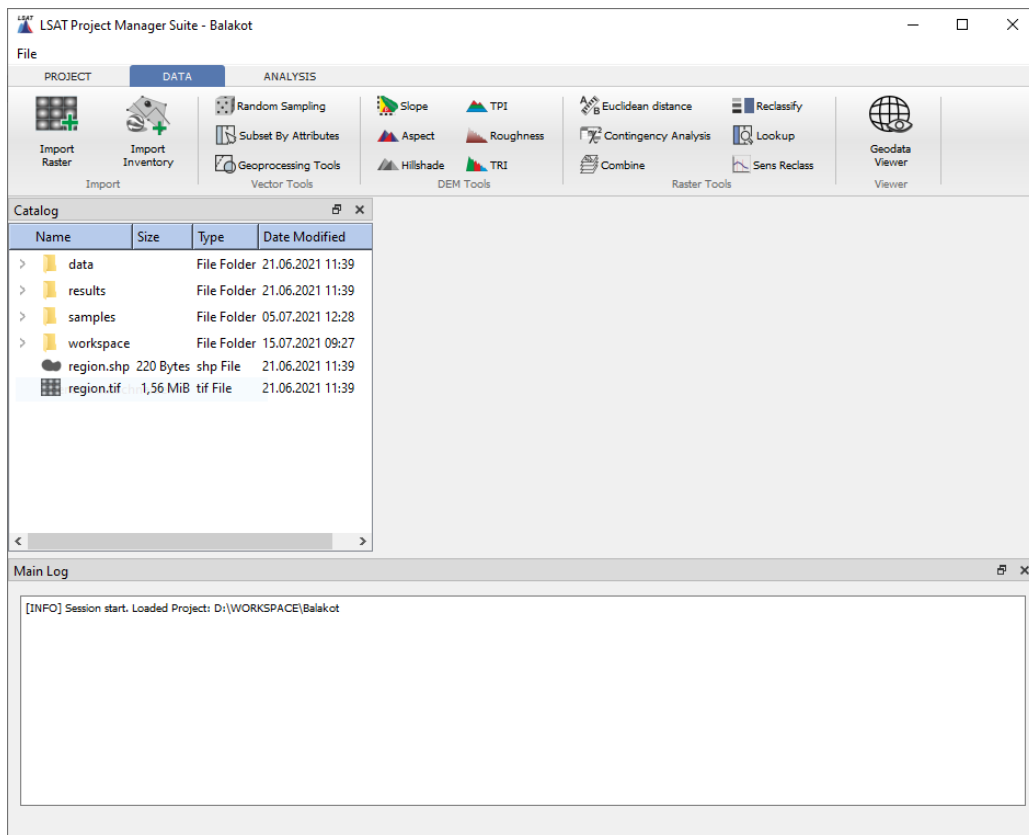
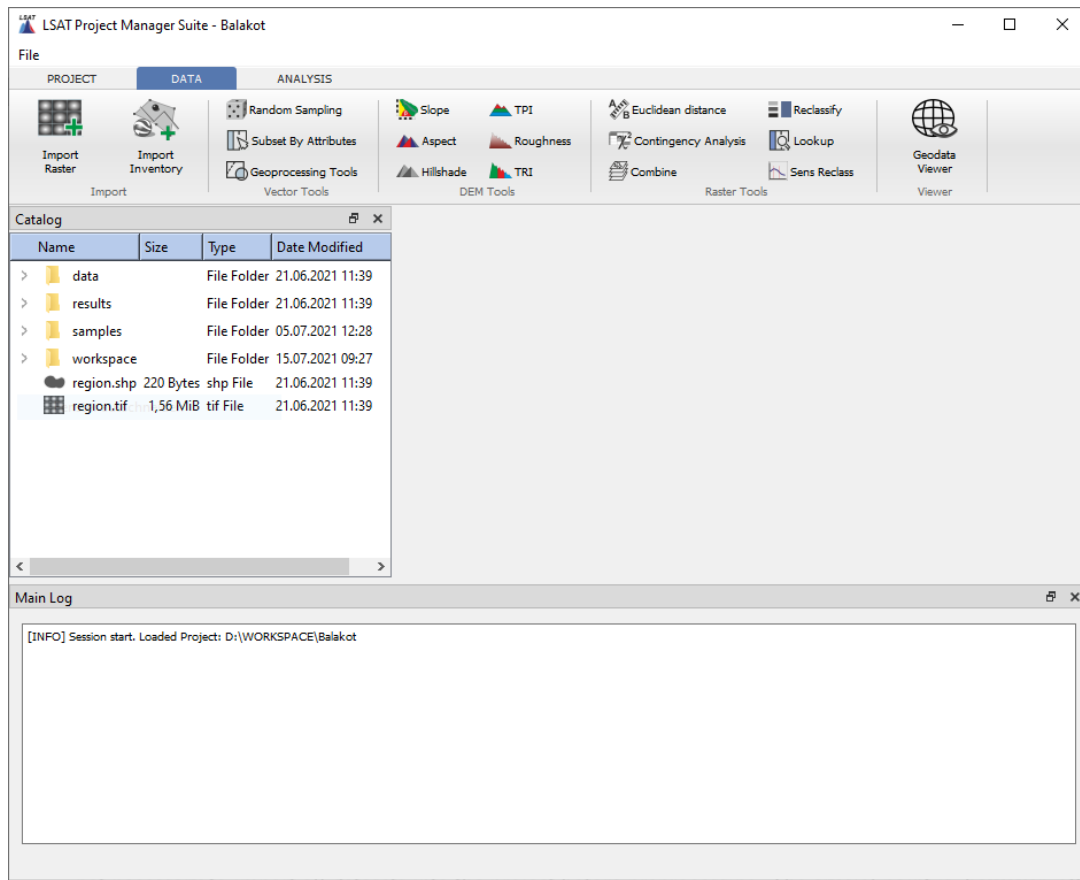



Figure 2: Main GUI with activated *Data* tab. The *Data* tab ~~exhibits~~contains tools for data import, vector data processing, DEM tools, tools for raster processing, and a simple data viewer.

4~~3~~.3 Data exploration and preprocessing

In the preprocessing step, we derive parameters and prepare the data according to the requirements of the upcoming analysis. ~~So the~~ *Vector,* *DEM* ~~in the~~, and *Raster Tools* of LSAT PM (Fig. 2) aid this purpose.

Data subsetting is an essential technique to evaluate data-driven models using a test dataset ~~serves~~not involved in the model's training, also known as cross-validation (e.g., Xu and Goodacre, 2018; Petschko et al., 2014; Chung and Fabbri, 2008). LSAT PM provides a palette of *Vector Tools* (Fig. 2) that supports the generation of inventory subsets based on random subsampling or feature attributes, e.g., the date (temporal split). Using *Geoprocessing Tools* in the same tool domain, the user can also subset the inventory based on spatial features (spatial split).

Digital elevation models (DEM) serve as a basis for morphometric parameters such as~~,~~ e.g., slope, ~~topographic position index, or~~ aspect~~. Using the combine~~, or TPI. *DEM Tools* (Fig. 2) derive basic morphometric features and generate raster data outputs from DEM raster datasets.

Raster Tools help perform basic raster operations and spatial analyses. Using the *Combine* tool, the user can combine several discrete ~~raster~~ datasets to generate a new ~~raster~~ dataset exhibiting unique conditions of higher complexity. ~~Alternatively, the user can convert~~ Linear and point vector data (e.g., tectonic features, roads, streams, or point locations) are usually unsuitable as direct input for spatial analysis ~~such as tectonic features or roads into meaningful parameters depicting~~. However, their possible spatial ~~effect, e.g., over the~~ effects are considerable via distance~~, by generating a euclidean~~ maps. The *Euclidean distance* ~~raster. In this step, the user also can generate datasets more suitable for the analysis type. For example, the implemented~~ tool generates a distance raster dataset based on input vector data.

Raster reclassification is a standard procedure in GIS analyses applied as value replacement in discrete datasets or binning of continuous datasets. Because WoE ~~analysis~~ utilizes discrete data only~~. Thus~~, continuous raster data such as slope or distance rasters need ~~to be classified. In LSAT PM, the user can perform the raster classification task by the tools~~ a binning in the data preparation process. The *Reclassify* ~~or the so-called~~ tool offers different classifiers such as equal intervals, quantiles, defined intervals, and user-defined values. Additionally, the *Sensitivity* ~~reclassification, which serves as~~ *Reclassification (Sens Reclass)* tool provides a sensitivity analysis to find optimal ~~class boundaries~~ cutoff thresholds (e.g., Torizin et al., 2017).

~~Also, preprocessing of inventory data is available. Depending on the study design, it might be necessary to subset the data. LSAT PM offers tools for random sampling and subsetting by attributes allowing subdividing datasets by spatial or temporal criteria. Other vector geoprocessing tools allow operations such as, e.g., erase, intersect, or union datasets.~~

~~These preprocessing tools apply to datasets in any location. Thus, it is not obligatory to import all datasets into the project before preprocessing. However, the crucial point is to use the LSAT PM's built in import tools to import preprocessed data afterward.~~

~~Contingency analysis~~ The *Contingency Analysis* tool performs the Chi-square-based contingency analysis on raster-based categorical data, estimating the associations between the datasets based on Pearsons C, Cramer's V, and Phi. It is the only tool ~~in the tool domain *Raster data*, which~~ that produces ~~an~~ output ~~file that stores the calculated contingency tables in the~~ in the subfolder *statistics* of the projects' results folder. ~~This tool represents an analysis helping to explore the dependencies (associations) in the discrete datasets rather than a preparation step.~~ The user can view the output contingency table via the *Show results* option from the *Catalog's* context menu.

~~4~~All the above-presented tools apply to datasets in any location. Thus, the user can perform preprocessing steps before the data import.

3.4 Analyses

As already introduced in Sect. 2, LSAT PM implements heuristic and data-driven methods for LSA representing different categories (e.g., bivariate and multivariate). All of the methods have different levels of complexity, which need to be accounted for when choosing a specific analysis method. Table 3 briefly summarizes the corner points of the approaches such as category, supported data types, and complexity. The introduced complexity is a subjective measure that we assigned based on our experience evaluating criteria such as needed mathematical background, the complexity of data preparation, model structure

and computation algorithm (transparency), and interpretability of the results. ~~The~~ Due to the high degree of automation, the more complex methods allow running the analysis with its default settings without any adjustment and, therefore, might appear ~~easier to apply depending on the automation degree because the user has to do only a few steps.~~more straightforward. However, this ~~deception~~first impression will vanish once the user encounters the advanced settings of the methods.

255 **Table 1 Analyses included in LSAT PM and their specifications**

Analysis name	Category	Supported data types	Complexity
Weights of Evidence	data-driven, bivariate	discrete	moderate
Logistic regression	data-driven, multivariate	continuous, discrete	high
Artificial Neural Network	data-driven, multivariate	continuous, discrete	very high
Analytic Hierarchy Process	heuristic	discrete	low

~~4~~3.4.1 Weights of Evidence

WoE is a bivariate statistical approach estimating the association between the observational data (dependent variable represented by the training landslide inventory) and a potential controlling ~~factors~~factor (independent ~~variables~~variable
260 represented by, e.g., geological conditions~~), analyzing them separately.~~). The analysis output is a raster of the specific controlling factor containing logarithmic log-likelihood weights, which characterize the relationship of discrete factor classes with a landslide occurrence. ~~Afterward, individually~~Individually weighted factors are ~~overlain~~then overlaid into a linear model to obtain the overall landslide susceptibility pattern (Torizin et al., 2017).

WoE in LSAT PM offers three different analysis modes~~.~~: simple cross-validation, on-the-fly subsampling, and sampling with
265 predefined samples. The default option is simple cross-validation~~.~~ (presupposes inventory split into training and test datasets). With this option, the ~~analysis~~model weights estimation runs once with the entire training dataset~~. The two other options are the so-called on-the-fly subsampling and~~ (no further subsampling ~~with predefined samples.~~). For on-the-fly subsampling, the weight estimation runs for several user-defined iterations taking random samples (without replacement) of user-defined size~~.~~ from the training inventory. The estimated weights are mean values from all iterations. The analysis with predefined samples
270 utilizes predefined sample datasets in a specified folder location. ~~The~~These predefined samples must be created beforehand by any subsetting algorithm introduced above~~. Likewise to the random subsampling, the~~ (Sect 3.3). The computed weights are mean values from all iterations, identical to the random subsampling.

After the training, the result table and the weighted raster are automatically exported into the corresponding result folders. The results can be visualized immediately after the analysis through the WoE widget or later by calling the result viewer from the
275 *Catalog's* context menu.

The model generation process is performed in the next step using the LSAT PM model builder module (see Sect. ~~4~~3.5).

4~~3~~.4.2 Logistic regression

Logistic regression (LR) is a multivariate statistical classification method to estimate relationships between the dependent variable and independent controlling factors. In contrast to WoE, LR analyzes the ~~relationships~~associations for all controlling factors at once and can utilize both continuous and discrete data ~~for~~as independent variables.

In LSAT PM, LR runs more automated than WoE. ~~For the parameters, the~~The user has to determine if the parameter is a discrete or continuous variable~~. The~~ in the beginning. After that, the data preparation process runs automatically. The continuous datasets are scaled using a ~~so called~~ min-max-scaler to the value range between 0 and 1; discrete datasets are transformed into binary dummy variables. All setting options for the logistic regression, e.g., regularisation or solver algorithm, implemented in the *sklearn* library, are adjustable upon the user's needs in the advanced settings GUI. After the training, the result table and the prediction raster are automatically exported into the corresponding result folders.

Unlike WoE, ~~where an additional modeling step is required to overlay the weighted parameter raster to a model,~~ the analysis output from multivariate LR already provides ~~the~~a multi-parameter landslide susceptibility model.

4~~3~~.4.3 Artificial Neural Network

ANNs are computer models inspired by biological neural networks. They consist of ~~a network of~~ artificial neurons ordered in a network structure to simulate information processing, ~~information~~ storage, and learning. The structure of an ANN usually consists of an input layer, one or more hidden layers, and an output layer. The number of hidden layers determines the depth of an ANN (e.g. Schmidhuber, 2015, Hernández-Blanco et al., 2019). This structure is also known as the Multi-Layer Perceptron (MLP). The layers are composed of neurons in which the information processing takes place. Theoretically, the number of hidden layers and their neurons is unlimited. Thus, the network design is strongly dependent on the complexity and non-linearity of the task and the available processing capacity. Most ANNs utilized in LSA are ~~the so called~~ feedforward networks, which have a multi-layer perceptron (MLP) structure with usually one hidden layer (e.g., Ermini et al., 2005; Lee and Evangelista, 2006; Alimohammadlou et al., 2014).

The implemented module for artificial neural network (ANN) has an experimental status and runs comparable to the LR. ~~The routine allows the user to flag the variables as~~ANN can utilize discrete ~~or~~and continuous data. Additionally, the user can specify the ANN properties. Among others, these settings comprise the number of neurons in the layers, the number of hidden layers, the activation function, and the solver to use. The design of the network and the activation function selection is one of the most sensitive steps of the analysis with ANN. Despite reviewing numerous studies, we do not have a straightforward, practical recipe for designing the network yet. Therefore, the default settings do not represent the best practical approach but rather the defaults delivered with the *sklearn* library.

Nevertheless, the GUI and included data preprocessing provides easy and fast access to the capabilities given by the *sklearn* library and allow exploring the ANN performance for LSA. ~~For more~~More technical information~~, we refer to the LSAT PM~~ is obtainable from documentation of LSAT PM and ~~corresponding references for the~~ *sklearn* library.

The data scaling process runs identically to the LR module. After the training process, the output raster defining the spatial probability of landslide occurrence is exported to the corresponding result folder.

Comparable to LR, the resulting model raster can be imported to the model builder tool, compared with other models, or evaluated with other inventories.

43.4.4 Analytical Hierarchy Process

As a heuristic approach, the AHP is different from the data-driven applications explained above. The user must specify the weights, usually based on his general expertise (knowledge of geological processes) and specific knowledge about the investigation area. The weighting process is a pairwise comparison of the inputs at different hierarchical levels. For LSA, the AHP typically takes two hierarchical levels. The first level controls the class priorities inside multi classmulticlass parameters (e.g., raster values), the second sets the priorities between the multi classmulticlass parameters (e.g., raster datasets).

The pairwise comparison is a complex task when working with parameters exhibiting a large number ofmany classes. The human's ability to compare is limited to approximately seven objects, plus or minus two (Miller, 1956). Saaty (1977) considered this proposing rank values between 1 and 9, specifying the factor's importance. Therefore, in the preparation process, it is advisable to reduce the number of classes by generalization or subdivision in different hierarchical groups (e.g., Balzer et al., 2020). The latter will make the hierarchy more complex.

The advantage of pairwise ranking compared to the simple ranking is the ability to verify the logical consistency of the decision mathematically. Therefore,AHP uses the AHP introduces a consistency ratio (CR that indicates) to indicate whether the introduced ranking is a logical inference or a random guess. Saaty (1980) recommends that thea CR value should not exceedunder 0.1 for consistent assessment.

Notable is that some studies applying AHP for LSA use a hybrid approach combining bivariate methods with AHP (e.g., Kamp et al. 2008, 2010). In the hybrid approach, a data-driven bivariate approach applies to the first hierarchical level is assessed using a data driven bivariate approach, e.g., WoE. Afterward, additional expert-based weights derived from the AHP priority vector are applied to overlay the parameters to the model. Such an approach can preserve the crude generalization of the patterns in the first hierarchical level making the analysis applicable withto more detailed datasets. For the AHP part, the method becomes more applicable by involving the expert weights at a higher hierarchical level, which benefits more from general process understanding than detailed local knowledge.

Conversely, it also has implications for the bivariate analysis part. Using conditionally dependent parameters becomes less critical since experts adjust the parameter's contribution in the upper hierarchical level. However, the hybrid approach is only possible if a sufficient number of observations for the first data-driven step is available.

Currently, theThe implemented AHP is a pure expert-based tool supporting two hierarchy levels. The user has to perform the pairwise ranking for single parameter classes in the first step and the parameters in the second step.

3.5 Model Builder

The ~~output consists of the landslide susceptibility model. Like LR and ANN, the results~~ Model Builder (MB) is, in simple terms, a raster ~~can be evaluated in the model builder if a landslide inventory is available. In the absence of observational data in urbanized regions, local geotechnical models or existing structural measures such as slope reinforcement may provide appropriate~~calculator with an integrated evaluation ~~points (e.g., Balzer et al., 2020).~~

~~4.5 Model builder~~

~~The model builder is the core~~ module ~~for generating models. It can combine WoE outputs to a model in an additive procedure (e.g., Torizin et al., 2017) and evaluate all model types generated within LSAT PM and foreign models in supported data formats.~~

. The algorithm behind the model evaluation is ~~basing on the computation of~~ the receiver operating characteristics (ROC) curve. ~~The ROC curve is~~ - a technique to visualize and evaluate the classifier's performance (e.g., Fawcett 2006) by depicting the ratio of the True Positive Rate (Sensitivity) and the False Positive Rate (1-Specificity). The area under the ROC curve (AUC or AUROC) provides a quantitative measure to compare the goodness of different models. ~~We extended~~

At the start, MB automatically searches and imports analysis results and existing models to the corresponding weighted layer and model collections, as shown in Fig. 3 (b and e). The user can specify which layers to include in the weighted overlay model shifting the layers into the model layer collection (Fig. 3c). The model generating expression is adjustable on demand. Default model generating expression is a simple additive overlay of selected weighted layers.

MB performs the overlay according to the data generating expression and uses the ROC curve to evaluate the model based on the specified landslide inventory dataset (Fig. 3a). Moreover, the evaluation procedure ~~with the ROC curve by~~can be performed based on iterative random subsampling ~~mode to assess~~or predefined sample groups. The user can select these options in the advanced settings of MB.

Using the training inventory, the user evaluates the model fit. For the test dataset, the user evaluates the predictivity of the model on new data. Using iterative subsampling, the user gets an idea about the variance of the model based on different samples that allows him to evaluate sampling errors~~, as introduced by~~ (e.g., Torizin et al~~. (.,~~ 2018~~,; Torizin et al.,~~ 2021). The latter might be a helpful feature for interpreting evaluation results if the test datasets are comparably small. Exploring the variance of the training dataset in conjunction with the test ROC curve helps to understand whether the uncertainty of the model is the property of ~~the~~insufficient data (sampling error) or ~~the~~model accuracy~~.~~ (bias).

Moreover, the user can mix the model input layers (Fig. 3c) and adjust the generating expression (Fig. 3d). This flexibility allows the generation of hybrid models and model ensembles. Hybrid models combine different classifiers in one model, e.g., WoE for discrete data and LR for continuous data; model ensembles consist of different homogeneous models, e.g., LR and ANN. Here it is essential to note that weighted layers or models generated by different classifiers may exhibit different value

ranges and need transformations when used together in a hybrid model or model ensemble. Through the model building expression, the user can implement those transformations instantly.

MB's model collection (Fig. 3e) lists all generated models and provides essential management to export models to GeoTiff format and visualize the corresponding ROC curves (Fig. 3f). All generated models include metadata with information on
375   model input layers and applied model generating expression, making the results more reproducible. *Model Info* function provides access to the model's metadata.

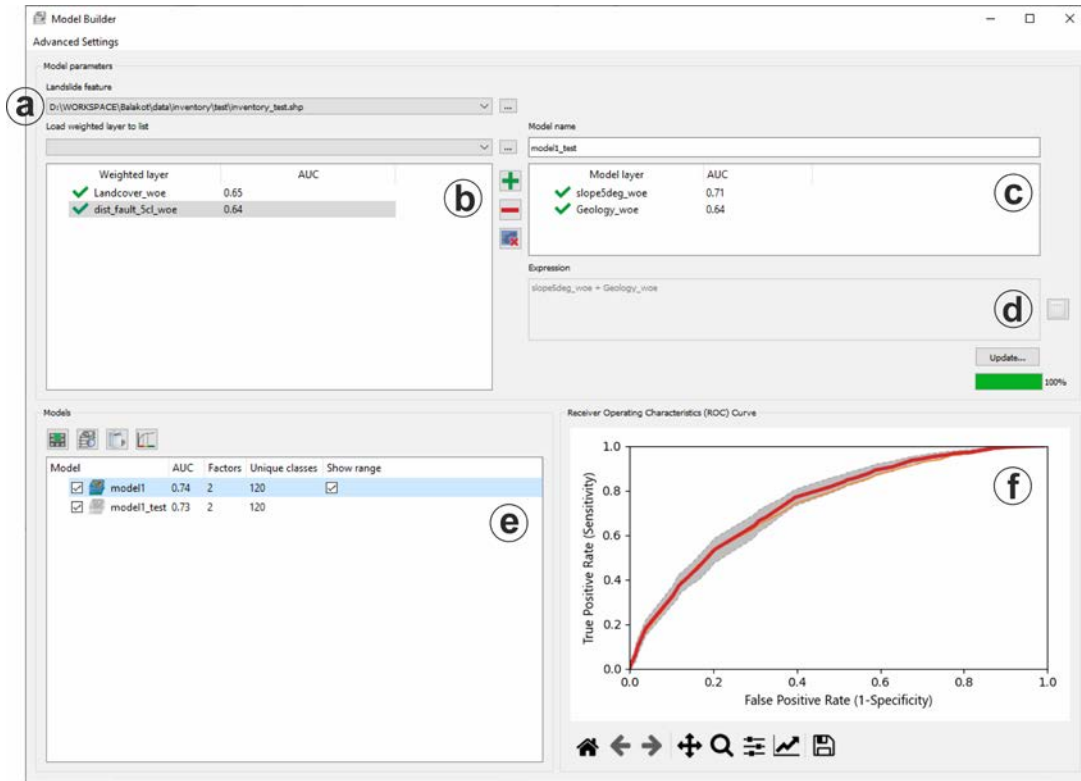



Figure 3~~The user can also generate hybrid models or model ensembles combining outputs from different classifiers. Therefore, the model builder supports the generation of complex mathematical expressions. The model metadata includes the expression~~
380   ~~used to build the model, providing the necessary transparency.~~
4: Model Builder GUI with its integral parts: a – Landslide inventory collection; b – weighted layer collection; c – model layer collection; d – model generating expression; e – models collection; f – ROC curves.

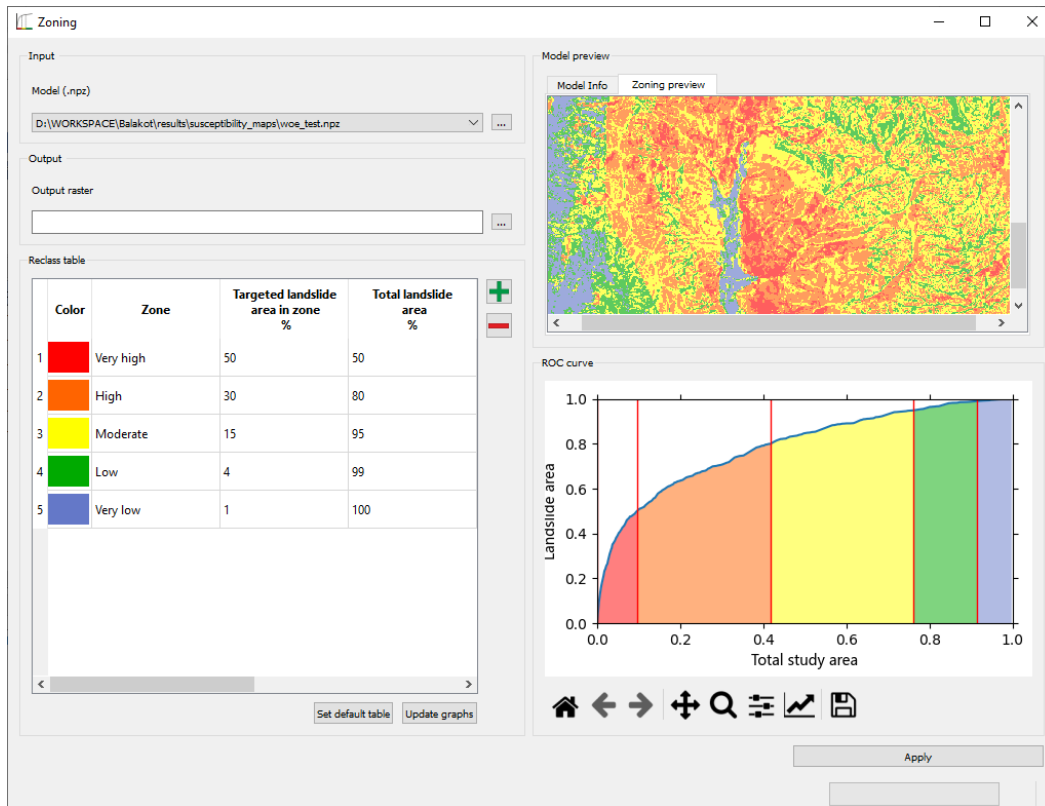
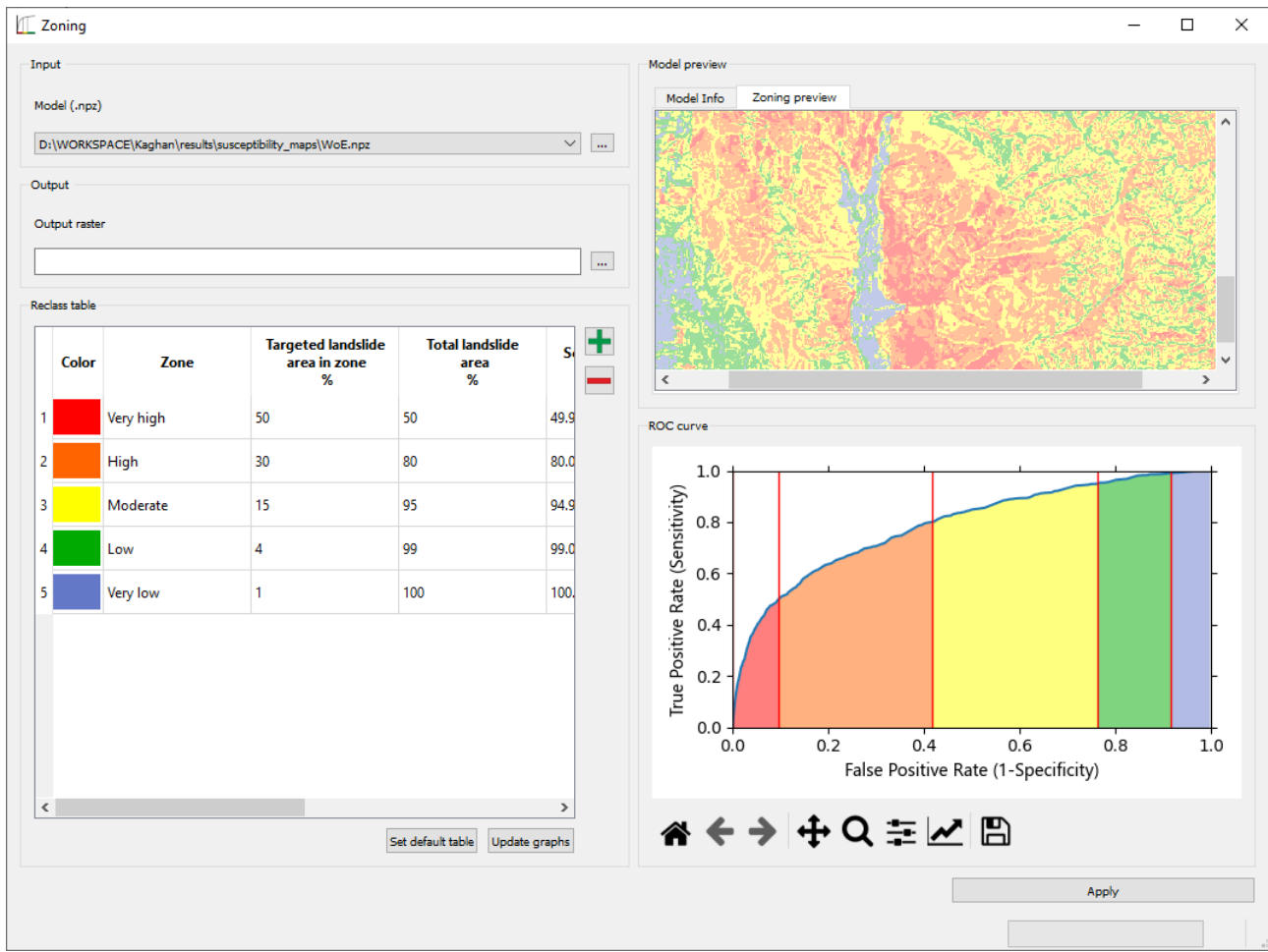
3.6 Zoning

The ~~zonation~~zoning procedure applies to all models generated or evaluated with LSAT PM. It uses the model's ROC curve to
385   aggregate the model output with many different landslide susceptibility index (LSI) values to a legible map with few susceptibility zones ~~(e.g., Chung and Fabbri 2003). Therefore, the user must specify the proportions of landslide areas that should fall within the specific zone.~~. The zoning procedure follows the general concept proposed by Chung and Fabbri (2003).

The basic idea is to specify class boundaries using cumulative landslide area over ranked unique condition classes representing the cumulative study area. Chung and Fabbri (2003) proposed to use the success rate depicting cumulative landslide area over the ranked cumulative area considered susceptible. However, it also partly applies to ROC curves since the y-axis representing the True Positive Rate corresponds to the cumulative landslide area. The x-axis in the ROC curve is the False Positive Rate depicting cumulative study area without landslide areas. In other words, areas that have been regarded as susceptible but do not contain landslide areas. Thus, the cumulative sum over the x-axis is only an approximation of the total study area, which is sufficiently accurate if the landslide areas are neglectable compared to the total area (e.g., when working with point data inventories in large regions). However, it also means that the accurate zone proportions cannot be directly estimated from the ROC curve graph if landslide areas are considerably large. Therefore, the *total area* values in the reclass table represent only approximations for zone area proportions.

Nevertheless, this discrepancy does not affect the implemented classification because we restrict the input to the proportion of cumulative landslide areas within a zone. The specified cumulative landslide area value relates to the rank position of the specific unique condition that exhibits a specific landslide susceptibility index (LSI). The LSI is finally used to set the classification threshold for the class boundary. The zone areas in the attribute table of the output zoning raster are computed directly from the raster values and represent accurate values. The other with success rate possible way to use the proportion of the total area to specify an LS zone, e.g., 10% of the total area in the zone *Very high* is suppressed.

There~~In our case studies (e.g., Torizin et al. 2017, Torizin et al.~~ 2018), we usually used the proportions of 50% of all landslides or landslide areas in the *Very high*, further 30% for the *High*, 15% for *Moderate*, 4% for *Low* and about 1% for the *very low* susceptibility zone (Fig. 4). The implemented tool uses these values as the default table. However, there are no well-established standards for the number of zones or the definition of zone boundaries. In our case studies (e.g., Torizin et al. 2017, Torizin et al. ~~Therefore, the user can freely specify the number of classes and the proportion of landslides in them~~2018), we used the proportions of 50% of all landslide pixels in the *Very high*, further 30% for the *High*, 15% for the *Moderate*, 4% for the *Low* and about 1% for the V*ery low* susceptibility zone (Fig. 4). These thresholds apply when the user selects the default table.

Alternatively, the user can use the ~~reclass~~Reclassify tool to aggregate the model to zones with ~~user specified~~customized thresholds directly on the model's LSI or probability values.

415 **Figure 4: Zoning tool with applied default classification** ~~on the WoE results of the Kaghan Valley dataset~~.

~~5 LSA with the test~~ 4 Application example

4.1 Test data

A dataset to test the functionalities of the software is available from https://doi.org/10.5281/zenodo.5109620. The example or test dataset is an excerpt of the data collected in the German-Pakistani technical cooperation project Georisk Assessment 420 Northern Pakistan (GANP), carried out by the Federal Institute for Geosciences and Natural Resources and Geological Survey of Pakistan (e.g., Torizin et al., 2017). The dataset covers about 664 square kilometers, including parts of the Kaghan and Siran Valleys in Khyber Pakhtunkhwa (KP), Northern Pakistan (Fig. 5a). Table 2 and Fig. 5 provide an overview of the test data.

Table 2: Overview of the datasets in the test data.

Data layer	Data Type	Source
Geology	raster, shapefile (.tif, .shp)	Calkins et al. (1975)
Landcover	raster, shapefile (.tif, .shp)	Fuchs et al. (2015)
AW3D30	raster (.tif)	JAXA (2017)
Confirmed faults	shapefile (.shp)	Calkins et al. (1975)
Landslides	shapefile (.shp)	Torizin et al. (2017)

425 A severe Kashmir Earthquake struck the region on 8 October 2005, with a moment magnitude of 7.8, triggering thousands of landslides. The collected landslide inventory results from the visual interpretation of optical satellite images available through Google Earth, which, during the data acquisition, consisted mainly of imagery from Quickbird (up to 0.60 m ground resolution), IKONOS (~4 m ground resolution), SPOT (SPOT5 about 5 m ground resolution), and Landsat (15 m ground resolution) (Torizin et al., 2017). In total, the landslide inventory includes 3819 events for the test area depicted as polygons.

430 Landslide sizes range from about 12 to about 88444 square meters representing the depletion area of the landslides (as far as it was possible to determine by visual interpretation of imagery).

The digital elevation model is the ALOS Global Digital Surface Model (AW3D30) (JAXA, 2017) with a ground resolution of approximately 30 m (Fig. 5a). The geological information and the tectonic features (faults) were derived from the geological map of Calkins et al. (1975) (Fig. 5d and Tab. 3). The landcover results from the supervised classification on Landsat imagery

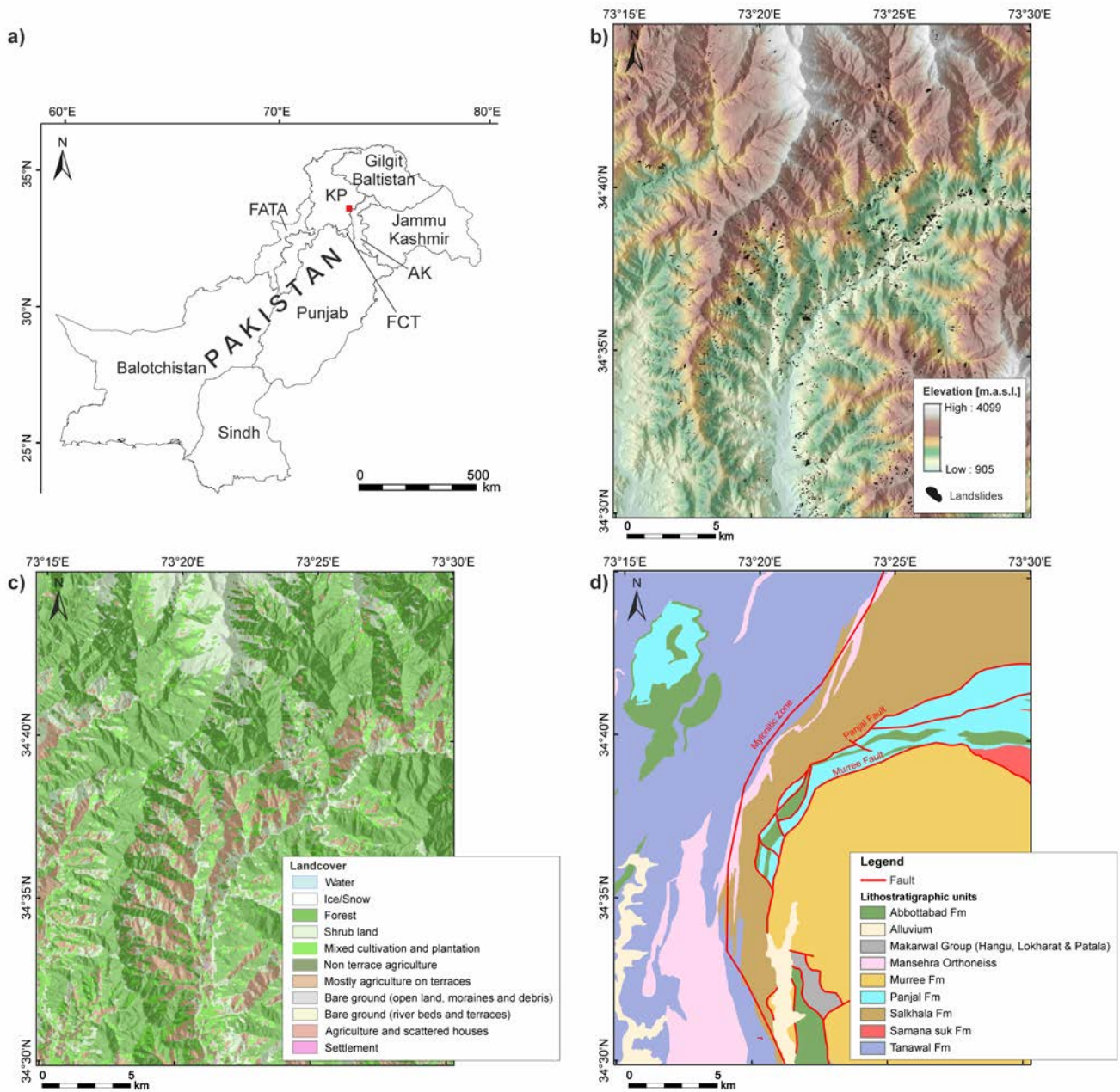
435 performed by Fuchs et al. (2015) (Fig. 5c). The test dataset contains geology and landcover in raster and vector data formats. Note that the vector formats for parameters cannot be directly used for analysis in LSAT PM yet. However, the vector datasets may help test the *Vector Tools* (e.g., subsetting landslides based on specific geology or landcover class).

440   **Figure 5**~~The included Kaghan Valley test dataset serves for exploration of the software.~~ **: Example dataset. a) AW3D30 with landslide inventory; b) Landcover; c) Geology with prominent (confirmed) faults.**

Table 3: Lithostratigraphic units of the geology layer with lithological description (after Calckins et al., 1975).

Lithostratigraphic unit	Lithological description
Abbottabad Fm	Dolomite, quartzite, phyllite, marble, and phosphate

Alluvium	Gravel, sand, silt, clay
Makarwal Group	Ferruginous oolitic sandstone, siltstone, and clay; massive nodular limestone with intercalations of marl and shales; shales and nodular limestone with coal seams
Mansehra Orthogneiss	Massive biotite granite, tourmaline-garnet granit gneiss and porphyroblastic granite gneiss
Murree Fm	Limestone and intercalated shales/marlstone
Panjal Fm	Carbonaceous slate, glassy quartzose, and agglomeratic sandstone
Salkhala Fm	Marble, graphite schist, quartz schist and qurtz-feildspathic gneiss
Samana suk Fm	Limestone and intercalated shales/marlstone
Tanawal Fm	Quartzose schist, quartzite, and schistose conglomerate

4.2 Analysis workflow

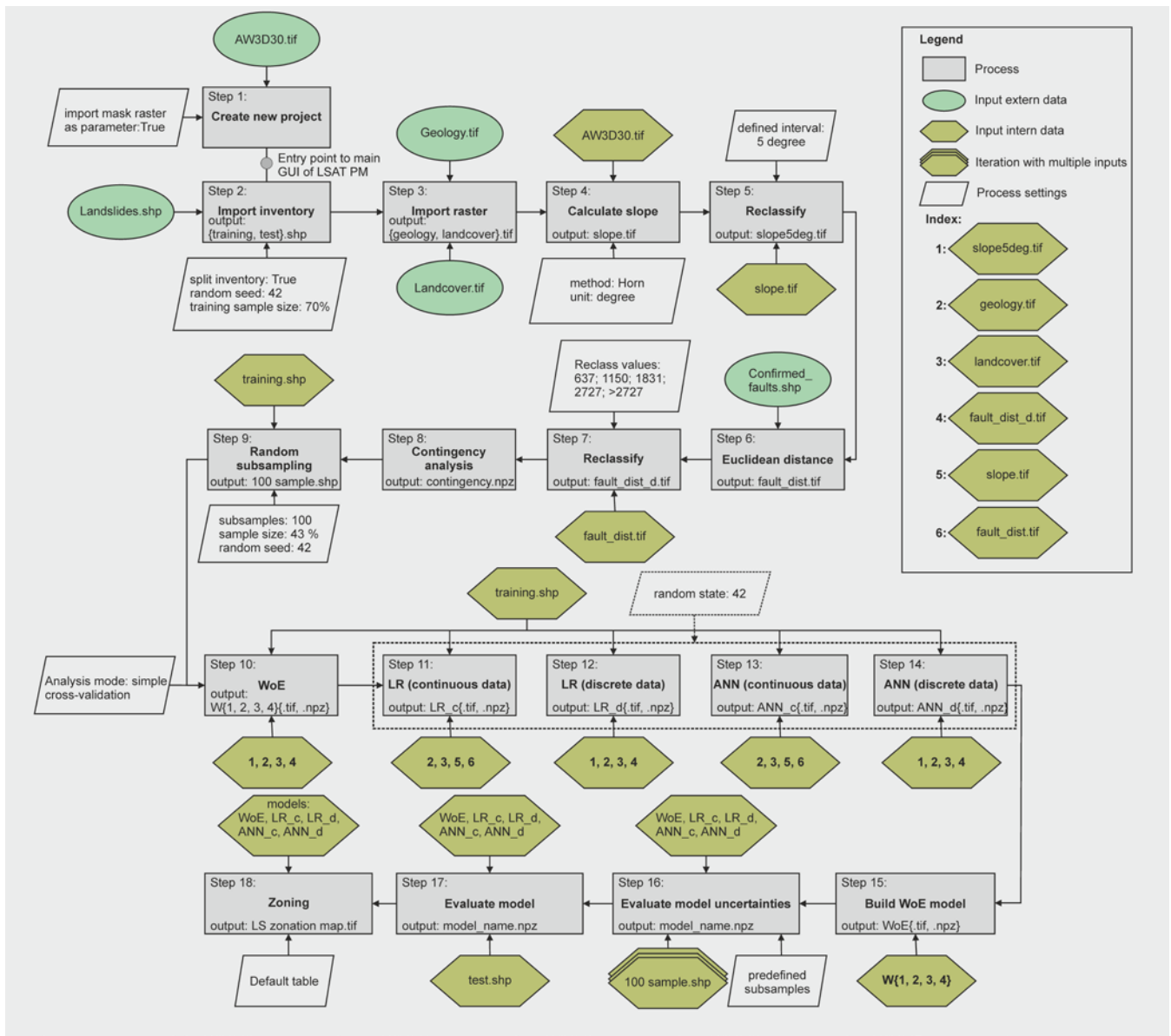
445   In the following example, we use the ~~Khagan Valley~~test dataset to showcase how to perform a simple LSA ~~based on~~with WoE, LR, and ANN in LSAT PM and compare the model outputs. We skip the analysis with AHP in this example due to its high subjectivity and our lacking detailed knowledge about the investigation area.

~~5.1 Step by step procedure~~

~~To get the results, we performed the following steps:~~

450   - ~~Create an LSAT PM project utilizing the test dataset AW3D30.tif as a project reference (mask)~~
   - ~~Import landslide inventory dataset (landslides.shp), setting the random seed to 42 and the training sample size to 70%~~
   - ~~Import raster datasets AW3D30.tif, geology.tif, and landcover.tif~~
   - ~~Calculate the slope from AW3D30.tif based on Horn (1981) algorithm using the included slope tool.~~
   - ~~Reclass generated slope raster to defined intervals of 5-degree using the reclass tool~~
455   - ~~Generate a euclidean distance raster based on the confirmed_faults.shp using the euclidean distance tool~~
   - ~~Reclass the euclidean distance raster, based on the sensitivity analysis with ten quantiles and the training inventory, into five classes [0-637, >637-1150, >1150-1831, >1831-2727, >2727] using the sensitivity reclass tool. Note that the reclassify tool currently provides better handling for reproducing the analysis results with given interval values.~~

~~The following two steps~~Fig. 6 shows the principal workflow of the performed steps. The first seven steps cover the project
460   creation, data import, and first preprocessing of the imported data, such as computation of the slope and euclidean distance and binning of continuous data in categories suitable for WoE. In step eight, the contingency analysis measures associations among discrete datasets based on Chi-square metrics. The utilization of strongly correlated datasets may lead to incorrect estimation of the factor's contribution and inflation of the estimated probability values (e.g., Agterberg and Cheng, 2002).

In step nine, we prepare the data ~~for evaluation of the~~ to evaluate model uncertainties. Therefore, we compared the size of the landslide training and test datasets. ~~The training dataset has~~In step two, the selected split option subdivided the imported landslide inventory into training containing 2674 events~~;~~ and the test dataset ~~has~~with 1146~~, which is about~~ events. The latter corresponds to approximately 43% of the training dataset. To estimate the sample size-dependent model variance, we generated 100 random subsamples from the training dataset the test dataset's size with the ~~size of the test dataset. Therefore, we used the random~~*Random* *sampling* tool. To make the results reproducible ~~also here~~, we set the random seed to 42. ~~The idea behind it is simple.~~ The training and test inventory are random subsamples from the same dataset ~~with a given spatial distribution.~~. Therefore, they represent the same spatial distribution but with a different mean sampling error (MSE) related to the sample size. Torizin et al. (2021) ~~show~~showed that evaluation of the model performance based on a test inventory of a smaller size than a training inventory must consider this for correct interpretation of the results. Evaluation with the 100 subsamples of the same size as the test inventory but derived from the training inventory dataset has two implications. First, all training events are known to the model and follow the same spatial distribution as the complete training inventory. Variations in model performance on these datasets define the MSE, which is expected to be in a similar range by evaluating the model with a test dataset of the corresponding size. ~~The~~Thus, the shape of the ROC curve and AUC value should fall within the MSE range if the model generalizes well without significant overfit.

480

- Figure 6Generate in the project data/inventory/ subfolder samples.
- Run the random sampling tool to generate 100 subsamples from the training inventory with random seed 42 and a sample size of 43%.

In: Workflow applied to the test dataset.

485 Steps 10 to 14 cover the analysis part, we. We calculated the models with WoE, LR, and ANN models in different ways to contrast the approaches. The WoE can utilize only discrete data; therefore, we used the classified and initially discrete data to generate the weighted layers. The LR and ANN support usage of both: continuous and discrete data. ThusTherefore, we used

the capability of both approaches to utilize discrete and continuous data and generated two models for LR and ANN, respectively. The first marked with a _c suffix utilizes both data types, the second marked by _d suffix utilizes only discrete datasets as used in WoE. ~~In detail, the steps are the following:~~

- ~~WoE analysis with discrete layers: geology, landcover, classified slope, and classified distance to faults~~
- ~~Build~~Step 15 is the generation of the WoE model in ~~model builder~~MB by adding single weighted layers ~~to the model.~~. To better compare the results from WoE with results from ANN and LR, we ~~transformed~~adjusted the model generating expression to transform the log-likelihoods of the WoE model into ~~probability space~~probabilities, applying the logistic function (see also ~~Sect. 7.1~~Appendix A for details).
- ~~LR_c with discrete layers geology and landcover and continuous layers slope and distance to faults (computation with default settings, except the random state, which we set as Integer, 42)~~
- ~~LR_d with discrete layers geology, landcover, classified slope, and classified distance to faults (computation with default settings, except the random state, which we set as Integer, 42)~~
- ~~ANN_c with discrete layers geology and landcover and continuous layers slope and distance to faults computation with default settings, except the random state, which we set as 42)~~
- ~~ANN_d with discrete layers geology, landcover, classified slope, and classified distance to faults computation with default settings, except the random state, which we set as 42)~~

~~For model comparison, we used the model builder. Therefore, in the advanced settings of the model builder tool, we selected the~~In step 16, models uncertainties related to sampling error were evaluated in MB using 100 predefined subsamples ~~option.~~ ~~The model performance is evaluated based on subsamples from the folder "samples"~~generated in ~~the steps above. The~~step ~~for~~nine. With this ~~is:~~

- ~~Compare the WoE, LR_c, LR_d, ANN_c, and ANN_d models in Model Builder computing,~~ the ROC ~~curves with the predefined subsets and~~ curve is iteratively computed for every subsample. In the consecutive step, the models were evaluated with the test ~~inventory.~~

dataset not involved in the training process before, thus, representing new data. Finally, we used the ROC curve from the test dataset to generate legible susceptibility maps consisting of five susceptibility zones (default table, see Sect. 3.6) using the *Zoning tool*. ~~For the definition of the susceptibility zones, the default table was used.~~

~~5.2~~4.3 Results and discussion

While the first steps of data import and preparation, such as reclassing usual GIS functionalities, are trivial and are partially described in Sect. 3, the contingency analysis in step eight is worth examining. The results of the contingency analysis are saved in *NumPy* archive format (.npz) in the folder statistics.

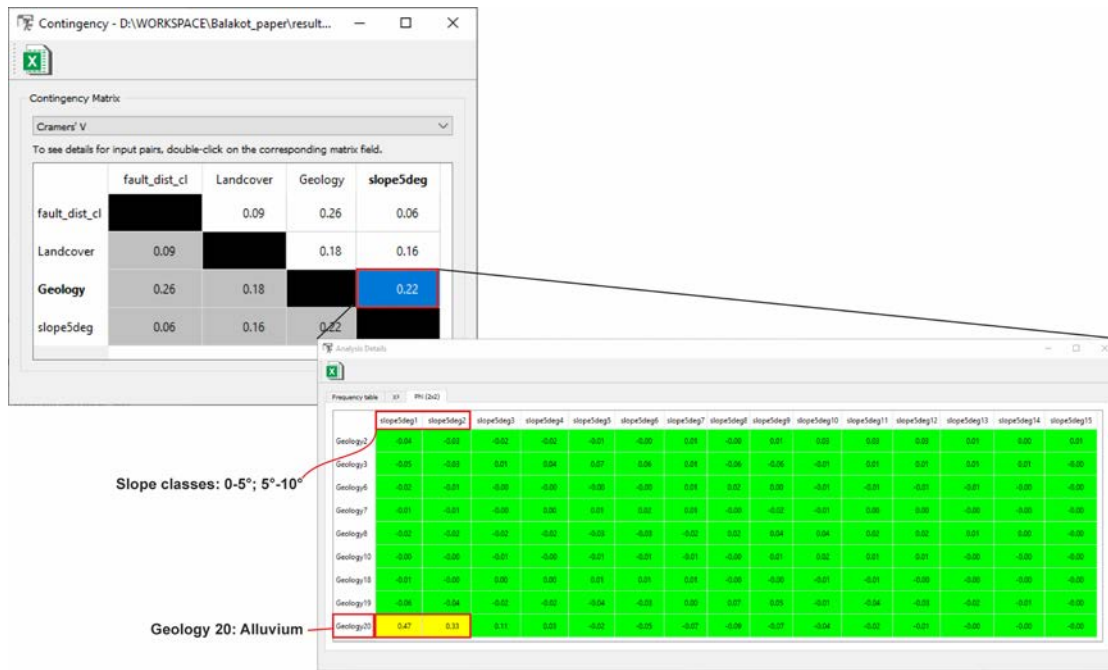



Figure 7 Example for the contingency result output: the first window represents an allover contingency matrix between multiclass parameters based on Cramer's V. On double-click, the detailed contingency between Geology and slope5deg is callable.

The analysis output consists of different tables (Fig. 7). The first result table shows the overview for all involved parameters based on Cramer's V and Pearsons C. Both metrics are estimates for the general association of the multiclass datasets. However, they do not allow to determine in detail whether, e.g., a moderate association comes from several moderately associated classes or as an average estimate from one strongly associated class pair among others not correlated discrete classes. Therefore, an additional Phi metric based on the 2x2 contingency table is callable on double-click on the specific matrix cell, providing more detail for the specific parameter pair. It highlights the pairwise association among single variable class pairs. The tables are colored to emphasize the strength of association: green for no to a marginal association, yellow for the moderate association, and red for the strong association. In Fig. 7, it is clear that while most classes of geology and classified slope layer are not correlated, a moderate association is present for alluvial deposits forming the valley fillings and slope between 0 and 10 degrees.

In the *WoE* tool, the user can explore the associations among the single factor and the occurrence of the events. The results viewer contains all relevant information on the modeling process. The result table represents discrete class area distribution, corresponding landslide pixel frequencies in those classes, computed weights, variance, standard deviation, posterior probability values, and expected observations. The default output raster with suffix _*woe* derives from the table column *weights*. The included graphical representation of the results provides a quick overview of class and landslide pixel distributions, weights, and corresponding ROC curves (Fig. 8). WoE results can be exported as an analysis report in MS Word format that concisely represents the relevant features.

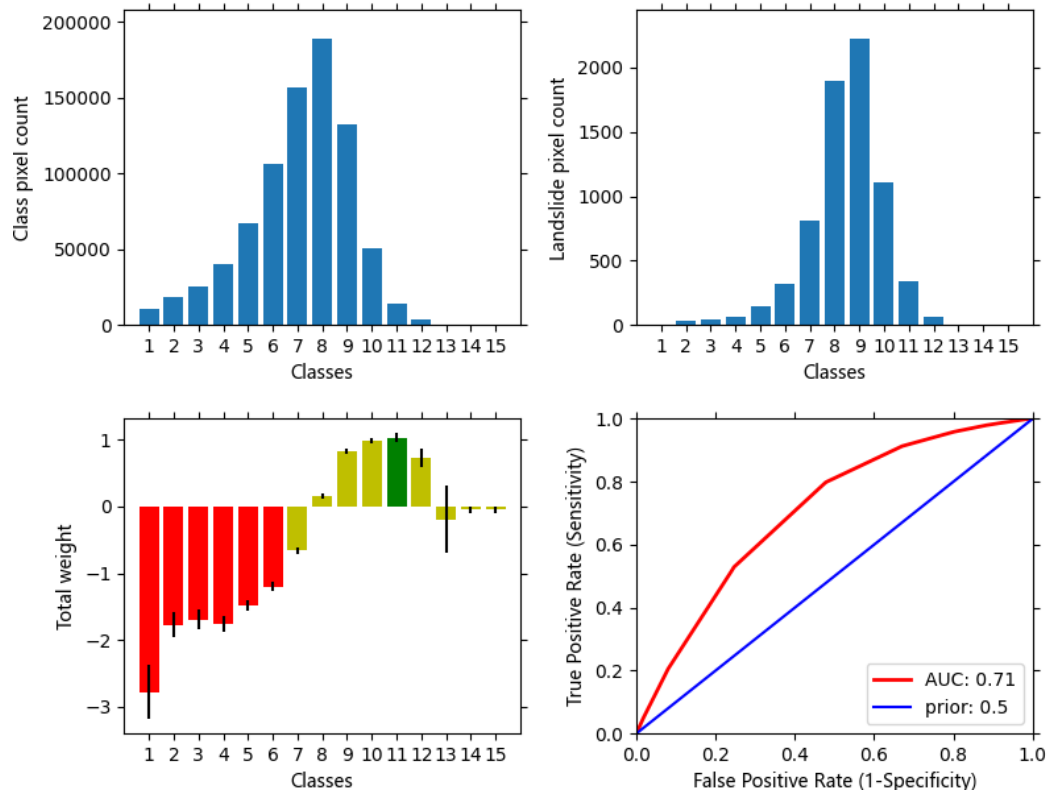



Figure 8~~All~~ Graphical result output of WoE analysis for the simple-cross validation mode. Note that bar plots change to box or violin plots for other analysis modes such as, e.g., on-the-fly subsampling.

Outputs from LR and ANN analyses are compact. Among the involved parameters and model settings, LR results highlight the estimated coefficients and some informational metrics helping to find a tradeoff between model complexity (number of explanatory variables) and explained variance such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), as well as the AUC of the corresponding ROC curve. The experimental ANN provides metadata on model inputs and settings, model score, and AUC value. For both, graphical result output and corresponding analysis reports are not implemented yet (see also Sect. 5).

The evaluation of model uncertainties in step 16 and model evaluation with test data in step 17 suggests that all three approaches deliver comparably good models. Although MB allows appropriate management for comparing the models (Fig. 9), we used customized scripts for the following figures 10, 11, and 13 (see Sect. 8) to showcase additional postprocessing possibilities based on LSAT PM outputs.

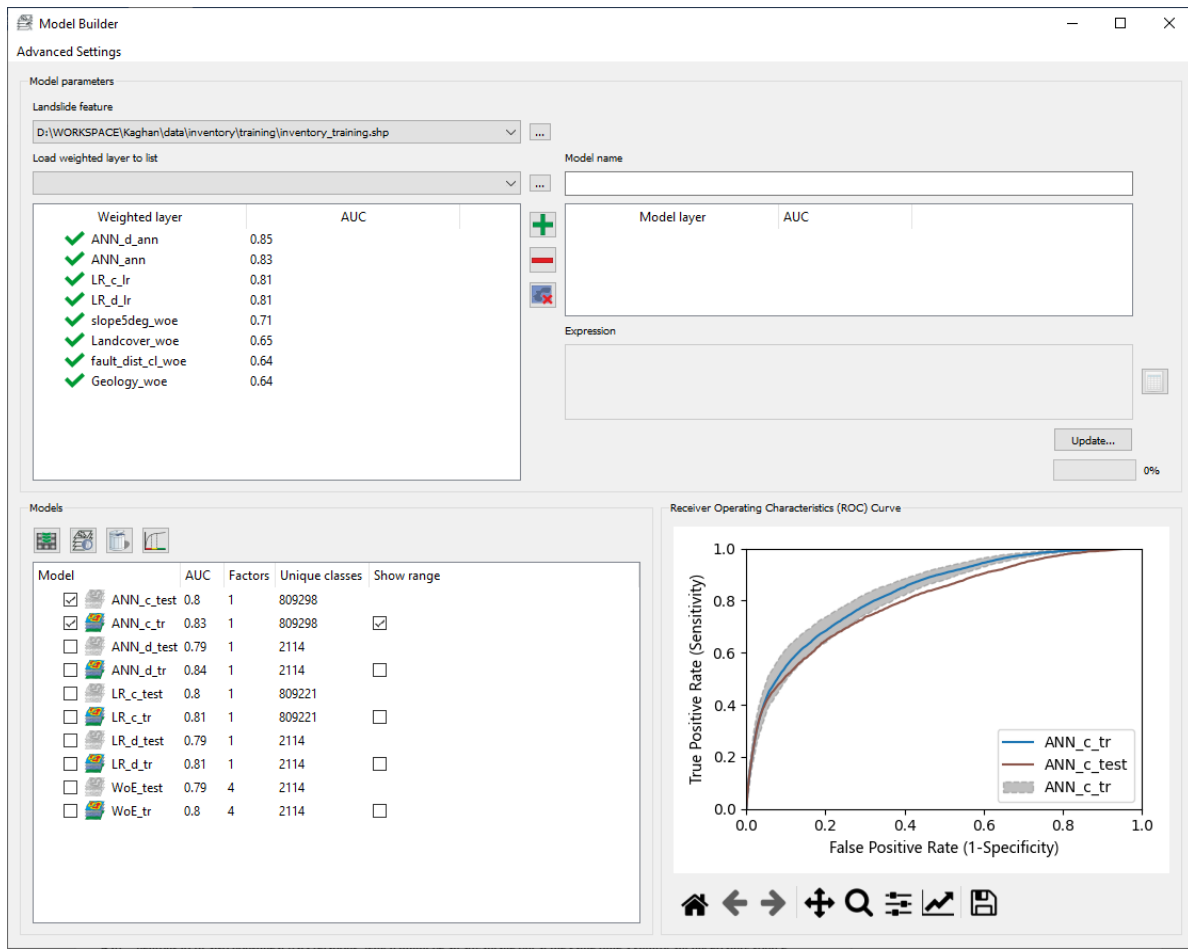



Figure 9 (Fig. 5). Figure 5 Evaluation and comparison of the models in MB.

To aggregate the results to a compact and legible figure (Fig. 10), providing some additional features not included in MB yet, we utilize the outputs of the MB. Figure 10 shows the performance of the models based on the predefined subsamples.

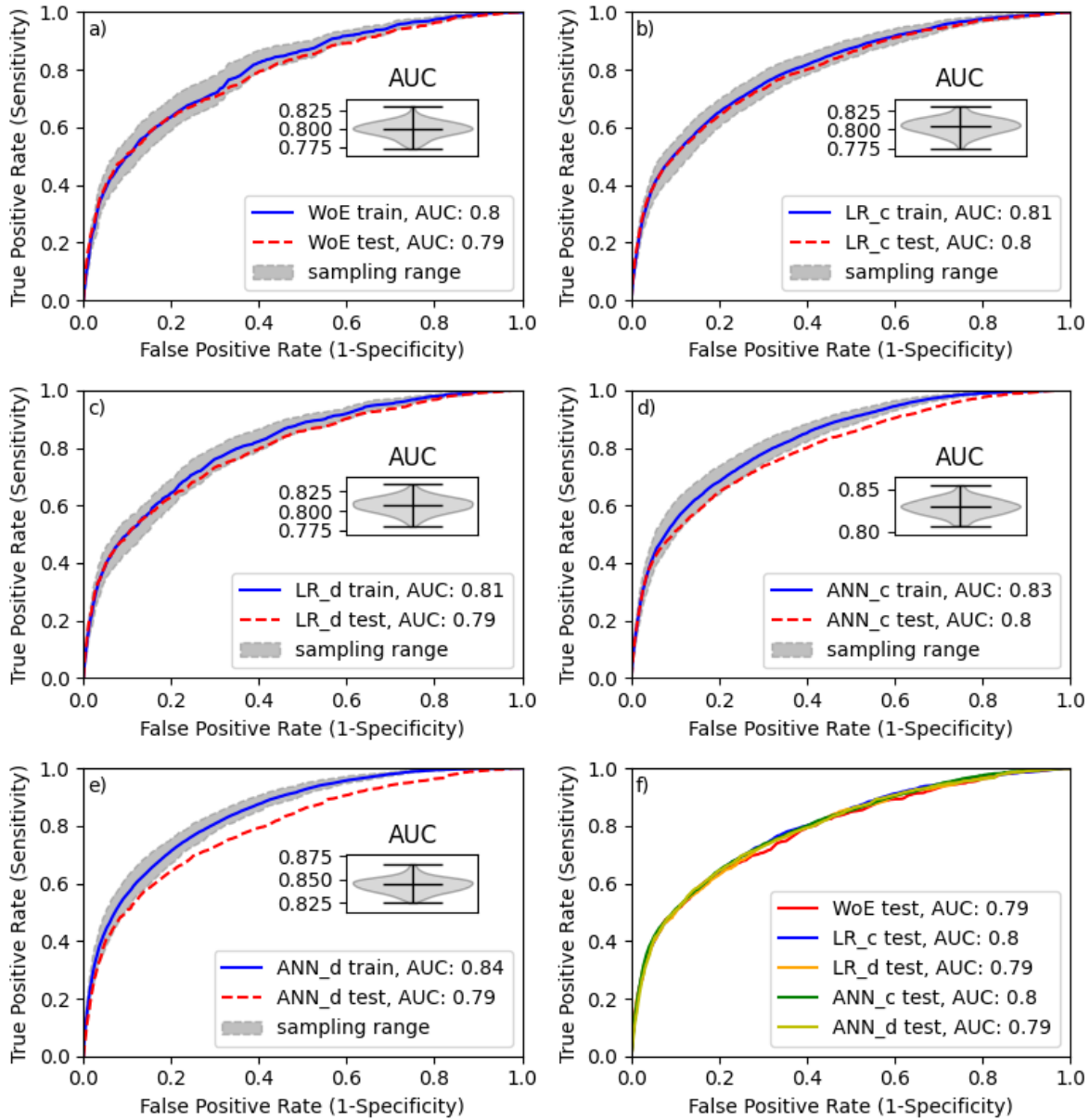



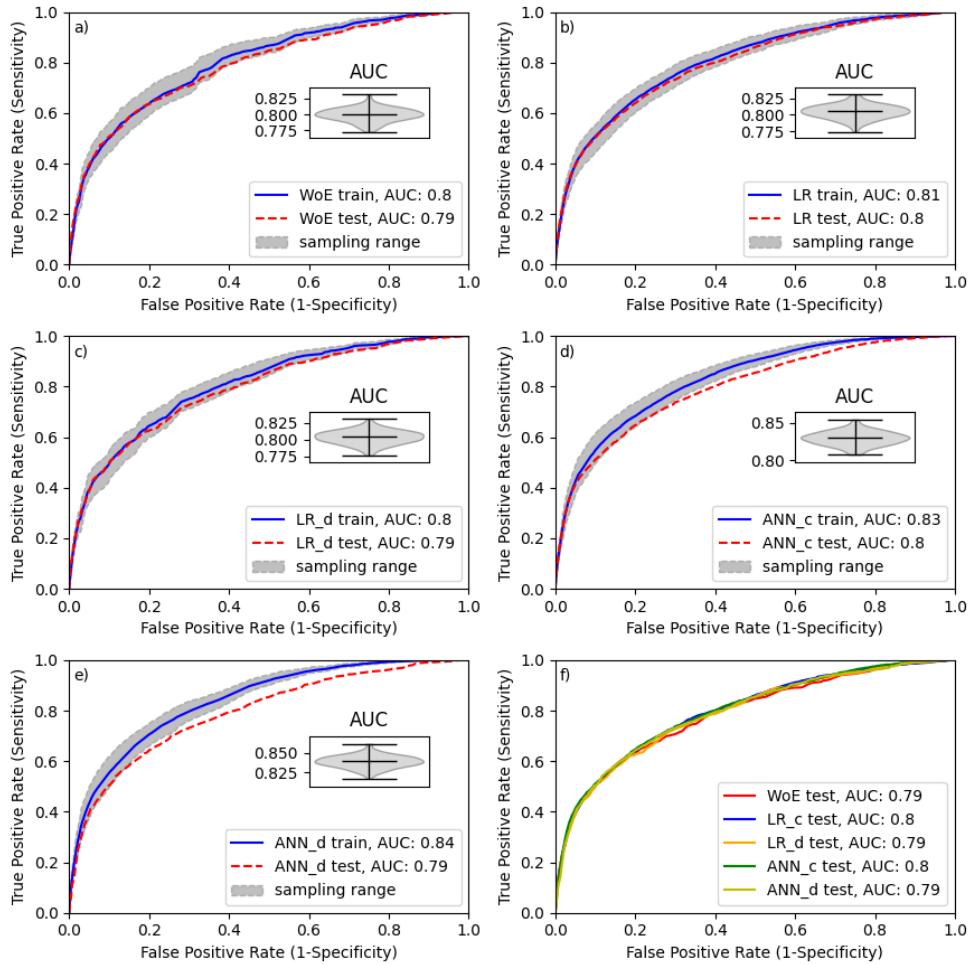
Figure 10 ROC curves for the different models. The greyish band marks the model uncertainty based on the MSE. The insert in a-e shows the corresponding distribution of AUC values for the utilized samples.

The greyish error band around the mean ROC curve indicates the MSE. As we can see from the violin plot inserts in Fig. ~~The greyish error band around the mean ROC curve indicates the MSE. As we can see from the violin plot inserts in Fig. 5a~~10a-e,

the variance of the corresponding AUC values shows the normal distribution. At first glance, ANN_d and ANN_c models have the best training performance with AUCs of 0.84 and 0.83, followed by LR_c with an AUC of 0.81. WoE and LR_d

model show ~~with AUC of 0.80~~ the worst training performance. ~~To see why ANN_d performed better than the other models using discrete data only, we have to look at~~ (AUC of 0.80). For the model evaluation on test ~~performance of the models. We~~data, we see that for the WoE, LR_d, and LR_c, the test ROC curve is within the expected MSE. For ANN_c and ANN_d, however, the test performance is significantly lower. In this case, we can interpret this as an overfit of the models. Using discrete variables in ANN_d, we introduced additional degrees of freedom compared to the ANN_c model; therefore, overfitting is more prominent in the ANN_d model. The reason is the flexibility of the ANNs with multiple neurons to fit also non-linear data relations, which might be an advantage but at the same time a significant uncertainty source. Thus, if we had aimed to optimize the susceptibility map with ANN, we would need to review the network design or the number of iterations in the network training process to prevent overfitting. At this point, it is worth noting that imbalanced samples can also cause comparable effects. When using polygon landslide data, the imbalance may develop from rare large landslides appearing only in the training or the test inventory. Although this affects all model types, flexible ANN might suffer more than, e.g., WoE. The solution would be to check the distributions of the training and test landslide datasets and if imbalances are present to generate balanced samples by, e.g., randomly drawing pixels from landslide areas instead of using them as a whole.

Because the interpretation of ANN results is not intuitive, we would generally recommend a parallel application of a multivariate linear model and ANN to see how much non-linearity is introduced by the ANN and how it affects the model generalization capabilities. ~~LSAT PM serves well for that purpose.~~

Further, looking at the test ROC curves of all models (Fig. ~~5f~~10f), we see that the predictivity of the models is comparable with minor advantages for models utilizing continuous datasets. Thus, given the simple study design and available data, the models are equivalent alternatives from the statistical point of view. However, although the ROC curve provides a quantitative measure for classifier performance, as any statistical measure, it is not suitable for evaluating the model's reliability (e.g., Rossi et al., 2010). As we can demonstrate here, models with equivalent AUCs can exhibit substantially different susceptibility patterns. How meaningful those patterns are is beyond the statistical analysis capabilities and has to be verified based on other sources of information.

585

~~To~~We compare the susceptibility models in a pair plot to evaluate the differences in obtained susceptibility patterns~~, we compare the susceptibility models in a pair plot.~~. The pair plot in Fig. ~~6~~11 visualizes the pairwise pixel-by-pixel comparison

590   of the model values. The matrix diagonal shows the distribution of the model values (marginal probabilities). In contrast, the scatter plots in the lower matrix corner of Fig. ~~6~~11 show the covariance of the value pairs overlain by linear regression to emphasize the trend. The pairwise comparison reveals general linear relation for all models with better comparability for multivariate models ANN and LR but substantial differences in detail. ~~In fact, dependent on the data composition and used approach, we generated five different susceptibility patterns.~~

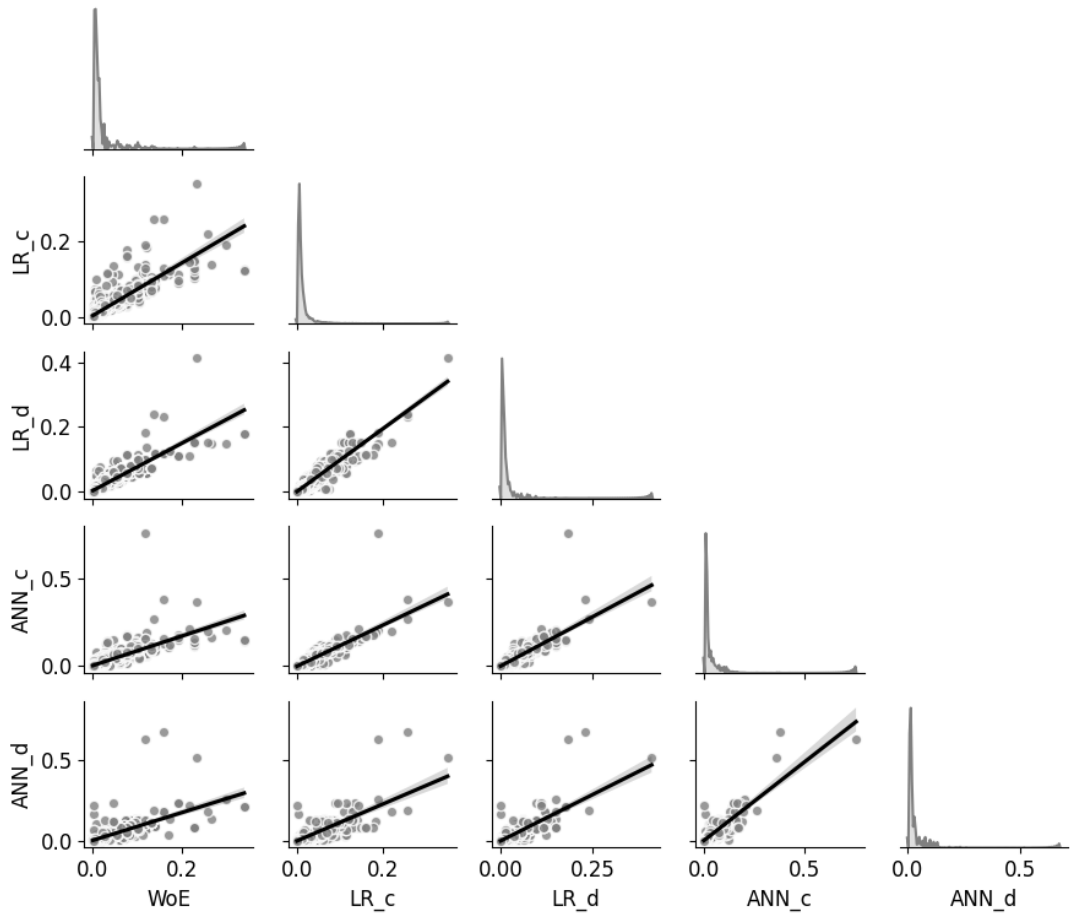
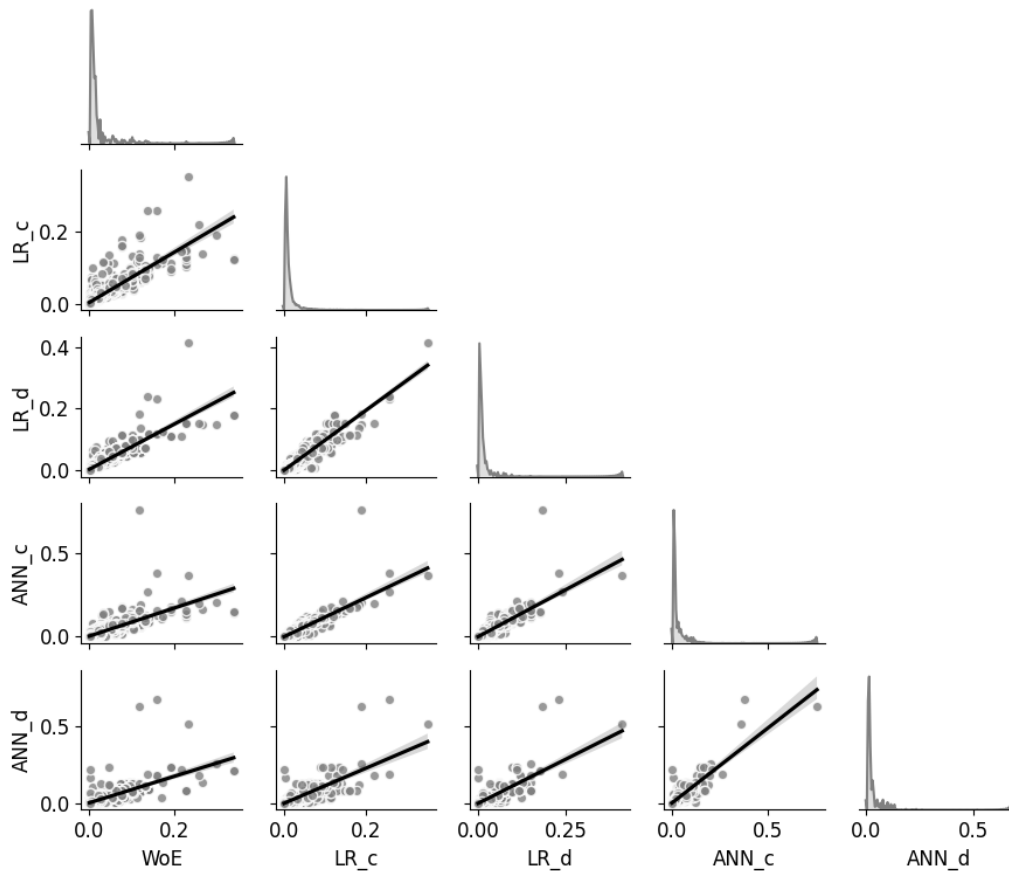



Figure 11 Pairwise comparison of the predicted landslide susceptibility patterns.

We wanted to see how these differences affect the final zonation and compared the models using simple class frequencies statistics after the zonation procedure. Fig. 11 shows the landslide susceptibility maps after zonation with default values introduced in Sect. 3. In Fig. 12In Fig. 7, the classified models are compared regarding their pixel counts within the susceptibility classes. While the highest susceptibility class containing 50% of all landslide areaspixels show minor differences, they become more significant in the lower susceptibility zones.

	Very high	High	Moderate	Low	Very low
WoE	78473	260900	279461	126897	69531
LR_c	82368	244282	264058	162475	62079
LR_d	80804	262245	267982	120293	83938
ANN_c	76968	246861	255649	143732	92052
ANN_d	78413	259443	259245	128639	89522

Landslide susceptibility zones

Very high | High | Moderate | Low | Very low

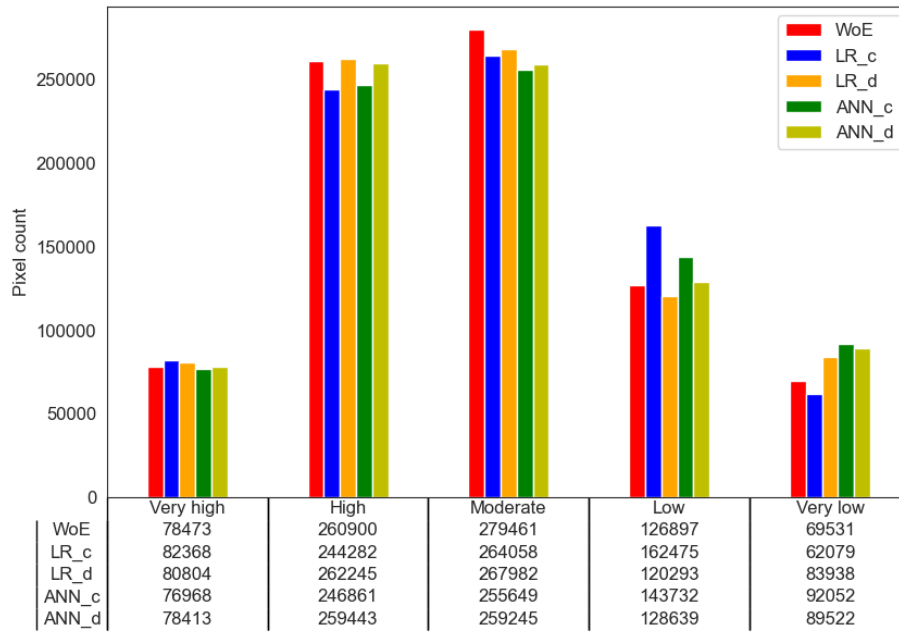
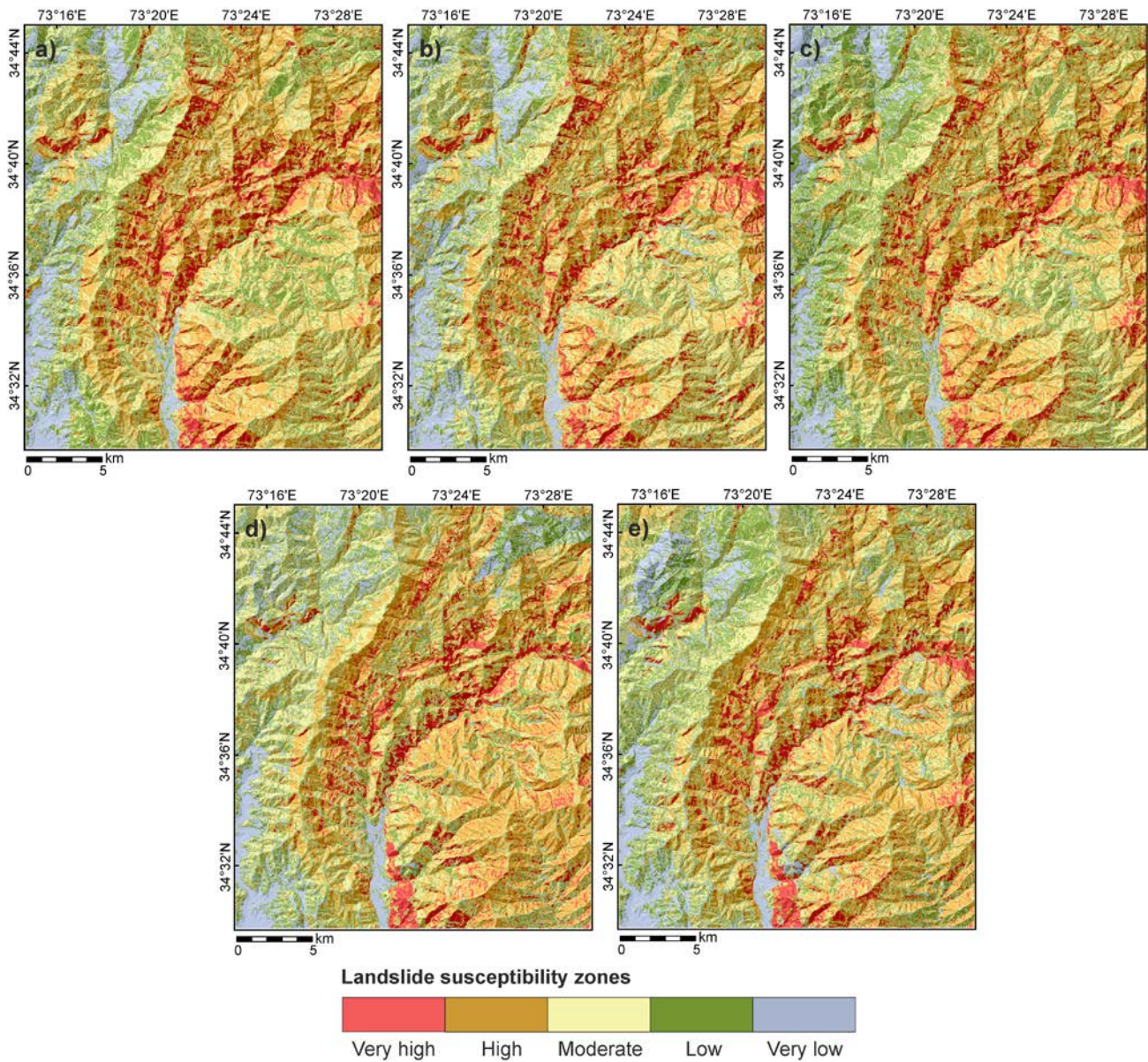



Figure 12 ~~Susceptibility zones distribution for the~~ Landslide susceptibility zones based on different models: a) WoE; b) LR_d; c) LR_c; d) ANN_d; e) ANN_c. *Very high* contains about 50%, *High* about 30%, *Moderate* about 15%, *Low* about 4% and *Very low* about 1% of all landslide ~~areas. The values in columns show the number of pixels within the zones~~pixels.

	Very high	High	Moderate	Low	Very low
WoE	78583	259787	283331	124828	68733
LR_c	79679	244125	266754	162469	62235
LR_d	82373	247177	281848	118160	85704
ANN_c	76969	249599	252910	143732	92052
ANN_d	77879	260056	266404	125981	84942

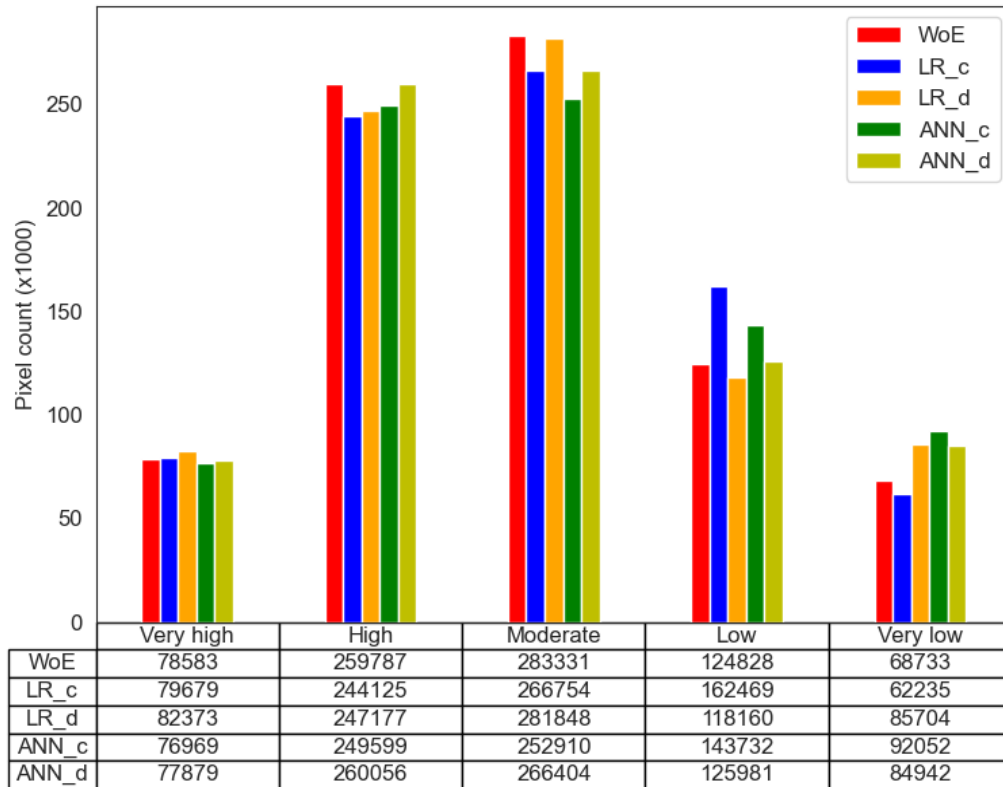



Figure 13~~6~~ Susceptibility zones distribution for the susceptibility models. The values in columns show the number of pixels within the zones.

5 Conclusions and Outlook

In this paper, we introduced LSAT PM ~~streamlines the application and evaluation of popular data and~~ that provides a framework for applying and evaluating knowledge- and data-driven spatial binary classification methods. ~~The tool further provides the ability to build~~ and ~~evaluate~~demonstrated in parts its capabilities based on a real-world test dataset. The given presentation is not exhaustive but provides a general idea of using the software.

Summarizing the capabilities of LSAT PM, we can emphasize that the project-based modular framework allows efficient data management at all steps of the LSA. The implemented logs on performed processing steps and metadata collection increase transparency and reproducibility and allow easy sharing of the modeling results, e.g., in working groups. At the same time, we tried to keep as many degrees of freedom as possible in the modeling procedures, providing the users not a fixed pipe but a toolkit that allows for flexible study designs. Thus, the preprocessing, analysis, and postprocessing steps are performable on the user's choice.

The MB module of LSAT PM implements functionalities allowing for practical evaluation of model uncertainties related to common sampling errors. Further, it provides a convenient way to compare models generated by the implemented algorithms and foreign models. The implemented raster calculator supports the generation of hybrid models and model ensembles and evaluate any other models. With the provided test dataset, users can .

LSAT PM targets a broad user profile. It provides access to LSA state-of-the-art methods to users beyond the academic community. Users with limited programming and scripting skills can perform analyses and explore different analysis techniques at different complexity the results via convenient GUI or export the results to other applications allowing further postprocessing. Skilled users can also benefit from implemented standards and quickly enhance the analysis outcomes stored in *NumPy* archive format by own scripts, as we have shown in Sect. 4. This versatility makes LSAT PM well suited for educational purposes at all levels.

Of course, there is always room for improvement. Therefore, the LSAT PM is subject to continuous further development. We intend to implement additional and improve existing methods, e.g., improve the AHP and machine learning workflows. Especially for the latter, we intend to implement additional features to visualize the results and increase the interpretability of the model outputs. To We intend to implement GPU support to perform better on massive datasets in machine learning methods, we intend to implement TensorFlow (Tensor Flow Team, 2015), adding GPU computation capability. Also, we want to implement new some management features such as a plugin builder tool that would support the easy implementation of customized plugins into the LSAT PM framework. are under preparation. Software documentation is going to continuously updates with the ongoing development and will be extended by short video tutorials introducing the work with LSAT PM. With the open-source approach, we would like to encourage interested scientists to join the development, introducing and discussing new ideas and sharing experience in spatial modeling of landslides and scientific programming in general.

76 Appendices

Appendix A: Supporting Information for Weights of Evidence

WoE uses the Bayes rule to estimate the conditional probability of an event based on prior knowledge and a set of pieces of evidence. Prior knowledge in terms of prior probability usually represents the average expectation of an event given a study area. For the raster-based analysis, the prior probability calculates as the number of event pixels divided by the total pixels in the study area. Thus, the prior probability follows a uniform distribution over the entire study area. (e.g., Torizin 2016). We update the prior probability by weighted evidence factors, which we assume to be conditionally independent. For the spatial analysis, the factors characterize how much the probability value in a specific location is higher or lower than the prior probability. The updated probability is called posterior probability. (e.g., Teerarungsigul et al., 2015). The performed knowledge update cannot find new events but rather redistributes the probability density patterns conserving the total events. (e.g., Agterberg and Cheng, 2002; Torizin, 2016). Thus, given the conditional independence of evidential patterns, we should obtain (approximately) the number of initial event pixels when summing up all pixels of the posterior probability raster. In

practice, complete conditional independence of evidential patterns is seldom. Therefore, ~~the usage of~~using many factors may

655 cause inflation of posterior probability by occasional double counting of the effects (~~e.g.,~~Agterberg and Cheng, 2002). ~~This~~The latter sets a general requirement to perform a contingency analysis (e.g., based on chi-square metrics) to estimate possible associations between evidence patterns.

In WoE, the weights for specific evidence patterns derive from the Bayes rule's formulation in logarithmic odds notation (e.g., Bonham-Carter~~,~~ et al., 1989; Bonham-Carter, 1994), considering the evidence as a binary pattern for presence or absence of a

660 specific feature. For ~~multi-class~~multiclass datasets, the computation is done as if the dataset ~~had been disintegrated into~~consists of several binary dummy variables. However, the straightforward analysis allows the weight calculation for ~~multi-class~~multiclass datasets in one table without generating binary dummy variables explicitly. The weight calculation for a particular raster cell in a binary pattern distinguishes two cases. First, if the particular feature class C is present, then the logit is given by:

$$w^+ = log_e \left(\frac{P\{C|E\}}{P\{C|\bar{E}\}} \right). \tag{1}$$

else, the logit is given by:

$$w^- = log_e \left(\frac{P\{\bar{C}|E\}}{P\{\bar{C}|\bar{E}\}} \right). \tag{2}$$

where $P\{C/E\}$ is the conditional probability of C (feature) given E (event ~~E;~~); $P\{C/\bar{E}\}$ is the conditional probability of C given $\bar{E}$ (no event);

670 $P\{\bar{C}/E\}$ is the conditional probability of $\bar{C}$ (no feature) given E; $P\{\bar{C}/\bar{E}\}$ is the conditional probability of $\bar{C}$ ~~(no feature)~~ given $\bar{E}$.

The weight notations w^+ and w^- do not represent ~~not~~ the mathematical sense of the values but the feature class presence (positive) and absence (negative) in the given raster cell.

With this formulation, positive logit values suggest a positive effect of the given variable, negative logits indicate a negative

675 effect, and logits with a zero value indicate no effect. Thus, the latter does not modify the prior probability.

~~For a specific number of weighted layers, the~~The posterior logit z is obtainable from weighted layers w_i (e.g., Barbieri and Cambuli, 2009; Torizin, 2016) as:

$$z = PriorLogit + \sum_{i=1}^{n} w_i. \tag{3}$$

where n is the number of weighted layers (evidence), and w_i is the ith weighted layer, and the *PriorLogit* is:

$$PriorLogit = log_e \left(\frac{\frac{total\ landslide\ pixel}{total\ area\ pixel}}{1 - \frac{total\ landslide\ pixel}{total\ area\ pixel}} \right). \tag{4}$$

To convert the logit formulation to posterior probability, we use the logistic function:

$$p_{post} = \frac{1}{1+e^{-z}}.\tag{5}$$

The sum of the weighted layers from Eq. 3 is the default output in the LSAT PM model builder for WoE models. It is already sufficient to obtain the relative susceptibility pattern needed for evaluation ~~with the ROC curve.~~. To obtain a model with probability values, the user should first compute the prior logit and modify the model generating expression in the model builder according to Eq. 5. Necessary information on the total number of landslide pixels and the total number of pixels in the study area is obtainable from the result table of any weighted layer. We recommend exporting the result table to Excel and conducting the simple side calculation as shown in Eq. 4.

~~8~~7 Code Availability

The current version of LSAT PM is available from the project website: https://github.com/BGR-EGHA/LSAT under the GNU GPL v3.0 license. The exact version of LSAT PM used to produce the results used in this paper is archived on Zenodo: https://doi.org/10.5281/zenodo.~~5511768~~5909726. The LSAT PM documentation is available separately from https://github.com/BGR-EGHA/LSAT-Documentation under the CC BY-SA 4.0 license. The documentation for the exact version is archived on Zenodo: https://doi.org/10.5281/zenodo.5909744. The scripts used for postprocessing of the results in Sect. 4.3 are available from https://doi.org/10.5281/zenodo.5913626~~https://doi.org/10.5281/zenodo.5511920.~~.

~~9~~8 Data availability

The corresponding test dataset is archived on Zenodo: https://doi.org/10.5281/zenodo.5109620 under the CC BY-SA 4.0 license.

~~10~~9 Author contribution

Jewgenij Torizin developed the theoretical concept~~,~~ and designed and coded LSAT PM. Nick Schüßler designed and coded parts of LSAT PM and migrated the code from Python 2.7 to Python 3. Michael Fuchs contributed with theoretical concepts and testing of the application through all stages of the development.

~~11~~10 Competing interests

The authors declare that they have no conflict of interest.

1211 Acknowledgements

We developed parts of the LSAT PM in the framework of a scientific-technical cooperation project between the Federal Institute for Geosciences and Natural Resources (BGR) and the China Geological Survey (CGS). This project was co-funded by the German Ministry of Economic Affairs and Energy (BMWi) and the Ministry of Natural Resources of the People's Republic of China. We also sincerely thank all colleagues who tested the prototypes of LSAT PM in its different development stages, helping us improve the software.

1312 References

Agterberg, F. P., and Cheng, Q.: Conditional independence Test for Weight-of-Evidence Modeling. Natural Resources Research 11 (4): 249-255, https://doi.org/10.1023/A:1021193827501, 2002.

Aleotti, P. and Chowdhury, R.: Landslide hazard assessment: summary review and new perspectives, Bull. Eng. Geol. Envir., 58(1), 21–44, https://doi.org/10.1007/s100640050066, 1999.

Alimohammadlou, Y., Najafi, A., and Gokceoglu, C.: Estimation of rainfall-induced landslides using ANN and fuzzy clustering methods: A case study in Saeen Slope, Azerbaijan province, Iran. Catena, 120, 149-162, https://doi.org/10.1016/j.catena.2014.04.009, 2014.

Balzer, D., Dommaschk, P., Ehret, D., Fuchs, M., Glaser, S., Henscheid, S., Kuhn, D., Strauß, R., Torizin, J., Wiedenmann, J.: Massenbewegungen in Deutschland (MBiD) – Beiträge zur Modellierung der Hangrutschungsempfindlichkeit. Ein Kooperationsprojekt zwischen den Staatlichen Geologischen Diensten der Bundesländer Baden-Württemberg, Bayern, Nordrhein-Westfalen, Sachsen und der Bundesanstalt für Geowissenschaften und Rohstoffe im Auftrag des Direktorenkreises der Staatlichen Geologischen Dienste in Deutschland. – Abschlussbericht, Augsburg, Freiberg, Freiburg, Hannover und Krefeld, 2020.

Barbieri, G., Cambuli, P.: The weight of evidence statistical method in landslide susceptibility mapping 424 of the Rio Pardu Valley (Sardinia, Italy). 18th World IMACS/MODSIM Congress, Cairns, Australia, 2009.

Bonham-Carter, G. F., Agterberg, F. P. and Wright, D. F.: Weights of evidence modelling: a new approach to mapping mineral potential. Stat. Appl. Earth. Sci. Geol. Survey Can. Paper, 89–9, 171–183, 1989.

Bonham-Carter, G. F.: Geographic information systems for geoscientists: Modelling with GIS. Ottawa: Pergamon Press, 1994.

Bouziat, A., Schmitz, J., Deschamps, R., and Labat, K.: Digital transformation and geoscience education: New tools to learn, new skills to grow. European Geologist, 50, https://doi.org/10.5281/zenodo.4311379, 2020

Brabb, E. E.: Innovative approaches to landslide hazard and risk mapping. Proceedings of the 4th International Symposium on Landslides, Toronto, vol. 1, 307–324, 1985.

Bragagnolo, L., da Silva, R. V., and Grzybowski, J. M. V.: Artificial neural network ensembles applied to the mapping of landslide susceptibility. Catena, 184, 104240, https://doi.org/10.1016/j.catena.2019.104240, 2020a.

Bragagnolo, L., da Silva, R. V., and Grzybowski, J. M. V.: Landslide susceptibility mapping with r.landslide: A free open-source GIS-integrated tool based on Artificial Neural Networks. Environmental Modelling & Software, 123, 104565, https://doi.org/10.1016/j.envsoft.2019.104565, 2020b.

Brenning, A.: Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. In J. Boehner, T. Blaschke and L. Montanarella (eds.), SAGA - Seconds Out (= Hamburger Beitraege zur Physischen Geographie und Landschaftsoekologie, vol. 19), 23-32, 2008.

Budimir, M. E. A., Atkinson, P. M., Lewis, H. G.:A systematic review of landslide probability mapping using logistic regression. Landslides, 12, 419–436, https://doi.org/10.1007/s10346-014-0550-5, 2015.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., Varoquaux, G.: API design for machine learning software: experiences from the scikit-learn project, ECML PKDD Workshop: Languages for Data Mining and Machine Learning, 108-122, 2013.

Calkins, J. A., Offield, T. W., Abdullah, S. K. M., Ali, T.: Geology of the Southern Himalaya in Hazara, Pakistan, and Adjacent Areas. Geological Survey Professional Paper 716-C, United States Government Printing Office, Washington, 1975.

Chung, A. J. F., Fabbri, A. G.:Canny, S.: python-docx - A Python library for creating and updating Microsoft Word (.docx) files, retrieved on 4 May 2018 from https://pypi.org/project/python-docx.

Chung, C.-J., Fabbri, A. G.: Predicting landslides for risk analysis – Spatial models tested by a cross-validation technique, Geomorphology 94(3–4), 438-452, https://doi.org/10.1016/j.geomorph.2006.12.036, 2008.

Chung, C.-J., Fabbri, A. G.: Validation of spatial prediction models for landslide hazard mapping. Nat Hazards, 30, 451–472, 2003.

Crozier, M. J.: Deciphering the effect of climate change on landslide activity: A review. Geomorphology 124(3-4): 260-268. https://doi.org/10.1016/j.geomorph.2010.04.009, 2010.

Ermini, L., Catani, F., and Casagli, N.: Artificial Neural Networks applied to landslide susceptibility assessment. Geomorphology, 66, 327-343, https://doi.org/10.1016/j.geomorph.2004.09.025, 2005.

Fawcett, T.: An introduction to ROC analysis. Pattern Recogn Lett, 27(8): 861–874. doi: 10.1016/j.patrec.2005.10.010, 2006.

Froude, M. J., Petley, D. N.: Global fatal landslide occurrence from 2004 to 2016. Nat. Hazards Earth Syst. Sci., 18: 2161–2181. https://doi.org/10.5194/nhess-18-2161-2018, 2018.

Fuchs, M., Khalid N.: Land Cover Map for the Districts of Mansehra & Torghar, Province Khyber Pakhtunkhwa, Islamic Republic of Pakistan, Final Report, 44 p., Islamabad/Hannover, 2015.

Gazoni, E. and Clark, C.: openpyxl - A Python library to read/write Excel 2010 xlsx/xlsm files, retrieved on 11. June 2018 from https://openpyxl.readthedocs.io.

GDAL/OGR contributors: GDAL/OGR Geospatial Data Abstraction software Library. Open Source Geospatial Foundation. https://gdal.org, 2021.

770 George A. Miller: The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review, 63, 81–97, https://doi.org/10.1037/h0043158, 1956.

GitHub Myanmar_Landslide_Models [code], https://github.com/katharineegan/Myanmar_Landslide_Models, last access: 17. December 2021a.

GitHub Shallow_landslide_susceptibility_analysis_using_Random_Forest_method_in_Val_D"Aosta_D'Aosta_Valley [code],

775 https://github.com/MattiaRaffa/RF-VDA-landslide-map, last access: 17. December 2021b.

GRASS Development Team: Geographic Resources Analysis Support System (GRASS) Software, Version 7.2.1 Open Source Geospatial Foundation. http://grass.osgeo.org, 2021.

Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P.: Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology, 31:181-216. https://doi.org/10.1016/s0169-555x(99)00078-

780 1, 1999.

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

Hall-Wallace, M. K.: Integrating Computing Across a Geosciences Curriculum Through an Applications Course. Journal of Geoscience Education 47(2): 119-123, https://doi.org/10.5408/1089-9995-47.2.119, 1999.

785 Harris, C.R., Millman, K.J., van der Walt, S.J. et al.: Array programming with NumPy. Nature 585, 357–362, https://doi.org/10.1038/s41586-020-2649-2, 2020.

Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., Navarro-Colorado, B.: A Systematic Review of Deep Learning Approaches to Educational Data Mining. Complexity, Article ID 1306039, https://doi.org/10.1155/2019/1306039, 2019.

Hunter, J. D.: Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90–95,

790 https://doi.org/10.1109/MCSE.2007.55, 2007.

JAXA: ALOS Global DSM AW3D30 Dataset Product Format Description for V 1.1, available on http://www.eorc.jaxa.jp/ALOS/en/aw3d30/aw3d30v11_format_e.pdf, 2017.

Jebur, M. N., Pradhan, B., Shafri, H. Z. M., Yusoff, Z.M., Tehrany, M.S. An integrated user-friendly ArcMap tool for bivariate statistical modelling in geoscience applications. Geosci Model Dev 8:881–891. https://doi.org/10.5194/gmd-8-881-2015,

795 2015.

Kamp, U., Growley, B. J., Khattak, G. A., Owen, L. A.: GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. Geomorphology 101: 631-642, https://doi.org/10.1016/j.geomorph.2008.03.003, 2008.

Kamp, U., Owen, L. A., Growley, B. J., Khattak, G. A.: Back analysis of landslide susceptibility zonation mapping for the 2005 Kashmir earthquake: an assessment of the reliability of susceptibility zoning maps, Natural Hazards 54: 1-25, 2010.

800 Keefer, D. K.: Investigating Landslides Caused by Earthquakes – A Historical Review. Surveys in Geophysics, 23:473-510, https://doi.org/10.1023/A:1021274710840, 2002.

Lee, S., Evangelista, D. G.: Earthquake-induced landslide susceptibility mapping using an artificial neural network. Natural Hazards and Earth System Sciences 6:687–695, https://doi.org/10.5194/nhess-6-687-2006, 2006.

Lee, S.: Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data International Journal of Remote Sensing, 26:7, 1477-1491. https://doi.org/10.1080/01431160412331331012, 2005.

Lombardo, L., Mai, M. P.: Presenting logistic regression-based landslide susceptibility results. Engineering Geology 244: 14-24, https://doi.org/10.1016/j.enggeo.2018.07.019, 2018.

Makkawi, M. H., Hariri, M. M., Ghaleb, A. R.: Computer Utilization in Teaching Earth Sciences: Experience of King Fahd University of Petroleum and Minerals. International Education Journal 4(2): 89-97, 2003.

Mathew, J., Jha V. K, Rawat, G. S.: Weights of evidence modelling for landslide hazard zonation mapping of Bhagirathi Valley, Uttarakhand. Current Science 92 (5): 628-638, 2007.

Merghadi, A.: An R Project for landslide susceptibility mapping in Mila basin (v1.1.0), Zenodo [code], https://doi.org/10.5281/zenodo.1000431, 2018.

Merghadi, A.: An R Project for landslide susceptibility mapping in Sihjhong basin, Taiwan (1.0.0), Zenodo [code], https://doi.org/10.5281/zenodo.3238689, 2019.

Moghaddam, M. H. R., Khayyam, M., Ahmadi, M., Farajzadeh, M.: Mapping susceptibility Landslide by using Weight-of-Evidence Model: A case study in Merek Valley, Iran. Journal of Applied Sciences 7(22): 3342-3355, https://doi.org/10.3923/jas.2007.3342.3355, 2007.

Neuhäuser, B., Damm, B., Terhorst, B.: GIS-based assessment of landslide susceptibility on the base of the Weights-of-Evidence model. Landslides 9(4): 511-528, https://doi.org/10.1007/s10346-011-0305-5, 2012.

Osna, T., Sezer, E.A., Akgun, A.: GeoFIS: an integrated tool for the assessment of landslide susceptibility. Comput Geosci 66:20–30, https://doi.org/10.1016/j.cageo.2013.12.016, 2014.

Panchal, S., Shrivastava, A. K.: Application of analytic hierarchy process in landslide susceptibility mapping at regional scale in GIS environment. Journal of Statistics and Management Systems 23(2):199-206, https://doi.org/10.1080/09720510.2020.1724620, 2020.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-Learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830, 2011.

Petley, D. (2012) Global patterns of loss of life from landslides. Geology 40: 927–930, https://doi.org/10.1130/G33217.1.

Petschko, H., Brenning, A., Bell, R., Goetz, J., and Glade, T.: Assessing the quality of landslide susceptibility maps – case study Lower Austria, Nat. Hazards Earth Syst. Sci., 14, 95–118, https://doi.org/10.5194/nhess-14-95-2014, 2014.

Polat, A.: An innovative, fast method for landslide susceptibility mapping using GIS-based LSAT toolbox. Environ Earth Sci 80, 217, https://doi.org/10.1007/s12665-021-09511-y, 2021.

Polemio, M., Petrucci, O.: Rainfall as a Landslide Triggering Factor: an overview of recent international research. In Landslides in Research, Theory and Practice: Proceedings of the 8th International Symposium on Landslides; Bromhead, E., Dixon, N., Ibsen, M.-L., Eds.; Thomas Telford: London, UK, pp. 1219–1226, 20202000.

Pradhan, B., Lee, S. (2010) Landslide susceptibility assessment, and factor effect analysis: back propagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environmental Modelling and Software 25: 747-759, https://doi.org/10.1016/j.envsoft.2009.10.016, 2010.

R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/, 2013

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

Rossi, M., Reichenbach, P.: LAND-SE: a software for statistically-based landslide susceptibility zonation, version 1.0. Geoscientific Model Development 9: 3533-3543, https://doi.org/10.5194/gmd-9-3533-2016, 2016.

Rossi, P. H., Guzzetti, F., Reichenbach, P., Mondini, A. C., Perruccacci, S.: Optimal landslide susceptibility zonation based on multiple forecasts. Geomorphology, 114(3),129–142, https://doi.org/10.1016/j.geomorph.2009.06.020, 2010.

Saaty, T. L.: A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology 15: 234-281, 1977.

Saaty, T. L.: The analytic hierarchy process, New York: McGraw, 1980.

Sahin, E. K., Colkesen, I., Acmali, S. S., Akgun, A., Aydinoglu, A. C.: Developing comprehensive geocomputation tools for landslide susceptibility mapping: LSM tool pack. Computers & Geosciences, 104592, https://doi.org/10.1016/j.cageo.2020.104592, 2020.

Schmidhuber, J.: Deep Learning in neural networks: An overview. Neural Networks, 61, 85-117, 2015.

Senger, K., Betlem, P., Grundvåg, S-A., Horota, R., Buckley, S., Smyrak-Sikora, A., Jochmann, M., Birchall, T., Janocha, J., Ogata, K., Kuckero, L., Johannessen, R., Lecomte, I., Cohen, S., Olaussen, S.: Teaching with digital geology in the high Arctic: opportunities and challenges. 10.5194/gc-2021-6, 2021.

Tanyaş, H., Allstadt, K. E. and van Westen, C. J.: An updated method for estimating landslide-event magnitude. Earth surface processes and landforms, 43(9), 1836–1847, https://doi.org/10.1002/esp.4359, 2018.

Teerarungsigul, S., Torizin, J., Fuchs, M., et al.: An integrative approach for regional landslide susceptibility assessment using weight of evidence method: a case study of Yom River Basin, Phrae Province, Northern Thailand. Landslides. 13(5): 1151-1165, https://doi.org/10.1007/s10346-015-0659-1, 2015.

Tensor Flow Team: Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Schuster, M., Monga, R., Moore, S., Murray, D., Olah, C., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng X.: TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org., 2015

Thiery, Y., Malet, J-P., Sterlacchini, S., Puissant, A., Maquaire, O.: Landslide susceptibility assessment by bivariate methods at large scales: Application to a complex mountainous environment. Geomorphology, 92(1-2), 38-59, ff10.1016/j.geomorph.2007.02.020, 2007.

Tian, T., Balzer, D., Wang, L., Torizin, J., Wan, L., Li, X., Chen, L., Li, A., Kuhn, D., Fuchs, M., Lege, T., Tong, B.: Landslide hazard and risk assessment Lanzhou, province Gansu, China - Project introduction and outlook. In: Mikoš, M., Tiwari, B., Yin, Y., Sassa, K. (eds) Advancing culture of living with landslides, pp. 1027–1033. WLF 2017. Springer, Cham. https://doi.org/10.1007/978-3-319-53498-5_116, 2017.

Titti, G., Sarretta, A., Lombardo, L.: CNR-IRPI-Padova/SZ: SZ plugin (v1.0). Zenodo [code], https://doi.org/10.5281/zenodo.5693351, 2021.

Torizin, J., Fuchs, M., Awan, A. A., et al.: Statistical landslide susceptibility assessment of the Mansehra and Thorgar districts, Khyber Pakhtunkhwa Province, Pakistan. Natural Hazards 89(2): 757-784, https://doi.org/10.1007/s11069-017-2992-2, 2017.

Torizin, J., Fuchs, M., Balzer, D., Kuhn, D., Arifianti, Y., Kusnadi: Methods for generation and evaluation of landslide susceptibility maps: a case study of Lombok Island, Indonesia. Proceedings of 19th Conference on Engineering Geology, Munich, pp 253–258, 2013.

Torizin, J., Fuchs, M., Kuhn, D., Balzer, D., Wang, L.: Practical Accounting for Uncertainties in Data-Driven Landslide Susceptibility Models. Examples from the Lanzhou Case Study. In: Guzzetti F., Mihalić Arbanas S., Reichenbach P., Sassa K., Bobrowsky P.T., Takara K. (eds) Understanding and Reducing Landslide Disaster Risk. WLF 2020. ICL Contribution to Landslide Disaster Risk Reduction. Springer, Cham. https://doi.org/10.1007/978-3-030-60227-7_27, 2021.

Torizin, J., Wang, L. C., Fuchs, M., et al.: Statistical landslide susceptibility assessment in a dynamic environment: A case study for Lanzhou City, Gansu Province, NW China. Journal of Mountain Science 15(6), https://doi.org/10.1007/s11629-017-4717-0, 2018.

Torizin, J.: Elimination of informational redundancy in the weight of evidence method: An application to landslide susceptibility assessment. Stochastic Environmental Research and Risk Assessment 30(2): 635-651, https://doi.org/10.1007/s00477-015-1077-6, 2016.

Torizin, J.: Landslide Susceptibility Assessment Tools for ArcGIS 10 and their Application. In: Proceedings of 34th IGC, Brisbane, 5-10 August 2012, p. 730, 2012.

Van Rossum, G., Drake, F. L.: Python 3 Reference Manual. Scotts Valley, CA: CreateSpace, 2009.

Van Westen, C. J., Castellanos Abella, E. A., & Sekhar, L. K.: Spatial data for landslide susceptibility, hazards and vulnerability assessment: an overview. Engineering Geology, 102(3-4), 112-131, https://doi.org/10.1016/j.enggeo.2008.03.010, 2008.

Van Westen, C., Van Asch, T. W. and Soeters, R.: Landslide hazard and risk zonation – why is it still so difficult? Bulletin of Engineering geologyGeology and the Environment, 65(2), 167–184, 2006.

Varnes, D. J.: Landslide hazard zonation: a review of principles and practice, Natural Hazards, 3. UNESCO. Paris. 63 pp., 1984.

Xu, Y., Goodacre, R.: On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. J. Anal. Test. 2, 249–262, https://doi.org/10.1007/s41664-018-0068-2, 2018.