ORIGINAL PAPER

Landslide development and susceptibility along the Yunling–Yanjing segment of the Lancang River using grid and slope units

Hong Wen[1] · Siyuan Zhao[2] · Yuhang Liang[3] · Sen Wang[4] · Ling Tao[1] · Jiren Xie[5]

Abstract

Selecting appropriate mapping units is vital for landslide susceptibility mapping as well as pattern investigation, given that various units-based analyses extensively control the prediction performances. This paper investigated the landslide development through the interpretation and field surveys along the Yunling–Yanjing segment of the Lancang River in southeastern Tibet, and fulfil LSM with the consideration of 15 conditioning factors. Two grid unit methods (single-point and multi-point patterns) and two slope unit methods were comparatively analyzed for model training and mapping of landslide susceptibility via machine learning algorithms. The results suggest that the landslides are preferentially distributed in an elevation range of 2000–4000 m, in a slope range of 20–40°, a local relief range of 1000–2500 m, and southwest-oriented slopes. The data extracted by the multi-point method denotes a higher representation of landslide development features. All models possessed positive prediction ability for landslide susceptibility, and the multi-point method based on grid unit performed the best with the AUC exceeding 0.9. The best-performing models indicated that zones of high and very high susceptibility mainly distributed adjacent to the mainstem and some tributaries of the Lancang River. Furthermore, the distribution of "safety islands" (the slopes less prone to landslides) along National Highway G214 was reasonably illustrated as well, which provides a hazard predictability for such an important transportation corridor along the deeply-incised valley in the Lancang River. This study demonstrates a theoretical basis for the regional disaster prevention and mitigation for human activity, and provides methodological references for landslide susceptibility evaluation in similar mountainous areas over the world.

Keywords Landslide · Landslide susceptibility assessment · Grid unit · Slope unit · Tibetan Plateau

1 Introduction

Landslide susceptibility mapping (LSM) is a measure of the spatial probability of landslides occurring in an area as determined by geoenvironmental conditions (Reichenbach et al. 2018; Merghadi et al. 2020; Shao and Xu 2022). Regional landslide susceptibility

mapping has become an important topic in the field of landslides in current context of the persistent landslide risk in hazard-prone areas in the world (Yong et al. 2022). The susceptibility mapping extracts the values of the conditioning factors of landslide and non-landslide samples according to certain statistical units, and use statistical methods or machine learning algorithms to predict the landslide susceptibility indexes of the evaluative units in the study area (Kavzoglu et al. 2019; Zhao and Chen 2020). Various parts of the LSM process are featured by uncertainty, which substantially controls the accuracy and applicability of evaluation results (Huang et al. 2022a; Yang et al. 2022; Zhao et al. 2023). Many researches has applied plenty of algorithms for hazard susceptibility modeling (Merghadi et al. 2020; Jin et al. 2022; Wen et al. 2022), and a increasing number of algorithms have been continuously proposed to train models with high accuracy (Bui et al. 2020; Chen et al. 2020; Chen and Chen 2021; Xi et al. 2022). However, selection of statistical and evaluative units introduces a large amount of indeterminacies in landslide pattern analysis and susceptibility assessment (Huang et al. 2022c). Revealing the impact of these indeterminacies and comparing preferred unit solutions can better serve landslide prevention.

In previous studies related to LSM, the main statistical and evaluative units included grid units, slope units, unique condition units, terrain units, geo-hydrological units, topographic units, and administrative units (Kavzoglu et al. 2019; Chang et al. 2023). Among them, the grid unit is one of the most adopted evaluative units, which could easily acquire data for various conditioning factors (Liao et al. 2022; Liu et al. 2023). Slope units that are divided according to the real topography and geomorphology improve the representation of regional geoenvironmental features and can properly reflect the disaster-formation of landslides (Ma et al. 2023; Rolain et al. 2023). Various optimized division approaches of slope units have been proposed, e.g., the curvature watershed method (Yu and Chen 2020) and multiscale segmentation method (Chang et al. 2023). Many efforts on comparative studies of grid and slope units have been made, and the statistical and evaluative units directly affect the performance of landslide susceptibility modeling. For this reason, the statistical unit affects the nonlinear correlation between landslide susceptibility and conditioning factors and significantly affects the modeled spatial dataset (Huang et al. 2022b; Ling et al. 2022). The prerequisite for achieving reliable susceptibility modeling is to extract representative landslide conditioning factors, which is crucial to the efficiency of landslide susceptibility mapping for extremely complex geoenvironmental regions, such as the margin areas of the Tibetan Plateau. However, the unit selection for LSM in such high-relief mountainous terrains with intense erosion remains poorly understood.

To fill the gap in knowledge about the uncertainties of the statistical and evaluative units and to explore a preferred unit solution, this study first identified 609 landslides in the Yunling–Yanjing segment along the Lancang River in southeastern Tibet via remote sensing interpretation and field validation and considered 15 conditioning factors. We established a grid unit-based sample set of multi-points (MP) within landslide polygons with a spacing of 90 m, and sampled the center grid cell within landslides as the single-point (SP) sample set; two slope unit-based sample sets of landslide polygon (PLG) were also constructed. The data extracted from those sample sets were comparatively analyzed for regularity, and three machine learning algorithms, support vector machine (SVM), random forest (RF), and deep neural network (DNN), were applied for landslide susceptibility modeling. Finally, their performances were evaluated and quantitatively compared, and the uncertainty of various methods is discussed from the perspective of landslide formation mechanisms and field investigations.

2 Study area

The Lancang River originates in the northeastern part of the Tanggula Mountains in China's Qinghai Province and flows through the Tibet Autonomous Region and Yunnan Province. It is called the Mekong River downstream of the Chinese national border, as it flows through Myanmar, Laos, Thailand, Cambodia, and Vietnam before merging into the sea. As the largest international river in Southeast Asia, its entire mainstem is 4880 km long, with a total catchment area of $\sim 810 \times 10^3$ km^2; in particular, the length of the Chinese segment is 2161 km, with a catchment area of $\sim 167 \times 10^3$ km^2.

The study area is located in the middle segment of the Lancang River from Yunling to Yanjing; it is bounded by watersheds to the east and west, with a mainstem length of 139.7 km and an area of approximately 3371 km^2. This river segment is attributed to Three Rivers region (defined by drainages of the Jinsha, Lancang and Nu Rivers) in the Hengduan Mountains of the southeastern Tibetan Plateau (Fig. 1a). The study area characterized by a typical alpine and deep canyon landscape, with large topographic relief and a narrow valley, which is a representative area with complex regional geoenvironment. The riverbed elevations are ~ 1850–2300 m, and the relative height between watershed and valley floor ranges in 3000–4000 m, with a maximum of 4686 m (Fig. 1b). The mountain peaks on both sides of the river valley in this segment
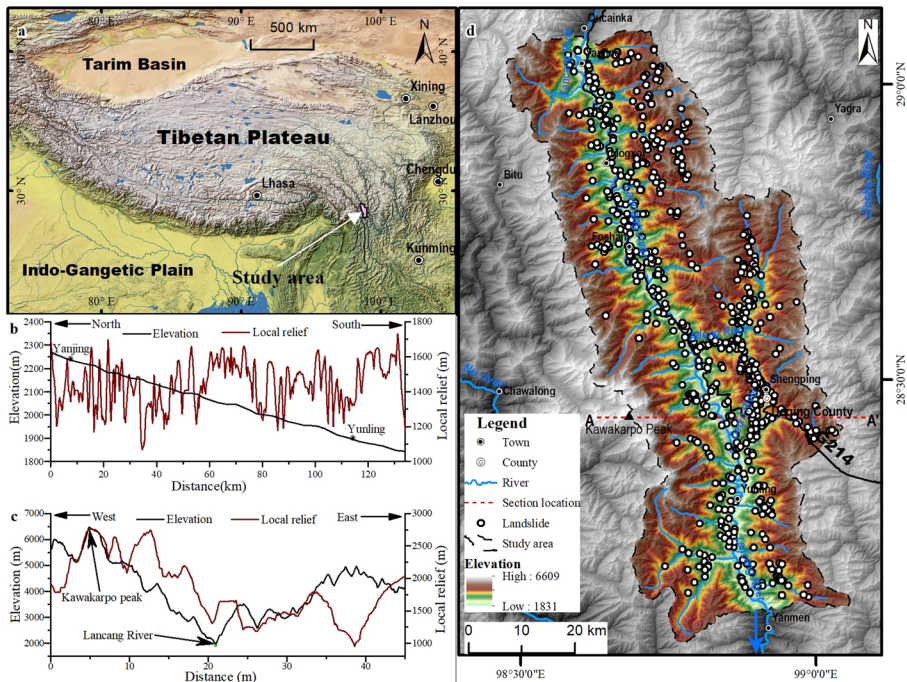

Fig. 1 Location and geomorphology of the study area. **a** Location of the study area. **b** Profiles of elevation and local relief along the mainstem from Yunling to Yanjing segment of the Lancang River. **c** Profiles of elevation and local relief along the cross-section A–A'. Local relief is defined by relative elevation between the highest and lowest points within a radius of 2.5 km for each cell. **d** Landslide inventory on a topographic map

are greater than 5500 m asl, of which the Kawagabo Peak is the highest (Fig. 1c) in an elevation of 6740 m asl. This segment is also the narrowest river valley in the entire Lancang River; especially nearby Yunling Town, the horizontal distance between the watersheds on both sides of the river are only 20–25 km. The study area is located in the Yunnan-Tibet geological tectonic convergence zone with high-frequency geohazards (Wang et al. 2023; Yan et al. 2023), and exposes rock strata from upper Paleozoic to Cenozoic, with the Triassic strata being the most widely distributed (Pan et al. 2012). The lithologies mainly contains limestone, dolomite, sandstone, slate, and phyllite. This region is characterized by a cold temperate climate of mountain monsoon, specifically with the features of dry-hot valleys (He and Zhang 2005; Li and Xiao 2020). The average annual rainfall at the Deqin meteorological station within the study area is~642 mm, with the rainy season (from May to October) accounting for 77.5% of the annual precipitation (Yuan et al. 2023). The average annual temperature is 4.7 °C, with an extreme annual maximum temperature of 25.1 °C and a minimum temperature of−27.4 °C. Engineering and human settlements in the study area are mainly distributed along the river, with some villages located on some topographic platforms on the hillside.

3 Materials and methods

3.1 Landslide inventory and statistical units

The landslide inventory was constructed mainly based on remote sensing interpretation and field surveys (Fig. 1d). The landslide polygons in the study area were identified through preliminary interpretation of Google Earth images (Fig. 2a and b), and source and depositional areas were key markers for recognizing landslides by remote sensing (Fig. 2c and d). A field survey was then conducted for more than two months to verify landslides and slope deformations in comparison with the remote sensing results (Fig. 2e and f). Finally, 609 landslides were compiled, as shown in Fig. 1d.

Elevation, slope, aspect, local relief, roughness, and other raster layers were further produced through the 12.5 m digital surface model (DSM) provided by the Advanced Land Observing Satellite (ALOS). In addition, rivers, roads, faults, lithology, and other layers overlapped in GIS were used to analyze the spatial preference of landslide development. For data extraction, the SP method based on grid unit was firstly used to extract the data at center point or representative point within the landslide polygon (Fig. 3a). The MP method was proposed to establish multiple points with a certain spacing within landslide polygon, in which we adopted a spacing of 90 m and generated 12,185 points (Fig. 3b). Then the sampling slope units of the PLG method extract the data directly from the landslide polygon by taking the general average value (Fig. 3c). Hence, the models trained with data extracted by the SP and MP methods were evaluated with grid units, and those with data extracted by the PLG method were evaluated with slope units. Herein, two PLG methods, respectively with hydrologic (PLGH) and multiscale segmentation (PLGMS) patterns, were used to produce the evaluative slope units. The PLGH method utilized the positive and negative topography to extract valley lines and ridgelines, and divided slope units according to the reverse catchments (Li et al. 2021). A total of 8008 slope units were generated by the PLGH method in the study area. The PLGMS method selected the aspect and hillshade raster layers as the basic input images to divide the slope units (Wang and Niu
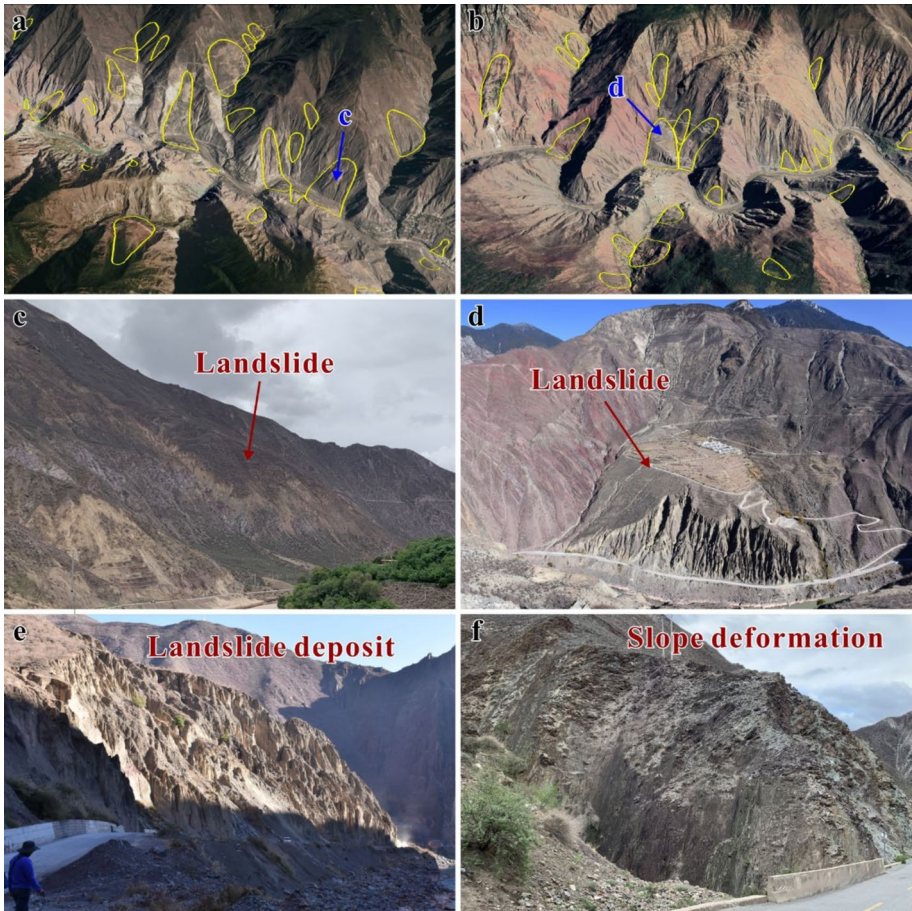
Fig. 2 Interpretation and compilation of landslide. **a–b** Remote sensing interpretation on satellite images. **c–f** Field verification in surveys
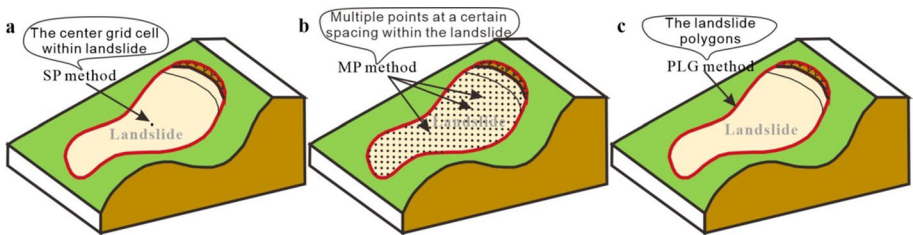

Fig. 3 Schematics of the sampling methods. **a** SP method. **b** MP method. **c** PLG method

2010), and a total of 10698 units were classified by this method. It is noted that the two PLG methods used the same sampling method shown in Fig. 3c.

3.2 Landslide conditioning factors

Landslide formation is affected by many factors, mainly including topography, stratigraphic lithology, climatic conditions, and human engineering activities (Hungr et al. 2018). The selection of landslide conditioning factors should take into account the representativeness of landslide-formation pattern, and their quantitative expression is essential (Costanzo et al. 2012; Merghadi et al. 2020). Therefore, the following 15 quantifiable extraction factors were selected: elevation, slope, aspect, standard curvature, profile curvature, plan curvature, roughness, local relief, distance to river, distance from road, distance to fault, lithology, rainfall, land cover, and normalized difference vegetation index (NDVI). The Peak Ground Acceleration (PGA) in the study area involves only 2 values that do not contribute well to the characterization of landslide developmental differentiation, which wasn't adopted. The data sources for these factors are detailed in Table 1. The selection of conditioning factors also needs to take into account the multicollinearity among the multiple factors. Multicollinearity refers to the relationship between conditioning factors due to the existence of precise correlation or high correlation, which makes the results less objective and accurate (Thompson et al. 2017). In this paper, the variance inflation factor (VIF) was used to test the multicollinearity between conditioning factors and to select more reasonable conditioning factors. The formula of the VIF is as follows (Thompson et al. 2017):

$$VIF = \frac{1}{1 - R_i^2} \tag{1}$$

where R_i is the negative correlation coefficient of the independent variables for regression analysis of the remaining independent variables. The larger the VIF is, the greater the possibility of multicollinearity between the independent variables. Generally, there is serious multicollinearity between the conditioning factors when the VIF exceeds 10. It is acceptable when the VIF is less than 10 (Thompson et al. 2017; Hearn and Hart 2019).

3.3 Landslide susceptibility modeling

A total of three machine learning algorithms, support vector machine (SVM), random forest (RF) and deep neural network (DNN), are applied in LSM. The SVM is a supervised machine learning algorithm for binary classification based on the structural risk minimization principle (Merghadi et al. 2020). The input variables are first transformed to x high-dimensional feature space by a kernel function, then the best hyperplane that can maximize the category spacing is found, and finally, the linear classification of the output variables is achieved (Chen et al. 2020). The RF is an efficient ensemble classifier that has been used to solve many nonlinear problems. It utilizes a randomly selected subset of variables to construct multiple decision trees for landslide susceptibility prediction (Chen et al. 2017; Dou et al. 2019). The DNN, as an improved algorithm of artificial neural networks (ANNs), shares a similar network structure with ANNs and can handle nonlinear problems well (Xi et al. 2022). Therefore, DNNs can more adequately map complex data and further explore the relationships between landslide susceptibility and conditioning factors (Bui et al. 2020). These algorithms are standard and can be programmatically invoked from the "Scikit-learn" library (Pedregosa et al. 2011) via the Jupyter Notebook with Python (3.6.5).

In this study, the ROC (receiver operating characteristic) curve is used to quantitatively analyze the modeling accuracy. The area under the ROC curve (AUC) is a specific

Table 1 Conditioning factors and data sources

| Conditioning factor	Data source	Spatial resolution
Elevation, slope, aspect, standard curvature, profile curvature, plan curvature, roughness, local relief	ALOS DEM	12.5 m
Distance to river, Distance to road	1: 250 000 Basic Geographic Information from National Catalogue Service for Geographic Information of China https://www.webmap.cn/	Vector
Distance to fault, Lithology	1: 200 000 Regional geological map http://dcc.ngac.org.cn/	Resample as 10 m
Rainfall	National Meteorological Science Data Center of China http://data.cma.cn/	Interpolated as 10 m
Land cover	European Space Agency https://viewer.esa-worldcover.org/worldcover/	10 m
NDVI	Sentinel 2 images from European Space Agency https://scihub.copernicus.eu/	10 m
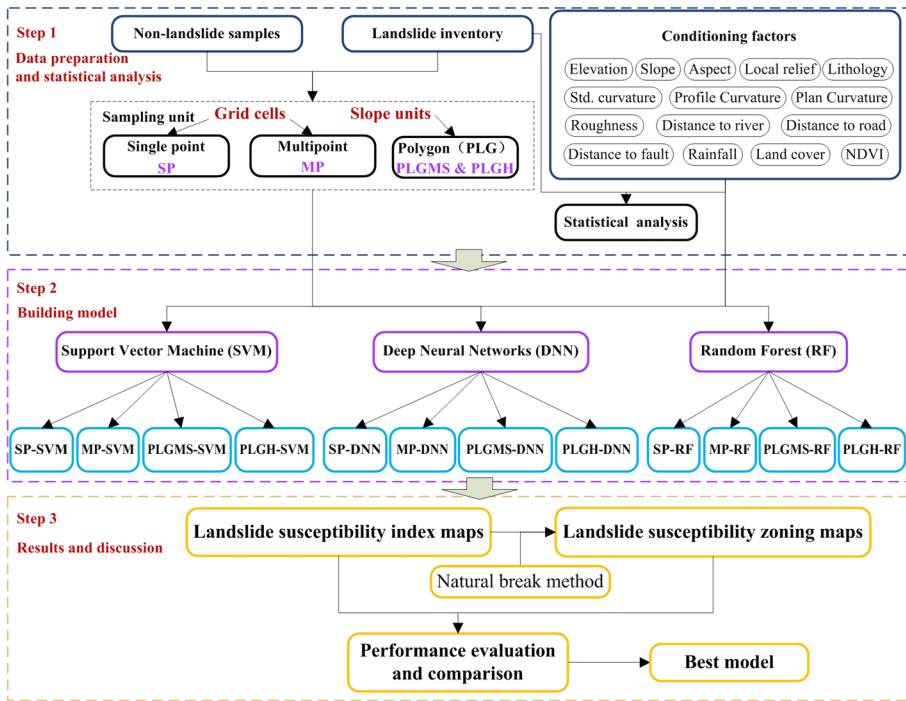
Fig. 4 Flowchart of the methodology for statistical analysis and susceptibility prediction of landslide

quantitative index used to test the performance of the model. The closer the value is to 1, the higher the accuracy of the model. The formula for calculating the AUC is shown below (Mandrekar 2010):

$$\text{AUC} = \frac{\sum_{i=1}^{n_0} r_i - n_0(n_0+1)/2}{n_0 n_1} \qquad (2)$$

where n_0 is the number of negative samples, n_1 is the number of positive samples, and r_i is the positional order of the i'th negative sample in the whole validation sample.

The methodological flow of the overall study is shown in Fig. 4. The data preparation and statistics in this figure were carried out on the GIS platform. The machine learning algorithm and landslide susceptibility modeling are implemented by coding in Jupyter Notebook. In addition, the uncertainties of the various methods are discussed with respects to field investigation and perspective of the landslide-formation mechanism.

4 Results

4.1 Statistical analysis of conditioning factors

On the basis of the landslide inventory and conditioning factors, we collected 609 sets of data for the SP and PLG methods, and generated 12,185 samples for the MP method. Their

descriptive statistics are shown in Table 2. Among the 15 conditioning factors, lithology and land cover are discrete data, and no descriptive statistics were performed. The statistical results illustrated that elevation, local relief, and NDVI between SP and PLG data were relatively similar, while the patterns of other topographic factors were not outstanding, and standard curvature, plan curvature, and profile curvature had positive and negative values. With regards to distances to rivers, faults and roads, the statistical results of the SP and MP methods were closer, while the PLG method collected data through buffer distances, which were a little bit different from the real distances. The rainfall layer was interpolated from the collected meteorological data, and its value variation were not significant. The coefficient of variation after normalization shows that the distribution of landslides was more dependent on the factors of standard curvature, profile curvature, plan curvature, slope, local relief, and NDVI. Through the multicollinearity test, the VIF values of each conditioning factor extracted by the SP and PLG methods were all less than 10, which satisfied the requirements of landslide susceptibility modeling. On the other hand, the VIF values of standard, profile and plan curvatures were much greater than 10, which had dropped below 2 if we eliminated standard curvature (as shown in column "MP-2" of Table 2). Thus, standard curvature was not involved in the MP-relevant models in the subsequent landslide susceptibility modeling.

Based on the extracted data, statistical mapping analysis was carried out, as shown in Fig. 5. The elevations of the total study area range from 1480 to 6610 m, and the elevations ranging from 3000 to 5000 account for over 55% for relative frequency. The elevations of landslide area extracted by the three methods mainly distributed in the range of 2000–4000 m, and each interval of 500 m in this range makes up more than 10%. The highest frequency in the intervals of elevation is 35.2% in the 2500–3000 interval (Fig. 5a). The slopes in the total study area range from 0 to 85.2°, and the slopes of landslide area are mostly distributed between 20° and 40°. Specifically, the landslide slopes extracted by the PLG method have a tendency in the 30–35° interval, with a relative frequency of 39.9%, while the landslide slopes of the SP and MP methods distributed in the same interval with their frequencies of 25.2% and 23.1%, respectively (Fig. 5b). In general, the slope orientations in the study area are uniformly distributed in all aspects; the slope orientations extracted by the three methods with frequencies of over 10% are oriented to the east, southeast, south, southwest, and west, and demonstrates a highest preference on the southwest direction (Fig. 5c). However, the PLG results illustrates no frequency on the north direction (337.5–360° and 0–22.5° intervals), and this is related to averaging in the statistical process, which is unreasonable. There is a notable difference in the frequency distribution of standard, profile and plan curvatures extracted by the three methods, and the roughness statistics exhibit a variation in the comparison as well (Fig. 5d–g). The local relief in the study area ranges from 575 to 3170 m, and all the intervals of 500 m make up more than 10% between 1000 and 2500 m, and the 1500–2000 m interval accounts for the highest frequency. Similarly, the frequency distributions of local relief within landslide areas have a tendency in the range of 1000–2500 m, with a highest proneness in the 1500–2000 m interval (Fig. 5h). In terms of lithology, rainfall, distance to faults, rivers and roads, the landslide distribution pattern is generally consistent with the total study area (Fig. 5i–m). All land cover types are present in the study area, among which cropland (C) and grassland (GL) account for the largest area; the land cover types within landslide areas extracted by the three methods are also prone to cropland and grassland. Among the values extracted by the PLG method, forest (F) and shrubland (SL) occupy larger area, while the overall study area has small areas of forest (F) and shrubland (SL) (Fig. 5n), which is also related to the averaging in the statistical process. The NDVI extracted by the three methods exhibited

Table 2 Descriptive statistics and VIF test of the conditioning factors

| Conditioning factor	Mean			Standard deviation			Minimum			Maximum			Coefficient of variation after normalization			VIF			
	SP	MP	PLG	SP	MP	PLG	SP	MP	PLG	SP	MP	PLG	SP	MP	PLG	SP	MP	MP-2	PLG
Elevation	3036	2909	3032	635	570	634	1960	1877	1947	4660	4720	4664	0.407	0.398	0.408	3.105	3.076	3.076	2.531
Slope	30	29	30	8	9	5	5	0	10	52	69	41	0.267	0.299	0.159	6.517	5.606	5.603	7.696
Aspect	183	182	182	89	90	77	2	−1	35	359	359	331	0.482	0.492	0.424	1.069	1.034	1.034	1.080
Std. curvature	−0.008	−0.104	−0.183	1.352	1.507	0.215	−5.128	−13.46	−2.194	7.050	11.54	0.465	0.049	0.055	0.008	1.256	1789.785	/	1.290
Profile curvature	0.014	0.060	0.119	0.838	0.895	0.164	−6.658	−7.685	−0.329	3.870	8.659	1.855	0.045	0.048	0.009	1.256	653.816	1.297	1.227
Plan curvature	0.006	−0.043	−0.063	0.739	0.849	0.127	−2.141	−7.036	−0.796	2.210	6.498	0.440	0.083	0.096	0.014	6.611	561.232	1.296	7.962
Roughness	1.178	1.166	1.175	0.097	0.109	0.058	1.004	1.0	1.021	1.613	2.751	1.402	0.547	0.654	0.330	1.469	5.592	5.590	1.407
Local relief	1589	1639	1590	252	245	250	719	720	726	2319	2378	2291	0.248	0.230	0.246	1.607	1.412	1.412	1.401
Distance to river	1015	1010	1305	1011	945	1080	16	0	500	5934	6087	6050	0.996	0.936	0.828	1.361	1.580	1.580	1.160
Distance to fault	945	935	1236	834	826	968	4	0	500	5421	5707	10,000	0.883	0.883	0.783	1.148	1.388	1.388	1.118
Lithology	Discontinuous data												0.614	0.756	0.602	1.249	1.169	1.169	1.178
Rainfall	684	691	685	146	145	146	515	515	515	1143	1146	1143	0.510	0.495	0.509	1.963	1.245	1.245	1.692
Distance to road	1732	1476	2714	2060	1719	3988	1	0	500	11,077	11,106	15,000	1.190	1.165	1.469	2.539	1.986	1.986	4.218
Land cover	Discontinuous data												0.955	0.924	0.786	2.522	2.315	2.315	4.053
NDVI	0.241	0.232	0.244	0.117	0.123	0.108	0.023	−0.118	0.026	0.551	0.604	0.517	0.215	0.229	0.197	3.105	2.324	2.323	2.531
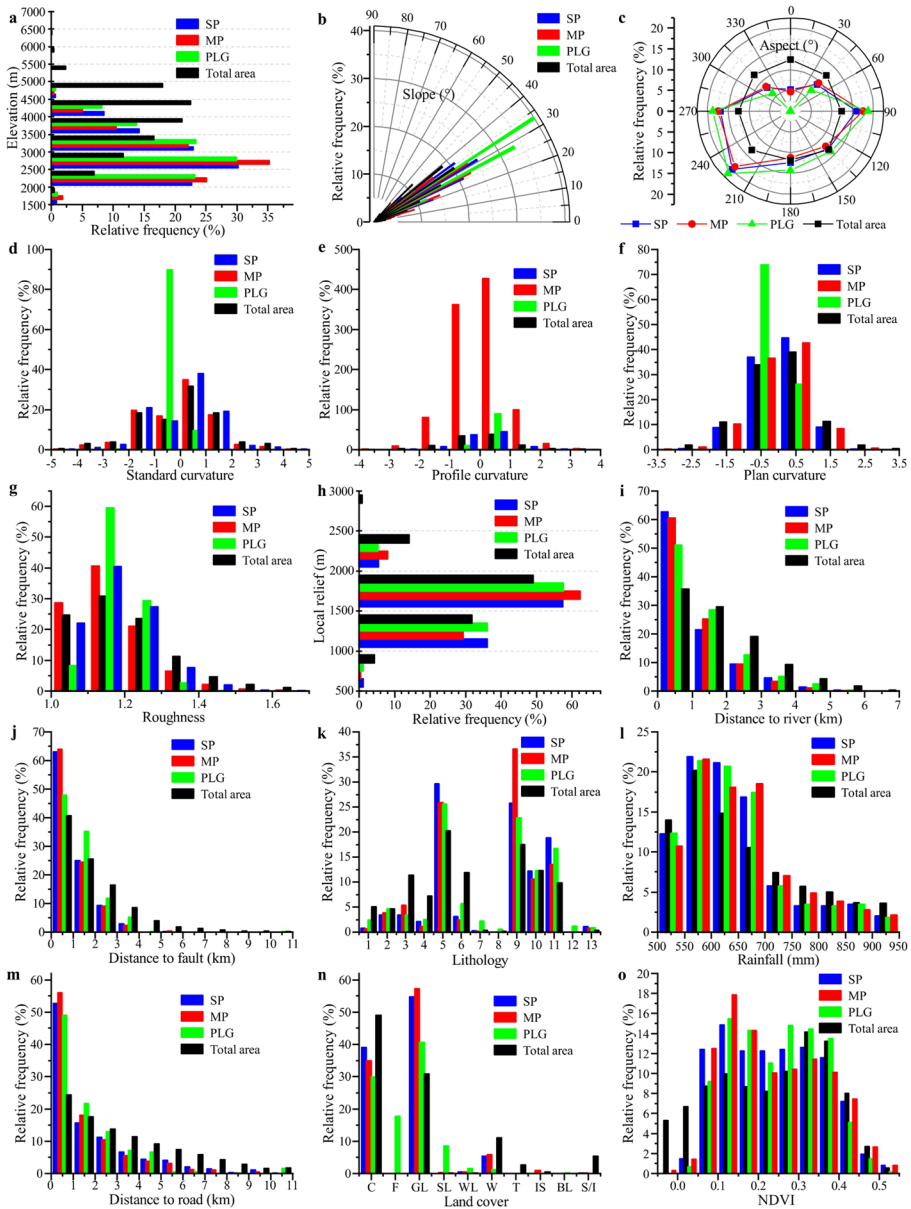
Fig. 5 Landslide frequency of various conditioning factors. **a–j** Plots of landslide frequency versus elevation, slope, slope aspect, standard curvature, profile curvature, plan curvature, roughness, local relief, distance to river, and distance to fault. **k** Plot of landslide frequency versus lithology. The rock types include: Proterozoic schist, quartzite and limestone (1), Silurian shale, siltstone and limestone (2), Carboniferous metasandstone, slate and phyllite (3), Permian sandstone, siltstone, limestone and shale (4), Triassic intermediate-acidic volcanic rock (5), Triassic sandstone, mudstone and limestone (6), Jurassic sandstone, mudstone and siltstone (7), Cretaceous sandstone, conglomerate and mudstone (8), Paleogene sandstone, conglomerate and mudstone (9), Triassic diabase and gabbro (10), Triassic granodiorite (11), Jurassic biotite granite (12), and serpentinite (13). **l–m** Plots of landslide frequency versus rainfall and distance to road. **n** Plot of landslide frequency versus land cover. The types include: cropland (C), forest (F), grassland (GL), shrubland (SL), wetland (WL), water (W), tundra (T), impervious surface (IS), bareland (BL), and snow/ice (S/I). **o** Plots of landslide frequency versus NDVI
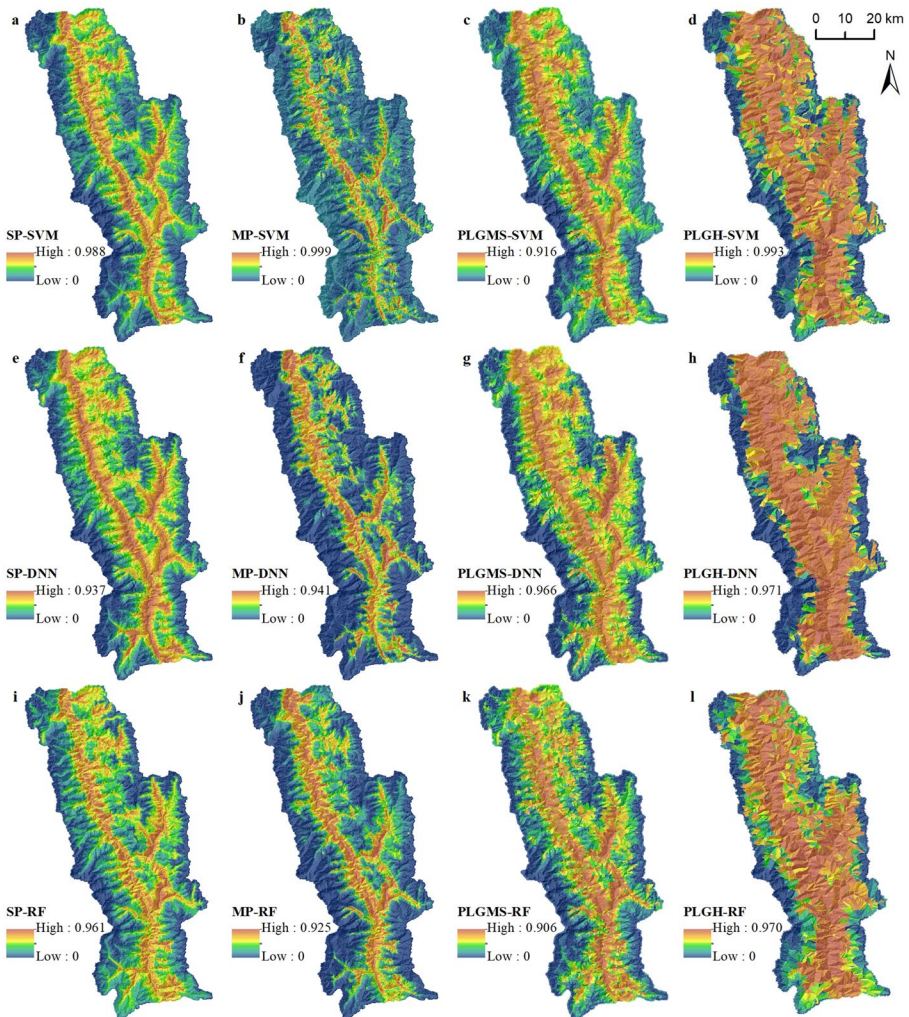
Fig. 6 Landslide susceptibility maps produced by various models

relatively concentrated values compared to the background values of the total area. The concentrated areas of MP and SP showed slightly lower values of NDVI than those in PLG (Fig. 5o). In general, the data extracted by the PLG method are distorted for the factors of discontinuous data and aspect due to averaging in the statistical process; the data extracted by the MP method can better represent the characteristics of landslide development if discontinuous factors are considered.

4.2 Landslide susceptibility assessment

Landslide susceptibility modeling and evaluation were performed in Jupyter Notebook with Python (3.6.5). The amount ratio of training set to validation set in the modeling was defined

as 8:2, and the parameters involved were optimized using the grid search technique. After the computation, the classification tendency scores were imported into GIS platform, and were exported as a raster, i.e., the landslide susceptibility index maps (Fig. 6). All the SVM (Fig. 6a–d), DNN (Fig. 6e–h) and RF (Fig. 6i–l) models showed positive predictive ability for landslide susceptibility. Overall, the susceptibility index values are higher on the bank slopes on both sides of the mainstem and some tributaries adjacent to the river base level, and the susceptibility are lower near the mountain ranges at higher elevations. The spatial distribution of susceptibility indices varied across the results of different models. Among the results based on grid units, the results of SP are relatively conservative, with high susceptibility ratings for all the bank slopes nearby the mainstem of the Lancang River (Fig. 6a, e and i), while some portion of areas with low susceptibility distributed along the mainstem in the results of the MP method (Fig. 6b, f and j). In the results based on slope units, the susceptibility index values are relatively higher along the mainstem of the Lancang River. Among them, the slope units divided by the PLGH method indicate larger areas of high landslide proneness compared to the PLGMS method (Fig. 6c, d, g, h, k and l), and most of the study area is covered by high susceptibility index values in the PLGH results.

5 Discussion

5.1 Comparison of the advantage between grid and slope units-based methods

The AUC results of the training samples in the modeling and validation datasets are shown in Fig. 7. As for statistical metrics, the validation uses the same dataset to sample all results and calculate the AUC values. In general, an AUC greater than 0.5 indicates that the prediction is meaningful, and an AUC greater than 0.8 is considered excellent (Mandrekar 2010). Overall, the models have different predictive abilities. The MP method performs the highest accuracy, and the AUC values of all results exceed 0.9. Specifically, the MP-SVM model achieves an AUC of 0.95 on the validation dataset. The PLGH method processed by the three algorithms fulfils high accuracy through parameter optimization in training with the AUCs greater than 0.9. However, the AUCs calculated from the validation dataset hover around 0.7, indicating lower accuracy. The results of the SP method and the PLGMS method present moderate prediction accuracy. The SP method employed here extracts data
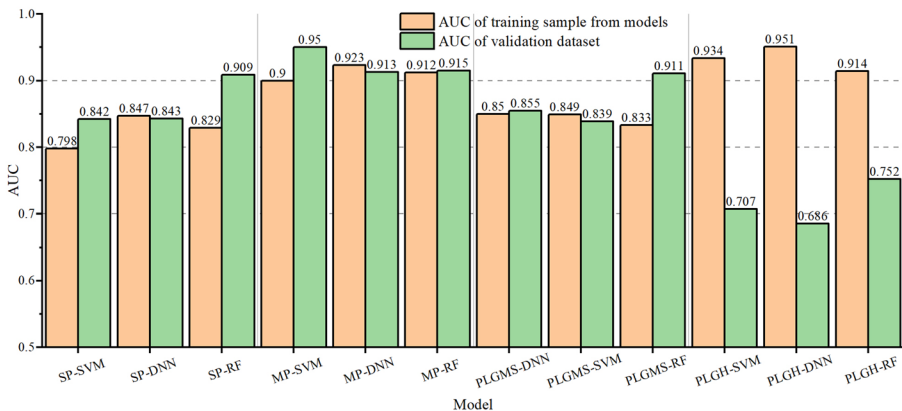

Fig. 7 The AUCs of the modeling and validation datasets

solely from the center point within the landslide polygon, and some of the points could be situated on a gentle surface of the landslide body in some cases, which inadequately characterize landslide development (Dou et al. 2020). In consideration of the reasons for the subpar performance of the PLGH method, the averaging in the statistical process can lead to distortion in the extracted data, affected by data discontinuities. Although good accuracy can be accomplished during modeling, overfitting can be seen in the results, and the predictions are barely satisfactory. The reason that the PLGMS method outperforms the PLGH method on AUC could be attributed to that the PLGMS method delineates a larger number of slope units within the study area, portraying more closely to the actual landslide polygons.

In combination with the verification in field investigation, the results of the SP, PLGH and PLGMS methods imply that all bank slopes along the mainstem of the Lancang River have high susceptibility indices (Fig. 8a, b and c), and there is no "safety island" (a slope that is less prone to landslides), which does not correspond to the actual situation. A road (e.g., National Highway G214 in the study area) constructed aside a deeply-incised river valley is subjected to landslide disasters (Zhao et al. 2022); however, not all hillslope regions are prone to landslides. On both sides of National Highway G214, there are relatively stable bedrock slopes with no failures, and a portion of slopes with low susceptibility indices distribute along the mainstem of the Lancang River in the MP results (Fig. 8d). Such situation further suggests that LSM job is of extensive significance for accurately recognizing the safe sites from disasters for human activities. Slopes with high susceptibility
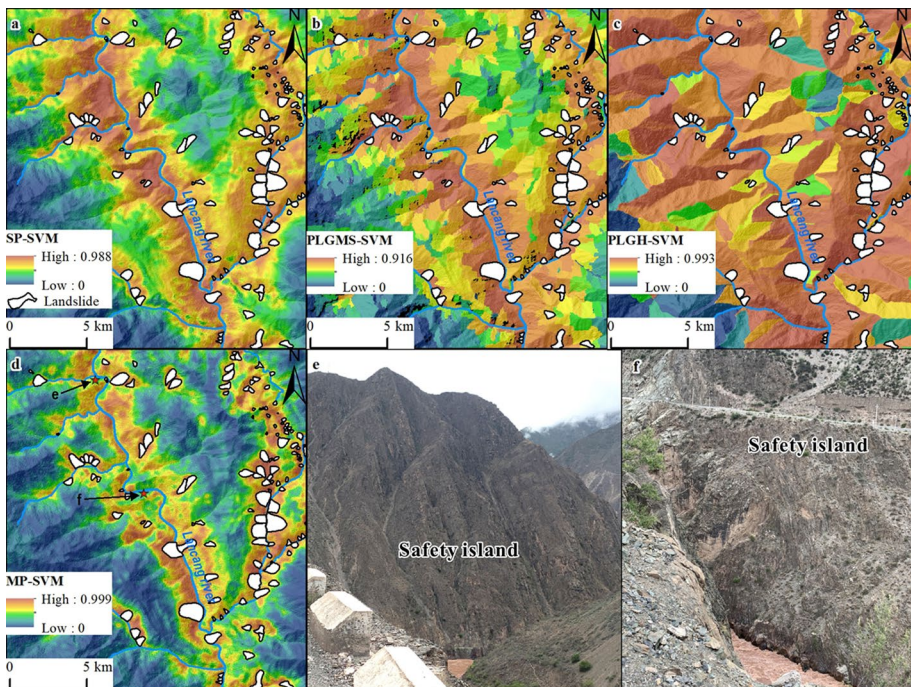

Fig. 8 Details and "safety islands" of the landslide susceptibility mapping. **a–d** Partial enlarged details of landslide susceptibility maps produced by the SP-SVM, PLGMS-SVM, PLGH-SVM, and MP-SVM (see Fig. 10 for the locations). **e–f** Field photographs of the "safety islands"

are commonly coincident with landslide areas, and slopes with low susceptibility indicate the distribution of "safety islands" (Fig. 8e and f); such results could highly benefit prevention and mitigation of landslide disaster (Reichenbach et al. 2018; Merghadi et al. 2020). Notably, the slopes that the geological structures strike normal or highly oblique to the river trend are typical "safety islands", which reasonably possess stabler conditions and low susceptibility in the MP results (Fig. 8e and f). Previous studies conservatively revealed the high-susceptibility areas in the form of bands or slices, and locating the "safety islands" according to field verifications should be paid more attention for practical disaster assessment. In summary, the MP sampling method based on grid units performs the best with higher accuracy for the reason that it appropriately characterizes landslide development, which is also accordant better with the understanding of landslide mechanism based on field investigation.

5.2 Landslide-prone areas

Each susceptibility result was categorized into five classes via the natural break classification: very low susceptibility (VLS), low susceptibility (LS), moderate susceptibility (MS), high susceptibility (HS), and very high susceptibility (VHL). The area percentage of each class for every result in the whole study area are shown in Fig. 9. Among them, the PLGH method obtains the largest percentage of very high-susceptibility zones, and the percentage of very high-susceptibility zones in the PLGH-DNN results attains 51.9%, which is deemed conservative for precisely discriminating the relatively stable hillslope like "safety island". The percentages of very high-susceptibility zones obtained by the MP method range from 8.3% to 13.6%, and very high-susceptibility zones of other models are in the range of 14.5%–51.9% (Chen et al. 2018a, b; Gao et al. 2021). This further suggests that the landslide susceptibility zoning results acquired by the MP method are more proper for an accurate disaster assessment.

The landslide susceptibility zoning based on the best-performing MP-SVM model is shown in Fig. 10a. The VLS, LS, MS, HS, and VHS zones account for 52.5%, 18.1%, 11.6%, 9.7% and 8.3% of the total area, respectively. The VHS and HS zones are mainly located on both sides of the mainstem and some tributaries of the Lancang River. The elevation difference of the slopes along the Lancang River exceeds 2500 m, providing favorable topographic conditions for landsliding. The region has undergone intense tectonic uplift and river erosion to form the present v-shaped valley (Pan et al. 2012). The formation
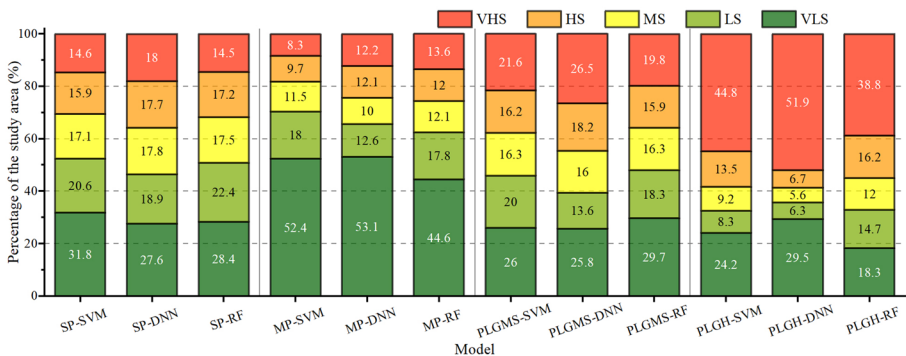

Fig. 9 Landslide susceptibility classes of the predictive models in percent

of terraces with large elevation difference and the development of hanging valleys and slope breaks record the distinctive phases of tectonic geomorphology in this area, and the continuous downcutting by the river has resulted in a wide distribution of large landslides (Zhao et al. 2023). The gravitational deformation of slopes in the area is still significant (Fig. 2f). The landslides that developed in the Lancang River valley are a serious engineering problem for National Highway G214, and a large number of engineering measures have been taken to mitigate and prevent these issues (Wang et al. 2023). There are still some relatively stable slopes along the highway, which are also referred to as "safety islands", which distribution could be clearly seen in the partially enlarged details of Fig. 10b and c. In addition, the relationships between landslide susceptibility zones and the factors of elevation, slope, and aspect (Fig. 10d, e and f) signify that the VHS and HS zones are mainly
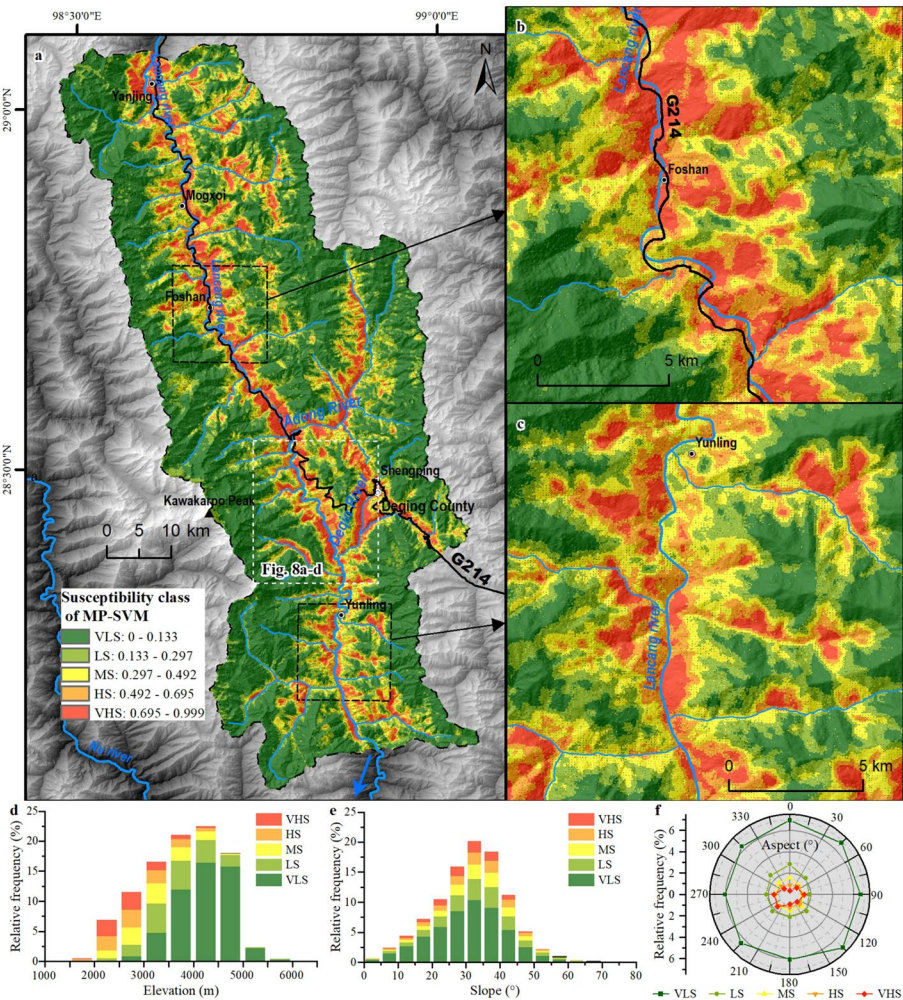

Fig. 10 Landslide susceptibility zoning map and correlation analysis based on MP-SVM. **a** Landslide susceptibility zoning map based on MP-SVM. **b–c** Partial enlarged details of landslide susceptibility zoning map based on MP-SVM. **d–f** Relationships between landslide susceptibility zones and slope, elevation, and slope orientation

distributed within an elevation range of 2000–3500 m and a slope range of 25–40°, and concentrate on the westward, southwestward and eastward orientations of hillslopes. This is similar to the statistical analysis of landslides distribution upon the hazard-forming factors (Fig. 5). The development of landslides in the study area is primarily controlled by to river downcutting and regional structural factors. To sum up, tectonic uplift and river incision macroscopically shape the fundamental topographic relief pattern and slope structure features in the study area (Pan et al. 2012). Lithology and fault collectively control the strength of slope materials, and continuous river undercutting or earthquake provide the triggers to slope destabilization (Zhao et al. 2019; He et al. 2023).

6 Conclusions

Through remote sensing interpretation and field investigation, 609 landslides were identified in the Yunling–Yanjing segment along the Lancang River. The landslides are preferentially distributed in an elevation range of 2000–4000 m, a slope range of 20–40°, a local relief range of 1000–2500 m, and demonstrate a tendency to be located on southwest-oriented slopes. The data extracted by the SP, MP, and PLG methods are slightly different. The PLG-related data are distorted for factors with discontinuous data, while the data extracted by the MP method can better depict the characteristics of landslide development.

All models in this study exhibit positive predictive ability for landslide susceptibility. Specifically, the spatial distribution of the susceptibility index varies among the models. The PLGH method in combination with the three algorithms (SVM, DNN and RF) was utilized to achieve high accuracy through parameter optimization in the training process, and all the AUCs attain values greater than 0.9. But the AUCs measured by the validation dataset are all approximately 0.7, presenting poor accuracy. In comparison, the results derived from the SP and PLGMS methods present moderate prediction accuracy. The MP sampling method based on grid units has the best performance, with an AUC exceeding 0.9. The sampling method that obtains more data samples can better portray the formation conditions of landslides, and the higher-accuracy prediction results are quite accordant with the understanding of landslide proneness acquired from the field investigation.

For the landslide susceptibility zoning based on the best-performing MP-SVM model, the VLS, LS, MS, HS, and VHS zones account for 52.5%, 18.1%, 11.6%, 9.7%, and 8.3% of the total area, respectively. The VHS and HS zones are mainly located on both sides of the mainstem and some tributaries of the Lancang River. The zoning results also indicate the distribution of "safety islands" along National Highway G214, an important transportation corridor located at the bottom of the Lancang River valley. The results of this study offer a rationale for hazard prevention and mitigation, and provide a reference for the safe site selection of important engineering projects.

Declarations

References

Bui DT, Tsangaratos P, Nguyen V-T, Liem NV, Trinh PT (2020) Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. Catena 188:104426. https://doi.org/10.1016/j.catena.2019.104426

Chang Z, Catani F, Huang F, Liu G, Meena SR, Huang J, Zhou C (2023) Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors. J Rock Mech Geotech Eng 15:1127–1143. https://doi.org/10.1016/j.jrmge.2022.07.009

Chen X, Chen W (2021) GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods. Catena 196:104833. https://doi.org/10.1016/j.catena.2020.104833

Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena 151:147–160. https://doi.org/10.1016/j.catena.2016.11.032

Chen W, Shahabi H, Shirzadi A, Hong H, Akgun A, Tian Y, Liu J, Zhu AX, Li S (2018a) Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. Bull Eng Geol Environ 78:4397–4419. https://doi.org/10.1007/s10064-018-1401-8

Chen W, Shahabi H, Zhang S, Khosravi K, Shirzadi A, Chapi K, Pham B, Zhang T, Zhang L, Chai H, Ma J, Chen Y, Wang X, Li R, Ahmad B (2018b) Landslide susceptibility modeling based on GIS and novel bagging-based kernel logistic regression. Appl Sci 8:2540. https://doi.org/10.3390/app8122540

Chen W, Chen Y, Tsangaratos P, Ilia I, Wang X (2020) Combining evolutionary algorithms and machine learning models in landslide susceptibility assessments. Remote Sens 12:3854. https://doi.org/10.3390/rs12233854

Costanzo D, Rotigliano E, Irigaray C, Jiménez-Perálvarez JD, Chacón J (2012) Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: application to the river Beiro basin (Spain). Nat Hazard 12:327–340. https://doi.org/10.5194/nhess-12-327-2012

Dou J, Yunus AP, Tien Bui D, Merghadi A, Sahana M, Zhu Z, Chen CW, Khosravi K, Yang Y, Pham BT (2019) Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Sci Total Environ 662:332–346. https://doi.org/10.1016/j.scitotenv.2019.01.221

Dou J, Yunus AP, Merghadi A, Shirzadi A, Nguyen H, Hussain Y, Avtar R, Chen Y, Pham BT, Yamagishi H (2020) Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. Sci Total Environ 720:137320. https://doi.org/10.1016/j.scitotenv.2020.137320

Gao Z, Ding M, Huang T, Liu X, Hao Z, Hu X, Xi C (2021) Landslide risk assessment of high-mountain settlements using Gaussian process classification combined with improved weight-based generalized objective function. Int J Disaster Risk Reduct 67:102662. https://doi.org/10.1016/j.ijdrr.2021.102662

He Y, Zhang Y (2005) Climate change from 1960 to 2000 in the Lancang River Valley, China. Mt Res Dev 25:341–348. https://doi.org/10.1659/0276-4741(2005)025[0341:CCFTIT]2.0.CO;2

He K, Xi C, Liu B, Hu X, Luo G, Ma G, Zhou R (2023) MPM-based mechanism and runout analysis of a compound reactivated landslide. Comput Geotech. https://doi.org/10.1016/j.compgeo.2023.105455

Hearn GJ, Hart AB (2019) Landslide susceptibility mapping: a practitioner's view. Bull Eng Geol Environ 78:5811–5826. https://doi.org/10.1007/s10064-019-01506-1

Huang F, Pan L, Fan X, Jiang S-H, Huang J, Zhou C (2022a) The uncertainty of landslide susceptibility prediction modeling: suitability of linear conditioning factors. Bull Eng Geol Environ 81:182. https://doi.org/10.1007/s10064-022-02672-5

Huang F, Tao S, Li D, Lian Z, Catani F, Huang J, Li K, Zhang C (2022b) Landslide susceptibility prediction considering neighborhood characteristics of landslide spatial datasets and hydrological slope units using remote sensing and GIS technologies. Remote Sens 14:4436. https://doi.org/10.3390/rs14184436

Huang F, Yan J, Fan X, Yao C, Huang J, Chen W, Hong H (2022c) Uncertainty pattern in landslide susceptibility prediction modelling: effects of different landslide boundaries and spatial shape expressions. Geosci Front 13:101317. https://doi.org/10.1016/j.gsf.2021.101317

Hungr O, Clague J, Morgenstern N, VanDine D, Stadel D (2018) A review of landslide risk acceptability practices in various countries. In: Thiebes E, Tomelleri A, Mejia-Aguilar M et al (eds) Landslides and engineered slopes. Experience, theory and practice. CRC Press, Napoli, Italy, pp 1121–1128

Jin T, Hu X, Liu B, Xi C, He K, Cao X, Luo G, Han M, Ma G, Yang Y, Wang Y (2022) Susceptibility prediction of post-fire debris flows in Xichang, China, using a logistic regression model from a spatiotemporal perspective. Remote Sens 14:1306. https://doi.org/10.3390/rs14061306

Kavzoglu T, Colkesen I, Sahin EK (2019) Machine learning techniques in landslide susceptibility mapping: a survey and a case study. In: Pradhan SP, Vishal V, Singh TN (eds) Landslides: theory, practice and modelling. Springer, Cham, pp 283–301

Li Z, Xiao Z (2020) Analyze on the contribution of the moisture sources to the precipitation over mid-low Lancang River nearby region and its variability in the beginning of wet season. Theoret Appl Climatol 141:775–789

Li Y, He J, Chen F, Han Z, Wang W, Chen G, Huang J (2021) Generation of homogeneous slope units using a novel object-oriented multi-resolution segmentation method. Water 13:3422. https://doi.org/10.3390/w13233422

Liao M, Wen H, Yang L (2022) Identifying the essential conditioning factors of landslide susceptibility models under different grid resolutions using hybrid machine learning: a case of Wushan and Wuxi counties, China. Catena 217:106428. https://doi.org/10.1016/j.catena.2022.106428

Ling S, Zhao S, Huang J, Zhang X (2022) Landslide susceptibility assessment using statistical and machine learning techniques: a case study in the upper reaches of the Minjiang River, southwestern China. Front Earth Sci 10:986172. https://doi.org/10.3389/feart.2022.986172

Liu S, Wang L, Zhang W, He Y, Pijush S (2023) A comprehensive review of machine learning-based methods in landslide susceptibility mapping. Geol J

Ma S, Shao X, Xu C (2023) Landslide susceptibility mapping in terms of the slope-unit or raster-unit, which is better? J Earth Sci 34:386–397. https://doi.org/10.1007/s12583-021-1407-1

Mandrekar JN (2010) Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 5:1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Merghadi A, Yunus AP, Dou J, Whiteley J, ThaiPham B, Bui DT, Avtar R, Abderrahmane B (2020) Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance. Earth Sci Rev 207:103225. https://doi.org/10.1016/j.earscirev.2020.103225

Pan G, Wang L, Li R, Yuan S, Ji W, Yin F, Zhang W, Wang B (2012) Tectonic evolution of the Qinghai-Tibet Plateau. J Asian Earth Sci 53:3–14. https://doi.org/10.1016/j.jseaes.2011.12.018

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

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

Rolain S, Alvioli M, Nguyen QD, Nguyen TL, Jacobs L, Kervyn M (2023) Influence of landslide inventory timespan and data selection on slope unit-based susceptibility models. Nat Hazards 118:2227–2244. https://doi.org/10.1007/s11069-023-06092-w

Shao X, Xu C (2022) Earthquake-induced landslides susceptibility assessment: a review of the state-of-the-art. Nat Hazards Res 2:172–182. https://doi.org/10.1016/j.nhres.2022.03.002

Thompson CG, Kim RS, Aloe AM, Becker BJ (2017) Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. Basic Appl Soc Psychol 39:81–90. https://doi.org/10.1080/01973533.2016.1277529

Wang X, Niu R (2010) Landslide intelligent prediction using object-oriented method. Soil Dyn Earthq Eng 30:1478–1486. https://doi.org/10.1016/j.soildyn.2010.06.017

Wang S, Ling S, Wu X, Wen H, Huang J, Wang F, Sun C (2023) Key predisposing factors and susceptibility assessment of landslides along the Yunnan-Tibet traffic corridor, Tibetan plateau: comparison with the LR, RF, NB, and MLP techniques. Front Earth Sci 10:1100363. https://doi.org/10.3389/feart.2022.1100363

Wen H, Wu X, Ling S, Sun C, Liu Q, Zhou G (2022) Characteristics and susceptibility assessment of the earthquake-triggered landslides in moderate-minor earthquake prone areas at southern margin of Sichuan Basin, China. Bull Eng Geol Environ 81:346. https://doi.org/10.1007/s10064-022-02821-w

Xi C, Han M, Hu X, Liu B, He K, Luo G, Cao X (2022) Effectiveness of Newmark-based sampling strategy for coseismic landslide susceptibility mapping using deep learning, support vector machine, and logistic regression. Bull Eng Geol Environ 81:174. https://doi.org/10.1007/s10064-022-02664-5

Yan Y, Tang H, Hu K, Turowski JM, Wei F (2023) Deriving debris-flow dynamics from real-time impact-force measurements. J Geophys Res Earth Surf 128:e2022JF006715. https://doi.org/10.1029/2022JF006715

Yang Z, Liu C, Nie R, Zhang W, Zhang L, Zhang Z, Li W, Liu G, Dai X, Zhang D, Zhang M, Miao S, Fu X, Ren Z, Lu H (2022) Research on uncertainty of landslide susceptibility prediction - bibliometrics and knowledge graph analysis. Remote Sens 14:3879. https://doi.org/10.3390/rs14163879

Yong C, Jinlong D, Fei G, Bin T, Tao Z, Hao F, Li W, Qinghua Z (2022) Review of landslide susceptibility assessment based on knowledge mapping. Stoch Env Res Risk Assess 36:2399–2417. https://doi.org/10.1007/s00477-021-02165-z

Yu C, Chen J (2020) Landslide susceptibility mapping using the slope unit for southeastern Helong City, Jilin Province, China: a comparison of ANN and SVM. Symmetry 12:1047. https://doi.org/10.3390/sym12061047

Yuan S, Jiang Y, Zhao Z, Cui M, Shi D, Wang S, Kang M (2023) Different trends and divergent responses to climate factors in the radial growth of Abies georgei along elevations in the central Hengduan Mountains. Dendrochronologia. https://doi.org/10.1016/j.dendro.2023.126114

Zhao X, Chen W (2020) Optimization of computational intelligence models for landslide susceptibility evaluation. Remote Sens 12:2180. https://doi.org/10.3390/rs12142180

Zhao S, Chigira M, Wu X (2019) Gigantic rockslides induced by fluvial incision in the Diexi area along the eastern margin of the Tibetan Plateau. Geomorphology 338:27–42. https://doi.org/10.1016/j.geomorph.2019.04.008

Zhao S, He Z, Deng J, Li H, Dai F, Gao Y, Chen F (2022) Giant river-blocking landslide dams with multiple failure sources in the Nu River and the impact on transient landscape evolution in southeastern Tibet. Geomorphology 413:108357. https://doi.org/10.1016/j.geomorph.2022.108357

Zhao S, Dai F, Deng J, Wen H, Li H, Chen F (2023) Insights into landslide development and susceptibility in extremely complex alpine geoenvironments along the western Sichuan-Tibet Engineering Corridor, China. Catena 227:107105. https://doi.org/10.1016/j.catena.2023.107105

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Hong Wen[1] · **Siyuan Zhao[2]** · **Yuhang Liang[3]** · **Sen Wang[4]** · **Ling Tao[1]** · **Jiren Xie[5]**

✉ Siyuan Zhao
zhaosiyuan@scu.edu.cn

1 School of Architecture and Civil Engineering, Xihua University, Chengdu 610039, China

2 State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resources and Hydropower, Sichuan University, No. 2 South Section 1, Yihuan Road, Chengdu 610065, Sichuan, China

3 School of Emergency Management, Xihua University, Chengdu 610039, China

4 Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China

5 School of Civil Engineering, Central South University, Changsha 410075, China