



Uncertainties of landslide susceptibility prediction: influences of different study area scales and mapping unit scales

Faming Huang¹ · Yu Cao¹ · Wenbin Li^{1,2} · Filippo Catani³ · Guquan Song¹ · Jinsong Huang⁴ · Changshi Yu⁵

Received: 6 August 2023 / Revised: 4 January 2024 / Accepted: 5 March 2024
© The Author(s) 2024

Abstract

This study aims to investigate the effects of different mapping unit scales and study area scales on the uncertainty rules of landslide susceptibility prediction (LSP). To illustrate various study area scales, Ganzhou City in China, its eastern region (Ganzhou East), and Ruijin County in Ganzhou East were chosen. Different mapping unit scales are represented by grid units with spatial resolution of 30 and 60 m, as well as slope units that were extracted by multi-scale segmentation method. The 3855 landslide locations and 21 typical environmental factors in Ganzhou City are first determined to create spatial datasets with input-outputs. Then, landslide susceptibility maps (LSMs) of Ganzhou City, Ganzhou East and Ruijin County are produced using a support vector machine (SVM) and random forest (RF), respectively. The LSMs of the above three regions are then extracted by mask from the LSM of Ganzhou City, along with the LSMs of Ruijin County from Ganzhou East. Additionally, LSMs of Ruijin at various mapping unit scales are generated in accordance. Accuracy and landslide susceptibility indexes (LSIs) distribution are used to express LSP uncertainties. The LSP uncertainties under grid units significantly decrease as study area scales decrease from Ganzhou City, Ganzhou East to Ruijin County, whereas those under slope units are less affected by study area scales. Of course, attentions should also be paid to the broader representativeness of large study areas. The LSP accuracy of slope units increases by about 6%–10% compared with those under grid units with 30 m and 60 m resolution in the same study area's scale. The significance of environmental factors exhibits an averaging trend as study area scale increases from small to large. The importance of environmental factors varies greatly with the 60 m grid unit, but it tends to be consistent to some extent in the 30 m grid unit and the slope unit.

✉ Wenbin Li
351113619004@email.ncu.edu.cn

¹ School of Civil Engineering and Architecture, Nanchang University, Nanchang 330031, China

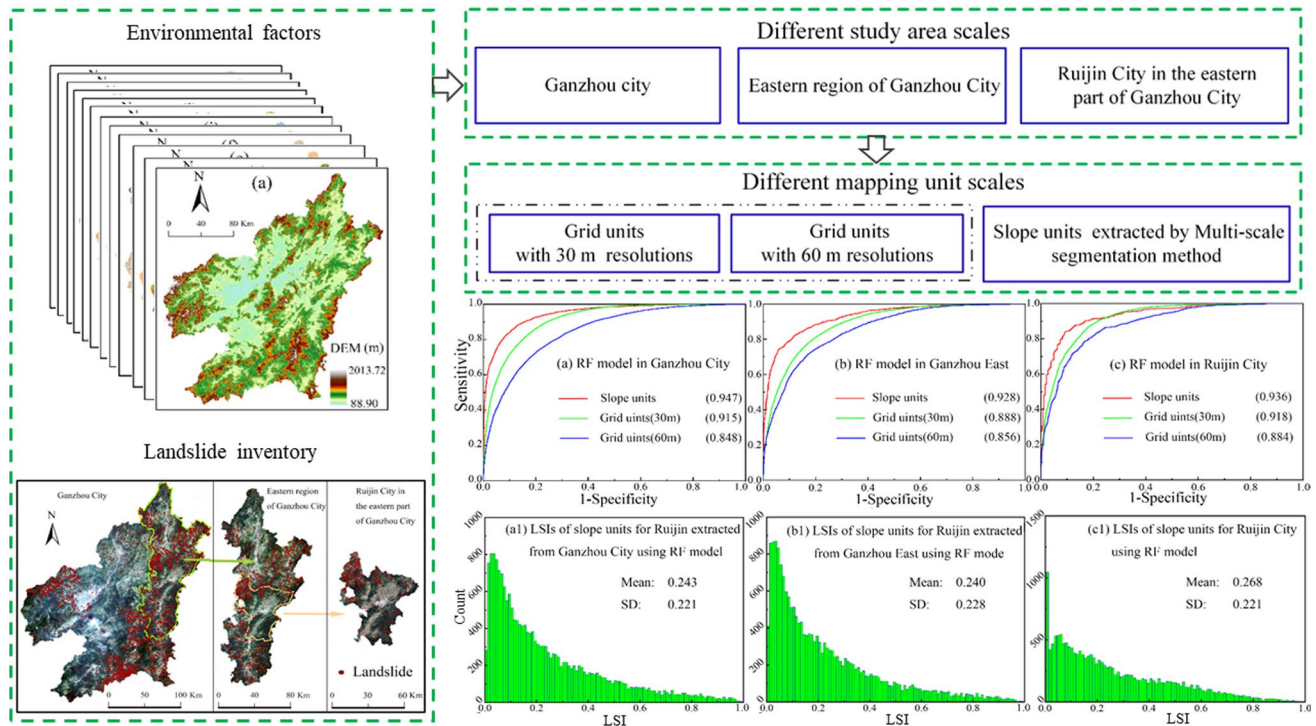
² Hebei International Joint Research Center for Remote Sensing of Agricultural Drought Monitoring/School of Land Science and Space Planning, Hebei GEO University, Shijiazhuang 050031, China

³ Department of Geosciences, University of Padova, Padua, Italy

⁴ Discipline of Civil, Surveying and Environmental Engineering, Priority Research Centre for Geotechnical Science and Engineering, University of Newcastle, Callaghan, NSW, Australia

⁵ School of Information Engineering, Nanchang University, Nanchang 330031, China

Graphic abstract



Highlights

- Influences of study area scale and mapping unit scale on the LSP uncertainties are explored;
- LSP accuracy increases while uncertainty decreases from a large to a smaller area scale under grid units;
- LSP accuracy and uncertainty under slope units are less affected by study area scales;
- Importance of environmental factors shows an averaging trend from a small to a larger area scale;
- Slope units are shown to be superior to grid units for LSP modeling.

Keywords Landslide susceptibility prediction · Uncertainty analysis · Study areas scales · Mapping unit scales · Slope units · Random forest

Abbreviations

| | |
|------|--|
| LSP | Landslide susceptibility prediction |
| LSM | Landslide susceptibility map |
| LSIs | Landslide susceptibility indexes |
| DEM | Digital Elevation Model |
| RF | Random forest |
| SVM | Support vector machine |
| AUC | Area under the receiver operation characteristic curve |

1 Introduction

Landslides are regarded as complex geological phenomena, and they pose a significant risk to human life, property and living conditions (Froude and Petley 2018; Haque et al.

2019; Wu et al. 2022). High-precision landslide susceptibility maps (LSMs) are helpful for the rational allocation of land resources and the reduction of pertinent decision-making risks.

Nowadays, Landslide susceptibility prediction (LSP) is still a significant research topic around the worldwide (Xu et al. 2012). Numerous uncertainty issues, including study area scales (Thi Ngo et al. 2021), mapping units scales (such as grid units and slope units) (Huang et al. 2021a), the quantity of landslide samples (Huang et al. 2022), the quality of original data sources (Chen et al. 2020b; Schlögel et al. 2018), the choice and combination of environmental factors (Huang et al. 2021b) and the selection of LSP models (such as data-driven models) (Cavazzi et al. 2013), all have an effect on the LSP modeling process (Zhao et al. 2024). These

uncertainty issues are viewed as errors in a broader sense and have an impact on our comprehension of LSP modeling.

Among the above uncertainty issues, the spatial patterns of landslides exhibit glaring differences and further transformed to the LSP results under different study area scales and mapping unit scales, although few attentions have been paid (Miller et al. 2015; Yu and Gao 2020). LSP modelling requires the preliminary selection of a suitable scale, such as county scale, watershed scale and/or large regional scale. The study scale also includes mapping unit scales that maximizes internal homogeneity and between-units heterogeneity (Calvello et al. 2013). Both scale characteristics can be found in remote sensing images, DEMs and other data sources used to describe topographic and geological characteristics, as well as land surface characteristics. Meanwhile, the combinations of environmental factors and LSP models have varying degrees of adaptability under various study areas scales and mapping unit scales. The reliability and accuracy of LSP results may be lowered, if the uncertainty issues related to both scales are poorly taken into account and the global and local study areas are generalized (Miller et al. 2015; Zhu et al. 2018a). This is because there will be no "adaptation" to local conditions for different scales (Huang et al. 2023b). Furthermore, few studies have explored how to jointly optimize various study area scales with landslide mapping unit scales and their various spatial resolutions. Grid units, for instance, are suitable for LSP under various spatial resolution conditions, while slope units, based on high-resolution Digital Elevation Model (DEM) data, are more suitable for both LSP and risk warning in a variety of study areas (Hodasová and Bednarik 2021).

To sum up, should LSP modeling be done at what study area scale and/or mapping unit scale? And do both scales affect how environmental factors are mapping for landslide evolution? These two problems suggest that joint consideration of study area scale and mapping unit scale is very crucial for LSP modeling. This study aims to investigate the effects of different study area scales and mapping unit scales on the uncertainty rules of LSP. The Ganzhou City, Ganzhou East, and Ruijin County in China were selected to represent different study area scales. Different mapping unit scales are represented by grid units with spatial resolution of 30 and 60 m, as well as slope units that were extracted by multi-scale segmentation method, so as to realize modeling uncertainty analysis under the combined working conditions of different scales.

2 Review of related studies

The identification of uncertainty issues is crucial to LSP modeling. This paper primarily examines these uncertainty issues, including the determination of study area scales

and mapping unit scales, as well as the environmental factors selection.

2.1 Influence of environmental factors selection on LSP

In the past 30 years, countless environmental factors, such as slope, lithology, aspect, hydrology, river, curvature, et al., have been proposed based on mapping units in various geological and climatic settings (Yang et al. 2023). By using synonyms and environmental factors with similar descriptors but not necessarily identical meanings, Reichenbach et al. (2018) reclassified a large number of environmental factors into 23 categories. The 23 identified categories were then divided into five subject groups, including geology, hydrology, land cover, landform, and other variables. According to an analysis of the literature database, researchers prefer "simple" (direct) measures of landform, such as DEM, relief, slope, aspect, and curvature, and the majority of articles used environmental factors that were primarily related to landform (Gaidzik and Ramirez-Herrera 2021). We observe that slope has consistently shown to be the most useful environmental factor for LSP (Loche et al. 2022). From 5 to 22 environmental factors, with an average of 9 variables, were used for each individual SLP modeling, which may have been constrained by data sources and spatial analysis techniques. Most studies only include more than ten types of easily measurable environmental factors. For instance, Huang et al. (2020a) chose 13 different categories of environmental factors, including lithology, topographic wetness, slope, and aspect. Hong et al. (2018) chose 16 environmental factors for LSP, including lithology, topography, hydrology, and average annual rainfall. The accuracy of the LSP modeling will inevitably be impacted by the use of these few categories of environmental factors, which make it difficult to accurately reflect landslide characteristics (Huang et al. 2022).

Furthermore, it can be challenging to assess the significance of different environmental factors on LSP modeling in a study area (Huang et al. 2020b; Huang et al. 2023c). As a result, in order to accurately represent the development characteristics of landslides from various angles, it is necessary to gather a relatively rich category of environmental factors in the early stages of modeling (Youssef and Pourghasemi 2021). The slope stability characteristics under heavy rainfall, engineering slope cutting, land cover, slope body structure, and soil mechanical properties can be taken into consideration based on the evolution characteristics of rainfall-type landslides in Ganzhou City of China. The analysis of annual rainfall, rainstorm frequency, topography, highway density, vegetation cover, soil clay/sand content, accumulation layer thickness, and rock weathering intensity can be focused on as the appropriate environmental factors (Li et al. 2021). Landform

and lithology environment factors (DEM, slope, aspect, plan curvature, profile curvature, slope forms, topography relief, DEM variation, cutting depth, surface roughness, lithology and fault), hydrological and land cover environment factors, and other environmental factors are all categorized into 21 classes that correspond to four thematic clusters in Ganzhou City (terrain wetness indexes, modified normalized differential water index, drainage density, average rainfall, normalized differential vegetable index, normalized differential building index, road density and population density).

2.2 Review of influences of study area scales on LSP modelling

There is a lot of subjectivity and uncertainty because most studies do not take into account the spatial correlation of the study area with its surroundings (Shirzadi et al. 2019). In general, soil types are similar between regions with similar landforms, climates, lithologies, and other conditions. The corresponding LSP results ought to be similar even though these locations are discontinuous in space because they typically have comparable environmental factors or combinations of environmental factors. Geographical features are closer together in environments that are more similar to one another (Zhu et al. 2018b). Landslide disasters are a type of environmental geological problem with specific spatiotemporal properties, and they contain relative properties at various scales. They are natural phenomena that occur at specific times and locations (Shou and Lin 2016; Zhu et al. 2015a). Additionally, because it is a regional natural condition, its scale has an impact on how environmental factors are characterized and evolve (Fressard et al. 2014; Kuan-Tsung et al. 2019; Zhu et al. 2015b). The dominant environmental factors affecting landslide evolution are different as a result of different geographical environments, climatic conditions, and human activities at various scales within a study area, which causes different spatial distribution of landslides (Miller et al. 2015).

The weight of spatially associated environmental factors and different sample quantities used for LSP model building can both be directly impacted by study areas with different scales. The first and most important step in LSP is the selection of environmental factors in the study area (Shi et al. 2018). In study areas with various scales, these landslide-related environmental factors, particularly terrain and hydrological factors, have a significant impact on LSP (Du et al. 2020; Zhu et al. 2018b). The actual characteristics of environmental factors are weakened or confused by a large study area (Yu and Gao 2020). Environmental factors can reflect changes in landslides in micro-regional environments in small study area (Kang

et al. 2016; Zhu et al. 2018a). Additionally, the primary environmental factors vary by study areas and scale.

2.3 Review of influences of mapping unit scales on LSP modelling

A mapping unit is a section of the land surface that is distinguished from neighboring units by a unique set of ground conditions (Reichenbach et al. 2018). A mapping unit of landslides and associated environmental factors for LSP can be thought of as an observation window at the scale of the analysis, though it is difficult to choose the best observation window (Loche et al. 2022). All of the commonly employed mapping units for LSP that have been suggested in the literature can be categorized into one of the following groups: I grid units, II slope units, III small watershed units, IV topographic units, V political or administrative units (Alvioli et al. 2016; Drăguț et al. 2010). Grid units and slope units are thought to be the two primary types of object units used for LSP, according to related literature (Huang et al. 2021a).

The most common mapping unit was by far grid units, which were distinguished by their constant shape and regular grid (Li et al. 2021). At the same time, we found that most LSP models used grid units with the same resolution for environmental and landslide factors, while very few used different resolutions for the mapping unit and the DEM (Gaidzik and Ramirez-Herrera 2021). The microtopographic morphometric signature can be captured in great detail by using small grid units (5 m × 5 m or less), but this may have little to or no geomorphological, geological, or geomorphic climatic significance for large study scale areas, where a coarser resolution DEM would be more effective (Salciarini et al. 2016). Furthermore, the size of the spatial sampling unit, which directly affects the number of samples required to build the model and the overall number of grids in the study area, is the essence of spatial resolution (Guo et al. 2022).

Slope units have received a lot of attention in the last 15 years because they accurately reflect the physical relationships between landslides and morphological features, but the extraction of slope units has always baffled cartographers (Wang et al. 2018). Meanwhile, the slope unit is derived from the grid units, and it should meet the following three conditions: (1) continuous closed geographical space; (2) uniform slope and aspect; (3) no jump in surface slope. Based on the above requirements, an object-oriented multi-scale image segmentation method is used to extract the slope units in the study area (Liu et al. 2023). This method is applicable in various study areas with different landforms, and is more efficient and accurate comparing to other slope unit extraction methods (Chang et al. 2023; Ma et al. 2023). The image pixel is transformed into the slope through multi-scale segmentation, which collects the pixels with the same

spatial and spectral characteristics into a "homogeneous and uniform" unit (Lim and Keles 2019). The size of the spatial sampling unit, which determines how many samples are needed to build the model and how many grids are present in the study area, is the key to understanding spatial resolution. The object units for LSP model building in this study are slope units obtained by multi-scale segmentation method and grid units with spatial resolution of 30 m and 60 m.

3 Summary

To determine LSP, a variety of methods have been put forth (Reichstein et al. 2019). Despite the variations, the reproducibility issue of "to what extent this conclusion can be applied to other regions" (Liu et al. 2022) affects all methods. The majority of geo-disaster studies lack a natural analytical scale, and the rules discovered in a study region depend on how the mapping units are divided, leading to variable mapping unit issues. The mapping unit issue is more apparent, particularly when analysis based on spatial extensibility environmental factors. The model is also impacted by the ambiguous geographic study area scales. Finding universal laws is made more difficult by the fact that various patterns can appear within the same study area (Liu 2022). The choice of the study area scale and the division of the mapping unit scale determine which rule is more accurate. The overall goal of this study is to conduct more in-depth analyses of LSP at various study area and mapping unit scales.

4 Methodologies

This study makes extensive use of multi-source remote sensing technology (Lissak et al. 2020), geographic information systems (Chen and Chen 2021), and data-driven models (Chen et al. 2020a) to construct LSP modelling in order to discuss the effects of study with various scales on LSP modeling. Three distinct study areas are chosen: Ganzhou City in China, Ganzhou City's eastern region (Ganzhou East), and Ruijin County in Ganzhou East's a region. To represent various mapping unit scales, grid units with 30 m and 60 m spatial resolutions, as well as slope units extracted using the multi-scale segmentation method, are used. In order to create landslide and non-landslide spatial datasets, the Ganzhou City landslide inventory data is first determined, and 21 landslide environmental factors are extracted.

The above spatial datasets are then trained and tested using SVM and RF models in order to obtain LSMs for Ganzhou City. In order to address LSP in Ganzhou East and Ruijin County, respectively, landslide information for Ganzhou East and Ruijin County is also extracted from Ganzhou City, followed by a repetition of the aforementioned modeling

process in Ganzhou City. The LSMs of Ganzhou East and Ruijin County are then extracted using masks from those of Ganzhou City, and the LSMs of Ruijin County are extracted using masks from that of Ganzhou East. In order to analyze the uncertainty rules of LSP modeling under study areas with different scales based on LSMs predicted and masked through the aforementioned 9 types of conditions, accuracy and distribution rule of LSIs are adopted. The specific flow chart used in this study is shown in Fig. 1.

4.1 Multi-scale segmentation method

In order to study regional LSP at various study area scales, grid units and slope units have chosen as mapping units (Alvioli et al. 2020). The ArcGIS 10.1 platform makes it relatively easy for grid units to implement LSP at various study area scales. The multi-scale segmentation method's extracted slope units are not a single pixel, but rather a homogeneous object with spectral features, spatial features, and shape features (Alvioli et al. 2022). Through iterative updating, adjacent pixels with similar spectral and shape features are combined into slope units with the same homogeneous properties based on the top-down region-growing segmentation algorithm of pixels (Fig. 2). The weighted sum of spectral and shape heterogeneity yields the heterogeneity index f (Eq. (1)) between slope units (Huang et al. 2021a).

$$f = w_{\text{color}}h_{\text{color}} + (1 - w_{\text{color}})h_{\text{shape}} \quad (1)$$

where w_{color} is the spectral weight, h_{color} represents the object spectral heterogeneity, which is determined by the scale and spectral standard deviation of the object, and h_{shape} represents the object shape heterogeneity, which is determined by the weighted sum of compactness and smoothness.

4.2 Machine learning models

The LSP models are also mainly divided into deterministic models, knowledge-driven models and machine learning models (Du et al. 2020; Huang et al. 2020a; Zheng et al. 2019). The deterministic model's reliance on many complex soil mechanical parameters and the knowledge-driven model's reliance on human subjectivity can be somewhat reduced by the data-driven model, as well as the LSP accuracy can be improved when there are gaps in the data or the data is of poor quality (Juliev et al. 2019; Pham et al. 2022; Zheng et al. 2023). To determine characteristics and discriminant rules of landslide occurrence, data-driven models primarily use the internal relationships between known landslides and different environmental factors (Hu et al. 2020; Sun et al. 2020). At present, remote sensing and machine learning models have been widely used in the regional LSP and a series of research results have been achieved (Pradhan

Fig. 1 Flow chart of LSP under different scales of study areas (AUC is the area under the receiver operating characteristic curve, Terrain wetness indexes, modified normalized differential water index, normalized differential vegetable index, normalized differential building index)

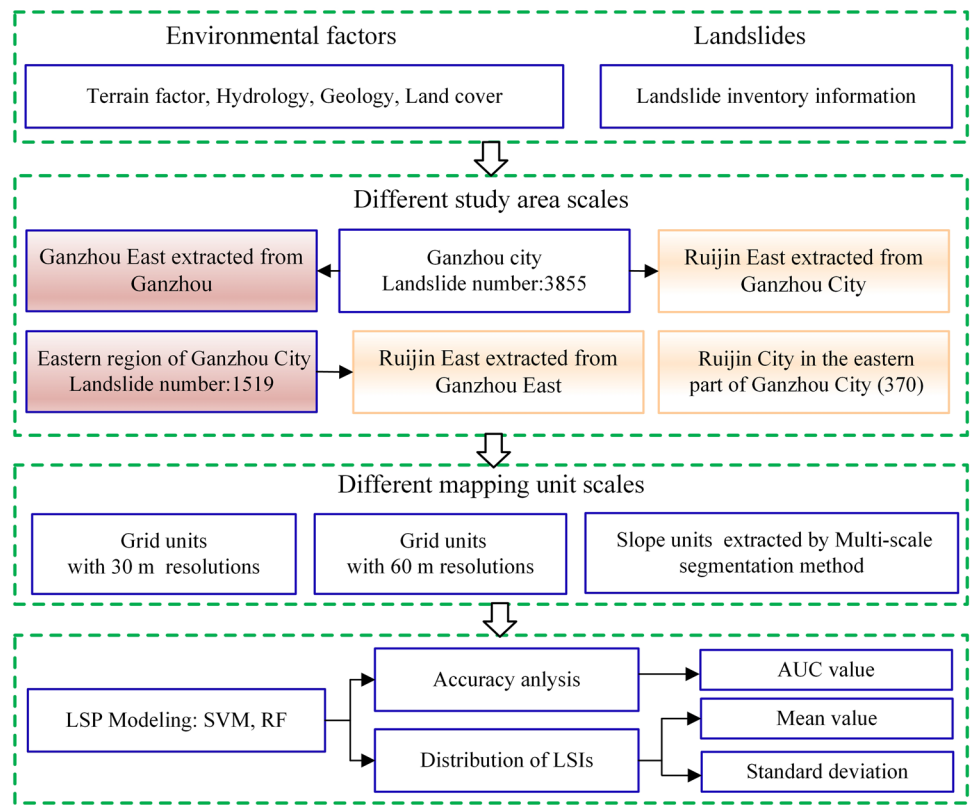
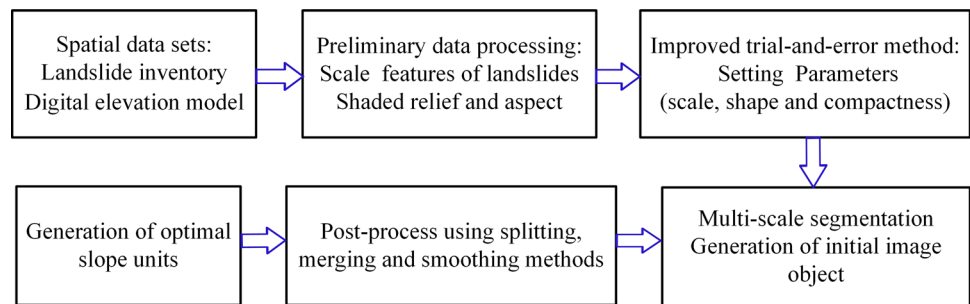


Fig. 2 Workflow of slope units extracted by multi-scale segmentation method



et al. 2023). Logistic regression, artificial neural networks, support vector machine (SVM), boosting algorithm, decision tree, random forest (RF), and deep learning models are typical examples of machine learning models (Huang et al. 2020a; Lombardo and Mai 2018; Lombardo and Tanyas 2020; Xiao and Zhang 2023; Zhao et al. 2019). Merghadi et al. (2020) have reviewed and compared various machine learning algorithms for landslide susceptibility studies. Among them, SVM and RF models are the most widely used, have the best prediction accuracy, and are not sensitive to the multicollinearity of environmental factors.

4.2.1 Support vector machine

SVMs are conducted to find the best hyperplane to use for modeling and to use support vectors on the hyperplane to

maximize the space between classes. By converting the input variables into an n -dimensional eigenspace using a kernel function, nonlinear data can be linearly separable. m_i represents each environmental factor for a set of linearly separable training vectors $m_i (i = 1, 2, \dots, n)$, with the corresponding output categories $y_i = \pm 1$ being landslides and non-landslides, respectively. In order to categorize landslides, the maximum distance through n -dimensional hyperplanes is determined. Between them, $\frac{1}{2} \|s\|^2$ is the widest spacing. Additionally, it uses relaxation variables ξ_i to control classification error for data with linear inseparability, and the corresponding constraint condition is $y_i((s \cdot m_i) + b) \geq 1 - \xi_i$. Furthermore, by introducing $\nu(0, 1)$, incorrect classification is taken into account. Equation (2) displays the hyperplane distance, where λ_i is the lagrange multiplier, b is the constant, $\|s\|$ is the norm of the normal hyperplane. The linear,

polynomial, radial basis, and Sigmoid functions are typically the main components of the SVM kernel function. This study (Xu et al. 2012) uses the radial basis kernel function, which has a good application in LSP.

$$L = \frac{1}{2} \|s\|^2 - \frac{1}{vn} \sum_{i=1}^n \xi_i \quad (2)$$

4.2.2 Random forest

To diversify the generated classification trees, the RF is to build various training datasets extracted by putting back and to pick various features at random (Youssef et al. 2016). A collection of various classification trees can more fully reflect factual findings than a single tree, improve model predictability, and prevent over-fitting. In addition, RF uses out-of-bag error to achieve unbiased generalization error estimation, and it gradually converges as the number of trees grows. The out-of-bag error can also serve as a proxy for the significance of each factor variable. When only one variable is changing in out-of-bag data, the importance of the factor variable is determined by the size of the out-of-bag error change. Mean decrease accuracy and mean decrease Gini are also used to gauge the significance of the input variable. Additionally, the number of trees and features in the model is a key factor in determining how well it predicts (Huang et al. 2022).

4.3 LSP results assessment

AUC accuracy evaluation and LSI distribution rule are the two factors that most clearly illustrate the uncertainties in LSP modeling. AUC value is a metric used to quantitatively assess the overall effectiveness of LSPs (Garosi et al. 2019). The receiver operation characteristic is the following: first, LSIs are calculated, and various samples in the test datasets are ranked; second, different cut-off points are chosen in this order; third, the samples are used one at a time as positive samples to predict; and finally, the true positive rate and false positive rate calculated in the current predictor each time are taken into consideration as vertical and horizontal axes in the receiver operation characteristic curve. Better LSP performance is suggested by a higher AUC. According to Eq. (3), the AUC value represents the likelihood that randomly selected positive samples will rank higher than randomly selected negative samples.

$$\text{AUC} = \frac{\sum_{i=1}^{n_0} r_i - n_0 \times (n_0 + 1) / 2}{n_0 \times n_1} \quad (3)$$

where n_0 represents the number of negative samples, n_1 represents the number of positive samples, r_i represents the order of the i th negative sample in entire test samples. The mean value and standard deviation, on the other hand, primarily reflect the different distribution rules of LSIs.

Additionally, the mean value depicts the LSIs' average level of distribution, and the standard deviation shows how widely distributed they are. The distribution rule of LSIs as a whole is analyzed using the mean value and standard deviation, which offers theoretical direction for LSP in the research area (Huang et al. 2020a).

5 Study area and materials

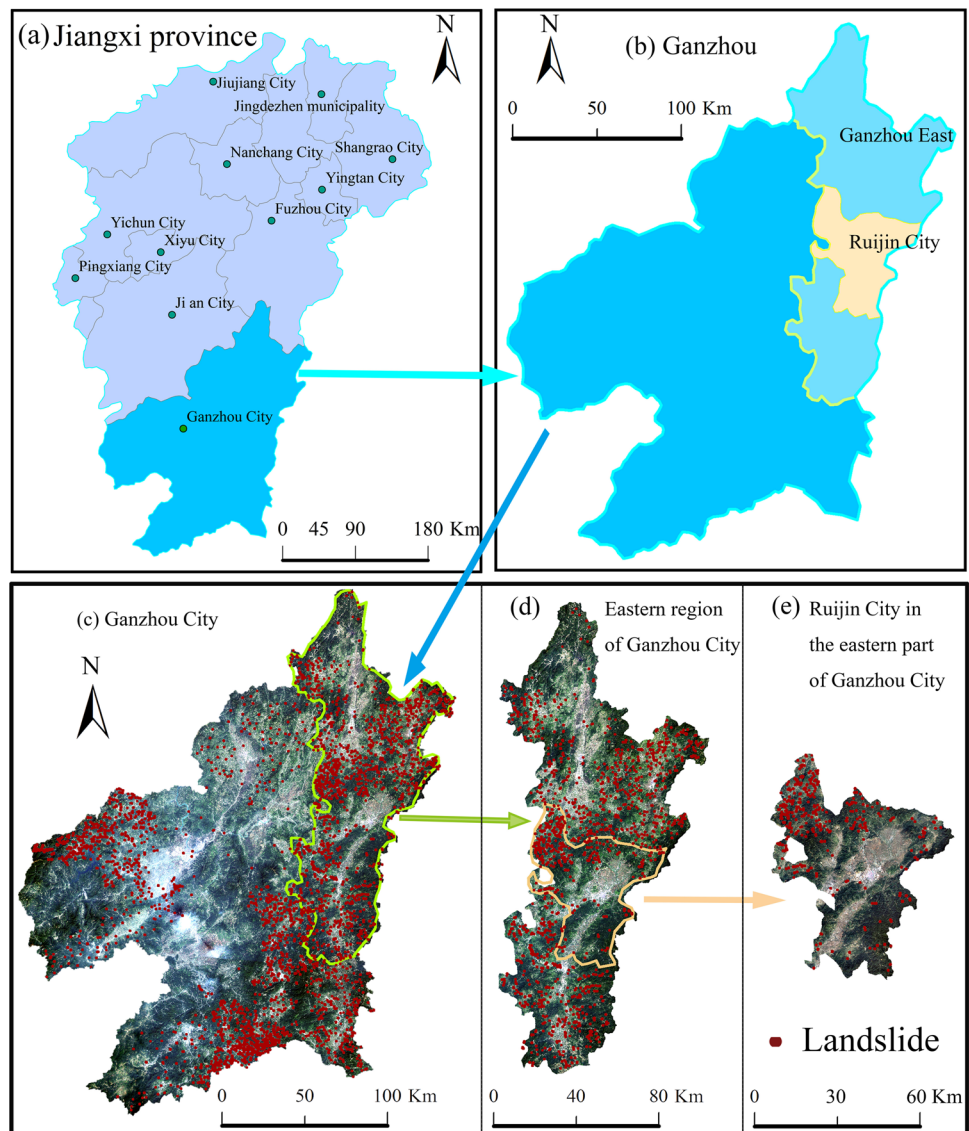
5.1 Study area

Ganzhou City is situated where the Jiangnan Hills and Lingnan Mountains meet (Figs. 3b, c). Cities like Ningdu, Shicheng, Ruijin, and Huichang are located in Ganzhou East (Fig. 3d). In the region, metamorphic rocks, such as Triassic strata, Cretaceous shale, siltstone, Cambrian strata, Sinian phyllite, and slate, as well as secondary intrusive magmatic rocks from each era, make up the majority of the exposed stratigraphic layers. There have been 2496 geological disasters overall, with a 17.59/100 km² average surface density. Eastern Ganzhou City's Ruijin County is situated at northern latitudes of 25°30' and 26°20' and eastern longitudes of 115°42' to 116°22'. Figures 3c, e show that rather than being next to one another, the study areas span from Ganzhou City to Ruijin County.

5.2 Landslide inventory information

Numerous environmental factors can cause landslides, and each environmental factor has a different weight and effect on the likelihood of a landslide (Abraham et al. 2020; Shi et al. 2018). The performance of the LSP in a study area is determined by the quality of the landslide data. Understanding the locations, movement types, triggering times, scales, and related evolution of geological environments of landslides from space is aided by information from landslide inventories (Franceschini et al. 2022; Segoni et al. 2018). Table 1 displays a series of data used for LSP.

The information is primarily derived from field surveys and high-resolution image interpretation of local geological disasters at 1:100000 scale. According to landslide inventory in Ganzhou City, there were 9555 disasters, including 3855 soil landslides with landslide density of 78/100 km²; by the end of 2014, there had also been a significant number of collapses, debris flows, and ground collapses in the study areas (Hung et al. 2014). In Ganzhou East, there were 2041

Fig. 3 Landslide location in the study area**Table 1** Data lists used for LSP

| Data types | Grid resolution | Data usage | Extraction method |
|---|----------------------|---|---|
| Landsat-8 MT | Multispectral 8/30 m | Normalized differential vegetable index; normalized differential building index; modified normalized differential water index | Automatic information extraction in ENVI 5.4 software |
| DEM | 30 m | Terrain factors, drainage density and Terrain wetness indexes | ArcGIS spatial analysis tools |
| Geological map | 1:100000 | Lithology and fault lines | Vectorization |
| The network (http://solargis.cn/imaps/) | 30 m | Obtaining population density and total radiation | Data collation, interpolation |
| Landsat image | | Obtaining road density | Automatic extraction in ArcGIS 10.3 software |
| Hydrological station | | Rainfall | Data collation, interpolation analysis |

disasters, of which 1519 were landslides, with a landslide density of 14.07/100 km², and they primarily occurred in the southern and eastern Ningdu City, the northern and north-eastern Shicheng City, the northwestern Ruijin County, and the southern Huichang City; Additionally, Ruijin County had 414 disasters, of which 370 were small- and medium-sized landslides, with a landslide density of 15.18/100 km², which occupies 89% of the total disaster amount and is mainly distributed in the northern mountainous area (Figs. 3c–e).

5.3 Landslide related environmental factors

There is no set standard for the selection of environmental factors plays, despite the fact that it has a huge impact on the quality of an LSM during the LSP process (Kang et al. 2016). Real-time rainfall (Fustos et al. 2020) and earthquakes are typically the main outside influence on initial landslides. The slope and basic environmental factors nearby also have an impact on the stability of landslides, particularly in small areas with similar external factors. Additionally, nearby basic environmental elements like topography, geology, hydrology, and vegetation conditions heavily influence landslide stability. Additionally, this study chose the 21 landslide environmental factors listed in Table 2 based on previous LSP studies (Chen and Li 2020) and an analysis of environmental traits in study areas.

5.3.1 Topographic environmental factors

The influence of geography on landslide evolution can be reflected in terrain factors. 10 different types of terrain factors are obtained in this study using DEM (Figs. 4a–j). Climate can be influenced by elevation (Fig. 4a), can low- and middle-altitude regions are frequently conducive to the formation of landslides, which are frequently cited as environmental factors (Shahabi et al. 2014). Due to its direct impact on the shear stress that leads to landslide instability and failure, slope is a significant factor in promoting the

occurrence of landslides (Hong et al. 2017a). According to statistics, the majority of the slopes in Ganzhou City where landslides have occurred are between 15° and 45° (Fig. 4b). However, there are some variations in the slope-generating landslides in various locations with various lithologies. For instance, landslides typically occur on slopes between 15° and 25° in hilly areas, whereas they typically happen on slopes between 30° and 45°, in middle and low mountainous areas. Figure 4c shows the variation in soil moisture content and the distribution of vegetation cover in all directions (Hong et al. 2017b). The plan curvature and profile curvature in Figs. 4d, e show, respectively, how the topographic gradient affects flow velocity and convergence (Chen et al. 2017). The geometries of various slopes are known as slope forms (Fig. 4f). Both the elevation variation coefficient (Fig. 4h) and the topographic relief (Fig. 4g) are macroscopic indices that reflect the degree of surface relief and fragmentation. For the purpose of analyzing soil loss and the development of surface erosion, surface roughness (Fig. 4i) and surface cutting depth (Fig. 4j) are crucial reference indices.

5.3.2 Geology factors

The permeability, matric suction, and shear strength of rock and soil are all different, which is reflected in the difference in lithological and physical properties (Hong et al. 2017b). The primary lithology types in Ganzhou City are clastic and carbonate rocks, which are sporadically distributed in counties, metamorphic rocks which are primarily found in the western and southeastern regions of the study area, and metamorphic rocks which are widely distributed near rivers and other bodies of water. Additionally, geological maps of Ganzhou City at a scale of 1:100000 are used to generate the lithologies in various subareas (Fig. 4k). The metamorphic strata with a high incidence of geological disasters, which account for 1623 landslides 42% of the total number of landslides, are followed by the clastic, magmatic, and carbonate strata.

Table 2 Environmental factor types

| Types | Environmental factors | Types | Environmental factors | |
|-----------------------|-----------------------|-------------------|--|---|
| Terrain factors | DEM | Hydrology factors | Terrain wetness indexes | |
| | Slope | | Modified normalized differential water index | |
| | Aspect | | Drainage density | |
| | Plan curvature | | Average rainfall | |
| | Profile curvature | Total radiation | | |
| | Slope forms | Topography relief | Land cover factors | normalized differential vegetable index; normalized differential building index |
| | | | | DEM variation |
| Surface cutting depth | | | | Population density |
| Surface roughness | | Geology factors | Lithology | |
| | | | Distance to faults | |

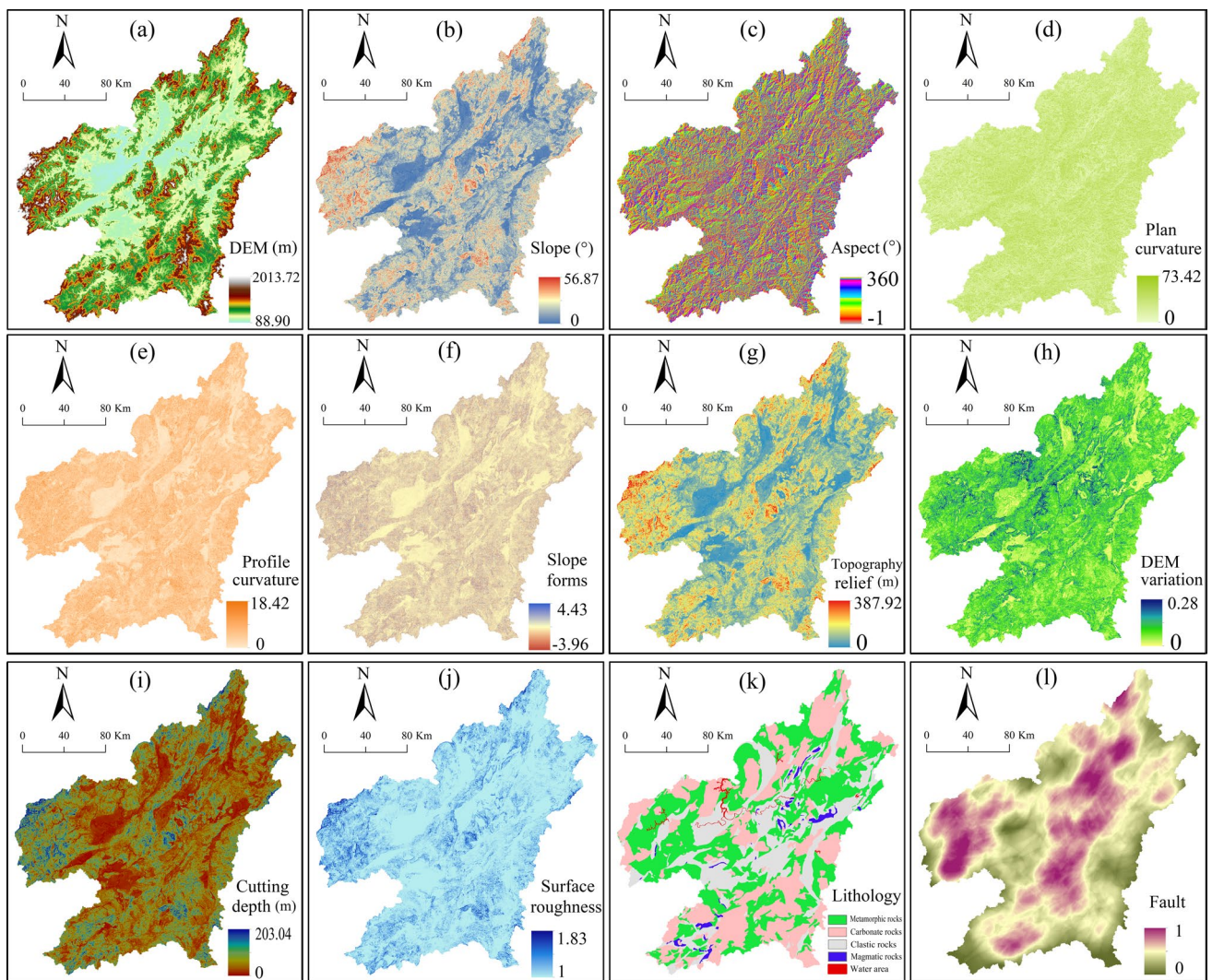


Fig. 4 Landform and lithology environment factor maps in Ganzhou City **a** DEM **b** Slope **c** Aspect **d** Plan curvature **e** Profile curvature **f** Slope forms **g** Topography relief **h** DEM variation **i** Cutting depth **j** Surface roughness **k** Lithology **l** Fault

In the geological structure (Fig. 4) also frequently has a negative effect on landslide stability, and tectonic movement is accompanied by differential lifting of faults and folds, which frequently creates a structural weak zone and lowers the stability of rock and soil mass. The deep Anyuan—Yingtang, Xunwu—Ruijin, Dayu—Nancheng and Quannan—Anyuan Faults are the main faults in Ganzhou City. The effect of geology on the evolution of landslides is represented in this study by lithology and the distance to faults.

5.3.3 Hydrological environmental factors

Topographic and remote sensing hydrological factors can be used to categorize hydrological environmental factors. The soil moisture content and groundwater distribution are shown by the terrain wetness indexes (Fig. 5a) (Xu et al.

2012). In the visible and near-infrared bands, there are clear distinctions between the spectral reflectance of water and vegetation. Similar to the modified normalized differential water index (Fig. 5b), which is primarily used to identify surface water, the terrain wetness indexes can effectively extract comprehensive water information, such as surface runoff and groundwater. Drainage density (Fig. 5c) is used to represent the impact of rivers on landslide occurrences because rivers' scour and erosion have significant detrimental effects on landslide stability.

Total radiation, which has a significant impact on surface vegetation, soil humidity, and atmospheric temperature is the sum of direct solar radiation and sky radiation received from horizontal surfaces and the main energy source of atmospheric circulation and water circulation (Huang et al. 2020c). Specifically, total radiation is the main energy source for

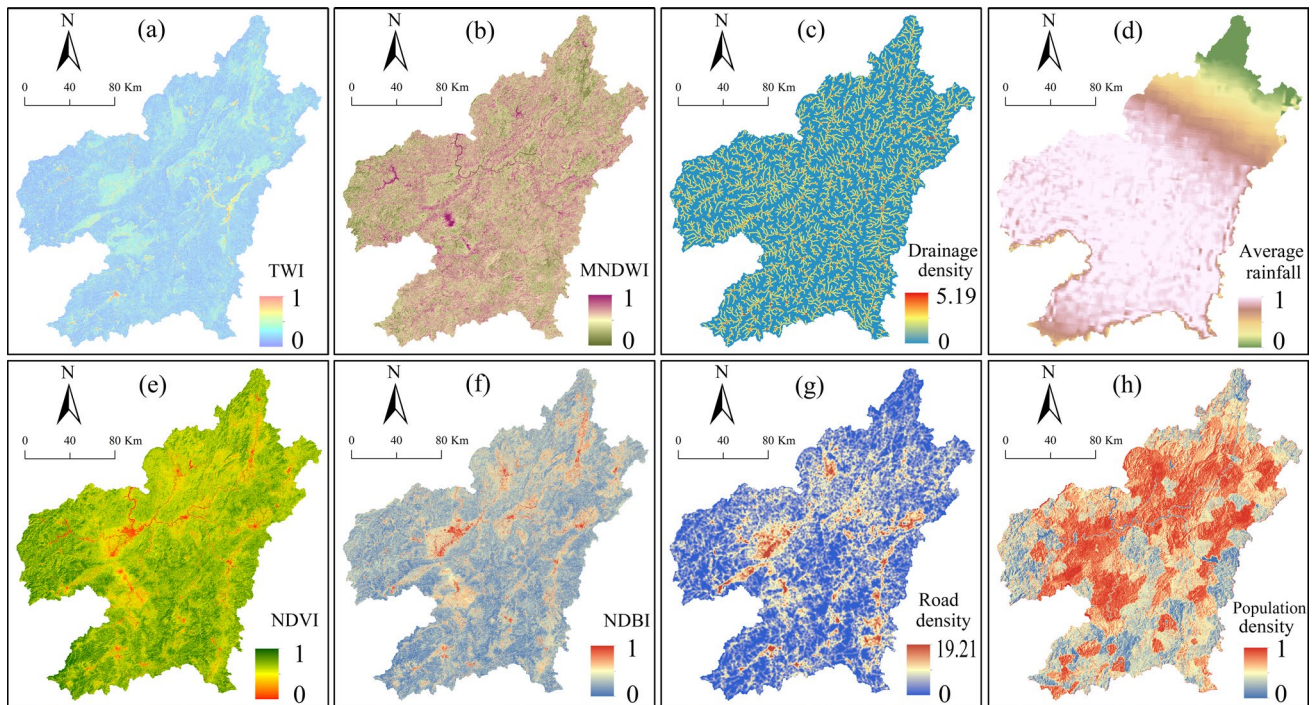


Fig. 5 Hydrological and land cover environment factor maps in Ganzhou City **a** TWI **b** MNDWI **c** Drainage density **d** Average rainfall **e** NDVI **f** NDBI **g** Road density **h** Population density

plants to carry out photosynthesis, and its intensity and duration will affect the evaporation and loss of soil water, and also change the temperature by affecting the surface energy balance. One of the primary factors causing landslide evolution in Ganzhou City is average annual rainfall (Fig. 5d). The weight of the landslide mass is greatly increased by rainfall infiltration, creating a weak zone. Additionally, as the landslide mass's pore water pressure rises, the effective stress decreases, which, in turn, lowers friction resistance in the sliding zone (Bai et al. 2020). In addition, factors affecting precipitation typically include total precipitation, rainfall days, and daily precipitation of more than 50 mm. The easiest way to quantify the impact of rainfall on landslide evolution is to use the annual average rainfall (Abraham et al. 2020).

5.3.4 Land cover factors

In addition to improving geotechnical physical properties through root action, vegetation also successfully lowers the erosion and infiltration effect of rainfall on landslides, affecting the stability of the landslide. The density and distribution of surface vegetation can be represented by the normalized differential vegetable index (Fig. 5e), which also prevents the evolution of landslides. In order to represent the distribution of residential building land and comprehend the dense local residential area, the normalized differential building

index (Fig. 5f) shows relatively common building information that has been extracted through remote sensing image data (Chang et al. 2020). Path/row 121/42 Landsat 8 images with a 30 m resolution taken on October 3, 2013 are used to calculate the normalized differential vegetable index and the normalized differential building index. The study area's population density can be measured using the population density (Fig. 5f), which is defined as the number of people per unit area. Landslides are negatively impacted by all phases of road construction and operation, and during the rainy season, artificially created high and steep slopes, excavation, blasting, and loading are common causes of landslides. To represent the impact of road and traffic facilities on landslide occurrences, the road density (Fig. 5g) is used.

5.4 Establishment of LSP modelling spatial dataset

5.4.1 Establishment of grid-unit dataset

As the mapping units for LSP in this study, related data sources with various resolutions are resampled to grid units with 30 m and 60 m resolution. This research area in Ganzhou City spans a vast 10,794.86 km² areas. Grid units of 45,525,924 (10,942,367), 9,310,512 (3,011,683), and 2,715,630 (687,633) are used to divide the study areas of Ganzhou City, Ganzhou East, and Ruijin County under 30 m (60 m) raster resolution, respectively. Ganzhou City

Table 3 Training datasets and test datasets of landslides in the study

| Study areas | Mapping unit | Number of Landslides | Units in study area | Landslide units | Landslide density | Training datasets/test datasets |
|--------------|-------------------|----------------------|---------------------|-----------------|-------------------|---------------------------------|
| Ganzhou City | Grid unit of 30 m | 3855 | 45,525,924 | 53,078 | 0.029% | 18719/8023 |
| Ganzhou East | | 1519 | 11,972,468 | 17,115 | 0.046% | 5977/2561 |
| Ruijing City | | 366 | 2,715,630 | 5482 | 0.052% | 1954/838 |
| Ganzhou City | Grid unit of 60 m | 3855 | 10,942,367 | 13,371 | 0.122% | 18719/8023 |
| Ganzhou East | | 1519 | 3,011,683 | 4499 | 0.149% | 6299/2699 |
| Ruijing City | | 366 | 6,87,633 | 1396 | 0.203% | 1954/838 |
| Ganzhou City | Slope unit | 3855 | 2,96,003 | 3855 | 1.302% | 3855 |
| Ganzhou East | | 1519 | 78,565 | 1519 | 1.933% | 1519 |
| Ruijing City | | 366 | 17,989 | 366 | 2.035% | 366 |

saw a total of 3855 landslides, which were divided into 53,078 (13,371), 17,115 (4499), and 5482 (1396) landslide grid units, respectively. Of these, 1519 landslides happened in Ganzhou East and 366 landslides happened in Ruijin County. In addition, a random number generator chooses the same number of non-landslide grids as corresponding landslide grids in the study areas. By randomly selecting non-landslides in non-landslide area, we can ensure that non-landslides are not concentrated (Chang et al. 2023). Landslides and non-landslides are output variables with assigned values of 1 and 0, which build training datasets and test datasets with a random partition of 7:3. The original 21 environmental factors are extracted as model input variables. To calculate LSIs, the trained model's original grid unit values from the entire study area are substituted in, and the model is then divided into 5 classes using the natural break point method (Li et al. 2019).

The input datasets of landslides used for modelling are shown in Table 3, and all polygonal landslide surfaces are converted into point data in ArcGIS 10.1 software. Through resampling and correlation analysis of DEM data with 30 m resolution and other relevant environmental factors, the environmental factors in various study areas are discovered. Figures 4 and 5 display the environmental factor maps of Ganzhou City.

5.4.2 Establishment of slope-unit datasets

In this study, slope units are extracted and a dataset is established using the multi-scale segmentation method. The software eCognition 8.7 performs object segmentation on the shaded relief and terrain aspect images using a multi-resolution segmentation algorithm. Scale parameters, shape weight, and compactness weight should be set when performing multi-scale segmentation. The landslide

morphology and scale characteristics are combined with the literature's trial-and-error method (Huang et al. 2020b), and the parameters of scale, shape, and compactness are set as 20, 0.8, and 0.8, respectively, to extract the ideal slope unit.

Based on the multi-scale segmentation method, the study areas, including Ganzhou City, Ganzhou East, and Ruijin County, are divided into slope units of 296,003, 78,565, and 17,989, respectively. As positive sample data for slope unit research, Ganzhou City experienced a total of 3855 landslides, of which 366 occurred in Ruijin County and 1519 in Ganzhou East. The results are displayed in Table 3 and are consistent with grid units' method for the selection of non-landslide sample data and the division of training and test datasets.

6 Results of LSP under different scales of study area and mapping unit

This study addresses RF modelling of spatial datasets at various study area scales using the random forest package of the R programming language (Youssef and Pourghasemi 2021). The number of the factor features and trees, which can be obtained by automatic parameter screening and out-of-bag errors for the best parameters (Hong et al. 2019), is what primarily controls the RF's accuracy. The finding indicate that the number of factors features of 5 and the number of trees of 800 are the RF model's ideal parameters in Ganzhou City. Only a representative set of the parameters of other working conditions are presented because they do not significantly differ from those of this group. Then, using these parameters at various study area scales, LSIs for slope units and grid units with 30 m and 60 m resolution are predicted. By using the natural break point method, we finally categorize them into five classes of susceptibility zones (Li et al. 2021).

6.1 Landslide susceptibility results under different study area scales

To represent the various study area scales, Ganzhou City, Ganzhou East (the eastern portion of Ganzhou City), and Ruijin County in Ganzhou East were chosen. Along with that, LSMs from Ganzhou East and Ruijin County are also extracted by mask from that of Ganzhou City, as are LSMs from Ruijin from Ganzhou East. The rule of the aforementioned six different conditions is used to analyze the results of the study area scales used to determine the susceptibility to landslides. As an illustration, consider the results of RF

models for land-slide susceptibility based on slope units and grid units with a 30 m resolution. Figures 6 and 7 demonstrate that (1) As the size of the study area gradually shrinks, so do the areas of the low- and very low-susceptibility zones. (2) In addition, there are notable differences between the LSMs of Ruijin County extracted from the mask and those that were predicted. (3) Furthermore, the high susceptibility zones are primarily found in low and medium altitude regions below 400 m, have slopes of 8°–20°, have topographic relief of 3–12, and have a profile curvature of 3–18, according to overlay analysis between environmental factor maps and LSMs obtained by the RF model.

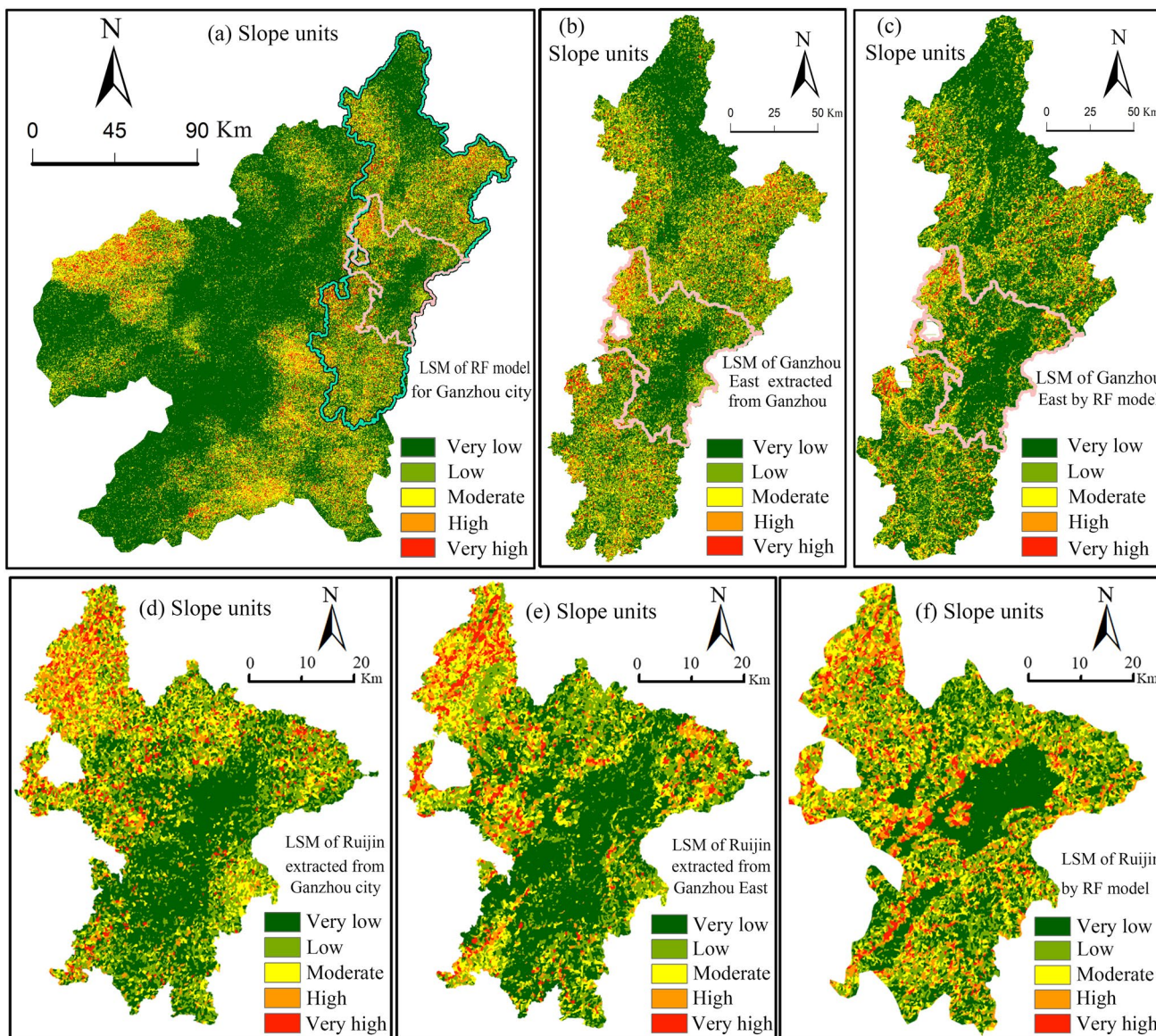


Fig. 6 LSMs under different study area scales **a** Ganzhou City of slope units and RF model **b** Ganzhou East extracted from Ganzhou City **c** Ganzhou East of slope units and RF model **d** Ruijin extracted

from Ganzhou City **e** Ruijin extracted from Ganzhou East **f** Ruijin County of slope units and RF model

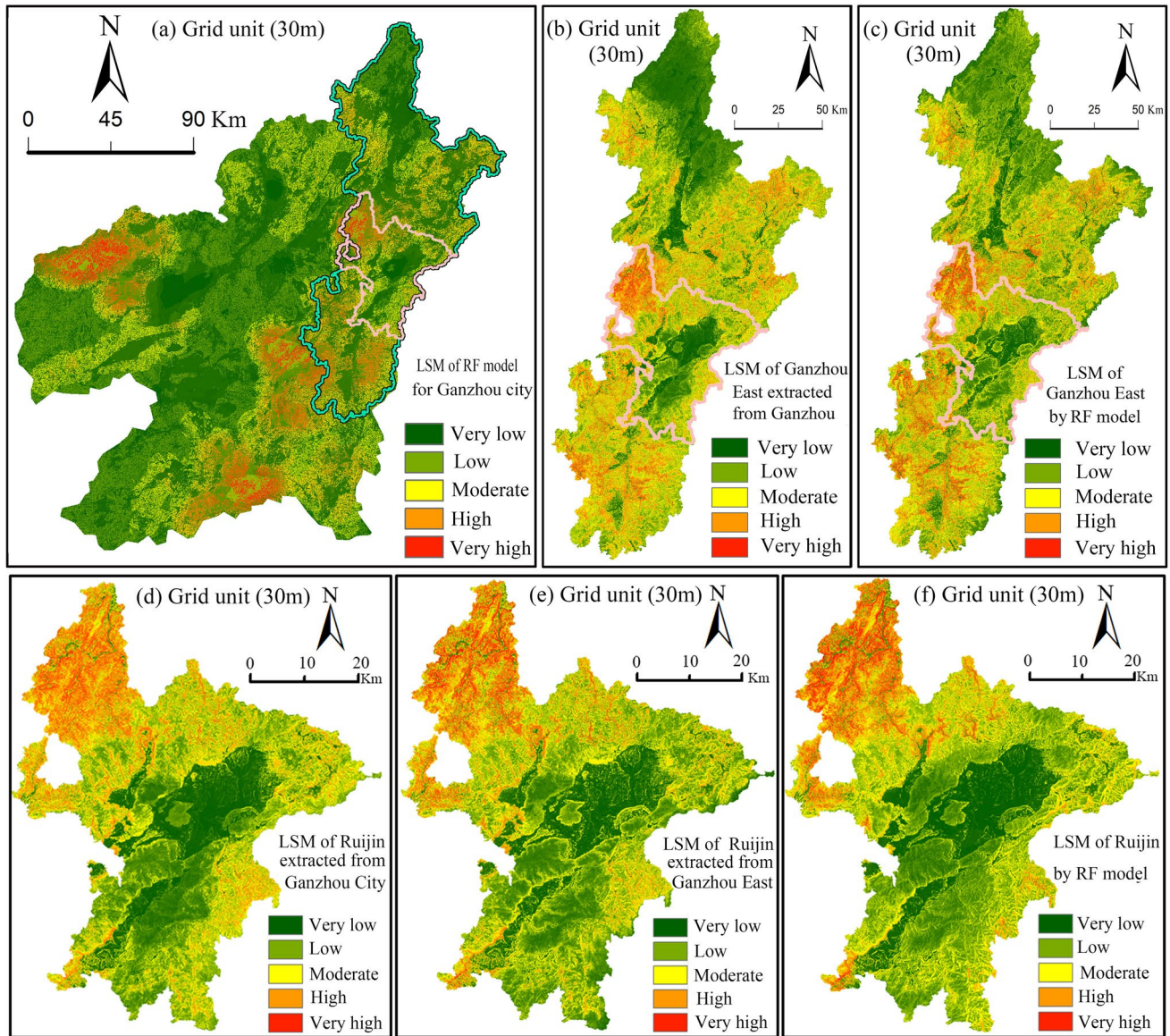


Fig. 7 LSM under different study area scales **a** Ganzhou City of 30 m grid units and RF model **b** Ganzhou East extracted from Ganzhou City **c** Ganzhou East of 30 m grid units and RF model **d** Ruijin

extracted from Ganzhou City **e** Ruijin extracted from Ganzhou East **f** Ruijin County of 30 m grid units and RF model

6.2 Landslide susceptibility results under different mapping unit scales

The different mapping unit scales are represented by the grid units with 30 m and 60 m resolutions as well as the slope units. The rules of the aforementioned various conditions are adopted to analyze the landslide susceptibility results under various mapping unit scales, using the results for Ruijin County under RF models, Ruijin County extracted from Ganzhou East, and Ruijin County extracted from Ganzhou City as examples. Figure 8 demonstrates that (1) the areas of the low- and very low-susceptibility zones grow as the

resolution of the grid units gradually declines. (2) In the same study area, very low to moderate susceptibility zones based on slope units are smaller than those based on grid units. (3) In addition, the slope units and grid units in the LSMs of Ruijin County differ significantly.

6.3 Uncertainties of LSP results under different scales

Accuracy and LSI distribution rule are used to express LSP uncertainties. The mean value of LSIs declines and the standard deviation of LSIs increases along with an increase

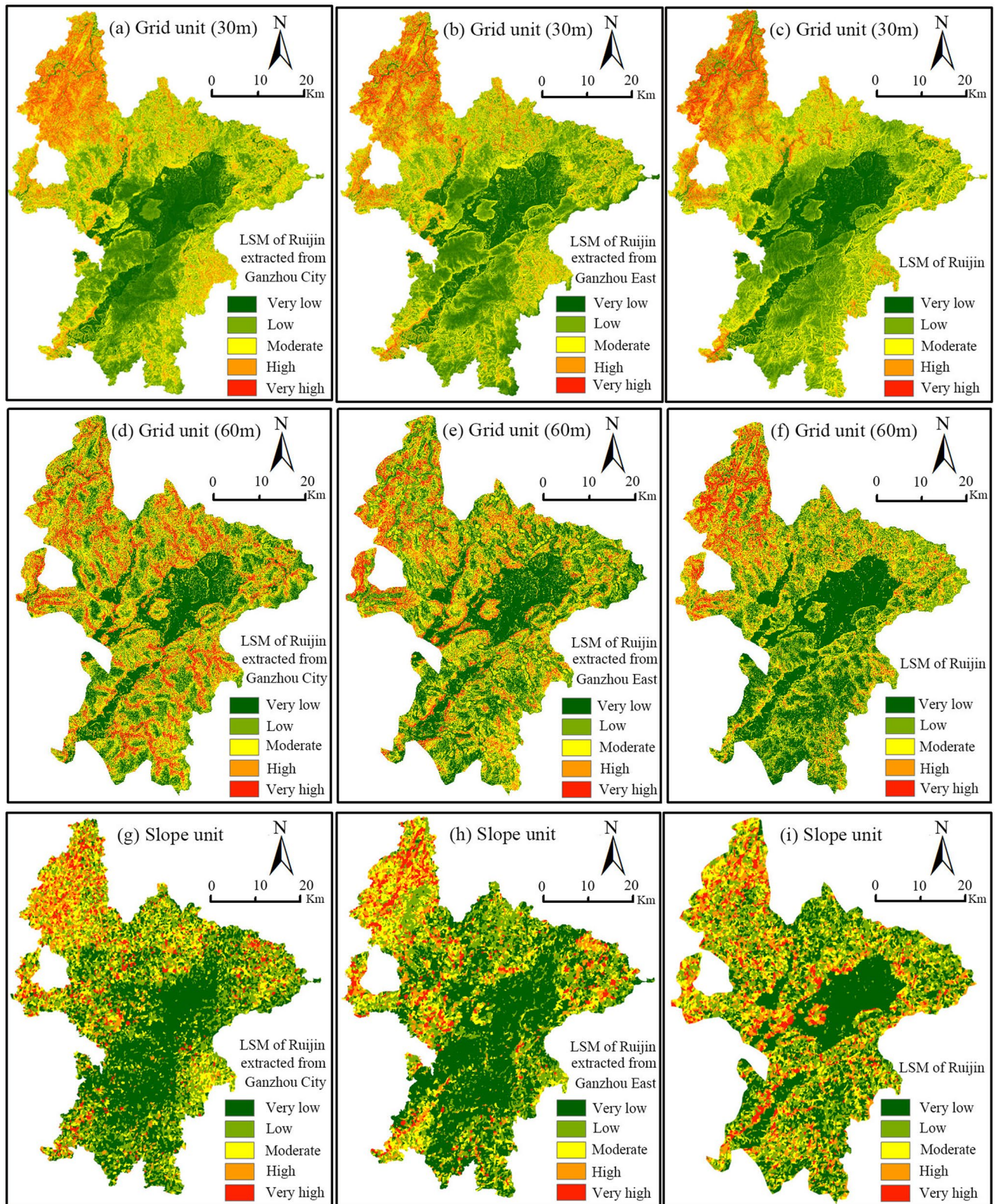


Fig. 8 LSMs of Ruijin County based on different mapping units under RF **a–c** Ruijin extracted from Ganzhou City, Ruijin extracted from Ganzhou East, and Ruijin County based on 30 m Grid units **d–f** Ruijin extracted from Ganzhou City, Ruijin extracted from Gan-

zhou East, and Ruijin County based on 60 m Grid units **g–i** Ruijin extracted from Ganzhou City, Ruijin extracted from Ganzhou East, and Ruijin County based on Slope units

Table 4 AUC accuracies of slope (grid) units in different study area scales based on RF (SVM) model

| AUC | SVM model | | | RF model | | |
|---|-------------|-------------------|-------------------|-------------|-------------------|-------------------|
| | Slope units | Grid units (30 m) | Grid units (60 m) | Slope units | Grid units (30 m) | Grid units (60 m) |
| Ganzhou City | 0.932 | 0.875 | 0.822 | 0.947 | 0.915 | 0.848 |
| Ganzhou East | 0.889 | 0.861 | 0.839 | 0.928 | 0.888 | 0.856 |
| Ganzhou East extracted from Ganzhou | 0.899 | 0.820 | 0.819 | 0.915 | 0.860 | 0.831 |
| Ruijin County | 0.934 | 0.882 | 0.873 | 0.936 | 0.918 | 0.884 |
| Ruijin East extracted from Ganzhou East | 0.932 | 0.854 | 0.831 | 0.939 | 0.880 | 0.838 |
| Ruijin East extracted from Ganzhou City | 0.946 | 0.848 | 0.820 | 0.949 | 0.881 | 0.839 |

in accuracy and efficiency, while the level of uncertainty in LSPs decreases. The key to assessing the impact of uncertainties on LSP is the availability of LSIs with high accuracy, low mean, and high standard deviation.

LSP's success depends on the evaluation of the modeling quality. The corresponding AUC values are used, and Table 4 displays the AUC values for slope and grid units (30 m/60 m) based on LSP modeling in study areas of various scales. With the reduction in study area scale from Ganzhou to Ruijin County, there is a tendency for the LSP accuracy of slope and grid (30 m/60 m) units in various study area scales to increase. For instance, the RF model's AUC values for slope units are 0.936 in Ruijin County, 0.928 in Ganzhou East, and 0.947 in Ganzhou City, respectively. Grid units (30 m) in the RF model have AUC values of 0.918. 0 in Ruijin County, 0.888 in Ganzhou East, and 0.915 in Ganzhou City. Grid units (60 m) in the RF model have the following AUC values: 0.884 in Ruijin County, 0.856 in Ganzhou East, and 0.848 in Ganzhou City. Additionally, the results for the LSP accuracy of Ruijin extracted from Ganzhou City, Ruijin extracted from Ganzhou East, and Ruijin County based on the mapping units under RF model are all the same.

The distribution rule of LSIs in the RF model is reflected in study areas with various scales and units using the mean value and standard deviation (Li et al. 2020). The mean value is used to measure the overall bias of LSIs, and the standard deviation shows how widely distributed LSIs are. In study areas with various scales, Table 5 displays the average values and standard deviation of slope and grid units (30 m/60 m) based on LSP modeling.

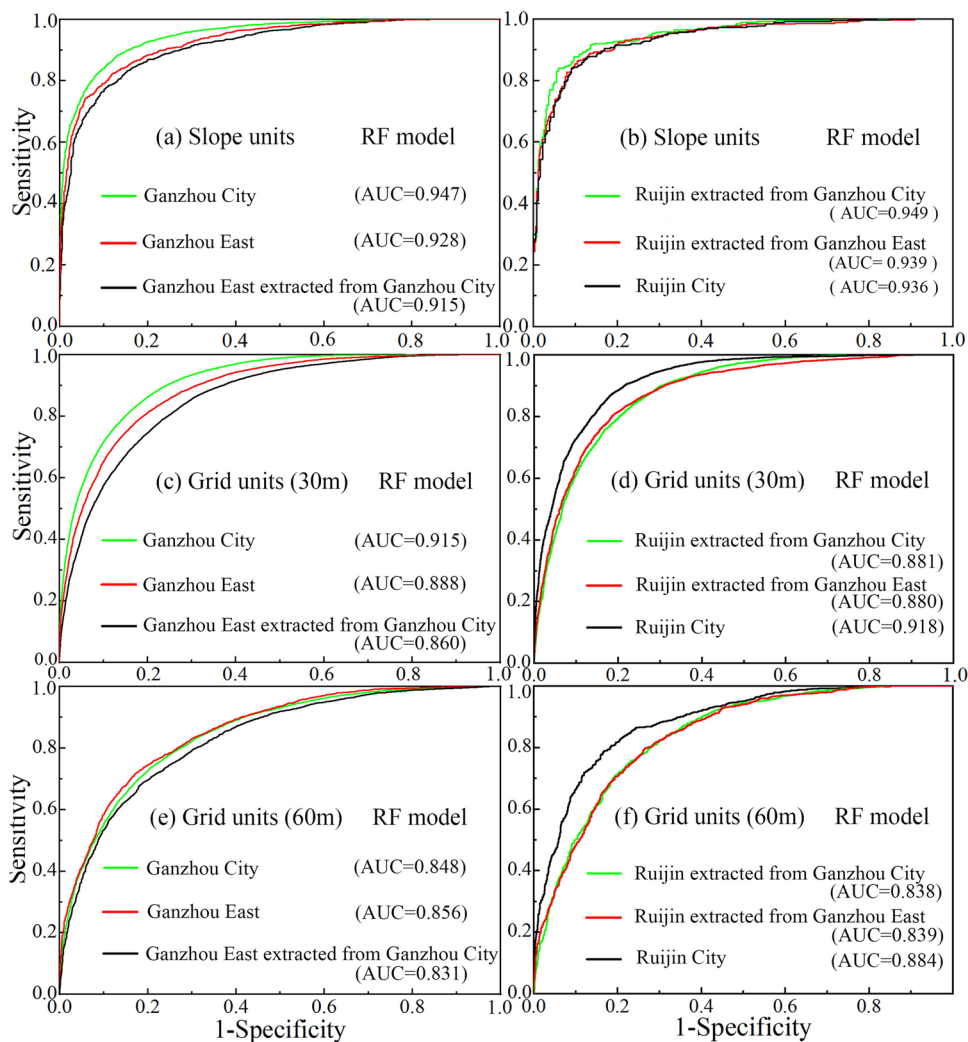
6.3.1 Accuracy assessment under different study area scales

The accuracy of LSIs predicted by the RF in study areas with different scales has an increasing tendency as the scale of the area decreases from Ganzhou to Ruijin County (Fig. 9). According to Figs. 9c and e, the AUC values for grid units (30 m/60 m) in the RF model are 0.918 (0.884) in Ruijin County, 0.888 (0.856) in Ganzhou East, and 0.915 (0.848) in Ganzhou City. The AUC values in Ganzhou East and City are 0.888 (0.860) and 0.856 (0.831), respectively. There is a tendency for the LSP accuracy to decrease between Ganzhou East and that extracted by masking from Ganzhou City. Ruijin County has the highest AUC accuracy,

Table 5 Mean and standard deviation values of mapping units in different study area scales based on RF and SVM

| Mean SD | SVM model | | | RF model | | |
|---|-------------|-------------------|-------------------|-------------|-------------------|-------------------|
| | Slope unit | Grid units (30 m) | Grid units (60 m) | Slope unit | Grid units (30 m) | Grid units (60 m) |
| Ganzhou City | 0.190/0.196 | 0.362/0.276 | 0.342/0.310 | 0.150/0.163 | 0.307/0.231 | 0.346/0.294 |
| Ganzhou East | 0.217/0.182 | 0.335/0.234 | 0.304/0.232 | 0.174/0.172 | 0.330/0.212 | 0.264/0.226 |
| Ganzhou East extracted from Ganzhou | 0.200/0.202 | 0.357/0.265 | 0.315/0.223 | 0.159/0.172 | 0.363/0.234 | 0.333/0.207 |
| Ruijin County | 0.300/0.282 | 0.280/0.272 | 0.243/0.258 | 0.268/0.221 | 0.279/0.273 | 0.226/0.216 |
| Ruijin East extracted from Ganzhou East | 0.273/0.219 | 0.333/0.275 | 0.323/0.254 | 0.240/0.228 | 0.325/0.227 | 0.276/0.211 |
| Ruijin East extracted from Ganzhou City | 0.350/0.293 | 0.361/0.285 | 0.336/0.231 | 0.243/0.221 | 0.368/0.250 | 0.340/0.201 |

Fig. 9 AUC values of different study area scales based on slope units and grid units with 30/60 m resolution under RF model **a–b** Slope units **c–d** Grid units with 30 m resolution **e–f** Grid units with 60 m resolution



followed by Ruijin extracted from Ganzhou East and Ruijin extracted from Ganzhou City (Figs. 9b, d and f). The LSP accuracy of Ruijin County and those extracted by masking from Ganzhou East and Ganzhou City are significantly different.

6.3.2 Accuracy assessment under different mapping unit scales

The effects of slope units and grid units with 30 m and 60 m resolution on the accuracy of LSP modeling are discussed.

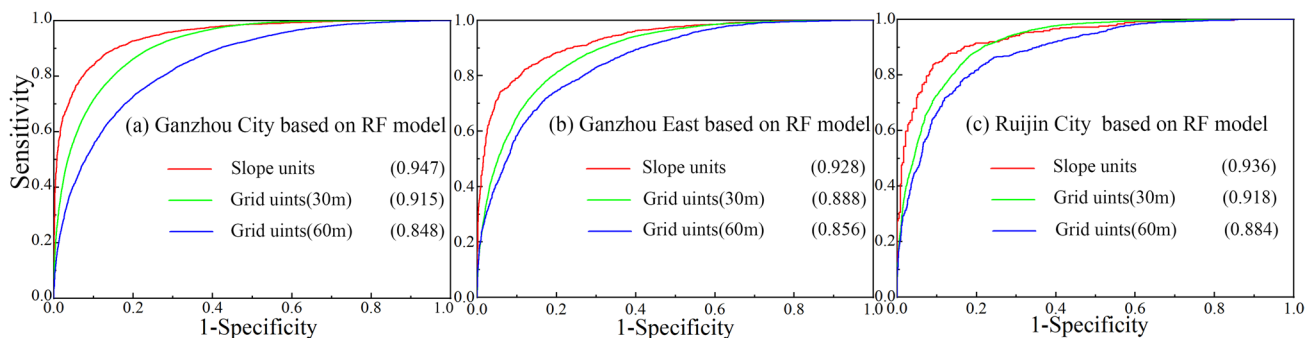


Fig. 10 AUC values of different mapping units in different study area scales based on RF model **a** Ganzhou City **b** Ganzhou East **c** Ruijin County

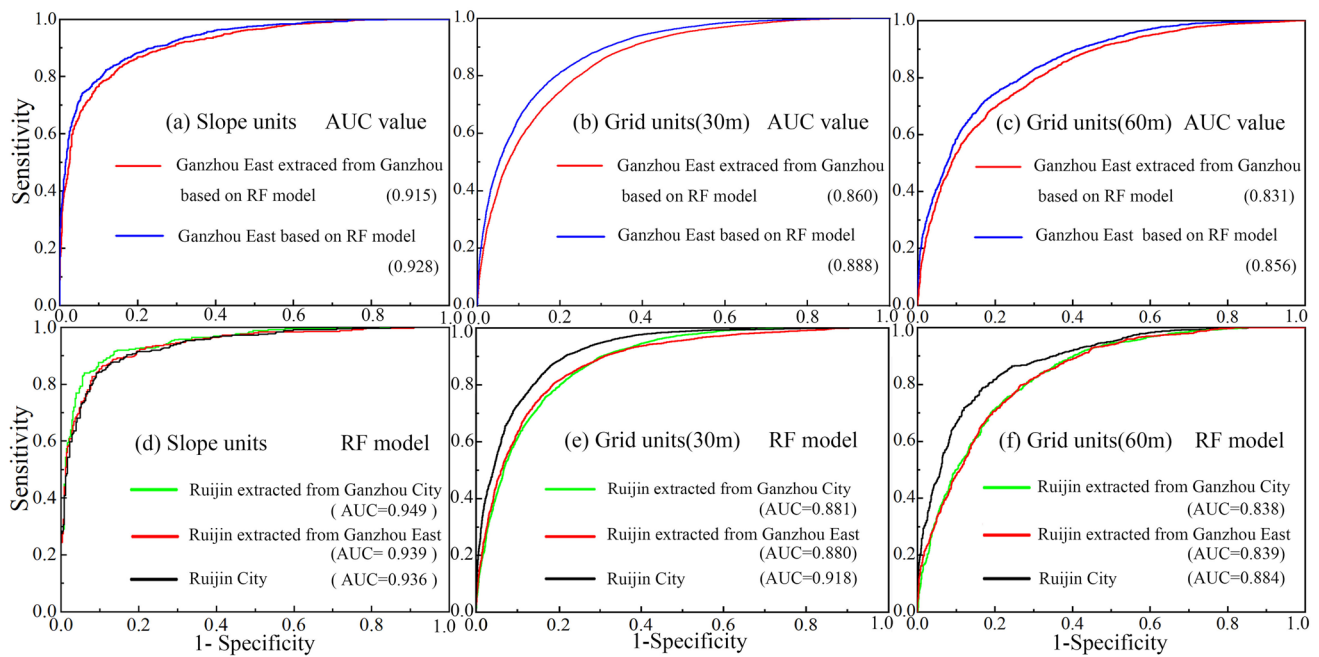


Fig. 11 AUC values of different mapping units in Ganzhou East and Ruijin County based on RF model **a, d** Slope units **b, e** Grid units with 30 m resolution **c, f** Grid units with 60 m resolution

These are the findings:

- (1) The LSP accuracies of slope units extracted by Multi-scale Segmentation method are higher than those of grid units, and the efficiency of LSIs has significantly increased, along with a reduction in the scale of the region from Ganzhou to Ruijin County. The AUC values for the slope-based RF model (grid units with a 30 m resolution) are 0.936 (0.918) in Ruijin County, 0.928 (0.888) in Ganzhou East, and 0.947 (0.915) in Ganzhou City, respectively (Fig. 10). The LSP accuracy of slope units based on RF model is higher than that of grid units in Ruijin County and that extracted by masking from Ganzhou East and Ganzhou City (Fig. 10), and the LSP accuracy of slope units in Ganzhou East and that extracted by masking from Ganzhou City is also better (Fig. 11).
- (2) There are significant differences between LSP accuracy of grid units with 30 m and 60 m resolutions in Ruijin County and that extracted by masking from Ganzhou East and Ganzhou City. Grid units (30 m/60 m) in the RF model have AUC values that are 0.918 (0.884) in Ruijin County, 0.888 (0.856) in Ganzhou East, and 0.915 (0.848) in Ganzhou City (Fig. 11). The AUC values in Ganzhou East and Ganzhou City are 0.888

(0.860) and 0.856 (0.831), respectively. There is a tendency for the LSP accuracy to decrease between Ganzhou East and that extracted by masking from Ganzhou City (Fig. 11). The LSP accuracy tends to decline in Ruijin County and that which was extracted by masking from Ganzhou East and Ganzhou City.

In conclusion, the LSP modeling shows that the LSP accuracy significantly rises as the size of the study areas decreases from Ganzhou City to Ruijin County. The LSP accuracy in Ruijin County that was extracted by masking from Ganzhou City is, in turn, less accurate than that in Ganzhou East and that in Ruijin County that was directly predicted. Furthermore, Ganzhou East's LSP accuracy, which was obtained by masking Ganzhou City, is lower than it is for Ganzhou East with direct prediction, indicating LSP accuracies that are comparable to those of Ruijin County. The RF and SVM models in Ruijin County have the highest LSP accuracy, with the LSP accuracy in study areas with different scales tending to increase as study area scales are reduced. In addition, slope units extracted using the multi-scale segmentation method have higher LSP accuracy than grid units with a resolution of 30 m.

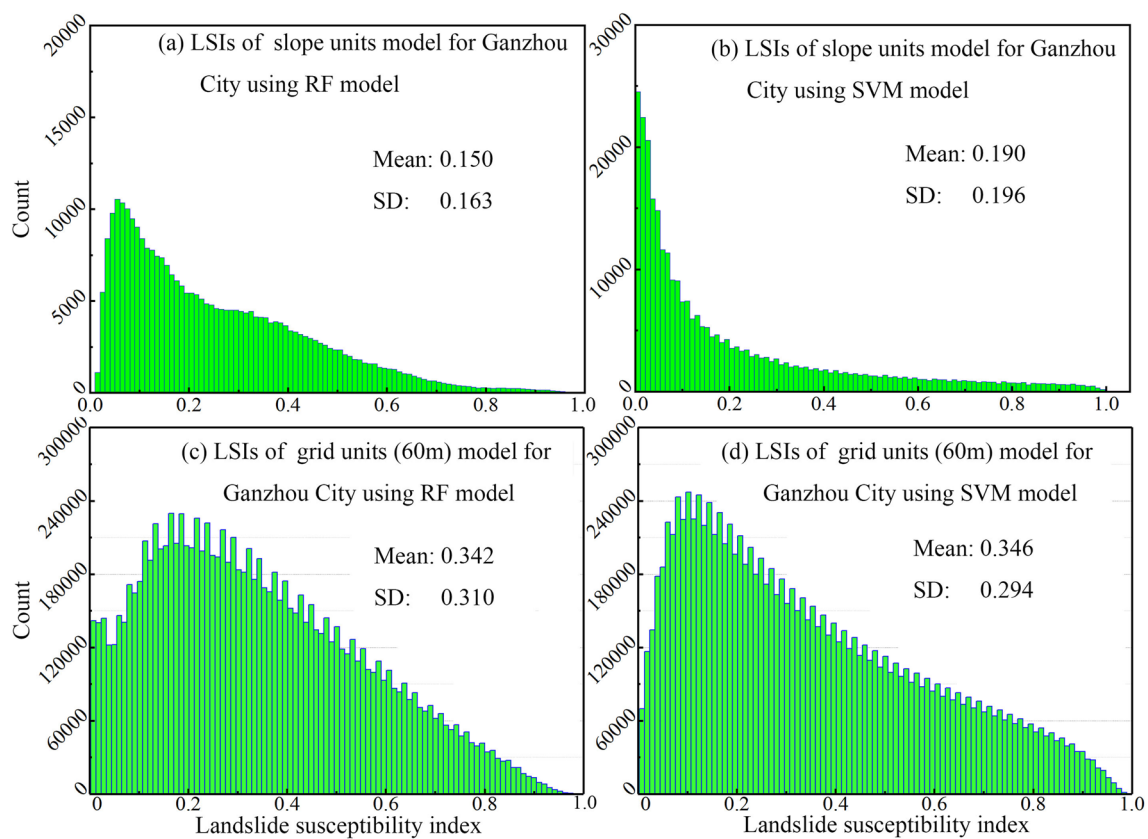


Fig. 12 The distribution of LSIs with RF (SVM) model in Ganzhou City

6.3.3 Distribution of LSIs under different study area scales

In comparison to those extracted by masking from Ganzhou City and Ganzhou East, the distribution rule of the LSIs directly predicted in Ruijin County is significantly different. Slope units, for instance, have mean values of 0.243 in Ruijin County, extracted from Ganzhou City, 0.240 in Ruijin County, extracted from Ganzhou East, and 0.268 in Ruijin County, directly predicted by Ruijin County (Figs. 12 and 13). Additionally, the standard deviation illustrates how, under the same circumstances, the LSI distribution differs from the mean value. The distribution of LSIs obtained by direct prediction in Ruijin County is more consistent with the actual distribution of landslide susceptibility, and there are significant differences between the extracted and directly predicted LSIs there. In contrast to that extracted by masking from Ganzhou City, which is consistent overall, the distribution rule of LSIs in Ganzhou East with direct prediction is more consistent with the actual distribution of landslide susceptibility.

6.3.4 Distribution of LSIs under different mapping unit scales

The mean value of LSIs tends to decrease as the scale of study area decreases for grid units, whereas slope units show the opposite trend in accordance with the influences of study areas with different scales on the distribution rule of LSIs (Table 5). Take grid units with 60 m resolution as example (Fig. 12). In study areas with various scales, the mean values of the LSIs in the RF model are 0.342 in Ganzhou City, 0.264 in Ganzhou East, and 0.226 in Ruijin County. The standard deviation, however, exhibits the mean's opposite trend. When the scale of the study area is reduced for grid units, the mean value of the LSIs rises while the standard deviation falls. The trend for slope units is the opposite, and they are more accurate, have a higher mean value, and are more effective than grid units (Fig. 12).

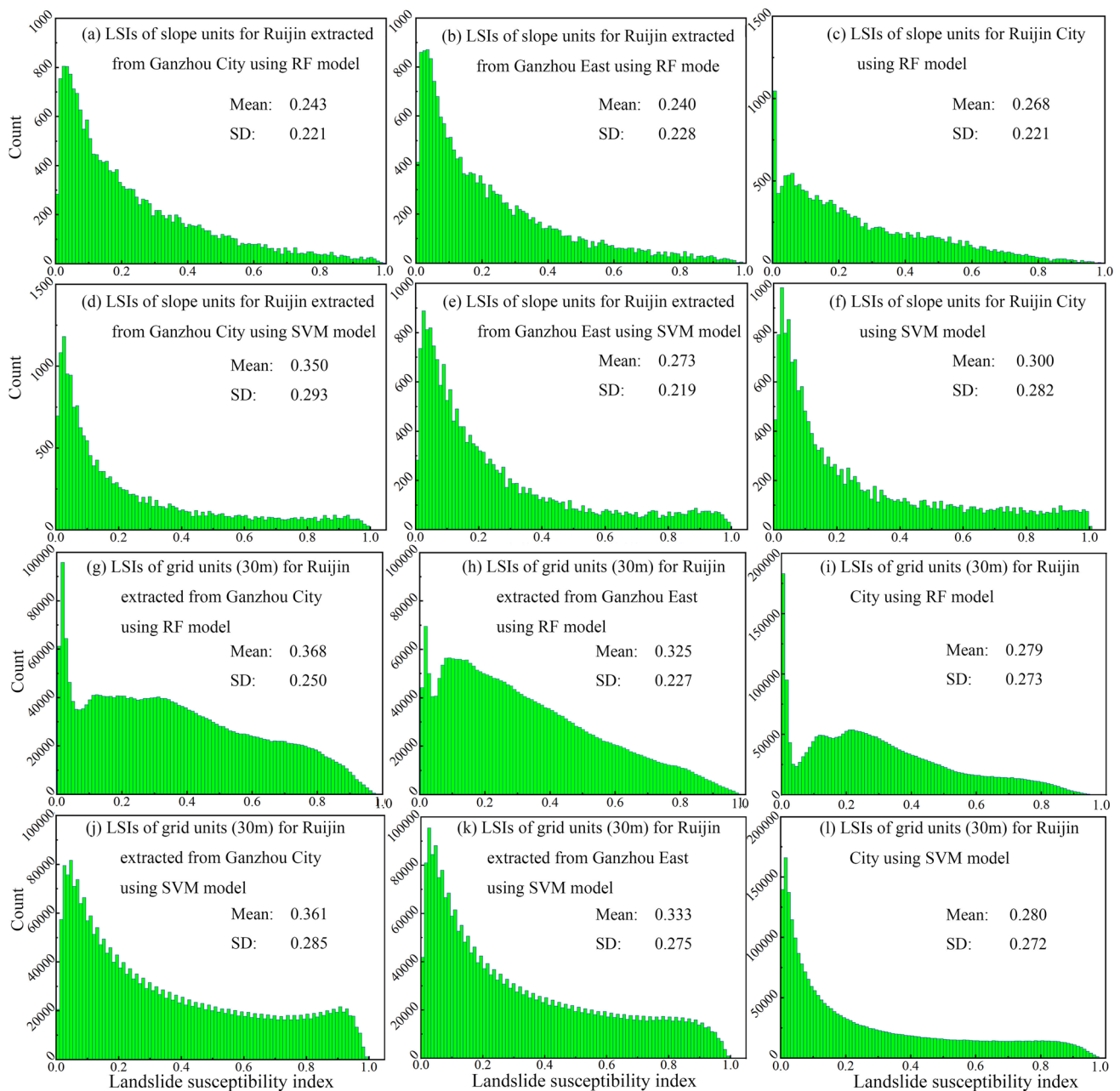


Fig. 13 The distribution of LSIs with RF (SVM) model in Ruijin County in study areas with different scales **a–f** Slope units **g–l** Grid units with 30 m resolution

7 Discussion

In general, the statistical results of big data in a larger study area scale should be more consistent with the actual situation because there are more landslide samples, and the results of LSP can better reflect the general rule (Araújo et al. 2022). LSP results under various scales using SVM model are set as the second case to demonstrate the generalizability of

small study area scale and slope unit as mapping units with better LSP accuracy and smaller uncertainty in LSP modeling, and small study area scale can better reflect the specific law in the local range. This is because the accuracy on the resulting model is inconsistent due to the need for more detailed data for a larger study area scale (Kulsoom et al. 2023). Therefore, it is expected to provide higher modeling accuracy and effectiveness in large study areas scale with

detailed landslide samples and diverse environmental factors (Gupta and Shukla 2023). For more accurate environmental factors and better LSP accuracy, the study area must be scaled appropriately at this stage (Pellicani et al. 2017). The importance of environmental factors and the accuracy of the LSP are discussed in this study from the perspectives of study areas at various scales.

7.1 LSP results under different scales using SVM model

For SVM modeling, SPSS modeler 18.0 software is also used. The SVM's kernel function is chosen to be the popular radial basis kernel function (Huang and Zhao 2018). Debugging is used to determine the best regular parameter (C) and kernel parameter (γ), with the default values being applied to the other parameters. The trained SVM is used to acquire LSIs in the study area, and the spatial superposition analysis in ArcGIS is used to determine the distribution of LSIs in study areas with different scales. Additionally, the natural break point method categorizes

LSIs into five classes: very low, low, moderate, high, and very high susceptibility zones.

7.1.1 Landslide susceptibility results under different study area scales

The LSMs of Ganzhou City, Ganzhou East, and Ruijin County in Ganzhou East using SVM models built on grid units with a 60 m resolution are considered as an illustration. Test results for landslide susceptibility at various study area scales revealed that as the study area's scale gradually shrank, so did the areas of the low- and very low-susceptibility zones. With a decrease in scale from Ganzhou to Ruijin County, the LSP accuracy of grid (60 m) units in different study area scales based on SVM model tends to increase (Fig. 14). Grid (60 m) units in Ruijin County that were extracted by masking Ganzhou City have a lower LSP accuracy than grid (60 m) units directly predicted in Ruijin County and extracted by masking Ganzhou East. Table 5 illustrates variations in the LSI distribution rule predicted by the RF and SVM models at various study area scales.

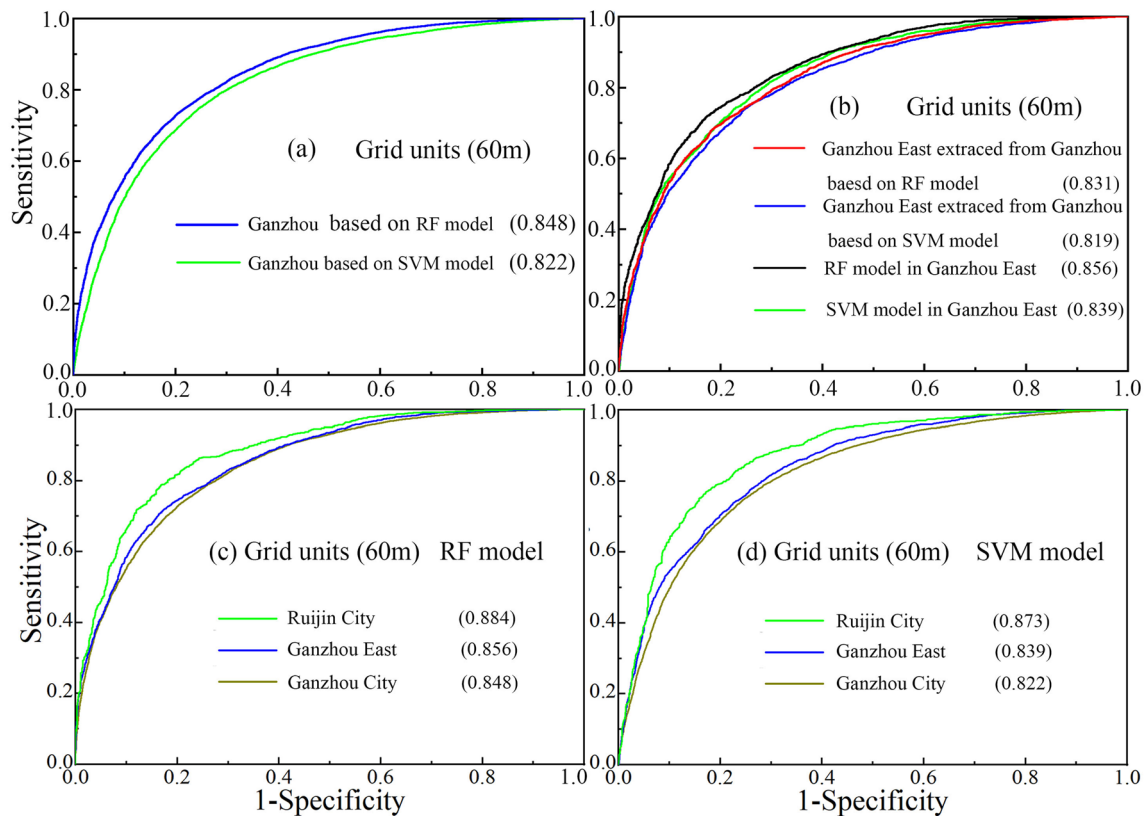
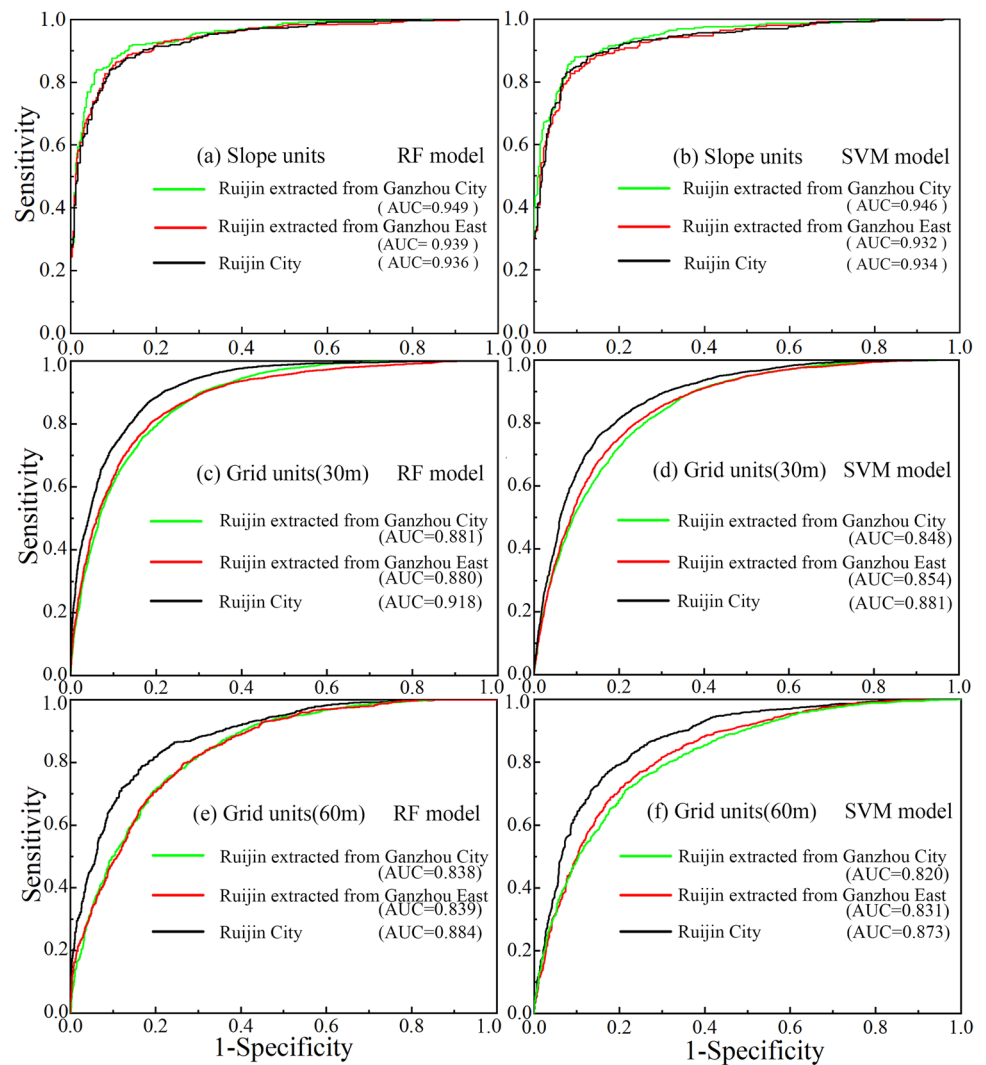


Fig. 14 AUC values of grid (60 m) units in different study area scales based on RF and SVM model **a** Ganzhou City **b** Ganzhou East and Ganzhou East extracted from Ganzhou City **c** Ruijin County, Gan-

zhou East and Ganzhou City based on RF model **d** Ruijin County, Ganzhou East and Ganzhou City based on SVM model

Fig. 15 AUC values of different mapping units in Ruijin County **a, c, e** RF model **b, d, f** SVM model



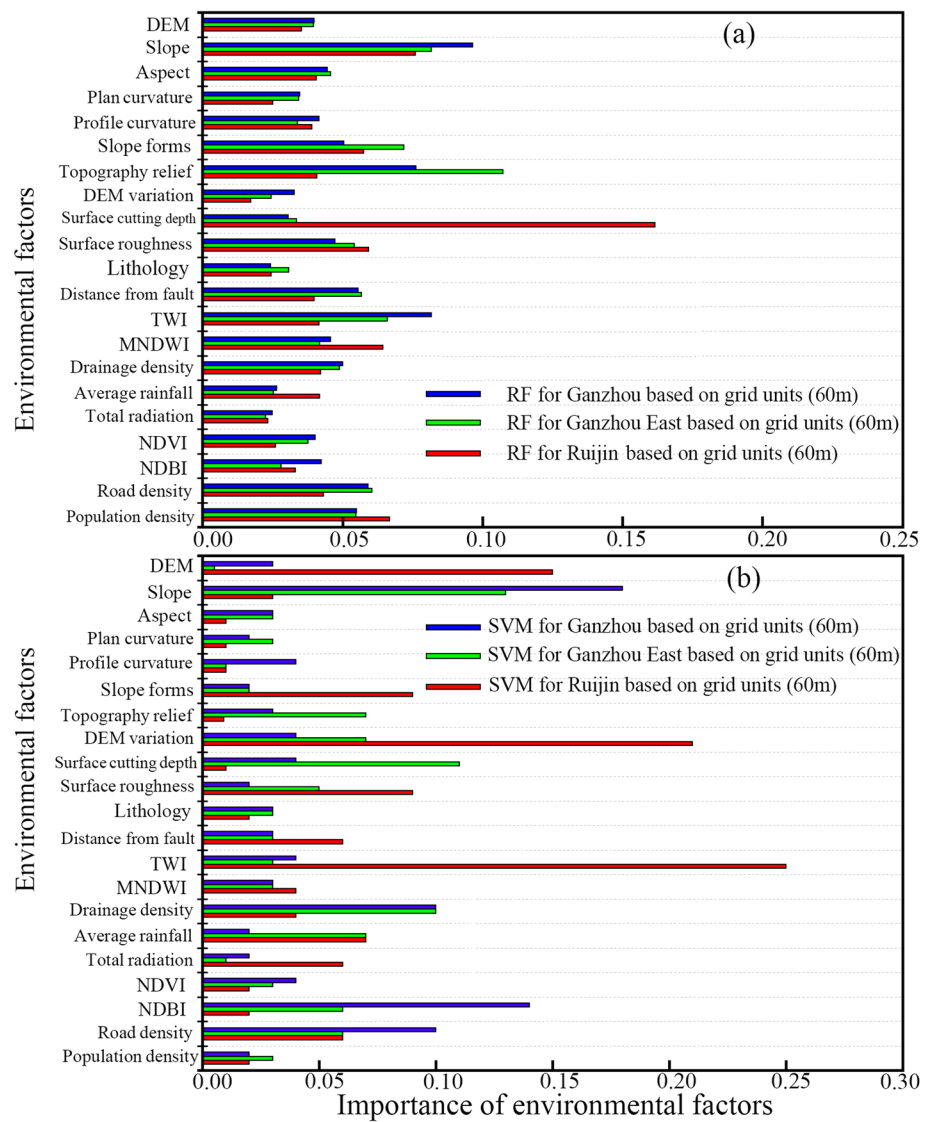
7.1.2 Landslide susceptibility results under different mapping unit scales

The slope units in Ruijin extracted from Ganzhou City, Ruijin extracted from Ganzhou East, and Ruijin County under SVM models as examples, as well as the LSMs of grid units with 30 m and 60 m resolutions. The results of the SVM model's analysis of the susceptibility to landslides at various mapping scales revealed that the very low to moderate susceptibility zones in the same study area are smaller when based on slope units than when based on grid (30 m/60 m) units. In the same study area, the LSP accuracy based on the SVM model gradually improves from grid units (60 m) and grid units (30 m) to slope units (Fig. 15). The results in the RF model and the distribution rule of the LSIs based on the SVM model under mapping unit scales are shown in Fig. 13.

7.2 Importance ranking of landslide related environmental factors

The correlation analysis tool in the SPSS 24.0 software is used to calculate the correlation coefficients of the 21 environmental factors. According to the findings, there are weak correlations between these environmental factors, as evidenced by correlation coefficients that are less than 0.5 and a significance level of less than 0.05 (Huang and Zhao 2018). Each environmental factor is also subject to a collinearity diagnosis. The multicollinearity issues among the environmental factors by removing highly correlated variables and estimating the variance inflation factor (VIF) by maintaining the threshold value (Zeng et al. 2023). A regression line was fitted between each predictor variable and the other predictor variables to obtain the VIF by measuring the square of multiple correlation coefficients (Achu et al. 2023). The variance inflation factor and tolerance

Fig. 16 Environmental factors importance under different study area scales and 60 m grid units **a** RF **b** SVM

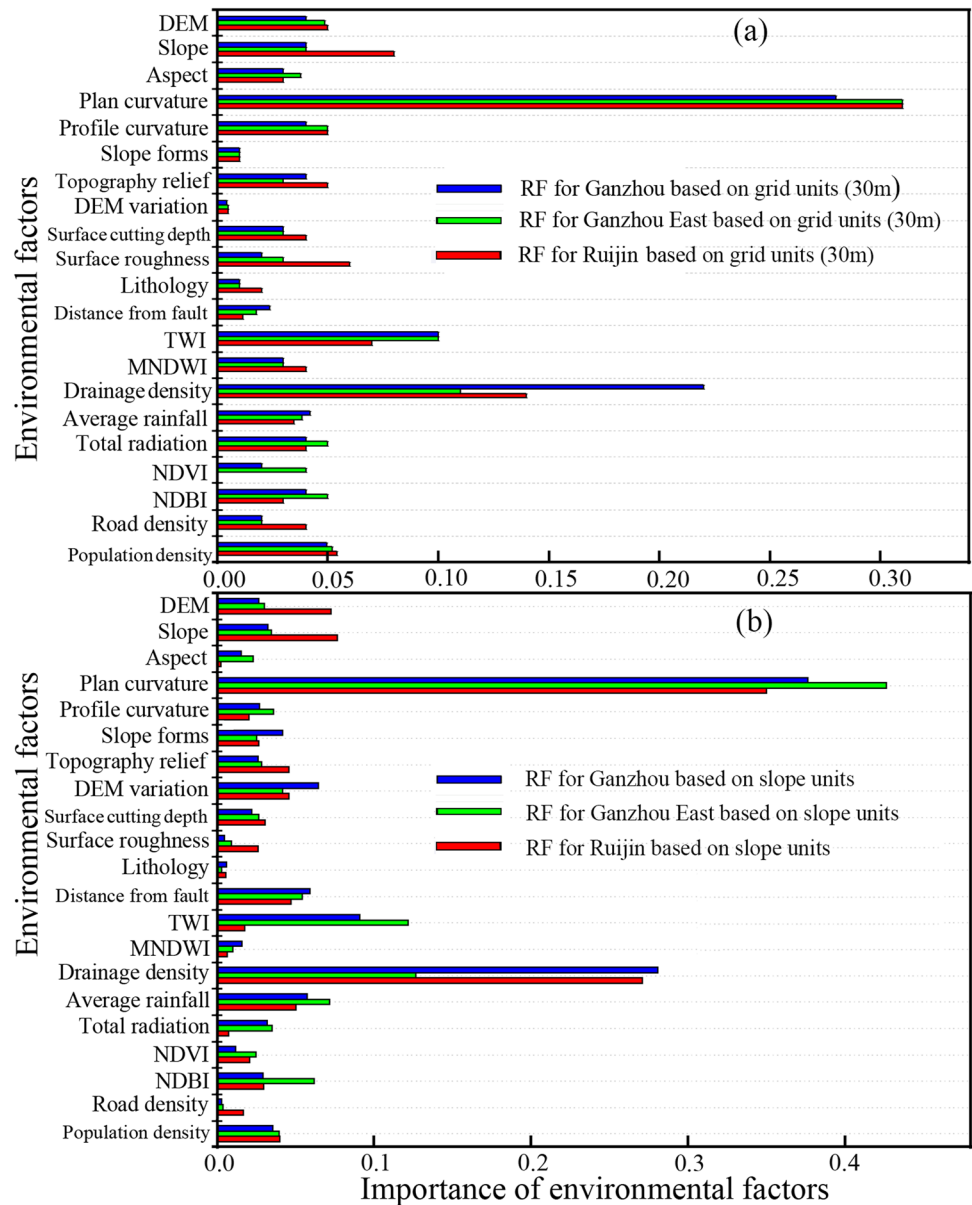


are used to determine the degree of collinearity between these environmental factors, with a serious level of collinearity being indicated by a variance inflation factor greater than 5 and/or a tolerance lower than 0.1. The findings demonstrate that there is no multi-collinearity among these environmental factors, with the maximum value of the variance inflation factor and the minimum value of the tolerance being 4.35 and 0.29, respectively. Therefore, all the 21 environmental factors can be used in this study for LSP modelling. Meanwhile, some studies indicate that error levels in environmental factors have a great influence on the LSP results. For example, Huang et al. (2023a) found that the greater the proportion of random error levels in environmental factors, the greater the uncertainty of LSP results. These original continuous environmental factors are processed by eliminating the random errors using low-pass filter method.

7.2.1 Importance of environmental factors in different study area scales

The importance prediction of the SVM and the mean decreasing accuracy of the RF model are used to rank the significance of environmental factors at various study area scales. Figure 16a illustrates the mean decreasing accuracy in the RF model based on grid units with a 60 m resolution. In Ganzhou City, topography relief has the highest importance value of 0.12, followed by slope with a value of 0.08 and TWI with a value of 0.07; in Ganzhou East, topography relief has the highest importance value of 0.16, followed by surface cutting depth with a value of 0.16. In the SVM based on 60 m resolution, Fig. 16b illustrates the significance of environmental factors in each study area. In Ganzhou City, the slope is of the utmost importance with a value of 0.18,

Fig. 17 Environmental factors importance under different study area scales and different mapping unit scales based on RF **a** Grid units (30 m) **b** Slope units



followed by the normalized differential building index with a value of 0.14, the road density, and drainage density with a value of 0.10; in Ganzhou East, the slope is of the utmost importance with a value of 0.13, followed by surface cutting depth with a value of 0.11, and drainage density with a value of 0.10; In Ganzhou East, slope has the highest importance with a value of 0.13, followed by surface cutting depth with a value of 0.11 and drainage density with a value of 0.10; In Ruijin City, terrain wetness indexes has the highest importance with a value of 0.25, followed by the elevation variation coefficient with a value of 0.21, average annual rainfall with a value of 0.16, road density with a value of 0.09, faults with a value of 0.07 and lithology with a value of 0.06.

7.2.2 Importance of environmental factors in different mapping-units scales

While it varies significantly from the 60 m resolution grid unit, the importance of environmental factors in the 30 m resolution grid unit and the slope unit tends to be consistent to some extent. Figure 17a, using the RF model as an illustration, demonstrates that plan curvature, which has the highest importance value of 0.30, is followed by drainage density, which has a mean value of 0.15, and TWI, which has a mean value of 0.08 at various study area scales based on grid units with 30 m resolutions. Meanwhile, plan curvature, drainage density, and TWI have respective mean

values of 0.38, 0.20, and 0.08 at various study area scales Fig. 17b depicts that there is no difference in the importance of environmental factors at different study area scales based on the grid units with 60 m resolutions. In Ganzhou City, topography relief, slope, and TWI have importance values of 0.12, 0.08, and 0.07, respectively, while surface cutting depth, slope, and elevation variation coefficient have importance values of 0.10, 0.08, and 0.07, respectively.

7.3 Discussions about influence of different scales on LSP modelling

7.3.1 Influence of study area scales on importance of environmental factors

The weights of environmental factors in study areas with various scales differ obviously when the same data-driven model and the same environmental factors are used in LSP modeling (Huang et al. 2021b). It is easy to cause the average phenomenon of environmental factors in the local area or ignore the local characteristics when the study area scale is very large. Data mining or statistical rules are prone to distortion with the increase of the study area scale. However, it will be more closely matched with local landslide characteristics. The results also prove that the greater the diversity of environmental factors, the more LSP accuracy of data-driven model. The weight of environmental factors with slow spatial variation, such as lithology and land cover, becomes minimal for the LSP in Ruijin County (Hurlimann et al. 2022). The importance of environmental factors is significantly different, and the weights of the terrain wetness index and elevation variation coefficient increase, which can lead to better LSP results.

However, in a large-scale study area like Ganzhou City, the weights of the terrain, road density, drainage density, and normalized differential building index increase while the DEM variation and surface roughness decrease, and the significance of environmental factors tends to be homogeneous. The scale of the study area directly affects how important landslide environmental factors are ranked in LSP modeling in study areas of various scales (Salciarini et al. 2016). The significance of environmental factors exhibits the phenomenon of "averageness" to take into account the characteristics of landslide evolution in a larger study area as the scale of the study area gradually increases (Grandjean et al. 2018). The differences between landslides and non-landslides in the training samples become more pronounced as the size of the study area shrinks, and there are more significant environmental factors or combinations of them that play a significant role in the evolution of landslides, which can produce better LSP results faster (Guo et al. 2022).

The importance of landslide environmental factors is directly impacted by the scale of the study areas, which shifts the environmental factors that were initially used to represent Ganzhou City's susceptibility in the surrounding areas of Ganzhou East and Ruijin County. The importance of environmental factors varies at this time, which helps to better reveal the spatial variation in LSP (Palau et al. 2022).

7.3.2 Influence of mapping unit scales on importance of environmental factors

A mapping unit, at the level of the analysis, is a geographical region that captures environmental factors (Liu 2022). The initial selection of a suitable mapping unit that can express information about environmental factors is necessary for LSP modeling (Kedron and Holler 2022). Different mapping unit scales are represented by grid units with spatial resolutions of 30 and 60 m, as well as slope units that were extracted using the multi-scale segmentation method from study areas with various scales. Secondly, it has been discovered that mapping unit scales significantly affect how environmental factors are expressed. Plan curvature, drainage density, and TWI weights based on slope units and grid units with 30 m spatial resolutions increase as study area scales change. The importance of environmental factors under grid units with 30 m spatial resolutions and slope units based on RF and SVM models tends to be somewhat consistent. The importance of environmental factors in various study areas has also been discovered to differ significantly between grid units with 30 m resolution and 60 m resolution. Based on grid units with a resolution of 60 m, environmental factors are equally important and clearly regulated at various study area scales. The excellent slope units make up for this since the grid resolution directly affects how important environmental factors are at various study area scales. The study of the mapping unit scale seeks to strike a balance between the maximum possible accuracy of LSP and the maximum possible efficiency of environmental factor expression. It also aims to determine the best way to make use of the mapping units' effect in order to maximize LSP's effectiveness.

In study areas using various scales and/or models, the differences in the significance of environmental factors are generally significant. Plan curvature, slope, and DEM significantly contribute to the RF and SVM models for Ruijin County, and plan curvature, terrain wetness indices, and drainage density significantly contribute to the models for Ganzhou East and Ganzhou City. The importance of environmental factors varies greatly between the different mapping units, but it is generally consistent in the slope unit and the 30 m resolution grid unit. Additionally, there is a gradual decline in the differences in the weights given to various environmental factors from Ruijin County to Ganzhou East

and Ganzhou City, indicating that there is a general trend toward an "average" weighting of environmental factors from Ruijin County to Ganzhou East and Ganzhou City.

7.3.3 Influence of study areas scales on LSP accuracy

A challenging issue that significantly hinders the applications of machine learning model to large scale areas is the data incompleteness in most landslide inventories, particularly the lack of detailed and accurate landslide inventory data that is a vital link between each landslide and its environmental factors (Sun et al. 2023; Xiao and Zhang 2023). Hence, we have to reduce the study area scale in exchange for the prediction accuracy of the local region (Kulsoom et al. 2023; Moeen Hamid Bukhari et al. 2023). The Ganzhou City landslide inventory data and 21 landslide environmental factors are used in modeling to produce the LSP results. In order to obtain LSP results, Ganzhou City also extracts the landslide and environmental factor data for Ganzhou East and Ruijin County. A smaller scale of the study area results in a higher LSP accuracy, and a lower mean value of the LSIs indicates a better LSP performance under the same mapping unit scale and environmental factors with three different scales of the study area (Deng et al. 2022; Kirschbaum et al. 2011). The grid units in the RF (SVM) model have AUC accuracies of 0.848 (0.822), 0.856 (0.839), and 0.884 (0.873) in Ganzhou City, Ganzhou East, and Ruijin County, respectively.

There is a sizable difference in LSP accuracy when small-scale areas are extracted from larger-scale areas using the same data-driven model and the same environmental factors. Ruijin County between direct prediction and mask extraction from Ganzhou City and Ganzhou East is taken as an illustration. The corresponding mean values of LSIs in the RF (SVM) model are 0.368 (0.361), 0.325 (0.333), and 0.279 (0.280), with an increasing standard deviation trend. The AUC values of grid units with resolution of 30 m in the RF (SVM) model are 0.881 (0.848) in Ruijin extracted from Ganzhou City, 0.880 (0.854) in Ruijin extracted from Ganzhou East, and 0.918 (0.882). The LSIs in Ruijin County under the three different study area scales differ significantly, and the LSIs with direct prediction in Ruijin County are more trustworthy. Additionally, the accuracy, LSI distribution rule, and significant difference in study areas with various scales in Ganzhou East are identical to those in Ruijin County. This demonstrates that as the scale of the study area decreases, the mean value of the LSIs decreases, and the standard deviation tends to increase, the LSP accuracy of grid units in study areas with different scales increases. The LSP accuracy of slope units in study areas with various scales is superior to grid units and has little effect on scale reduction (Alvioli et al. 2022; Huang et al. 2021a).

In order to increase the number of samples and lower the LSP's level of uncertainty, the study area was enlarged from Ruijin County to Ganzhou City. The findings show that, despite an increase in the number of landslide samples due to the study area expansion, the rule of spatial heterogeneity in geography is broken by incomplete sample information (Yu et al. 2022). On the other hand, the prediction performance is better in the smaller study area.

7.3.4 Influence of mapping unit scales on LSP accuracy

The accuracy and efficiency of slope units divided by high-resolution grid data are less affected by the scale of the study area. Whereas the trend for grid units is the opposite, the LSP accuracy of grid units increases by about 5% as the study area's scale decreases (Liu et al. 2023). In the RF (SVM) model, grid units with a resolution of 60 m have AUC accuracies of 0.848 (0.822), 0.856 (0.839), and 0.884 (0.873), respectively, while grid units with a resolution of 30 m have AUC accuracies of 0.915 (0.875), 0.888 (0.861), and 0.918 (0.882), respectively. Additionally, with an increasing standard deviation, the mean LSI values for grid units with a resolution of 30 m (60 m) in the RF model are 0.342 (0.346) in Ganzhou City, 0.264 (0.304) in Ganzhou East, and 0.226 (0.243) in Ruijin County. The AUC accuracies of slope units are better than grid units with resolution of 30 m and 60 m, with values of 0.947 (0.932) in Ganzhou City, 0.928 (0.889) in Ganzhou East, and 0.936 (0.934) in Ruijin County. The LSP accuracy of slope units in study areas with various scales is superior to grid units and has little effect on scale reduction (Alvioli et al. 2022; Huang et al. 2021a).

7.4 Future study prospects

In conclusion, a smaller study area scale from Ganzhou City to Ruijin County improves prediction accuracy. To better reveal the spatial variability of geographical phenomena, the next step is to study the township level, taking into account regional variations in towns and counties, shifting the environmental factors originally used to reflect global relationships to reflect each township area, and comparing the scale effect of township area (Gariano et al. 2017; Palau et al. 2022; Yu et al. 2022). To produce a regional LSM that is accurate to the objective reality, it is important to select an appropriate study area scale that takes into account both the spatial correlation and spatial heterogeneity of the geographic environment. Applying the combination of multi-scale segmentation and physically-based numerical modeling to suitable study area scales for disaster risk management will require additional work. The resolutions of 30 m and 60 m are representative for the grid

units in various study area scales. The uncertainty rules of LSP under grid units with 30 m and 60 m resolution can be well compared with those under the slope unit. Additionally, it is important to investigate the uncertainty rule of LSP modeling at various spatial resolutions, including mapping units with 10 m, 15 m, 30 m, 60 m, 90 m and 120 m resolutions.

8 Conclusions

In the present study, the effects of mapping unit scales and study area scales on the uncertainty rules of LSP are introduced. The study areas are Ganzhou City, the Eastern portion of Ganzhou City (Ganzhou East), and Ruijin County in Ganzhou East. Different mapping unit scales are represented by grid units with spatial resolution of 30 and 60 m, as well as slope units that were extracted using multi-scale segmentation method. The effects of combined consideration of the study areas at various scales on the uncertainty of LSP modeling are discussed. These are the conclusions:

- (1) LSP accuracies under grid units significantly increases as the study area scales decrease from Ganzhou City to Ruijin County, whereas the accuracy of LSP under the slope units is less affected by study area scales. However, the standard deviation of LSIs shows the opposite trend, the mean values of LSIs in Ruijin County that were extracted by mask from Ganzhou City are higher than those that were extracted from Ganzhou East and those that were directly predicted. Considering that a larger study area can provide more abundant model training and testing landslide samples, more studies are needed in research fields of different study area scales.
- (2) In the study areas with various scales, LSP modelling under slope units is more accurate, stable, and efficient than those under grid units with 30 m resolution and 60 m resolution. Slope units outperform grid units in terms of efficiency and mean value.
- (3) The environmental factors displayed an averaging trend at the larger study area scale, and their importance changes significantly with the decrease of the study area scales. The importance of environmental factors varies greatly between the different mapping units, but it is generally consistent in the slope units and the grid units with 30 m resolution.

Acknowledgements This research is funded by the Natural Science Foundation of China (41807285), Interdisciplinary Innovation Fund of Natural Science, NanChang University (9167-28220007-YB2107).

Author contribution FH: Conceptualization, Methodology. YC: Data curation, Investigation. WL: Visualization, Writing-Original draft preparation, Software. FC: Supervision. GS: Software, Validation. JH: Writing- Reviewing and Editing.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abraham MT, Satyam N, Pradhan B, Alamri AM (2020) Forecasting of landslides using rainfall severity and soil wetness: a probabilistic approach for Darjeeling Himalayas. *Water* 12(3):66
- Achu AL, Aju CD, Di Napoli M, Prakash P, Gopinath G, Shaji E, Chandra V (2023) Machine-learning based landslide susceptibility modelling with emphasis on uncertainty analysis. *Geosci Front* 14(6):66
- Alvioli M, Marchesini I, Reichenbach P, Rossi M, Ardizzone F, Fiorucci F, Guzzetti F (2016) Automatic delineation of geomorphological slope units with r. slopeunits v1.0 and their optimization for landslide susceptibility modeling. *Geosci Model Dev* 9(11):3975–3991
- Alvioli M, Guzzetti F, Marchesini I (2020) Parameter-free delineation of slope units and terrain subdivision of Italy. *Geomorphology* 358:66
- Alvioli M, Marchesini I, Pokharel B, Gnyawali K, Lim S (2022) Geomorphological slope units of the Himalayas. *J Maps* 66:1–14
- Araújo JR, Ramos AM, Soares PMM, Melo R, Oliveira SC, Trigo RM (2022) Impact of extreme rainfall events on landslide activity in Portugal under climate change scenarios. *Landslides* 19(10):2279–2293
- Bai S, Lu P, Thiebes B (2020) Comparing characteristics of rainfall- and earthquake-triggered landslides in the Upper Minjiang catchment, China. *Eng Geol* 268:66
- Calvello M, Cascini L, Mastroianni S (2013) Landslide zoning over large areas from a sample inventory by means of scale-dependent terrain units. *Geomorphology* 182:33–48
- Cavazzi S, Corstanje R, Mayr T, Hannam J, Fealy R (2013) Are fine resolution digital elevation models always the best choice in digital soil mapping? *Geoderma* 195:111–121
- Chang Z, Du Z, Zhang F, Huang F, Chen J, Li W, Guo Z (2020) Landslide susceptibility prediction based on remote sensing images and GIS: comparisons of supervised and unsupervised machine learning models. *Remote Sens* 12(3):502
- 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(5):1127–1143
- Chen X, Chen W (2021) GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods. *Catena* 196:66
- Chen W, Li Y (2020) GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models. *Catena* 195:66
- 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
- Chen W, Chen Y, Tsangaratos P, Ilia I, Wang X (2020a) Combining evolutionary algorithms and machine learning models in landslide susceptibility assessments. *Remote Sens* 12(23):66
- Chen Z, Ye F, Fu W, Ke Y, Hong H (2020b) The influence of DEM spatial resolution on landslide susceptibility mapping in the Baxie River basin, NW China. *Nat Hazards* 66:1–25
- Deng X, Sun G, He N, Yu Y (2022) Landslide susceptibility mapping with the integration of information theory, fractal theory, and statistical analyses at a regional scale: a case study of Altay Prefecture, China. *Environ Earth Sci* 81(13):346
- Drăguț L, Tiede D, Levick SR (2010) ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *Int J Geogr Inf Sci* 24(6):859–871
- Du J, Glade T, Woldai T, Chai B, Zeng B (2020) Landslide susceptibility assessment based on an incomplete landslide inventory in the Jilong Valley, Tibet, Chinese Himalayas. *Eng Geol* 270:66
- Franceschini R, Rosi A, Catani F, Casagli NL (2022) Exploring a landslide inventory created by automated web data mining: the case of Italy. *Landslides* 19(4):841–853
- Fressard M, Thiery Y, Maquaire O (2014) Which data for quantitative landslide susceptibility mapping at operational scale? Case study of the Pays d'Auge plateau hillslopes (Normandy, France). *Nat Hazard* 14(3):569–588
- Froude MJ, Petley DN (2018) Global fatal landslide occurrence from 2004 to 2016. *Nat Hazard* 18(8):2161–2181
- Fustos I, Abarca-del-Rio R, Moreno-Yaeger P, Somos-Valenzuela M (2020) Rainfall-Induced Landslides forecast using local precipitation and global climate indexes. *Nat Hazards* 102(1):115–131
- Gaidzik K, Ramirez-Herrera MT (2021) The importance of input data on landslide susceptibility mapping. *Sci Rep* 11(1):19334
- Gariano SL, Petrucci O, Rianna G, Santini M, Guzzetti F (2017) Impacts of past and future land changes on landslides in southern Italy. *Reg Environ Change* 18(2):437–449
- Garosi Y, Sheklabadi M, Conoscenti C, Pourghasemi HR, Van Oost K (2019) Assessing the performance of GIS-based machine learning models with different accuracy measures for determining susceptibility to gully erosion. *Sci Total Environ* 664:1117–1132
- Grandjean G, Thomas L, Bernardie S, The ST (2018) A novel multi-risk assessment web-tool for evaluating future impacts of global change in mountainous areas. *Climate* 6(4):66
- Guo Z, Torra O, Hürlimann M, Abancó C, Medina V (2022) FSLAM: A QGIS plugin for fast regional susceptibility assessment of rainfall-induced landslides. *Environ Model Softw* 150:66
- Gupta SK, Shukla DP (2023) Handling data imbalance in machine learning based landslide susceptibility mapping: a case study of Mandakini River Basin, North-Western Himalayas. *Landslides* 20(5):933–949
- Haiqing Sun WL, Scaioni M, Jing Fu, Guo X, Gao J (2023) Influence of spatial heterogeneity on landslide susceptibility in the trans-boundary area of the Himalayas. *Geomorphology* 433:108723
- Haque U, da Silva PF, Devoli G, Pilz J, Zhao BX, Khaloua A, Wilopo W, Andersen P, Lu P, Lee J, Yamamoto T, Keellings D, Wu JH, Glass GE (2019) The human cost of global warming: deadly landslides and their triggers (1995–2014). *Sci Total Environ* 682:673–684
- Hodasová K, Bednarik M (2021) Effect of using various weighting methods in a process of landslide susceptibility assessment. *Nat Hazards* 105(1):481–499
- Hong H, Chen W, Xu C, Youssef AM, Pradhan B, Dieu Tien B (2017a) Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy. *Geocarto Int* 32(2):139–154
- Hong H, Ilia I, Tsangaratos P, Chen W, Xu C (2017b) A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the Wuyuan area, China. *Geomorphology* 290:1–16
- Hong H, Kornejady A, Soltani A, Termeh SVR, Liu J, Zhu AX, Hesar AY, Bin Ahmad B, Wang Y (2018) Landslide susceptibility assessment in the Anfu County, China: comparing different statistical and probabilistic models considering the new topohydrological factor (HAND). *Earth Sci Inf* 11(4):605–622
- Hong H, Miao Y, Liu J, Zhu AX (2019) Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. *Catena* 176:45–64
- Hu Q, Zhou Y, Wang S, Wang F (2020) Machine learning and fractal theory models for landslide susceptibility mapping: Case study from the Jinsha River Basin. *Geomorphology* 351:66
- Huang Y, Zhao L (2018) Review on landslide susceptibility mapping using support vector machines. *Catena* 165:520–529
- Huang F, Cao Z, Guo J, Jiang S-H, Li S, Guo Z (2020a) Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena* 191:104580
- Huang F, Cao Z, Jiang S-H, Zhou C, Huang J, Guo Z (2020b) Landslide susceptibility prediction based on a semi-supervised multiple-layer perceptron model. *Landslides* 17(12):2919–2930
- Huang F, Zhang J, Zhou C, Wang Y, Huang J, Zhu L (2020c) A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. *Landslides* 17(1):217–229
- Huang F, Tao S, Chang Z, Huang J, Fan X, Jiang S-H, Li W (2021a) Efficient and automatic extraction of slope units based on multi-scale segmentation method for landslide assessments. *Landslides* 18(11):3715–3731
- Huang F, Ye Z, Jiang S-H, Huang J, Chang Z, Chen J (2021b) Uncertainty study of landslide susceptibility prediction considering the different attribute interval numbers of environmental factors and different data-based models. *Catena* 202:105250
- Huang F, Yan J, Fan X, Yao C, Huang J, Chen W, Hong H (2022) Uncertainty pattern in landslide susceptibility prediction modelling: Effects of different landslide boundaries and spatial shape expressions. *Geosci Front* 13(2):101317
- Huang F, Teng Z, Yao C, Jiang S-H, Catani F, Chen W, Huang J (2023a) Uncertainties of landslide susceptibility prediction: Influences of random errors in landslide conditioning factors and errors reduction by low pass filter method. *J Rock Mech Geotech Eng* 6:66
- Huang F, Xiong H, Yao C, Catani F, Zhou C, Huang J (2023b) Uncertainties of landslide susceptibility prediction considering different landslide types. *J Rock Mech Geotech Eng* 15(11):2954–2972
- Huang F, Xiong H, Chen S, Lv Z, Huang J, Chang Z, Catani F (2023c) Slope stability prediction based on a long short-term memory neural network: comparisons with convolutional neural networks, support vector machines and random forest models. *Int J Coal Sci Technol* 10(1):18. <https://doi.org/10.1007/s40789-023-00579-4>
- Hung O, Leroueil S, Picarelli L (2014) The Varnes classification of landslide types, an update. *Landslides* 11(2):167–194

- Hurlimann M, Guo Z, Puig-Polo C, Medina V (2022) Impacts of future climate and land cover changes on landslide susceptibility: regional scale modelling in the Val d'Aran region (Pyrenees, Spain). *Landslides* 19(1):99–118
- Juliev M, Mergili M, Mondal I, Nurtaev B, Pulatov A, Huebl J (2019) Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan. *Sci Total Environ* 653:801–814
- Kang X, Wang Y, Qin C, Cheng W, Zhao S, Zhu A, Zhangwenshi A (2016) A new method of landform element classification based on multi-scale morphology. *Geograph Res* 35(9):1637–1646
- Kedron P, Holler J (2022) Replication and the search for the laws in the geographic sciences. *Ann GIS* 28(1):45–56
- Kirschbaum DB, Adler R, Hong Y, Kumar S, Peters-Lidard C, Lerner-Lam A (2011) Advances in landslide nowcasting: evaluation of a global and regional modeling approach. *Environ Earth Sci* 66(6):1683–1696
- Kuan-Tsung C, Merghadi A, Yunus AP, Binh Thai P, Dou J (2019) Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. *Sci Rep* 9:66
- Kulsoom I, Hua W, Hussain S, Chen Q, Khan G, Shihao D (2023) SBAS-InSAR based validated landslide susceptibility mapping along the Karakoram Highway: a case study of Gilgit-Baltistan. *Pak Sci Rep* 13(1):3344
- Li D, Huang F, Yan L, Cao Z, Chen J, Ye Z (2019) Landslide susceptibility prediction using particle-swarm-optimized multilayer perceptron: comparisons with multilayer-perceptron-only, BP neural network, and information value models. *Appl Sci* 9(18):66
- Li W, Fan X, Huang F, Chen W, Hong H, Huang J, Guo Z (2020) Uncertainties analysis of collapse susceptibility prediction based on remote sensing and GIS: influences of different data-based models and connections between collapses and environmental factors. *Remote Sens* 12(24):4134
- Li W, Fan X, Huang F, Wu X, Yin K, Zhilu C (2021) Uncertainties of landslide susceptibility modeling under different environmental factor connections and prediction models. *Earth Sci* 46:3777–3795
- Lim LA, Keles HY (2019) Learning multi-scale features for foreground segmentation. *Pattern Anal Appl* 23(3):1369–1380
- Lissak C, Bartsch A, De Michele M, Gomez C, Maquaire O, Raucoules D, Roulland T (2020) Remote sensing for assessing landslides and associated hazards. *Surv Geophys* 41(6):1391–1435
- Liu Y (2022) Core or edge? Revisiting GIScience from the geography-discipline perspective. *Sci China Earth Sci* 65(2):387–390
- Liu Y, Guo H, Li H, Dong W, Pei T (2022) A note on GeoAI from the perspective of geographical laws. *Acta Geodetica et Cartographica Sin* 51(6):1062–1069
- Liu YFX, Zhu Q, Zeng H, Wu H (2023) An efficient and fine slope units division method with consideration of the regional geomorphological characteristics. *Sci Surv Mapp* 48(4):211–220
- Loche M, Alvioli M, Marchesini I, Bakka H, Lombardo L (2022) Landslide susceptibility maps of Italy: lesson learnt from dealing with multiple landslide types and the uneven spatial distribution of the national inventory. *Earth Sci Rev* 232:66
- Lombardo L, Mai PM (2018) Presenting logistic regression-based landslide susceptibility results. *Eng Geol* 244:14–24
- Lombardo L, Tanyas H (2020) Chrono-validation of near-real-time landslide susceptibility models via plug-in statistical simulations. *Eng Geol* 278:66
- 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(2):386–397
- 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* 207:66
- Miller BA, Koszinski S, Wehrhan M, Sommer M (2015) Impact of multi-scale predictor selection for modeling soil properties. *Geoderma* 239:97–106
- Moeen Hamid Bukhari PF, da Silva JP, Istanbuluoglu E, Görüm T, Lee J (2023) Community perceptions of landslide risk and susceptibility: a multi-country study. *Landslides* 20:1321–1334
- Palau RM, Berenguer M, Hurlimann M, Sempere-Torres D (2022) Application of a fuzzy verification framework for the evaluation of a regional-scale landslide early warning system during the January 2020 Gloria storm in Catalonia (NE Spain). *Landslides* 19(7):1599–1616
- Pellicani R, Argentiero I, Spilotro G (2017) GIS-based predictive models for regional-scale landslide susceptibility assessment and risk mapping along road corridors. *Geomat Nat Haz Risk* 8(2):1012–1033
- Pham BT, Phong TV, Nguyen-Thoi T, Parial KK, Singh S, Ly H-B, Nguyen KT, Ho LS, Le HV, Prakash IJGI (2022) Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers. *Geocarto Int* 37(3):735–757
- Pradhan B, Dikshit A, Lee S, Kim H (2023) An explainable AI (XAI) model for landslide susceptibility modeling. *Appl Soft Comput* 142:66
- Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. *Earth Sci Rev* 180:60
- Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat R (2019) Deep learning and process understanding for data-driven Earth system science. *Nature* 566(7743):195–204
- Salciarini D, Volpe E, Kelley SA, Brocca L, Camici S, Fanelli G, Tamagnini C (2016) Modeling the effects induced by the expected climatic trends on landslide activity at large scale. *Procedia Eng* 158:541–545
- Schlögel R, Marchesini I, Alvioli M, Reichenbach P, Rossi M, Malet JP (2018) Optimizing landslide susceptibility zonation: effects of DEM spatial resolution and slope unit delineation on logistic regression models. *Geomorphology* 301:10–20
- Segoni S, Piciullo L, Gariano SL (2018) A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides* 15(8):1483–1501
- Shahabi H, Khezri S, Bin Ahmad B, Hashim M (2014) Landslide susceptibility mapping at central Zab basin, Iran: a comparison between analytical hierarchy process, frequency ratio and logistic regression models. *Catena* 115:55–70
- Shi J, Yang L, Zeng C, Zhu A, Qin C, Liang P (2018) Selection of environmental variables and their scales in multiple soil properties mapping: a case study in Heilongjiang Heshan Farm. *Geogr Res* 37(3):635–646
- Shirzadi A, Solaimani K, Roshan MH, Kaviani A, Chapi K, Shahabi H, Keesstra S, Bin Ahmad B, Dieu Tien B (2019) Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. *Catena* 178:172–188
- Shou K-J, Lin J-F (2016) Multi-scale landslide susceptibility analysis along a mountain highway in Central Taiwan. *Eng Geol* 212:120–135
- Sun D, Wen H, Wang D, Xu J (2020) A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology* 362:66
- Thi Ngo PT, Panahi M, Khosravi K, Ghorbanzadeh O, Kariminejad N, Cerda A, Lee S (2021) Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran. *Geosci Front* 12(2):505–519

- Wang K, Zhang S, DelgadoTéllez R, Wei F (2018) A new slope unit extraction method for regional landslide analysis based on morphological image analysis. *Bull Eng Geol Env* 78(6):4139–4151
- Wu H, Chen Y, Lv H, Xie Q, Chen Y, Gu J (2022) Stability analysis of rib pillars in highwall mining under dynamic and static loads in open-pit coal mine. *Int J Coal Sci Technol* 9(1):38. <https://doi.org/10.1007/s40789-022-00504-1>
- Xiao T, Zhang L-M (2023) Data-driven landslide forecasting: Methods, data completeness, and real-time warning. *Eng Geol* 66:317
- Xu C, Dai F, Xu X, Lee YH (2012) GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* 145:70–80
- Yang JT, Xu C, Jin X (2023) Joint effects and spatiotemporal characteristics of the driving factors of landslides in earthquake areas. *J Earth Sci* 34(2):330–338
- Youssef AM, Pourghasemi HR (2021) Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Geosci Front* 12(2):639–655
- Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2016) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* 13(5):839–856
- Yu X, Gao H (2020) A landslide susceptibility map based on spatial scale segmentation: a case study at Zigui-Badong in the Three Gorges Reservoir Area, China. *PLoS ONE* 15(3):66
- Yu L, Hao G, Haifeng L, Weihua D, Tao PJA (2022) A note on GeoAI from the perspective of geographical laws. *Acta Geodaetica et Cartographica Sinica* 51(6):1062
- Zeng T, Wu L, Peduto D, Glade T, Hayakawa YS, Yin K (2023) Ensemble learning framework for landslide susceptibility mapping: different basic classifier and ensemble strategy. *Geosci Front* 14(6):66
- Zhao Y, Wang R, Jiang Y, Liu H, Wei Z (2019) GIS-based logistic regression for rainfall-induced landslide susceptibility mapping under different grid sizes in Yueqing, Southeastern China. *Eng Geol* 259:66
- Zhao J-S, Duan S-Q, Chen B-R, Li L, He B-G, Li P-X, Liu G-F (2024) Failure mechanism of rock masses with complex geological conditions in a large underground cavern: a case study. *Soil Dyn Earthq Eng* 177:108439
- Zheng Y, Chen C, Liu T, Song D, Meng F (2019) Stability analysis of anti-dip bedding rock slopes locally reinforced by rock bolts. *Eng Geol* 251:228–240
- Zheng Y, Chen C, Meng F, Fu X, Yuan W (2023) Smart and fast reinforcement design for anti-dip bedding rock slopes. *J Rock Mech Geotech Eng* 15(11):2943–2953
- Zhu J, Zhang L, Zhou X, Lu Y (2015a) Spatial effect analysis of regional geological hazards. *South-to-North Water Transf Water Sci Technol* 13(2):334–338
- Zhu J, Zhang L, Zhou X, Wang Q, Lu Y (2015b) Study of scale effect-based calibration on weight of factors influencing landslide. *Res Soil Water Conserv* 22(5):181–185
- Zhu A, Yang L, Fan N, Zeng C, Zhang G (2018a) The review and outlook of digital soil mapping. *Prog Geogr* 37(1):66–78
- Zhu AX, Lu G, Liu J, Qin CZ, Zhou C (2018b) Spatial prediction based on Third Law of Geography. *Ann GIS* 24(4):225–240

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