International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 1, Number 1, Year 2023
DOI: https://doi.org/10.62051/ijcsit.v1n1.16
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Evaluation of landslide disaster susceptibility in Shuangbai County based on random forest modeling

Haoyun Xie

Beijing University of Civil Engineering and Architecture, Beijing 102627, China

ABSTRACT

Shuangbai County, Chuxiong Yi Autonomous Prefecture, Yunnan Province, China, has a high incidence of geohazard events and is a key area for geologic risk prevention and control in China, with landslides accounting for the majority of all geohazard events, so understanding landslide-prone areas in the region is critical to reducing the impacts of landslide hazards. In this study, based on literature reference and background investigation of the study area, relevant data and existing research results were collected, and six types of evaluation factors including elevation, slope, slope direction, rainfall, vegetation type and vegetation index were selected to construct a random forest (RF) model to analyze the landslide susceptibility of Shuangbai County, Chuxiong Prefecture, Yunnan Province, and to obtain the evaluation map of landslide susceptibility in Shuangbai County. The results show that the landslide disaster high susceptibility areas in Shuangbai County are mainly distributed in the northern part of the county in Toudian Township and Dazhuang Township, the southeastern part of Anlongbao Township and the southwestern part of Ejia Township in the vicinity of the Hengduan Mountain Range, which is an important reference for the development of scientific and effective disaster prevention and mitigation strategies, and for the enhancement of the quality of life and the safety of the residents in the area of Shuangbai County.

KEYWORDS

landslides; Random forest; Susceptibility assessment.

1. STUDY AREA AND DATA SOURCES

1.1. Study area

Shuangbai County, Chuxiong Prefecture, is located in the northwestern part of China's Yunnan Province, in the mountainous region of western Yunnan, with a geographic coordinate range of latitude 25°14'N to 26°38'N, and longitude 101°48'E to 104°12'E. The terrain of the region is mountainous and undulating, including branches of the Hengduan Mountains and hilly areas, with complex and varied landforms covering high mountains, valleys, hills and other landform types. The climate is subtropical monsoon, with more rainfall in summer and relatively drier winters, with rainfall concentrated in summer and fall . The location distribution of Shuangbai County is shown in Fig.1.

As a mountainous area, Shuangbai County has mountainous terrain and steep terrain, which, together with the long-term erosive effect of rainwater, makes the area often threatened by landslides and other geologic disasters. The risk of landslides is further heightened especially during the onset of the rainy season when surface moisture increases and soil saturation increases. Therefore, it is important to understand the distribution characteristics, susceptibility evaluation factors and other geologic information of landslides in the area for disaster prevention and mitigation.

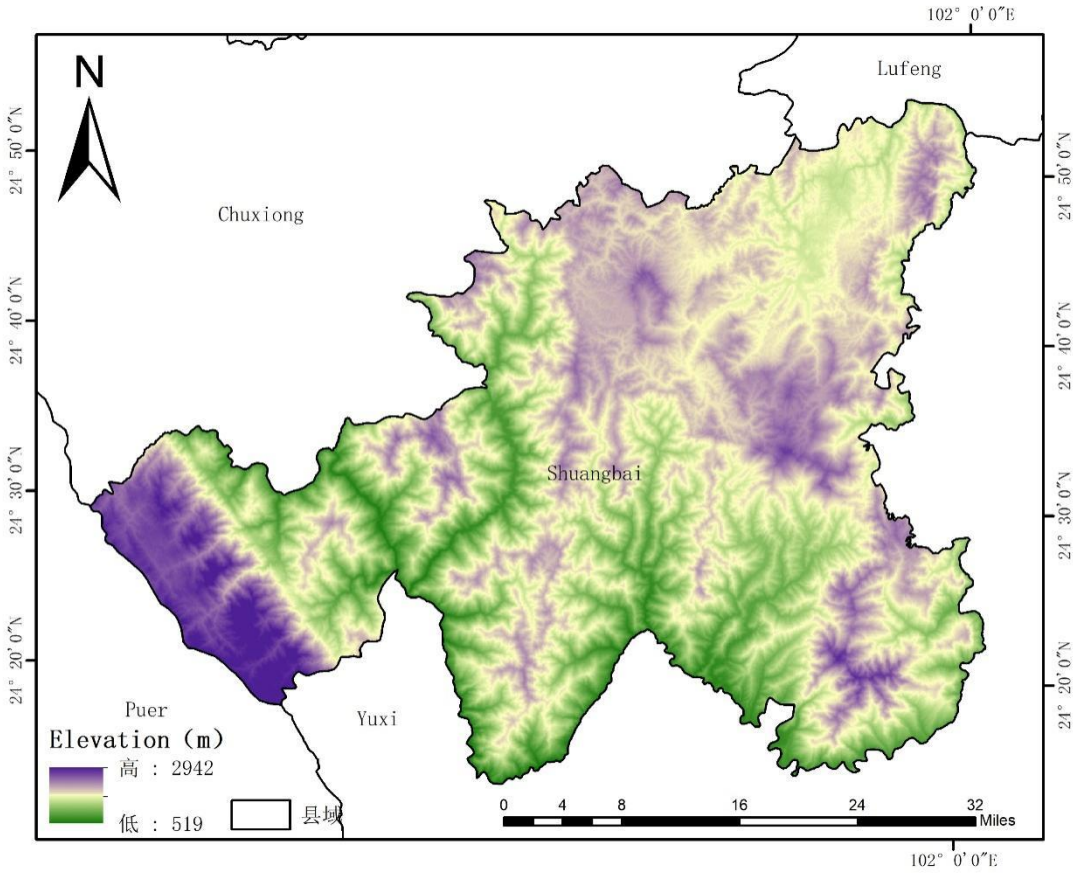


Fig.1 Location distribution of the study area in Shuangbai County

1.2. Data sources

Table 1. Data description and data sources

Data type	Factor	Data description	Data source
Terrain	Elevation	ALOS PALSAR 12.5 m DEM	NASA (https://search.asf.alaska.edu/#/)(https://search.asf.alaska.edu/#/)
	Slope	ArcGIS is used to process DEM data based on DEM data	-
	Aspect	ArcGIS is used to process DEM data based on DEM data	-
Ecology	Vegetation Index	Spatial distribution of Vegetation index (NDVI) in China at 100M/month	Resources and Environmental Sciences and Data Center, Chinese Academy of Sciences (https://www.resdc.cn/)
	Vegetation Type	Spatial distribution data of 1 million vegetation types in China	
Meteorology	Rainfall	Annual precipitation data of 30 m resolution in China from 1991 to 2020	Ecological remote sensing frontier (https://mp.weixin.qq.com/s/FmQJ1ai3W0C_f5lARKYxQA)

The historical landslide hazard data used in this paper were obtained from the Institute of Geoscience and Resources, Chinese Academy of Sciences (https://www.resdc.cn/), and there are 152 landslides in the landslide cataloging information. The landslide susceptibility evaluation indexes include

elevation, slope,aspect, vegetation index, vegetation type and rainfall in total 6 index factors to evaluate and analyze the landslide susceptibility in Shuangbai County. Table 1 shows the selected evaluation factors and their data sources.

2. EVALUATION FACTOR SCREENING

2.1. Pearson correlation coefficient

Since evaluation factors with less significance will reduce the predictive ability of the model, this paper firstly adopts Pearson Correlation Coefficient (PCC) to eliminate the indicator factors with high correlation. Pearson's correlation coefficient reflects the degree of linear correlation between two variables x and y. Its output range is [-1,1], with negative values indicating a negative correlation between the dependent variable x (independent variable) and y (dependent variable), positive values indicating a positive correlation between x and y, and 0 indicating that there is no correlation between the factors, i.e., the closer the absolute value is to 1, the higher the degree of correlation in the sample. The correlation division is shown in Table 2.

Table 2. Degrees of correlation assigned to factors.

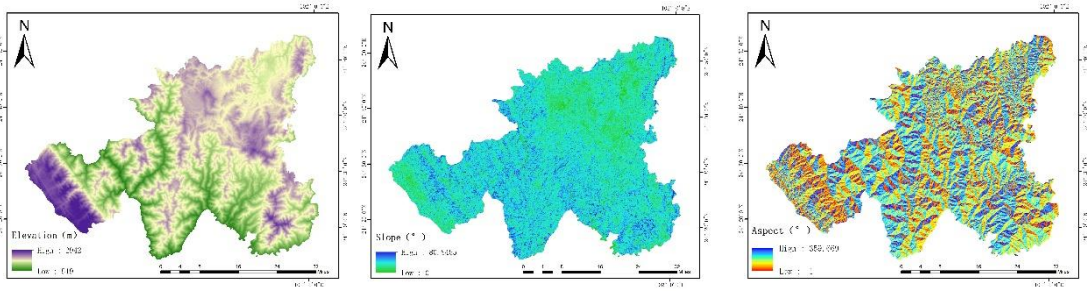
Random	Low correlation	Moderate correlation	High correlation
$p < 0.3$	$0.3 < p < 0.5$	$0.5 < p < 0.8$	$p > 0.8$

2.2. Multicollinearity check

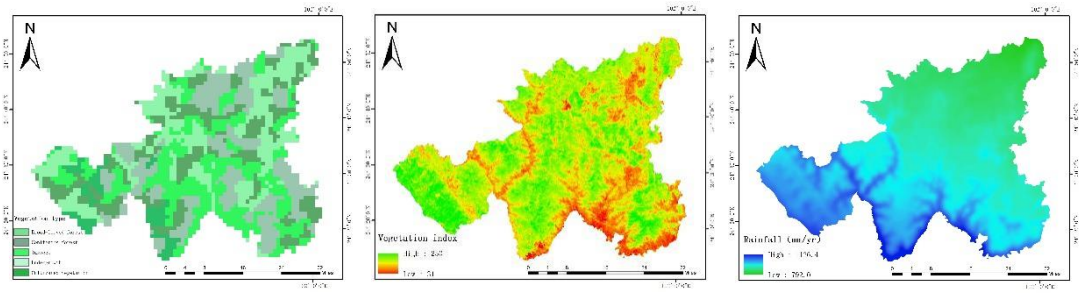
Multicollinearity refers to the presence of a correlation between two or more variables. The multicollinearity check results in two columns of values, one for Tolerance and the other for Variance Inflation Factor (VIF), which are the inverse of each other. In general, a factor is considered to be covariant if the variance inflation factor VIF value is >10, and the larger the VIF value, the more severe the covariance is considered to be. Therefore, eliminating factors with high covariance can improve the accuracy and reliability of the model.

2.3. Landslide susceptibility evaluation factors

In this paper, Pearson correlation coefficient and multiple covariance check are used to analyze the correlation between evaluation indexes, and the factors with Pearson correlation coefficient greater than 0.5 (ρ>0.5) or VIF value >10 are excluded to ensure the independence between factors. The six types of evaluation factors selected (elevation, slope, aspect, rainfall, vegetation type and vegetation index) all have a high degree of independence from each other, and all of them can participate in the subsequent experimental study. The results are shown in Fig.2.



(a)Elevation (b)Slope (c)Aspect



(d)Vegetation type (e)vegetation index (f)rainfall

Fig.2 Landslide susceptibility evaluation factors

3. LANDSLIDE VULNERABILITY ASSESSMENT

3.1. Assessment model

Random Forest (RF) is a bagging integration algorithm proposed by Breiman (Breiman, 2001) based on CART decision trees (Classification and Regression trees). The Random Forest algorithm is based on the integrated learning idea of combining single classifiers into multiple classifiers, sampling from the original data, building and combining multiple decision trees based on features, and arriving at the final prediction by voting. Random forest models are efficient and accurate in prediction compared to other algorithms, and have a good tolerance for both outliers and noise.

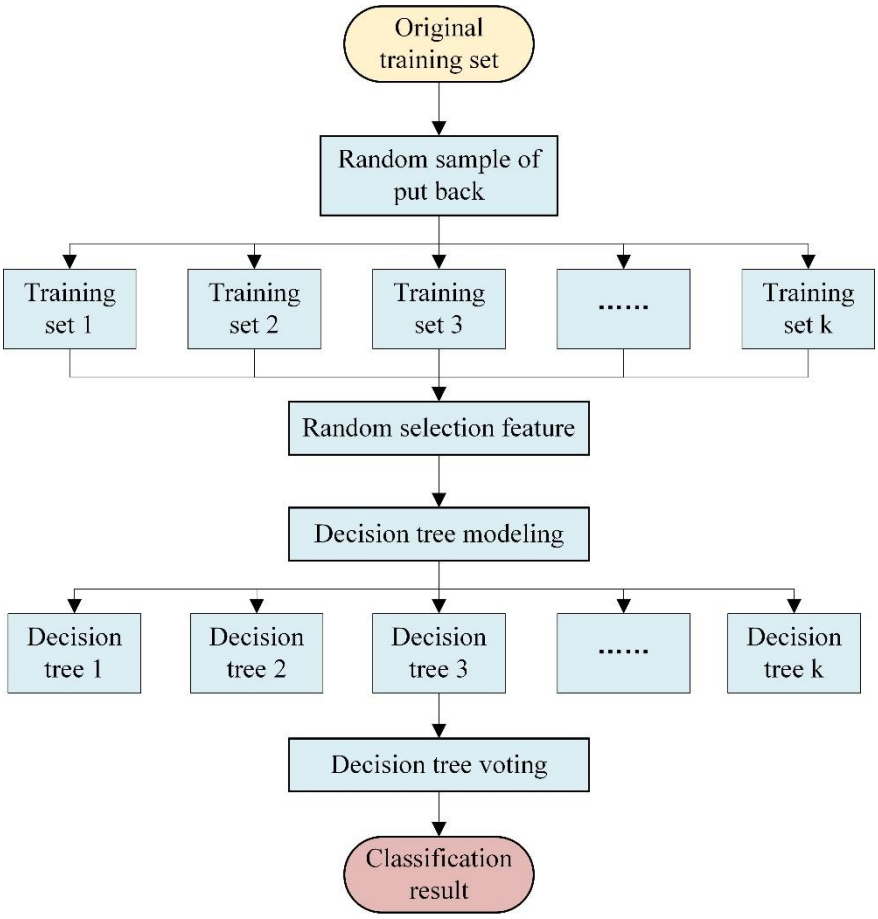


Fig.3 Random Forest flowchart

The main idea of the random forest model is shown in Fig.3. To ensure its randomness, random sampling with put back (bootstrap sample) is performed on the training set, n samples with sample

capacity consistent with the size of the original training set are drawn from the original training set, then the sample features are randomly selected and decision tree modeling is performed based on the features, and after obtaining the results of the n modeling results, the final classification is finally voted on by the vote.

In landslide susceptibility assessment, the random forest model shows high reliability and stability compared to other models in complex and noise-containing datasets.

3.2. Landslide susceptibility zoning

After obtaining the evaluation factors and normalizing the layers, the landslide susceptibility analysis of the study area in Shuangbai County was carried out based on the random forest model, and the susceptibility results were classified into five grades of very low susceptibility, low susceptibility, moderate susceptibility, high susceptibility and very high susceptibility by applying the natural discontinuity method, and the results are shown in Fig.4.

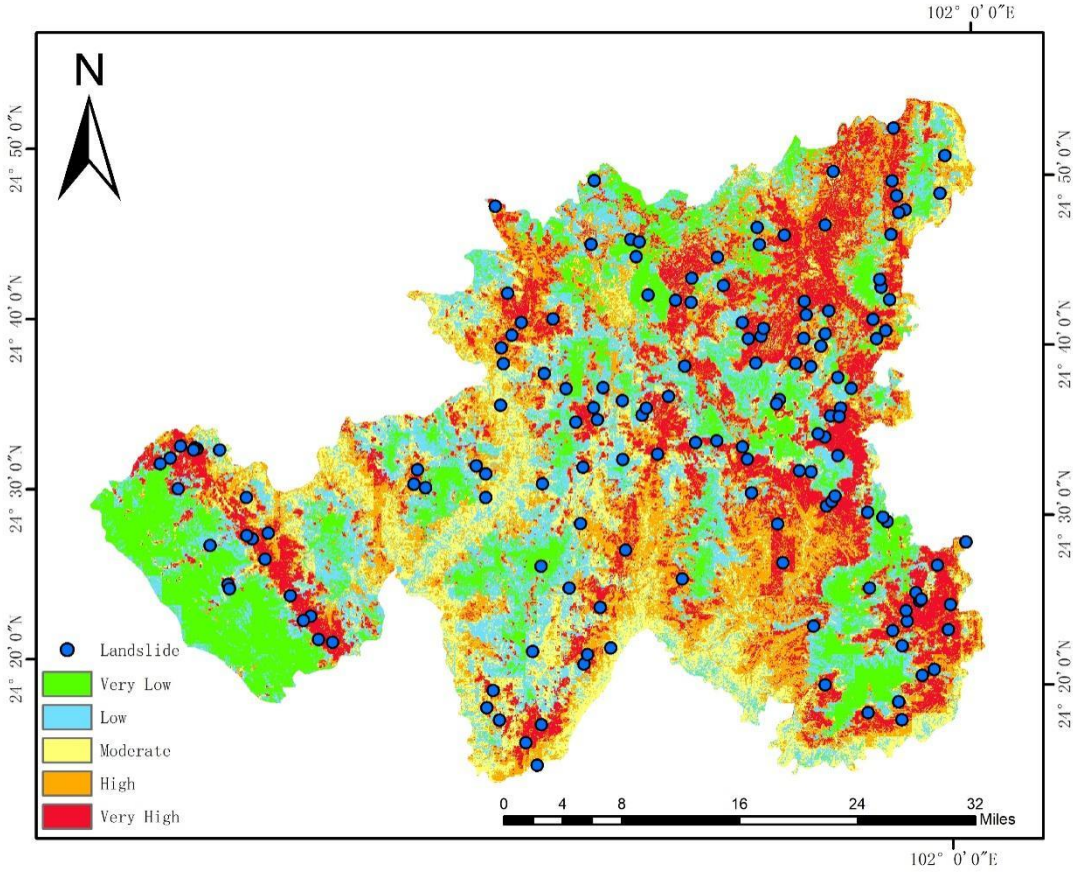


Fig.4 Landslide susceptibility zoning map of Shuangbai County

As shown in the figure, the landslides in Shuangbai County are mainly distributed in the very high susceptibility area and high susceptibility area, and basically no landslide disaster occurs in the very low susceptible area and low susceptibility area, in which the very low susceptible area accounts for 24.3% of the total area, low susceptibility area accounts for 20.8% of the total area, moderate susceptibility area accounts for 18.3% of the total area, high susceptibility area accounts for 17.3% of the total area, and very high susceptibility area accounts for 19.3% of the total area. According to the results of zoning, the landslide disaster high susceptibility area and very high susceptibility area are mainly distributed in the northern part of the county, Toudian Township and Dazhuang Township, the southeastern part of Anlongbao Township, and the southwestern part of Ejia Township, near the Hengduan Mountain Range.

4. CONCLUSION

(1) Taking Shuangbai County, Chuxiong Prefecture, Yunnan Province, as the research object, six indicators of elevation, slope, aspect, vegetation index, vegetation type and rainfall were selected as the evaluation factors of landslide susceptibility, and the Pearson's correlation test and multiple covariance check were used to exclude the factors that were strongly correlated, and the results showed that the six types of indicator factors were independent of each other, and the landslide susceptibility evaluation system was constructed for Shuangbai County on the basis of this.

(2) Based on the historical landslide data, the random forest model was used to draw the landslide susceptibility zoning map of Shuangbai County, and the natural discontinuity method was used to divide the study area into five grades: very low susceptibility, low susceptibility, moderate susceptibility, high susceptibility and very high susceptibility. The results of the zoning showed that the high susceptibility area of Shuangbai County to landslides is mainly located in the northern part of Toudian Township and Dazhuang Township, and in the southeastern part of Anlongbao Township, and in the southwestern part of Ejia Township, near the Hengduan Mountain Range.

REFERENCES

[1] Wu, Z. Study on the genesis factors and susceptibility of landslides and debris flows in Yunnan Province [D]. Beijing: China University of Geosciences, 2016.

[2] HUANG W, DING M, LI Z, et al. An Efficient User-Friendly Integration Tool for Landslide Susceptibility Mapping Based on Support Vector Machines: SVM-LSM Toolbox[J]. Remote Sensing, 2022, 14(14): 3408.

[3] SHARMA A, PRAKASH C, MANIVASAGAM V S. Entropy-Based Hybrid Integration of Random Forest and Support Vector Machine for Landslide Susceptibility Analysis[J]. Geomatics, 2021, 1(4): 399-416.

[4] BREIMAN L. Random Forests[J]. Machine Learning, 2001, 45(1): 5-32.

[5] GANAIE M A, TANVEER M, SUGANTHAN P N, et al. Oblique and rotation double random forest[J]. Neural networks : the official journal of the International Neural Network Society, 2022, 153: 496-517.

[6] Li, J., Wang, H., Wang, R. Landslide Susceptibility Assessment in Fuzhou Area Based on Machine Learning Models[J]. Journal of Water Resources and Architectural Engineering, 2022, 20(02), 146-152.