Journal of Geoscience and Environment Protection, 2023, 11, 87-105
https://www.scirp.org/journal/gep
ISSN Online: 2327-4344
ISSN Print: 2327-4336

Remote Sensing Landslide Monitoring Based on Machine Learning Method

Zhen Chen[1], **Yiyang Zheng**[2*]

[1]School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing, China
[2]Mudanjiang Natural Resources Survey Center, China Geological Survey, Mudanjiang, China
Email: *czscott@126.com

How to cite this paper: Chen, Z., & Zheng, Y. Y. (2023). Remote Sensing Landslide Monitoring Based on Machine Learning Method. *Journal of Geoscience and Environment Protection, 11,* 87-105.
https://doi.org/10.4236/gep.2023.1110008

Received: September 1, 2023
Accepted: October 17, 2023
Published: October 20, 2023

Abstract

The susceptibility evaluation of landslides has become one of the key environmental issues that people are concerned about. This study took the landslides in Xishuangbanna, Yunnan Province as the study object, and selected 10 evaluation factors such as digital elevation model (DEM), slope aspect, precipitation, land use, water system, roads, population density, lithology, faults, and NDVI. Different machine learning methods were compared and studied, and the ROC (receiver operating characteristics) curve verification revealed that the accuracy of the random forest evaluation model was high. In the prediction and evaluation of the susceptibility of landslides, five risk levels were divided. After the superimposed analysis, 87.26% of the disaster points fell in the first and second susceptibility areas. The spot analysis found that the distribution of hot spots is consistent with the distribution of disaster spots. In a word, the results of this study can provide better technical support for the evaluation and early warning of landslides in Southwest China.

Keywords

Landslide, Evaluation Index, Random Forest, Geological Disaster

1. Introduction

Landslides are geological phenomena, and landslides may cause the geological environment to be damaged to varying degrees under the action of internal and external geological forces or human activities, and brings serious harm to human life and property. Therefore, it is very important to carry out in-depth study on landslides.

A landslide is an important type of geological disaster, and investigating landslides is an important content. Susceptibility evaluation of landslide disasters is a quantitative prediction and evaluation of the possibility of landslide disasters in a spatial scale, which can provide substantive technical support for land and spacial planning (Balteanu, Micu, Jurchescu et al., 2020). In terms of evaluation models, the evaluation models with geological hazard susceptibility of landslides widely used by scholars mainly include deterministic models, mathematical statistical models and heuristic models (Salciarini, Godt, Savage et al., 2006; Liu, Wang, Zhou et al., 2020; Liu, Li, & Chen, 2018; Zhou, Yin, Xiang et al., 2015). The deterministic model is based on the geomechanical process of landslides. Generally, the susceptibility of geological landslides is evaluated by calculating the stability coefficient of the geological hazard body. The infinite slope model is a relatively typical model, but the application is difficult in hazard studies of large scale landslide geology, because of the need to determine the soil strength parameters of slide zones and accurate groundwater levels (Wang, Li, & Wang, 2012). Mathematical statistical models are based on the engineering geological analogy method, and primarily include models such as the volume of information and the weights of evidence. The heuristic model, also known as the expert experience model, mainly relies on the technical experience and professional knowledge of experts to establish an evaluation model, which has a large human subjectivity. At present, with the continuous development of machine learning and deep learning algorithms, more and more attention has been paid to the study of landslide disaster susceptibility evaluation models based on machine learning and deep learning algorithms, e.g., various neural network models, decision tree models and support vector machine model, etc. (Pradhan, 2013; Su, Wang, Wang et al., 2015; Dou, Yamagishi, Pourghasemi et al., 2015). Most of the above evaluation models can better reflect the complex nonlinear characteristics of landslides, but some still have problems such as the difficult interpretation of prediction results and overfitting.

In order to improve the prediction accuracy of the landslide disaster susceptibility evaluation model and avoid problems such as overfitting, the random forest model has received more and more attention. This model is an improved integrated learning method for decision tree, which has been widely used in many fields (Pourghasemi & Kerle, 2016; Liu, Di, Zhan et al., 2018). At present, the random forest model also has some application cases in the field of landslide evaluation. For example, Merghadi (Merghadi, Abderrahmane, & Tien Bui, 2018) took the Mira Basin in North Africa as the study area and compared the prediction capability of 5 geological landslide susceptibility evaluation models such as logistic regression, random forest, neural network, gradient boosting machine, and support vector machine. It was concluded after verification that the prediction performance of the random forest model was better. Goetz et al. (Goetz, Brenning, Petschko et al., 2015) compared and analyzed the prediction and evaluation effects of traditional geostatistical methods and various machine learning methods on the evaluation of landslide geological hazard susceptibility. Through experiments, it was concluded that among many machine learning methods, the random forest model has better prediction performance. Sun et al. (2020) carried out the landslide susceptibility evaluation in Fengjie County and established a Bayesian optimization algorithm for the random forest evaluation model for landslide susceptibility. The results show that the model has high prediction accuracy.

At present, the commonly used landslide evaluation models include spatial model, landscape ecological model and mathematical model. Most of these evaluation models are based on landscape ecology theory and comprehensively considering the overall situation of regional geological disasters (Wang, Cheng, & Qian, 2003).

The reason why this paper chose the random forest algorithm for landslide prediction and evaluation was that the random forest algorithm is an integrated learning algorithm based on decision trees, and its core idea is to collect several decision trees to obtain the optimal solution. The principle of random forest is to randomly select features and eigenvalues in each decision tree to divide the data, then each decision tree gives the prediction result, and it finally determine the final prediction result through the voting result. The advantages of the algorithm are stable, accurate prediction, and they can deal with missing values, and the calculation results are interpretable. The main parameters include decision tree number, feature selection strategy, minimum sample number of internal node subdivision, minimum sample number of leaf node, etc.

In this study, the Random Forest (RF) model and support vector machine (SVM) model in machine learning are selected to evaluate the susceptibility of landslides in the Xishuangbanna area. Through ROC curve verification, it was proved that the Random Forest model has high prediction accuracy. The evaluation of landslide susceptibility carried out in this study can provide a basis for the relevant authorities to establish landslide prevention and mitigation measures, and minimize various losses caused by landslides.

2. Geological Background and Data Sources

2.1. Geological Background

Xishuangbanna is located at 21˚10'-22˚40' north latitude and 99˚55'-101˚50' east longitude. It is located at the northern edge of the tropics south of the Tropic of Cancer, covering an area of 19, 124.5 square kilometers.

The strata in the Xishuangbanna area is distributed from the Paleoproterozoic to the Cenozoic, and the sequence and contact relationship of the strata is relatively clear. Among them, the Damenglongyan Formation is the oldest stratum, which is widely distributed on the west side of the Lancangjiang fault, and the lithology is a set of deeper metamorphic schist, gneiss, and granulite. The Lancang Formation is the most widely distributed in the study area, and the lithology is sericite-dolomitic quartz schist, sericite micromorphite, phyllite, and slate with certain metamorphism. The study area also possess a small amount of Meso-Cenozoic strata distributed in the basin and on the edge of the basin, and the lithology is mainly clastic rocks intercalated with volcanic rocks (Hu, 2019).

The structural trends are mainly NNE and NE. On this basis, NWW and NW faults are superimposed and developed in multiple circles. The main faults are large-scale NEE-trending Jinghong-Daluo fault and the NE-trending Lancangjiang fault (Yu, Li, Liu et al., 2013). There are two fold systems, namely the Gongshan-Tengchong and Tanggula-Chamdo-Lanping-Simao fold systems, and the boundary is the Lancangjiang fault (Bai, Meng, Lv, & Zhang, 2015). The Dehua complex syncline has a large area of Mesozoic red formations, and locally-distributed carbonate formations, coal-bearing clastic formations and molasse formations. The southeastern end of the Lincang-Menghai fold has the Damenglong metamorphic rock and the Proterozoic Lancang Formation outcropping, and the Varioxi-Indosinian Menghai granite base is widely distributed (Yunnan Provincial Bureau of Geology and Mineral Resources, 1990). The magmatic rocks in the study area can be roughly divided into two categories. The first is the Lincang granite, which is dominated by biotite monzogranite. The second is the complex distributed along the Lancang River fault. Medium acid, neutral and basic are distributed (**Figure 1**).

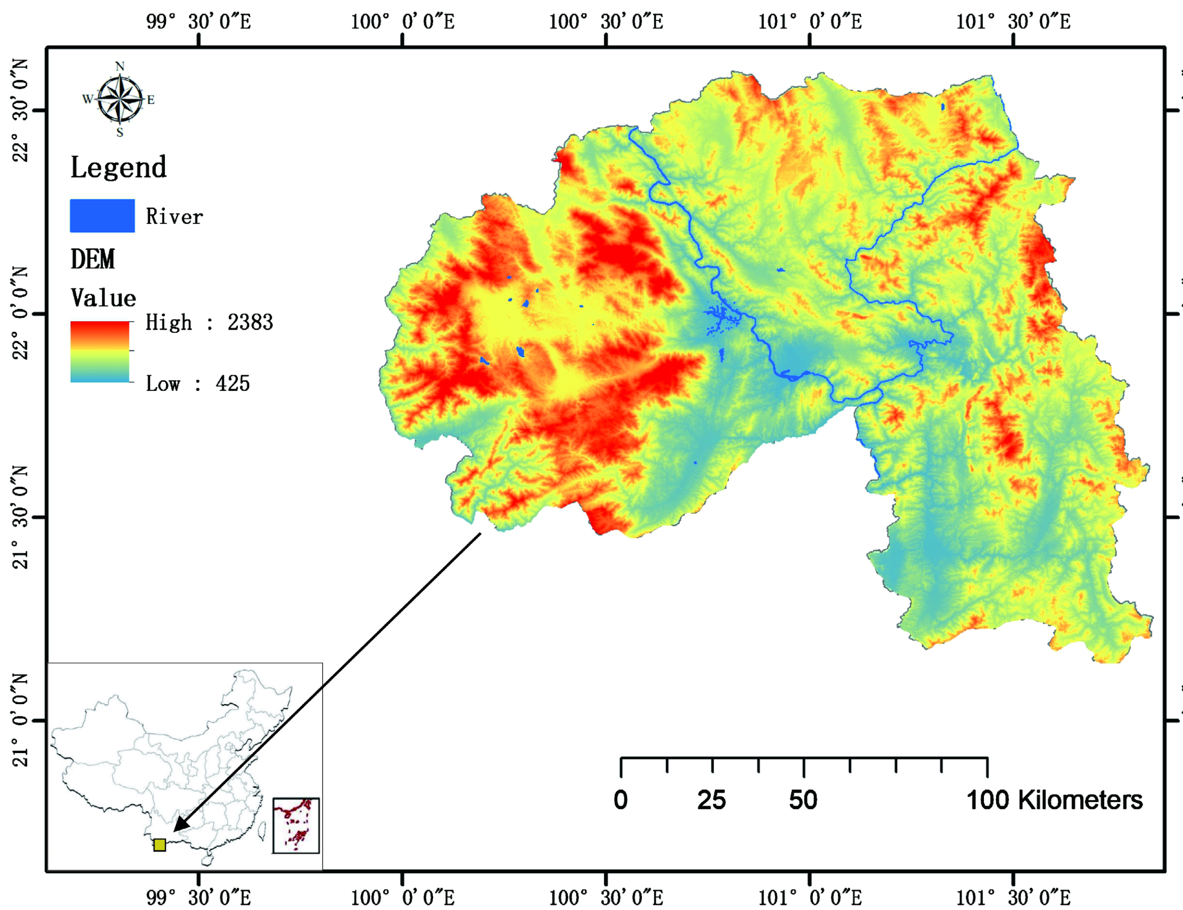


Figure 1. The map of Xishuangbanna area.

2.2. Data Sources

The study used data from: Geospatial Data Cloud website, Landsat8 remote sensing image. Source of geological disasters and landslide data: China Geological Survey Dataset of Detailed Survey of Geological Hazards. Source of rainfall data: Interpolation of point observation data from China Meteorological Data Website, http://data.cma.cn/site/index.html. Source of geological data: National Geological Archives, http://www.ngac.org.cn/. Land use data source: Resource and Environment Cloud Platform, Institute of Geography, Chinese Academy of Sciences, https://www.resdc.cn/. DEM data source: https://search.asf.alaska.edu/ (**Table 1**).

2.3. Selection of Evaluation Factors

In this study, 10 evaluation factors including a DEM, slope aspect, precipitation, land use, water system, road, population density, lithology, fault, and NDVI are selected (as shown in **Table 2**). These evaluation factors are directly related to landslide. Through correlation analysis, the information overlap of these evaluation factors, which meets the selection criteria of evaluation factors.

3. Method

The application of machine learning mainly includes four fields: classification, clustering, regression and data dimensionality reduction. In this study, the ideas of clustering idea and regression are combined to evaluate the landslide susceptibility

Table 1. Data source table.

Data name	Data type	Source of Data	Resolution	acquisition year
Landsat8	tif	http://www.gscloud.cn/	30 m	2021
landslide data	shp	China Geological Survey Dataset of Detailed Survey of Geological Hazards		2021
rainfall data	shp	http://data.cma.cn/site/index.html		2021
geological data	shp	http://www.ngac.org.cn/		2012
Land use data	shp	https://www.resdc.cn/		2020
DEM	tif	https://search.asf.alaska.edu/	12.5 m	2015

Table 2. Evaluation factors.

factor code	factor name	factor code	factor name
B1	DEM	B6	road
B2	slope	B7	Population density
B3	precipitation	B8	lithology
B4	landuse	B9	fault
B5	river	B10	NDVI

in the Xishuangbanna area. Algorithms and data analysis are implemented through the Python programming platform, and spatial analysis and mapping are completed by ArcGIS 10.2 software of ESRI company.

3.1. Random Forest Model

In the process of RF model documented in this study, the model was trained first and used to predict and classify the basalt tectonic backgrounds of the test samples. The importance of different elements was also calculated using this model. Random forest (Athey, Tibshirani, Wager et al., 2019) belonged to an ensemble learning (Choi, Gu, Chin et al., 2020) algorithm. The basic idea of this algorithm is to use the Bootstrap sampling method to perform a sampling operation with replacement from the original data set, then build a decision tree from the sampled original data subset, and finally combine multiple decision trees into one. The mean value of the built decision tree is finally used as the result of the random forest regression prediction. The specific steps of the algorithm include that each decision tree in the random forest algorithm model (**Figure 2**) contains a tree-like sequence of decision nodes. Based on this sequence, the tree is split into various branches until it reaches the end (leaf) of the tree. The prediction results of each decision tree are output through leaf nodes, and finally, the outputs of multiple decision trees are combined for prediction. The random forest algorithm has the advantages of fast training speed and avoiding overfitting. Python3.6 language, Sklearn and other function libraries are used in this study to program the random forest model (**Figure 2**).

For classification tasks, the most commonly used combination strategy is the voting method. Suppose the set of categories is $\{C_1, C_2, \cdots, C_N\}$, For the convenience of discussion, the predicted output of HI on sample x is expressed as an n-dimensional vector $\left(h_i^1(x), h_i^2(x), \cdots, H_i^N(x)\right)^{\mathrm{T}}$, where $h_i^j(x)$ represents the output of H_i on class C_j. Majority voting formula is as follows.

$$H(x)=\begin{cases} C_j, & \sum_{i=1}^{T} h_i^j(x) > 0.5\sum_{k=1}^{N}\sum_{i=1}^{T} h_i^k(x) \\ \text{reject}, & \text{other situations} \end{cases} \tag{1}$$

If a mark receives a majority of votes, it is predicted to be in that category; otherwise, the prediction is rejected. Relative majority vote (plurality voting),

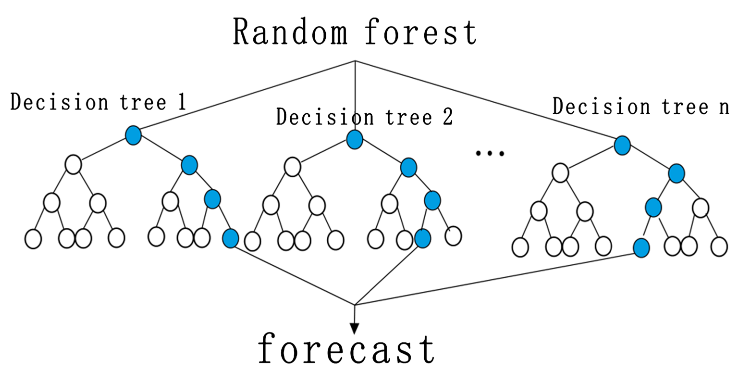


Figure 2. Random forest algorithm model (Choi, Gu, & Chin, 2020).

prediction for categories with the most votes, if there are multiple categories to get the highest votes at the same time, is to a randomly selected from it.

In this paper, the steps to implement the random forest algorithm are as follows: 1) Training random forest model by non-landslide samples obtained through remote sensing visual interpretation and landslide point. 2) The trained random forest model was used to input the feature vector (forecast data), and the voting results of each decision tree were counted to obtain the votes of each object belonging to landslide and non-landslide. 3) According to the number of votes of each object belonging to landslide and non-landslide, the landslide probability can be obtained.

3.2. Support Vector Machine (SVM)

Support Vector Machine (SVM for short) is a typical classification model, which is a typical classification model built on mathematical statistics. Unlike traditional mathematical statistics methods, support vector machines (SVMs) are an improvement of structural risk minimization methods (Wen, 2008). This algorithm was proposed by Vapnik in 1963. For linear problems, the algorithm has great advantages, but for nonlinear problems, there are still some difficulties. Later, with the introduction of the concept of kernel skills by Boser and Cuyon, the computational challenges of nonlinear support vector machines were resolved, allowing SVMs to be referenced in various application areas. The idea of the algorithm is to build an optimal classification hyperplane in high dimensions, and the hyperplane keeps the distance from the sample points on both sides to maximize, so as to realize sample classification (Huang, 2001). For the linearly separable support vector model, we first find a hyperplane that completely separates sample points of different types and maximizes the geometric interval. Converting this algorithm principle to a two-dimensional space can be represented as shown in **Figure 3**. In **Figure 3**, the cross and the solid circle represent two different categories (Tuo, 2014).

The maximum geometric interval classification hyperplane and constraints are shown in formulas (1)-(2):

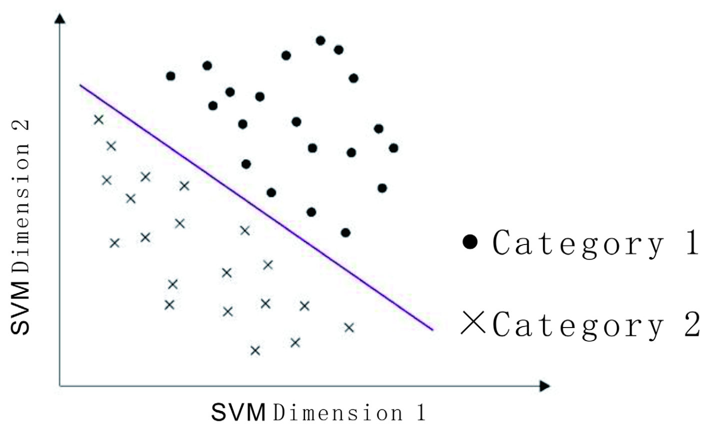


Figure 3. Two dimensional classification figure (Tuo, 2014).

$$\max_{w,b} \gamma \tag{2}$$

$$\text{s.t. } y_i\left(\frac{w}{\|w\|}\cdot x_i+\frac{b}{\|w\|}\right)\geq \gamma,\quad i=1,2,\cdots,N \tag{3}$$

Solve the maximum geometric interval γ value (unique solution), establish the Lagrangian function, and solve it based on the dual problem (w, b), $w^* \cdot x + b^* = 0$, and obtain the maximum geometric interval hyperplane: The separation decision function is shown in Equation (3):

$$f(x)=\operatorname{sign}\left(w^*\cdot x+b^*\right) \tag{4}$$

3.3. Hot Spot Analysis Method

Hot spot analysis is to calculates the $G_i^* d$ value for each evaluation unit in the evaluation model data set, that is, to calculate the locally extremely high cluster value of the evaluation factor (such as formula 4), and to determine the location where the high-value elements are clustered in space (Wang, Zhu, & Ze, 2021). The hot spot analysis function is to reflect the aggregation degree of high values of evaluation factors.

$$G_i^* d=\sum_{j=1}^{n}\omega_{ij}(d)x_j \Big/ \sum_{j=1}^{n}x_j \tag{5}$$

4. Data Processing

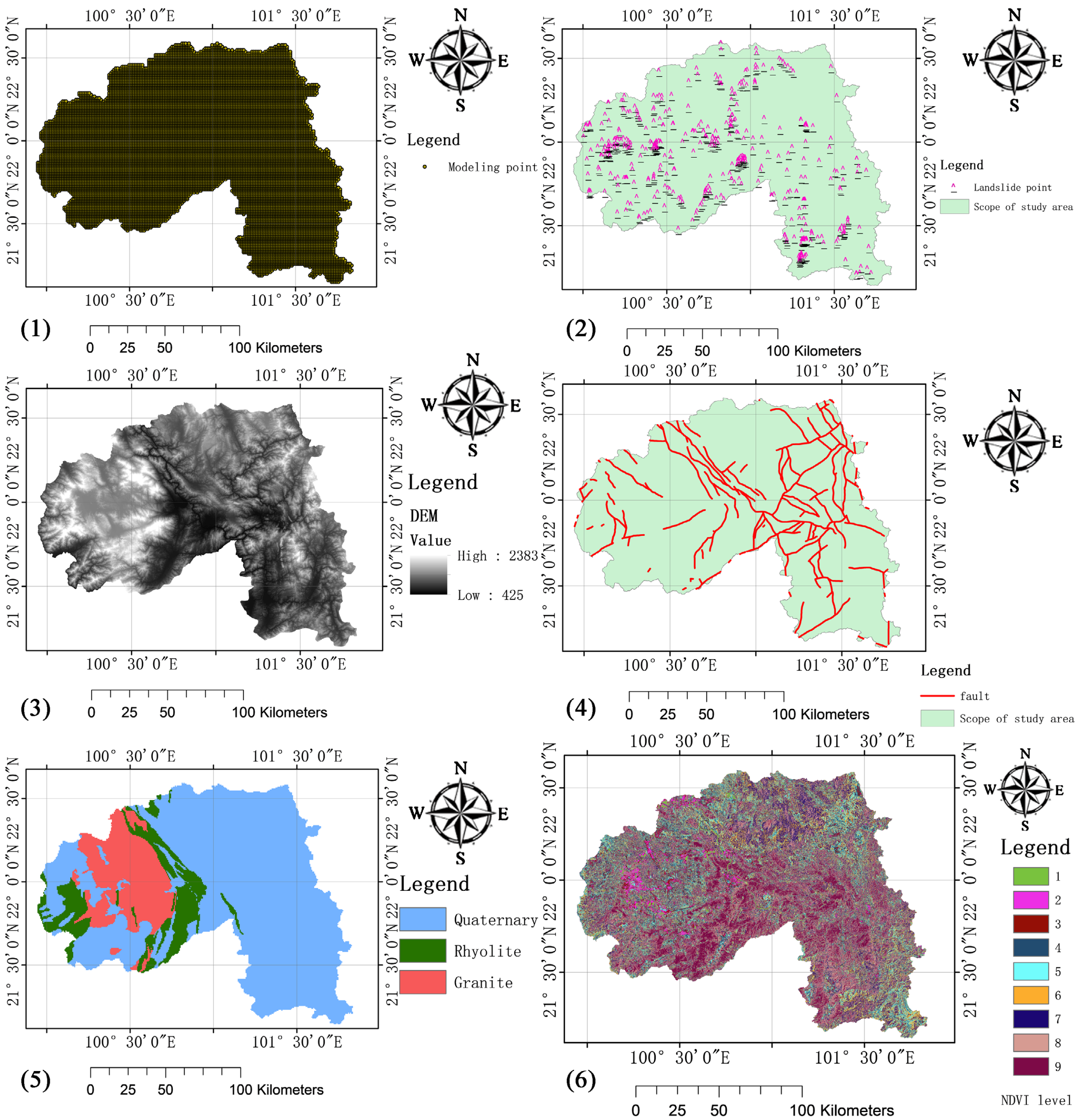
4.1. Data Preprocessing

Taking into consideration the general circumstances in the study area as well as the expert advice, 10 evaluation factors including DEM, slope aspect, precipitation, land use, water system, road, population density, lithology, fault and NDVI were selected. The fishnet tool in ArcGIS is used in this study to build a fishnet and establish an evaluation model, with a total of 4544 evaluation units. Evaluation models are built in preparation for implementing machine learning algorithms. The fishnet layer, the landslide point map, and the evaluation factor layers are shown in the following figures (**Figure 4(a)**, **Figure 4(b)**).

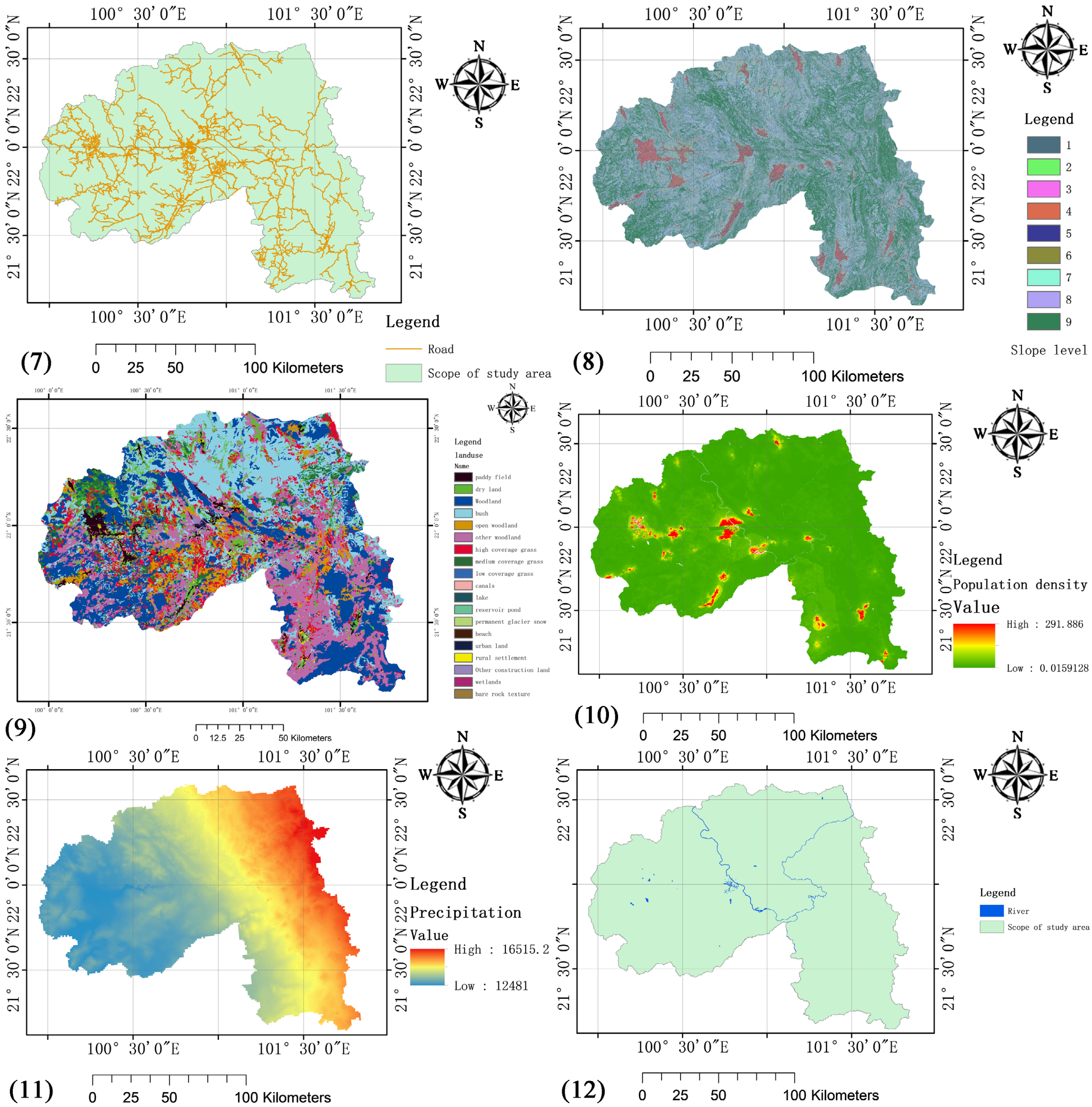
4.2. Sample Making

In this study, the geological disasters and landslide data obtained from the detailed survey data about landslides of the China Geological Survey are taken as the landslide point data, and the non-landslide points selected by remote sensing visual interpretation are taken as the non-landslide point data. The point data are made into training samples. All evaluation units in the study area were used as prediction samples. Furthermore, 214 landslide points were obtained from the detailed survey data set of geological hazards of the China Geological Survey, and 214 non-landslide points were obtained by visual interpretation.

After many attempts, make buffers of 500 m, 1000 m, 1500 m, and 2000 m for the rivers, roads, and faults obtained in Section 4.1, and then assign the attribute values of each evaluation factor layer to the fishnet layer, that is, the fishnet point file. Perform multi-attribute assignment, and finally export the multi-attribute table of the fishnet into an Excel table, and finally make training samples (Table 3) and prediction samples according to the above principles. Because this paper uses the classification algorithm of machine learning to evaluate the susceptibility of landslides, the data processing of the attribute values of each evaluation factor, such as normalization processing, principal component analysis, etc., is implemented in Python 3.6 using the corresponding functions.



(a)



(b)

Figure 4. (a) (1) Model map of fishing net evaluation (The fishing net size is kilometer grid net); (2) Distribution map of landslide points; (3) DEM map (4) Fault distribution map; (5) Lithology distribution map; (6) NDVI index distribution map; (b) (7) Road distribution map; (8) Slope distribution map; (9) Landuse distribution map; (10) Population distribution map; (11) Precipitation distribution map; (12) River distribution map.

4.3. Model Accuracy Verification and Model Selection

In this study, the model comparison test is implemented through SPSS22 software, and the accuracy of the random forest (RF) model and the support vector machine (SVM) model are compared. The two corresponding formations of ROC accuracy verification curves are shown in Figure 5.

Table 3. Training samples (intercepted part).

FID	longtitude	altitude	Label	river	NDVI	fault	road	landuse	rain	dem	popden	slope	lithology
86	101.64	21.21	1	0	6	2000	1000	24	7	3	3	6	1
119	101.66	21.22	1	0	7	0	500	31	7	4	4	8	1
460	101.31	21.29	0	0	7	0	500	12	4	1	1	5	1
487	101.58	21.29	1	0	5	0	500	24	6	4	4	8	1
567	101.32	21.31	0	0	4	0	500	52	4	1	1	2	1
618	101.30	21.32	0	0	6	0	500	31	4	1	1	5	1
622	101.34	21.32	0	0	4	0	1000	12	4	1	1	1	1
1133	101.32	21.41	0	0	5	0	500	12	4	1	1	4	1
1134	101.33	21.41	0	0	4	0	500	52	4	1	1	5	1

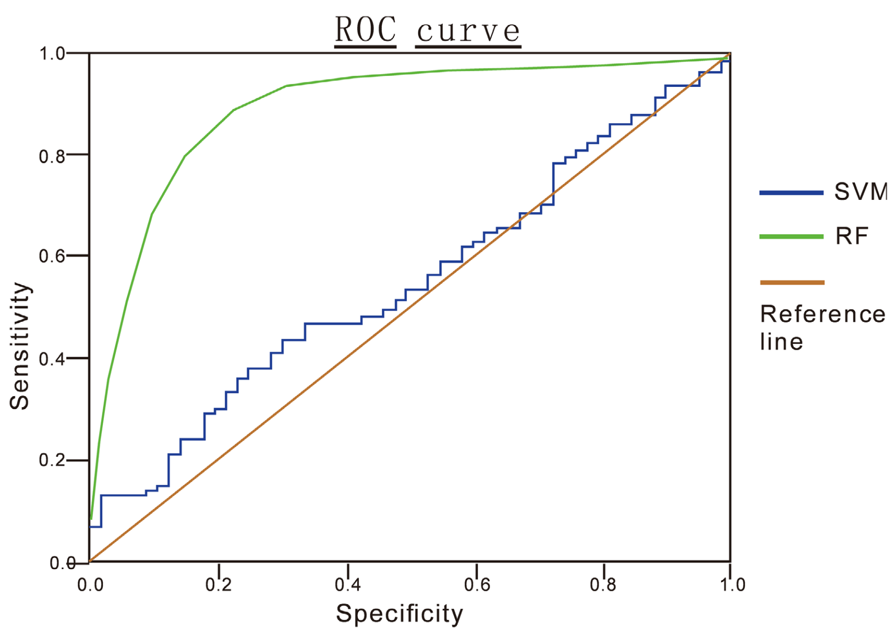


Figure 5. ROC curves of Random Forest (RF) and support vector machine (SVM) models.

Table 4. ROC curve parameters of random forest (RF) and support vector machine (SVM) models.

Test Outcome Variables	area	standard error	Asymptotic significance level b	Asymptotic 95% confidence interval	
				lower limit	upper limit
SVM	0.549	0.035	0.196	0.482	0.617
RF	0.872	0.031	0.000	0.701	0.822

Both **Figure 5** and **Table 4** shows that the AUC (area under the ROC curve) of the random forest (RF) is 0.872, and the AUC area under the ROC curve of the support vector machine (SVM) model is 0.549 (**Figure 5**). It can be seen that the random forest (RF) model has higher precision, and it is more suitable for application and promotion of landslide susceptibility evaluation.

4.4. Application of Random Forest Model

The Python 3.6 Programming language is used to implement the random forest (RF) model and the support vector machine (SVM) model. By repeatedly comparing the ROC curve characteristics of the random forest (RF) and support vector machine (SVM) models (Table 2), it is proved that the effect of the random forest model is better, and the accuracy of the landslide susceptibility evaluation is higher.

It can be seen from Figure 6 that the median of non-ground landslide points is higher than that of ground landslide points, indicating that there may be unknown ground landslide points hidden in non-ground disaster points, and the distribution of ground landslide points is relatively concentrated.

Figure 7 provides that the training of the random forest (RF) model has a faster convergence speed, and it avoids problems such as overfitting.

This paper uses python programming and principal component analysis algorithm to calculate the importance of each evaluation factor (Figure 8), and is ranked as follows, slope aspect, normalized vegetation index (NDVI), land use type, DEM, population density (popden), precipitation, fault, lithology, road and river.

5. Results

5.1. Evaluation Division of Landslide Susceptibility

The susceptibility probability predicted by the random forest model is displayed in 5 levels by the natural interval method (Figure 9). It can be seen from Figure 9 that the distribution area of the disaster points is consistent with the first-level susceptibility area. The landslide points are divided into two categories, the

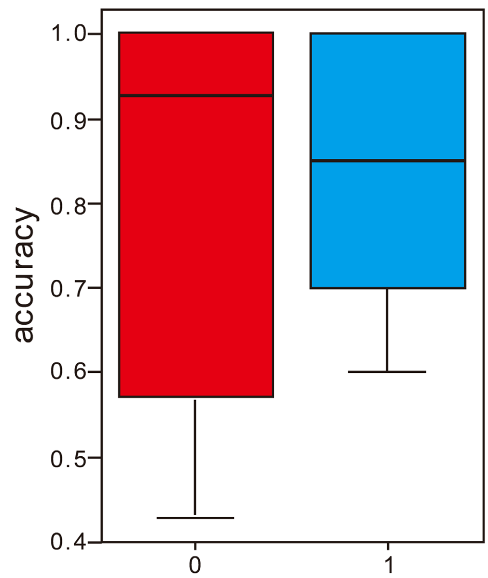


Figure 6. Box diagram (0 is the off-site disaster point, 1 is the ground disaster point).

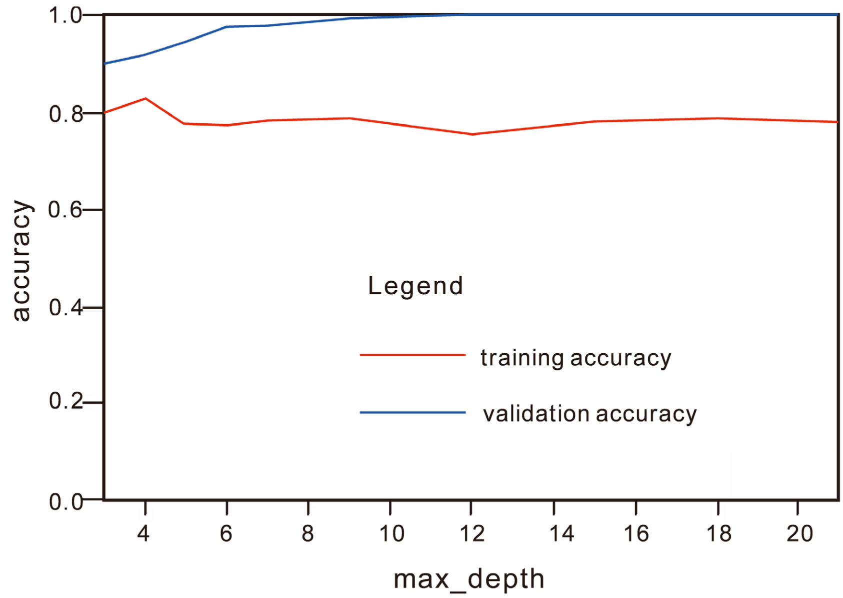


Figure 7. Random forest model training figure.

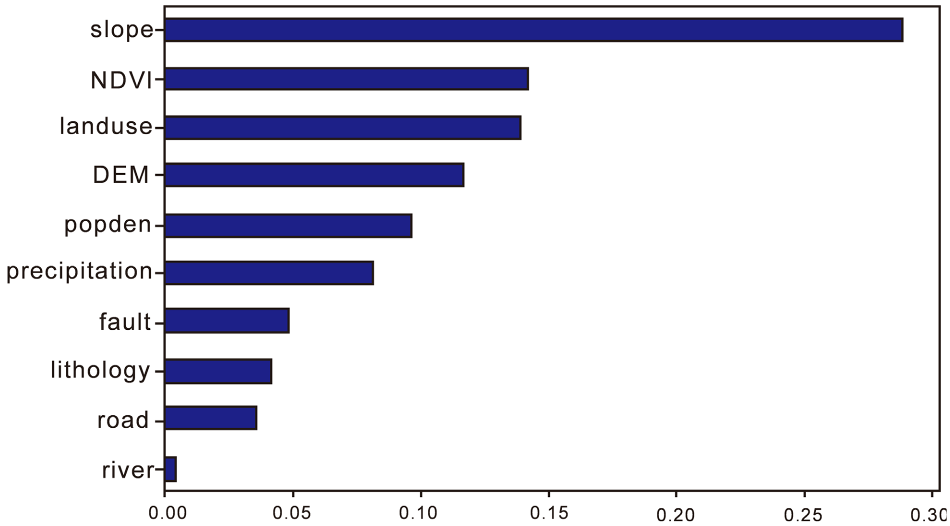


Figure 8. Ranking figure of the importance of evaluation factors.

Table 5. Frequency ratio table.

Disaster-prone area level	Evaluation units	Disaster points	Frequency ratio
1	321	146	9.748780344
2	783	29	1.06759199

first-level disaster-prone area (321 evaluation units) contains 146 disaster points, with a frequency ratio of 9.748780344. And the second-level disaster-prone area (783 evaluation units) contains 29 disaster points with a frequency ratio of 1.06759199 (Table 5). Moreover, after the superposition analysis, 87.26% of the disaster points fall in the 1st and 2nd susceptibility areas, which proves that the random forest (RF) model has a high prediction accuracy (Figure 9).

5.2. Hot Spot Analysis of Geological Hazards in Xishuangbanna Area

In this study, the probability value of landslide susceptibility predicted by the random forest model is used to achieve the hot spot analysis of landslide susceptibility in the Xishuangbanna area by utilizing the hot spot function module in ArcGIS 10.2 software. As it can be seen from Figure 10, that the hot spots of landslides are mainly distributed in the west and southwest of the study area, and the distribution of hot spots and landslides is substantially the same, which proves that the random forest model is applied to the evaluation of the susceptibility of landslides with high accuracy (Figure 10).

The NW and NW faults in the study area are developed in multiple circles superposition, and granite, rhyolite and dacite are widely distributed in the southwest of the study area. These geological phenomena are consistent with the large

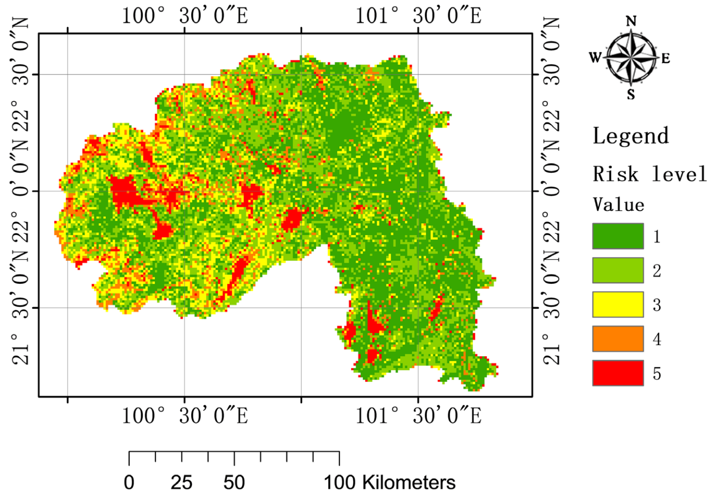


Figure 9. Landslide susceptibility classification map of Xishuangbanna.

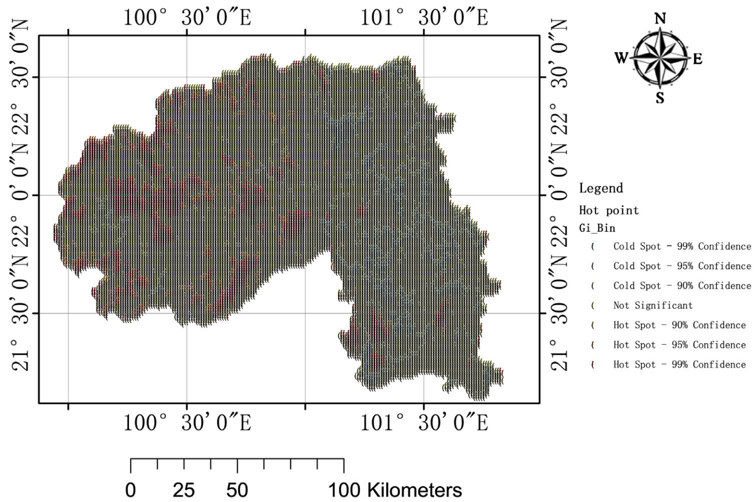


Figure 10. Hot spot analysis map of landslide susceptibility in Xishuangbanna.

distribution area of landslide hotspots in the southwest of the study area, indicating that the occurrence of landslide points is positively correlated with the distribution of faults and granite, rhyolite and dacite.

6. Discussion

Landslide is one of the most destructive typical geological disasters in mountainous areas (Achour et al., 2017; Pham et al., 2021). Therefore, landslide susceptibility mapping is crucial for economic and social development in mountainous areas (Ha et al., 2020).

At present, a variety of machine learning methods have been applied to landslide susceptibility mapping, including random forest (Chen et al., 2018), Support vector machines (Nhu et al., 2020), Decision Tree (Saito, Nakayama, & Matsuyama, 2009), Neural network (Wang et al., 2020) and Extreme learning Machine (Zhou et al., 2018). In addition, these methods are effective for solving classification and regression problems and dimensionality reduction of high-dimensional data (Trinh, Wu, Huang, & Azhar, 2020).

As can be seen from **Figure 5** and **Table 4** of this paper, the accuracy of random forest (RF) is 0.872, and the accuracy of support vector machine (SVM) model is 0.549. It can be seen that random forest (RF) model has higher precision and is more suitable for the application and extension of landslide susceptibility assessment and mapping in mountainous areas.

The random forest algorithm has the following advantages: 1) It can process high-dimensional data without feature selection, because the feature subset is randomly selected; 2) After training, it can determine the importance of features; 3) When creating random forests, unbiased estimation is used for generalization error, and the model has strong generalization ability; 4) Random forest has the data outside the bag), and there is no need to separate the cross-verification set: 5) The training between the trees is independent of each other, the training speed is fast, easy to make parallel method; 6) It is not sensitive to the missing value, if a large part of the features are lost, and it can still maintain accuracy.

In addition to the comparison between random forest algorithm and support vector machine algorithm, this paper also compares BP neural network, deep neural network. Random forest algorithm is the best for information extraction with strong random characteristics such as landslide geological disasters.

7. Conclusion

The core content of landslide susceptibility evaluation is the spatial probability of landslide occurrence in a certain location under certain conditions within the regional scope. As the basis of judging landslide hazard and risk, susceptibility evaluation is an indispensable work in disaster prevention and reduction. In this paper, the random forest model is used to predict the vulnerability probability of the study area, and the following results are obtained.

1) After a full analysis of overall circumstances in the study area, a total of 10

evaluation factors such as DEM, slope aspect, precipitation, land use, water system, road, population density, lithology, faults, and NDVI were selected to conduct a landslide analysis in the Xishuangbanna area.

2) The Random forest (RF) model and support vector machine (SVM) model were used to evaluate the susceptibility of landslides in the Xishuangbanna area, and the accuracy of the model was verified using the ROC curve. The AUC area of the random forest (RF) ROC curve was 0.761, and the AUC area of the ROC curve of the support vector machine (SVM) model is 0.549, which proves that the random forest (RF) has high accuracy, and it is suitable for the application and promotion of landslide susceptibility evaluation.

3) According to the susceptibility probability of each evaluation unit calculated by the random forest model, the study area is divided into 5 landslide risk levels using the natural discontinuity method in ArcGIS 10.2. It is concluded by using the overlay analysis function in GIS that 87.26% of the disaster points fall in the first- and second-level disaster-prone areas, which proves that the random forest (RF) model has a higher prediction accuracy.

4) The results of hot spot analysis are substantially consistent with the distribution of landslides. Hot spots and especially the hot spots of landslides are mainly distributed in the west and southwest of the study area, which is consistent with the actual situation in the study area.

Author Contribution

Zhen Chen: data collection, analysis, writing, investigation, methodology, writing—original draft, writing—review and editing, validation, project administration, resources, software, supervision, and visualization. Yiyang Zheng: conceptualization, data collection, formal analysis, supervision and resources, and data acquisition. All authors reviewed the results and approved the final version of the manuscript.

Data Availability

The datasets generated during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Achour, Y., Boumezbeur, A., Hadji, R., Chouabbi, A., Cavaleiro, V., & Bendaoud, E. A. (2017). Landslide Susceptibility Mapping Using Analytic Hierarchy Process and Information Value Methods along a Highway Road Section in Constantine, Algeria. *Arabian Journal of Geosciences, 10,* Article No. 194. https://doi.org/10.1007/s12517-017-2980-6

Athey, S., Tibshirani, J., Wager, S. et al. (2019). Generalized Random Forests. *Annals of Statistics, 47,* 1148-1178. https://doi.org/10.1214/18-AOS1709

Bai, L., Meng, Y., Lv, Q. L., & Zhang, W. W. (2015). Status and Regional Characteristics of Tourism Geological Resources in Xishuangbanna. *Journal of Anhui Agricultural Sciences, 43,* 250-253.

Balteanu, D., Micu, M., Jurchescu, M. et al. (2020). National-Scale Landslide Susceptibility Map of Romania in a European Methodological Framework. *Geomorphology, 371,* 22-44. https://doi.org/10.1016/j.geomorph.2020.107432

Chen, W., Shahabi, H., Zhang, S., Khosravi, K., Shirzadi, A., Chapi, K., & Ahmad, B. B. (2018). Landslide Susceptibility Modeling Based on GIS and Novel Bagging-Based Kernel Logistic Regression. *Applied Sciences, 8,* Article No. 2540. https://doi.org/10.3390/app8122540

Choi, J., Gu, B., Chin, S. et al. (2020). Machine Learning Predictive Model Based on National Data for Fatal Accidents of Construction Workers. *Automation in Construction, 110,* Article ID: 102974. https://doi.org/10.1016/j.autcon.2019.102974

Dou, J., Yamagishi, H., Pourghasemi, H. R. et al. (2015). An Integrated Artificial Neural Network Model for the Landslide Susceptibility Assessment of Osado Island, Japan. *Natural Hazards, 78,* 1749-1776. https://doi.org/10.1007/s11069-015-1799-2

Goetz, J. N., Brenning, A., Petschko, H. et al. (2015). Evaluating Machine Learning and Statistical Prediction Techniques for Landslide Susceptibility Modeling. *Computers & Geosciences, 81,* 1-11. https://doi.org/10.1016/j.cageo.2015.04.007

Ha, N. D., Sayama, T., Sassa, K., Takara, K., Uzuoka, R., Dang, K., & Van Pham, T. (2020). A Coupled Hydrological-Geotechnical Framework for Forecasting Shallow Landslide Hazard—A Case Study in Halong City, Vietnam. *Landslides, 17,* 1619-1634. http://link.springer.com/10.1007/s10346-020- 01385-8 https://doi.org/10.1007/s10346-020-01385-8

Hu, G. B. (2019). Interpretation of Geological Structures in Xishuangbanna Area, Yunnan Based on Multi-Source Remote Sensing Data. *Journal of Geology, 43,* 460-468.

Huang, R. Q. (2001). Information Technology for Geological Environment Management and Geological Hazard Assessment in the 21st Century (Serial I). *Science and Technology Management of Land and Resources, No. 3,* 30-34.

Liu, J., Li, S. L., & Chen, T. (2018). Landslide Susceptibility Evaluation Based on Optimized Random Forest Model. *Journal of Wuhan University (Information Science Edition), 43,* 1085-1091.

Liu, Y. Y., Di, B. F., Zhan, Y. et al. (2018). Debris Flow Susceptibility Evaluation Based on Random Forest Model—Taking the Hard-Hit Area of Wenchuan Earthquake as an Example. *Journal of Mountainous Sciences, 36,* 765-773.

Liu, Y., Wang, N. T., Zhou, C. et al. (2020). Landslide Susceptibility Assessment in Fengjie County of the Three Gorges Reservoir Area Based on the Integrated Model of ROC Curve and Deterministic Coefficient Method. *Safety and Environmental Engineering, 27,* 61-70.

Merghadi, A., Abderrahmane, B., & Tien Bui, D. (2018). Landslide Susceptibility Assessment at Mila Basin (Algeria): A Comparative Assessment of Prediction Capability of Advanced Machine Learning Methods. *International Journal of Geo-Information, 7,* 268-298. https://doi.org/10.3390/ijgi7070268

Nhu, V. H., Shirzadi, A., Shahabi, H., Singh, S. K., Al-Ansari, N., Clague, J. J., & Ahmad, B. B. (2020). Shallow Landslide Susceptibility Mapping: A Comparison between Logistic Model Tree, Logistic Regression, Naive Bayes Tree, Artificial Neural Network, and Support Vector Machine Algorithms. *International Journal of Environmental Research and Public Health, 17,* Article No. 2749. https://doi.org/10.3390/ijerph17082749

Pham, Q. B., Achour, Y., Ali, S. A., Parvin, F., Vojtek, M., Vojtekova, J., & Anh, D. T.

(2021). A Comparison among Fuzzy Multi-Criteria Decision Making, Bivariate, Multivariate and Machine Learning Models in Landslide Susceptibility Mapping. *Geomatics, Natural Hazards and Risk, 12,* 1741-1777. https://doi.org/10.1080/19475705.2021.1944330

Pourghasemi, H. R., & Kerle, N. (2016). Random Forests and Evidential Belief Function-Based Landslide Susceptibility Assessment in Western Mazandaran Province, Iran. *Environmental Earth Sciences, 75,* 185-202. https://doi.org/10.1007/s12665-015-4950-1

Pradhan, B. (2013). A Comparative Study on the Predictive Ability of the Decision Tree, Support Vector Machine and Neuro-Fuzzy Models in Landslide Susceptibility Mapping Using GLS. *Computers & Geo-Sciences, 51,* 350-365. https://doi.org/10.1016/j.cageo.2012.08.023

Saito, H., Nakayama, D., & Matsuyama, H. (2009). Comparison of Landslide Susceptibility Based on a Decision-Tree Model and Actual Landslide Occurrence: 750 The Akaishi Mountains, Japan. *Geomorphology, 109,* 108-121. https://doi.org/10.1016/j.geomorph.2009.02.026

Salciarini, D., Godt, J. W., Savage, W. Z. et al. (2006). Modeling Regional Initiation of Rainfall-Induced Shallow Landslides in the Eastern Um-Bria Region of Central Italy. *Landslides, 3,* 181-194. https://doi.org/10.1007/s10346-006-0037-0

Su, C., Wang, L. L., Wang, X. Z. et al. (2015). Mapping of Rainfall-Induced Landslide Susceptibility in Wencheng, China, Using Support Vector Machine. *Natural Hazards, 76,* 1759-1779. https://doi.org/10.1007/s11069-014-1562-0

Sun, D. L., Wen, H. J., Wang, D. Z. et al. (2020). A Random Forest Model of Landslide Susceptibility Mapping Based on Hyperparameter Optimization Using Bayes Algorithm. *Geomorphology, 362,* 28-42. https://doi.org/10.1016/j.geomorph.2020.107201

Trinh, T., Wu, D., Huang, J. Z., & Azhar, M. (2020). Activeness and Loyalty Analysis in Event-Based Social Networks. *Entropy, 22,* Article No. 119. https://doi.org/10.3390/e22010119

Tuo, Z. (2014). *Research on Image Steganalysis Algorithm Based on Image Restoration Technology*. Hunan University.

Wang, G. X., Cheng, G. D., & Qian, J. (2003). Some Problems in Ecological Security Evaluation. *Chinese Journal of Applied Ecology, No. 9,* 1551-1556.

Wang, S. H., Zhu, B. Q., & Ze, Z. (2021). *Rock Mass Discontinuity Classification Method Based on Improved Ideal Point Model* (Natural Science ed.). Journal of Northeast University.

Wang, Y., Fang, Z., Wang, M., Peng, L., & Hong, H. (2020). Comparative Study of Landslide Susceptibility Mapping with Different Recurrent Neural Networks. *Computers & Geosciences, 138,* Article ID: 104445. https://doi.org/10.1016/j.cageo.2020.104445

Wang, Z. W., Li, D. Y., & Wang, X. G. (2012). A Review of Spatial Prediction Methods for Regional Landslides. *Journal of Yangtze River Academy of Sciences, 29,* 78-85, 94.

Wen, N. (2008). *Research and Application of Support Vector Machine Classification Algorithm Based on Incremental Learning*. Kunming University of Science and Technology.

Yu, S. Y., Li, X. G., Liu, G. C. et al. (2013). *Regional Geological Survey Report of 1:250,000 Lancang, Jinghong, Menghai and Mengla Sheets in Yunnan Province*. Yunnan Geological Survey Institute.

Yunnan Provincial Bureau of Geology and Mineral Resources (1990). *Regional Geology of Yunnan Province*. Geological Press.

Zhou, C., Yin, K. L., Xiang, Z. B. et al. (2015). Quantitative Evaluation of Landslide Susceptibility in Chun'an County Based on GIS. *Safety and Environmental Engineering, 22,* 45-50, 55.

Zhou, C., Yin, K., Cao, Y., Ahmed, B., Li, Y., Catani, F., & R, H. (2018). Pourghasemi, Landslide Susceptibility Modeling Applying Machine Learning Methods: A Case Study from Longju in the Three Gorges Reservoir Area, China. *Computers & Geosciences, 112,* 23-37. https://doi.org/10.1016/j.cageo.2017.11.019