Landslide Hazard Assessment Based on Bayesian Optimization-Support Vector Machine in Nanping City, China

Wei Xie
JiangXi University of Science and Technology
Wen Nie (wen.nie@vip.tom.com)
JiangXi University of Science and Technology https://orcid.org/0000-0003-2623-0122
Pooya Saffari
Quanzhou Institute of Equipment, Haixi Institutes, China Academy of Sciences
Luis F. Robledo
Universidad Andres Bello
Pierre-Yves Descote
Universidad Andres Bello
Wenbin Jian
Fuzhou University

Manuscript

Keywords: Landslide hazard assessment, Bayesian optimization method, Support vector machine, GIS, machine learning

Posted Date: February 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-182891/v1

Abstract

Landslide hazard assessment is critical for preventing and mitigating landslide disasters. The tuning of model hyperparameters is of great importance to the accuracy and precision of one landslide hazard assessment model. In this study, Bayesian Optimization (BO) method was used to tune the hyperparameters of Support Vector Machine (SVM) model to obtain a high accuracy landslide hazard zoning map. 1711 historical landslide disaster points were obtained as landslide inventory in a case of Nanping City landslide hazard assessment. A total of 12 factors including elevation, slope, aspect, curvature, lithology, soil type, soil erosion, rainfall, river, land use, highway, and railway were selected as landslide conditional factors. The multicollinearity diagnosis was performed on factors using the Spearman correlation coefficient. 1711 landslides and 1711 non-landslides were collected as the dataset and divided into the same number of training dataset and testing dataset. The confusion matrix and receiver operating characteristic (ROC) curve were used to verify the models. The results of confusion matrix accuracy and the area under ROC curve (AUC) showed that BO-SVM (89.53%, 97%) performed better than only SVM (84.91%, 0.93), which indicated the superiority of the proposed method during this study.

1 Introduction

Landslides are cascading geological hazards in which rock, soil, or rock debris moves down the slope under the action of gravity (Fan et al. 2019; Hungr et al. 2013). Every year, landslide disasters caused a large number of casualties and economic losses worldwide (Kirschbaum et al. 2009). With aggravated climate change, it is unequivocal that the stability of natural and engineered slopes has been affected, resulting in a greater risk of landslides (Gariano and Guzzetti 2016). From 2004 to 2016 alone, in total 55,997 people were killed in 4,862 fatal landslide events (Froude and Petley 2018). Therefore, how to mitigate the serious threat of landslide disasters and prevent new landslides has become an increasingly important issue (Intrieri et al. 2019). Landslide hazard assessment is an effective measure to prevent landslide hazards, which can provide key information for disaster prevention, disaster mitigation, and disaster risk reduction (van Westen et al. 2008; Xu et al. 2012).

Geomorphological mapping, heuristic terrain, and susceptibility zoning, physically-based numerical modeling, and statistically-based classification methods have been proposed to ascertain landslide hazard assessment over the last few decades (Reichenbach et al. 2018). With the rapid development of geographic information system (GIS) and artificial intelligence (AI) technologies, a variety of advanced algorithms and models (logistic regression, support vector machines, Bayesian methods, decision tree methods, artificial neural networks, and so on) have been widely used in a large number of engineering projects and scientific research (Jafarian et al. 2019; Olen and Bookhagen 2018; Theron and Engelbrecht 2018; Violante et al. 2018; Wu et al. 2018). GIS provides a platform for collecting, organizing, and analyzing landslide events and landslide conditional factors, machine learning (ML) techniques provide more solutions for calculating the relationship between landslide conditional factors and landslide events.

Support Vector Machine (SVM) is one of the most used machine learning methods and performs satisfactorily in many studies (Reichenbach et al. 2018). Marjanovic et al. (2011) compared SVM, Decision Trees, and Logistic Regression in a specific area of the Fruska Gora Mountain (Serbia) indicating the SVM classifier outperformed the others. Chen et al. (2017) used maximum entropy (MaxEnt), SVM, and Artificial Neural Network (ANN) to find the "Spatial contraindication" pattern by their ensembles, SVM was found to be the most practical model with the highest spatial area in highly susceptible classes. Luo et al. (2019) applied ANN, SVM, and information value model (IVM) to assess one mine landslide sensitivity. ANN model and SVM achieved high prediction capability, proving their advantage of solving nonlinear and complex problems. Generally, SVM has good stability and performance in most cases.

However, the hyperparameter selection of SVM is often confusing, which affects the precision and generalization ability of the model (Dou et al. 2019). For SVM, the kernel function type is the most important hyperparameter, and the penalty factor (*C*) and *Gamma* also affect the performance of the model. Many studies used default hyperparameters in software or tools that result in less than optimal results (Abdollahi et al. 2018; Chen et al. 2017). Few studies used random search, grid search, or genetic algorithm to optimize the hyperparameters of the algorithm (Luo et al. 2019; Tang et al. 2019). Nevertheless, these optimization methods usually make the model cannot reach the balance between the overall and local optimal, thus affecting the generated landslide hazard zoning map.

To resolve the issues of hyperparameter selection of SVM, using Bayesian optimization (BO) method to optimize the hyperparameters of SVM, the new model is referred to as BO-SVM in the current work. To evaluate the optimization effect of BO method, this study compares the performance of common SVM and BO-SVM in Nanping City, China.

2 Study Area And Data

2.1 Study area and landslide inventory

The study area is Nanping city, located in the northwestern part of Fujian province, China. The geographic coverage of the study area is 17°46'31"-18°17'9"N latitude and 42°14'55 – 42°48'30"E longitude, with an area around 26,300 km^2 (Fig. 1). Nanping is administratively divided into 10 counties and districts including Jianyang, Yanping, Shunchang, Pucheng, Guangze, Songxi, Zhenghe, Shaowu, Wuyishan, and Jian'ou (Fig. 1).

The elevation ranges from 20m to 2150m above sea level, and the land above 1000 meters accounts for 12% of the total area. Surface topographic features in the study area are strongly influenced by tectonic movement. The tectonic and geomorphic features are quite obvious. The low hills are widely distributed in this whole area and high mountains are only placed in the northeast and southwest. Mountain basin valleys are distributed alternately along the river. Fault-block mountains are dominated by faults with steep peaks. The main rivers in the research area are the Minjiang River, the Jian River, and the Futun

River. The climate type of the study area is subtropical monsoon climate. Rainfall is concentrated in summer and tropical cyclones usually cause considerable precipitation.

Nanping is one of the most landslide-prone areas of southeastern China with numerous landslide events historically (Zhang and Huang 2018). The main factors causing landslide events in the study area are typhoon rainstorms and the interaction of anthropic activities and engineering geological conditions (Yin et al. 2013). The landslide inventory of the study area was collected by the field geological survey and as shown in Fig. 1. In this study, the term "landslide" includes slide, fall, and flow. The classification comes from the updated Varnes landslide classification (Hungr et al. 2013) and all landslides are represented by spatial points. There were 1008 cases of slides, 679 cases of falls, and 24 cases of flows. The information on landslide disaster points includes the type of disaster body, number of disaster body, name of disaster body, field number, geographical location, treatment situation, type of groundwater, etc.

2.2 Landslide conditional factors

According to the geographic environment information and data availability of the study area, we chose 12 landslide conditional factors: elevation, slope, aspect, curvature, lithology, soil type, soil erosion, precipitation, river, land cover, highway, and railway (Fig. 2). We have classified the input 12 thematic variables into five clusters: (□) morphological (4 variables), (□) geological (3 variables), (□) hydrological (2 variables), (□) land cover (1variables), and (□) human activities (2 variables) (Table 1). All factors were discretized into categorical variables. The aspect factor was classified according to the direction, and the remaining factors were classified according to the natural break method.

Table 1
Landslide conditional factors

Cluster	Factor	Description	Source
Morphological	Elevation	ASTER GDEM V2,30m resolution	http://www.gscloud.cn/
	Slope	30m resolution	Extracted by DEM
	Aspect	30m resolution	Extracted by DEM
	Curvature	30m resolution	Extracted by DEM
Geological	Lithology	Vector data	http://www.geodata.cn
	Soil type	Reclassify to 30m resolution	http://www.resdc.cn
	Soil erosion	Reclassify to 30m resolution	http://www.resdc.cn
Hydrological	Rainfall	Average rainfall from 1980 to 2015, interpolated from rainfall site data	http://data.cma.cn/
	River	Vector data	http://www.geodata.cn
Land cover	Land use	30m resolution	http://www.webmap.cn
Human activity	Highway	Vector data	http://www.webmap.cn
	Railway	Vector data	http://www.webmap.cn

Topography and landform play an important role in controlling the formation of landslides (Ambrosi et al. 2018). Digital Elevation Model (DEM) and slope degree were widely used factors and particularly effective in landslide predicting (Reichenbach et al. 2018). The elevation (Fig. 2a) has a direct effect on human engineering activities and other environmental factors, thereby affecting the stability of slope. The slope degree (Fig. 2b) affects the stability and overall movement rate of the unstable rock and soil on the slope (Lo et al. 2018), which is mainly 6°-20° in the study area. Aspect determines the direction of solar radiation and water flow (Fig. 2c); curvature (Fig. 2d) affects the acceleration and deceleration of flow, convergence, and dispersion (Youssef et al. 2015).

Geological structure and soil properties (Fig. 2e-g) directly determine whether the landslide would occur and how it would occur (Zezere et al. 2017). In stratigraphic lithology, landslides often occur in soft structural planes and weak rock layers. Proterozoic sedimentary rocks occupy about half of the area in the study area, followed by Granite and Mesozoic metamorphic rocks. For soil properties, looser soil and a greater degree of soil erosion are conducive to landslide breeding (Sorbino et al. 2009). Red soil and paddy soil are the main soil types in the study area, and most of the soil erosion types in the study area are hydraulic erosion Level 1. The soil erosion classification referred to the People's Republic of China industry standard SL190-96 "*Soil erosion classification and classification standards*".

The hydrological conditional factors and changes in land cover (Fig. 2h-j) are predisposing factors or direct factors of landslides (Phong et al. 2019). Dozens of landslides occur in the rainy season every year in the study area, which indicates that hydrology has a great influence on landslides. Continuous rainfall can directly cause landslides and erode the slope to provide conditions for its instability. For land cover, bare slopes are more prone to instability than slopes with lush vegetation, and forests with luxuriant root systems are more stable than grasslands. The forest coverage rate in the study area is 74.75%, and the remaining lands are mainly cultivated lands and grasslands.

The direct effects of human activities on slopes are increasing with the development of economic engineering. A large number of infrastructures and road constructions have destroyed the original structure of the hillsides, leading to the aggravation of slope instability. The construction of railways (Fig. 2j) and highways (Fig. 2j) in mountainous areas usually involves the excavation of tunnels and manual cutting of slopes. Therefore, areas along highways and railways are the worst-hit areas of landslides and the key areas for disaster reduction and prevention (Den Eeckhaut et al. 2010).

3 Methods

3.1 mapping units and dataset division

In most current studies, the mapping units commonly used in landslide susceptibility mapping and landslide hazard assessment are grid units, geomorphic units, administrative units, unique condition units, and slope units (Reichenbach et al. 2018). Compared with others, grid units usually perform better for complex calculations and simulation processes (Yang et al. 2019). Therefore, in this study, the grid units was selected as the basic mapping units. A 300m*300m grid was selected, and a total of 374,666 units were obtained. The number of landslide points falling into each unit was calculated separately. The unit with several 0 was recorded as 0, while the unit with a number rather than 0 was recorded as 1 to form a binary distribution. They are the dependent variables of the hazard assessment model. Then the attribute values of the landslide factors of each grid cell were spatially overlapped as the independent variables of the model.

In this study, 1,711 landslide disaster points fell into 1,653 units, which were used as positive samples in the data set. An equal number of non-landslides were chosen as negative samples in an area 300 meters away from the landslide points. Therefore, a total of 3306 samples were used for model training and testing. The number of training dataset and testing dataset was 50% and 50%, respectively.

3.2 Multicollinearity diagnosis

Feature selection is a necessary step in the process of machine learning modeling, which was used to eliminate redundant factors and retain useful factors. Multicollinearity is usually used as an indicator of feature selection, which means that might be correlations between multiple conditional factors (Lee et al. 2018). The existence of multicollinearity would make it difficult to capture useful information from the model, thereby affecting the evaluation results (Yanar et al. 2020). Multicollinearity diagnosis of factors

and elimination of redundant factors have a positive effect on the evaluation model. In this study, Spearman correlation analysis was used to analyze each factor in the study area, and the multicollinearity of factors was measured by the correlation coefficient R. The value range of R is [-1, 1] (when R > 0, the factors are positively correlated, |R| closer to 1, the higher the correlation; when R < 0, the factors are negatively correlated; when R = 0, there is no linear correlation).

3.3 Landslide hazard assessment model

SVM is a machine learning method based on statistical theory, which integrates multiple technologies such as relaxation variables, maximum interval hyperplane, kernel function, etc. It is suitable to solve the classification problems of small samples, nonlinearity, and high dimensionality (Cortes and Vapnik 1995). With the development of multidisciplinary integration, SVM was gradually applied to the field of natural disasters. The basic principle is to map the samples of the input space to a high-dimensional characteristic space through nonlinear transformation, and then determine the optimal classification plane that linearly separates the samples in the characteristic space (Chang and Lin 2011; Smola and Schölkopf 2004). In the studies of landslide susceptibility assessment and risk assessment, the occurrence of landslides fits well with the characteristics of the algorithm of solving binary classification problems (Ballabio and Sterlacchini 2012; Tien Bui et al. 2012a; Tien Bui et al. 2018).

The schematic diagram of SVM's principle is as shown in Fig. 3. The distance between the hyperplane and the nearest sample point is called the margin. The larger the margin, the higher the generalization ability of the classifier. Therefore, the purpose of SVM is to find the hyperplane that maximizes the margin, that is, the optimal hyperplane. All points on the hyperplane on both sides of the margin are called support vectors, and the classification boundary is determined only by the support vectors not by other data and the amount of data. Therefore, the adjustment of hyperparameters is extremely critical to the performance of SVM. The main hyperparameters involved in SVM are kernel type, *C*, and *gamma*. As mentioned above, the kernel maps the observations into some feature space. Hyperparameter *C* controls the trade-off between decision boundary and accuracy by adding a penalty for each misclassified data point. *Gamma* is a parameter related to *C* in some kernel types. If gamma is large, the effect of *C* becomes negligible. If gamma is small, *C* affects the model like it affects a linear model. In this study, we used the *scikit-learn* package based on Python to implementation SVM (Pedregosa et al. 2011).

3.4 Bayesian optimization algorithm

The process of implementing machine learning algorithms usually needs to consider the tuning of learning parameters and model hyperparameters (Snoek et al. 2012). The hyperparameters define the attributes of the model or the training process, which have a great influence on the final effect of the model (Greenhill et al. 2020). BO is a hyperparameter optimization (selection) method of general machine learning algorithms. BO algorithm is widely utilized in the field of cutting-edge artificial intelligence with obvious advantages over genetic algorithm, particle swarm optimization algorithm, or other algorithms (Greenhill et al. 2020; Kobliha et al. 2006). It builds a surrogate for the objective and quantifies the uncertainty in that surrogate using a Bayesian machine learning technique and Gaussian process

regression, and then uses an acquisition function defined from this surrogate to decide where to sample (Frazier 2018). Generally speaking, the problem scenarios that Bayesian optimization algorithm mainly faces are:

$$X^* = arg_{x \in S} \max f(x) \quad (1)$$

Where *S* is the candidate set of *x*. The goal is to choose x from *S* such that the value of *f(x)* is the smallest or largest. The hyperparameter value obtained by the BO algorithm replaces the original value. Finally, a new hybrid model (BO-SVM) is constructed. Package *hyperopt* on the Python platform was used to implement BO algorithm in this study.

3.5 Model evaluation and verification

3.5.1 Confusion matrix

The confusion matrix measures the accuracy of a classifier classification and it is also known as the error matrix. The confusion matrix is often used to evaluate the results of binary regression models such as logistic regression and SVM. This method can quantitatively express the correct rate of 0-value prediction, the correct rate of 1-value prediction, and the overall prediction rate in the model results (Yang et al. 2019).

3.5.2 ROC curve

The Receiver Operating Characteristic (ROC) curve is a comprehensive indicator of response sensitivity and specific variables (Tehrany et al. 2015). In the landslide risk assessment, the X-axis of the ROC curve is the specificity indicating the probability of misprediction of the non-disaster points. And the Y-axis is the sensitivity, representing the prediction success rate of the disaster point. The prediction accuracy of the model is expressed by the size of the area enclosed by the curve and the abscissa. The closer the curve is to the upper left corner, the higher the accuracy of the model. The area under the curve is called AUC and the range of AUC values is [0, 1]. The value of AUC closer to 1 indicates the higher accuracy of the model.

4 Results

4.1 Multicollinearity of factors

Generally, factors with high multicollinearity values should be removed or detected iteratively to ensure the reliability of the model. The multicollinearity diagnosis results among the 12 factors are given in Fig. 4. The correlation coefficient between each implemented factor is lower than 0.5, indicating that the low multicollinearity between factors. Therefore, in this study, all 12 conditional factors were retained. The value of collinearity among most factors is around 0, indicating that the correlation between them is extremely low. Moreover, the multicollinearity between land use factors and human activities factors has a higher value comparing others.

4.2 Verification and comparison of models

The hyperparameters corresponding to the most optimal evaluation value of BO proceeding were: the radial basis function (RBF) was used as the kernel, the penalty factor C was $1*10^{8.475}$, and the RBF gamma value was $2.895*10^{-7}$. These values would be set as the hyperparameter values of BO-SVM before modeling while the SVM used the default hyperparameters. The performance of SVM and BO-SVM models were verified and compared using the confusion matrix and ROC curve, respectively. ROC curve presses the predictive capabilities of the models, and the confusion matrix represents the details of the predictive ability of the model.

The results of the confusion matrix are shown in Table 2, where the accuracy of BO-SVM is 89.63%, which is 5% approximately higher than SVM with 84.91%. Compared with SVM, BO-SVM has higher prediction accuracy for landslides and non-landslides. The prediction accuracy of SVM for landslides and non-landslides are 88.64% and 81.18%, respectively, which are lower than BO-SVM, indicating that BO-SVM's prediction accuracy for negative and positive are relatively robust.

Table 2
The confusion matrix of the SVM and BO-SVM.

Method	Landslide occurred(Actual)	Prediction		Percent	Accuracy
		Yes	No		
SVM	Yes	1466	187	88.64%	84.91%
	No	311	1342	81.18%	
BO-SVM	Yes	1523	130	92.14%	89.53%
	No	216	1437	86.93%	

The ROC curve and the area under the ROC curve (AUC) are illustrated in Fig. 5. Generally, AUC values greater than 0.9 are considered excellent (Merghadi et al. 2020). In this case, both models have high AUC values of more than 0.9, and BO-SVM with 0.97 is 4% higher than SVM with 0.93. According to the results of the confusion matrix and ROC curve, it can be concluded that the performance of BO-SVM is better than that of SVM.

4.2 Landslide hazard map

According to the analysis results of the landslide hazard models, the spatial overlay analysis of each conditional factor was carried out to obtain the landslide hazard index in the study area. The results ranged from 0 to 1, and it was divided into four zones with an interval of 0.25, namely: low, moderate, high, and very high. The landslide hazard maps generated by SVM and BO-SVM are as shown in Fig. 6. The statistic results of map units and landslides in each hazard zone are as shown in Table 3. In the results of BO-SVM, the area of low hazard zone accounted for 54.98% of the study area and the number

of historical landslides only accounted for 5.96% of the total landslides. The area of very high and high hazard zone accounted for 14.83% of the total area and the number of historical landslides accounted for 64.87% of the total landslides. As a contrast, the SVM shows less accurate results, where the low zone and high zone are similar to that of BO-SVM, but the very high zone only includes 15.14% landslides. Besides, there are 39.45% landslides in the moderate zone, which demonstrate low confidence. Therefore, the effectiveness and reliability of BO-SVM are higher than those of SVM.

Table 3
Statistic result of two landslide hazard models.

Hazard zone	Model	Grid number	Area proportion	Landslide number	Landslide proportion
Low	SVM	548968	15.96%	79	4.62%
	BO-SVM	1891009	54.98%	102	5.96%
Moderate	SVM	2193576	63.78%	675	39.45%
	BO-SVM	1038155	30.19%	499	29.16%
High	SVM	602680	17.52%	698	40.79%
	BO-SVM	410526	11.94%	702	41.03%
Very high	SVM	93946	2.73%	259	15.14%
	BO-SVM	99,480	2.89%	408	23.85%

5 Discussion

In general, the improvement of the reliability and accuracy of the landslide hazard assessment results is concentrated in two parts: better data and stronger model. As the data improvement is limited by the availability of data and the actual situation of the research area, thus, the improvement of assessment models becomes particularly crucial. In many studies, researchers considered many algorithms to compare their performance while ignoring the improvement of a single model itself (Akgun 2011; Tien Bui et al. 2012b). By contrast, this study shows that the in-depth improvement of some commonly used models like SVM is also very helpful for landslide hazard assessment. A new model named BO-SVM based on BO algorithm was proposed in this study. In theory, replacing the empirical risk minimization principle in the traditional methods with the structural risk minimization principle, the BO algorithm obtains the overall optimal hyperparameters of the model through the Gaussian process to improve the performance of the model. In practice, in the case of the same input dataset, by optimizing the hyperparameters of the model through the BO algorithm, the result of the landslide hazard assessment

was improved. The prediction effect of the BO-SVM model is higher than that of the SVM model, the prediction accuracy and AUC value are increased by 5% and 4%, respectively.

In the entire study area, the hazard zones' spatial distribution of the two models is roughly similar (Fig. 6). The results of SVM and BO-SVM are roughly similar in most areas because the difference between the models is not large fundamentally based on dualistic statistics. The very high hazard zones are distributed in the southwest and northeast of the study area (Yanping District and Pucheng County), where they have suffered the most landslide disasters in history. The low hazard zones and moderate hazard zones are most widely distributed in the study area. There have been a few landslide disasters in these areas, but the results of the two models showed that they are far from reaching a high risk.

While, on a local scale, the results of the two models have some noticeable differences. The moderate zones area of the SVM model is significantly larger than BO-SVM. These moderate zones are usually called the uncertainty interval, indicating that the model has a low degree of confidence in the occurrence of the landslide (Sun et al. 2020). The result of SVM showed many central parts of the study area including most parts of Wuyishan County, Jianyang District, and Jianou County. However, there are some scars of high zone along the railways, which in fact there were few landslides reported in these areas. BO-SVM produced less uncertain intervals and maintained a high accuracy in the areas along the railways. The above results show that BO-SVM outperforms SVM and indicate the improvement of the model by BO.

In this study, the key step to run BO was to realize the Gaussian process regression algorithm and optimize the compute through the kernel trick of SVM. It is difficult for common optimization algorithms to make full use of all known results, and they may also ignore the potential relationships between the known results (Nhu et al. 2020). Considering BO based on probability, our method could be more effective. To some extent, this study still had some limitations. The entire model has a certain black-box nature, and the optimization of the process relies on the "trial and error" of the computer, rather than calculating towards a visible goal. Reducing the nature of the black box and improving the interpretability of the model is an important point for future research.

6 Conclusion

The findings of this study could provide a rational perspective for improving landslide hazard assessment. It aimed to use a hyperparameter optimization method (BO algorithm) to solve the problem of machine learning hyperparameter selection in landslide hazard assessment. In the case of BO-SVM and SVM as the assessment model and Nanping as the study area, the results of BO-SVM were better than SVM. For BO-SVM, the accuracy of confusion matrix and the AUC value of ROC curve were 89.53% and 0.97, respectively. In the landslide hazard zoning map generated by the two models, the BO-SVM is also more reliable, 65% of historical landslides fell in very high zone and high zone with an area of less than 15%. The novel model in this study is of certain significance to other landslide studies using

machine learning methods. Additionally, the results could be helpful for risk assessment and management of other natural disasters.

Declarations

Acknowledgments

We are grateful to the anonymous reviewers and the Editor for their constructive comments that helped us improve the quality of the paper. Acknowledgment for the data support from "National Earth System Science Data Center, National Science & Technology Infrastructure of China. (http://www.geodata.cn)", International Scientific & Technical Data Mirror Site, Computer Network Information Center, Chinese Academy of Sciences. (http://www.gscloud.cn).

Funding

The work was supported by the National Natural Science Foundation of China (No.41861134011) and (No.51874268).

Conflicts of interest

The authors declare no competing financial interests.

Availability of data and material

All data and material in this paper are available from the Internet and the URL where the data were obtained has been showed in the text.

Code availability

Code Non-Public.

References

1. Abdollahi S, Pourghasemi HR, Ghanbarian GA, Safaeian R (2018) Prioritization of effective factors in the occurrence of land subsidence and its susceptibility mapping using an SVM model and their different kernel functions. Bull Eng Geol Environ 78:4017-4034. doi:10.1007/s10064-018-1403-6
2. Akgun A (2011) A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey. Landslides 9:93-106. doi:10.1007/s10346-011-0283-7
3. Ambrosi C, Strozzi T, Scapozza C, Wegmüller U (2018) Landslide hazard assessment in the Himalayas (Nepal and Bhutan) based on Earth-Observation data. Eng Geol 237:217-228. doi:10.1016/j.enggeo.2018.02.020

4. Ballabio C, Sterlacchini S (2012) Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy. Math Geosci 44:47-70. doi:10.1007/s11004-011-9379-9
5. Chang C-C, Lin C-J (2011) Libsvm. ACM Trans Intell Syst Technol 2:1-27. doi:10.1145/1961189.1961199
6. Chen W, Pourghasemi HR, Kornejady A, Zhang N (2017) Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. Geoderma 305:314-327. doi:10.1016/j.geoderma.2017.06.020
7. Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20:273-297. doi:10.1007/BF00994018
8. Den Eeckhaut MV, Marre A, Poesen J (2010) Comparison of two landslide susceptibility assessments in the Champagne–Ardenne region (France). Geomorphology 115:141-155. doi:10.1016/j.geomorph.2009.09.042
9. Dou J et al. (2019) Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. Landslides 17:641-658. doi:10.1007/s10346-019-01286-5
10. Fan X et al. (2019) Earthquake-Induced Chains of Geologic Hazards: Patterns, Mechanisms, and Impacts. Rev Geophys 57:421-503. doi:10.1029/2018rg000626
11. Frazier PI (2018) A tutorial on bayesian optimization. arXiv preprint arXiv:180702811
12. Froude MJ, Petley DN (2018) Global fatal landslide occurrence from 2004 to 2016. Nat Hazards Earth Syst Sci 18:2161-2181. doi:10.5194/nhess-18-2161-2018
13. Gariano SL, Guzzetti F (2016) Landslides in a changing climate. Earth-Sci Rev 162:227-252. doi:10.1016/j.earscirev.2016.08.011
14. Greenhill S, Rana S, Gupta S, Vellanki P, Venkatesh S (2020) Bayesian Optimization for Adaptive Experimental Design: A Review. IEEE Access 8:13937-13948. doi:10.1109/access.2020.2966228
15. Hungr O, Leroueil S, Picarelli L (2013) The Varnes classification of landslide types, an update. Landslides 11:167-194. doi:10.1007/s10346-013-0436-y
16. Intrieri E, Carlà T, Gigli G (2019) Forecasting the time of failure of landslides at slope-scale: A literature review. Earth-Sci Rev 193:333-349. doi:10.1016/j.earscirev.2019.03.019
17. Jafarian Y, Lashgari A, Haddad A (2019) Predictive Model and Probabilistic Assessment of Sliding Displacement for Regional Scale Seismic Landslide Hazard Estimation in Iran. Bull Seismol Soc Amer 109:1581-1593. doi:10.1785/0120190004
18. Kirschbaum DB, Adler R, Hong Y, Hill S, Lerner-Lam A (2009) A global landslide catalog for hazard applications: method, results, and limitations. Nat Hazards 52:561-575. doi:10.1007/s11069-009-9401-4
19. Kobliha M, Schwarz J, Ocenasek J (2006) Bayesian optimization algorithms for dynamic problems. In: Rothlauf F (ed) Applications of Evolutionary Computing, Proceedings, vol 3907. Lecture Notes in Computer Science. pp 800-804

20. Lee J-H, Sameen MI, Pradhan B, Park H-J (2018) Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. Geomorphology 303:284-298. doi:10.1016/j.geomorph.2017.12.007
21. Lo C-M, Feng Z-Y, Chang K-T (2018) Landslide hazard zoning based on numerical simulation and hazard assessment. Geomatics, Natural Hazards and Risk 9:368-388. doi:10.1080/19475705.2018.1445662
22. Luo X, Lin F, Zhu S, Yu M, Zhang Z, Meng L, Peng J (2019) Mine landslide susceptibility assessment using IVM, ANN and SVM models considering the contribution of affecting factors. PLoS One 14:e0215134. doi:10.1371/journal.pone.0215134
23. Marjanović M, Kovačević M, Bajat B, Voženílek V (2011) Landslide susceptibility assessment using SVM machine learning algorithm. Eng Geol 123:225-234. doi:10.1016/j.enggeo.2011.09.006
24. Merghadi A et al. (2020) Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Sci Rev 207. doi:10.1016/j.earscirev.2020.103225
25. Nhu V-H et al. (2020) Effectiveness assessment of Keras based deep learning with different robust optimization algorithms for shallow landslide susceptibility mapping at tropical area. Catena 188:13. doi:10.1016/j.catena.2020.104458
26. Olen S, Bookhagen B (2018) Mapping Damage-Affected Areas after Natural Hazard Events Using Sentinel-1 Coherence Time Series. Remote Sens 10:19. doi:10.3390/rs10081272
27. Pedregosa F et al. (2011) Scikit-learn: Machine learning in Python. the Journal of machine Learning research 12:2825-2830
28. Phong TV et al. (2019) Landslide susceptibility modeling using different artificial intelligence methods: a case study at Muong Lay district, Vietnam. Geocarto Int:1-24. doi:10.1080/10106049.2019.1665715
29. Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. Earth-Sci Rev 180:60-91. doi:10.1016/j.earscirev.2018.03.001
30. Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199-222. doi:10.1023/b:Stco.0000035301.49549.88
31. Snoek J, Larochelle H, Adams RP Practical bayesian optimization of machine learning algorithms. In: Advances in neural information processing systems, 2012. pp 2951-2959
32. Sorbino G, Sica C, Cascini L (2009) Susceptibility analysis of shallow landslides source areas using physically based models. Nat Hazards 53:313-332. doi:10.1007/s11069-009-9431-y
33. 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 362:14. doi:10.1016/j.geomorph.2020.107201
34. Tang X, Hong H, Shu Y, Tang H, Li J, Liu W (2019) Urban waterlogging susceptibility assessment based on a PSO-SVM method using a novel repeatedly random sampling idea to select negative samples. J Hydrol 576:583-595. doi:10.1016/j.jhydrol.2019.06.058

35. Tehrany MS, Pradhan B, Mansor S, Ahmad N (2015) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. Catena 125:91-101. doi:10.1016/j.catena.2014.10.017
36. Theron A, Engelbrecht J (2018) The Role of Earth Observation, with a Focus on SAR Interferometry, for Sinkhole Hazard Assessment. Remote Sens 10:30. doi:10.3390/rs10101506
37. Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012a) Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree, and Naïve Bayes Models. Math Probl Eng 2012:1-26. doi:10.1155/2012/974638
38. Tien Bui D, Pradhan B, Lofman O, Revhaug I, Dick OB (2012b) Landslide susceptibility assessment in the Hoa Binh province of Vietnam: A comparison of the Levenberg–Marquardt and Bayesian regularized neural networks. Geomorphology 171-172:12-29. doi:10.1016/j.geomorph.2012.04.023
39. Tien Bui D et al. (2018) Landslide Detection and Susceptibility Mapping by AIRSAR Data Using Support Vector Machine and Index of Entropy Models in Cameron Highlands, Malaysia. Remote Sens 10:32. doi:10.3390/rs10101527
40. van Westen CJ, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. Eng Geol 102:112-131. doi:10.1016/j.enggeo.2008.03.010
41. Violante RA, Bozzano G, Rovere EI (2018) The Marine Environment: Hazards, Resources and the Application of Geoethics Principles. Ann Geophys 60:1-10. doi:10.4401/ag-7564
42. Wu D, Huang M, Zhang Y, Bhatti UA, Chen Q (2018) Strategy for assessment of disaster risk using typhoon hazards modeling based on chlorophyll-a content of seawater. EURASIP J Wirel Commun Netw 2018:12. doi:10.1186/s13638-018-1293-0
43. Xu C, Xu X, Lee YH, Tan X, Yu G, Dai F (2012) The 2010 Yushu earthquake triggered landslide hazard mapping using GIS and weight of evidence modeling. Environ Earth Sci 66:1603-1616. doi:10.1007/s12665-012-1624-0
44. Yanar T, Kocaman S, Gokceoglu C (2020) Use of Mamdani Fuzzy Algorithm for Multi-Hazard Susceptibility Assessment in a Developing Urban Settlement (Mamak, Ankara, Turkey). ISPRS Int Geo-Inf 9:114-139. doi:10.3390/ijgi9020114
45. Yang J, Song C, Yang Y, Xu C, Guo F, Xie L (2019) New method for landslide susceptibility mapping supported by spatial logistic regression and GeoDetector: A case study of Duwen Highway Basin, Sichuan Province, China. Geomorphology 324:62-71. doi:10.1016/j.geomorph.2018.09.019
46. Yin J, Yin Z, Xu S (2013) Composite risk assessment of typhoon-induced disaster for China's coastal area. Nat Hazards 69:1423-1434. doi:10.1007/s11069-013-0755-2
47. Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2015) 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 13:839-856. doi:10.1007/s10346-015-0614-1

48. Zezere JL, Pereira S, Melo R, Oliveira SC, Garcia RAC (2017) Mapping landslide susceptibility using data-driven methods. Sci Total Environ 589:250-267. doi:10.1016/j.scitotenv.2017.02.188
49. Zhang F, Huang X (2018) Trend and spatiotemporal distribution of fatal landslides triggered by non-seismic effects in China. Landslides 15:1663-1674. doi:10.1007/s10346-018-1007-z

Figures

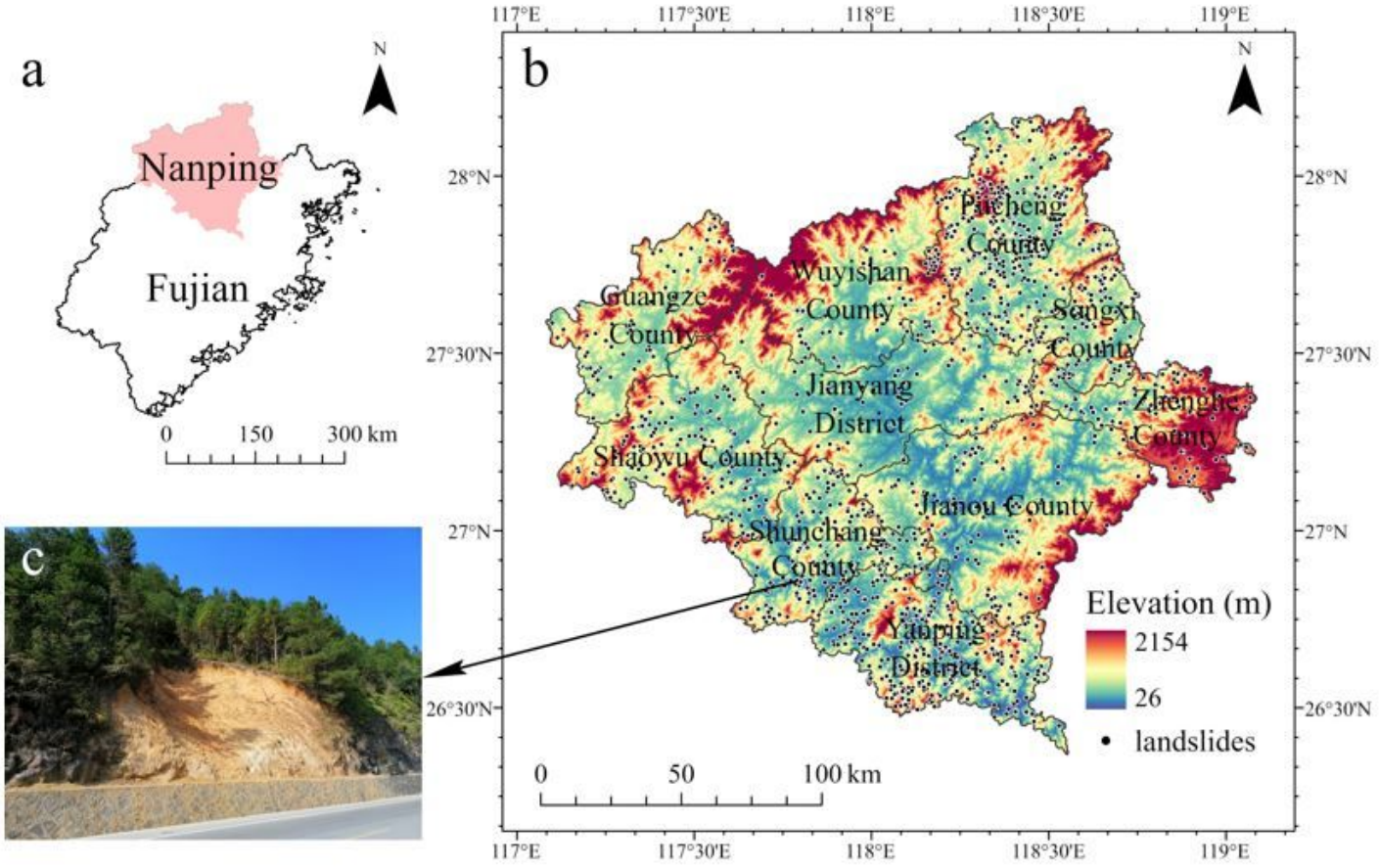

Figure 1

Location of the study area and landslide inventory. (a) Location of the study area, (b) landslide inventory of the study area, (c) a landslide case occurred in August 2018, Shunchang County. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

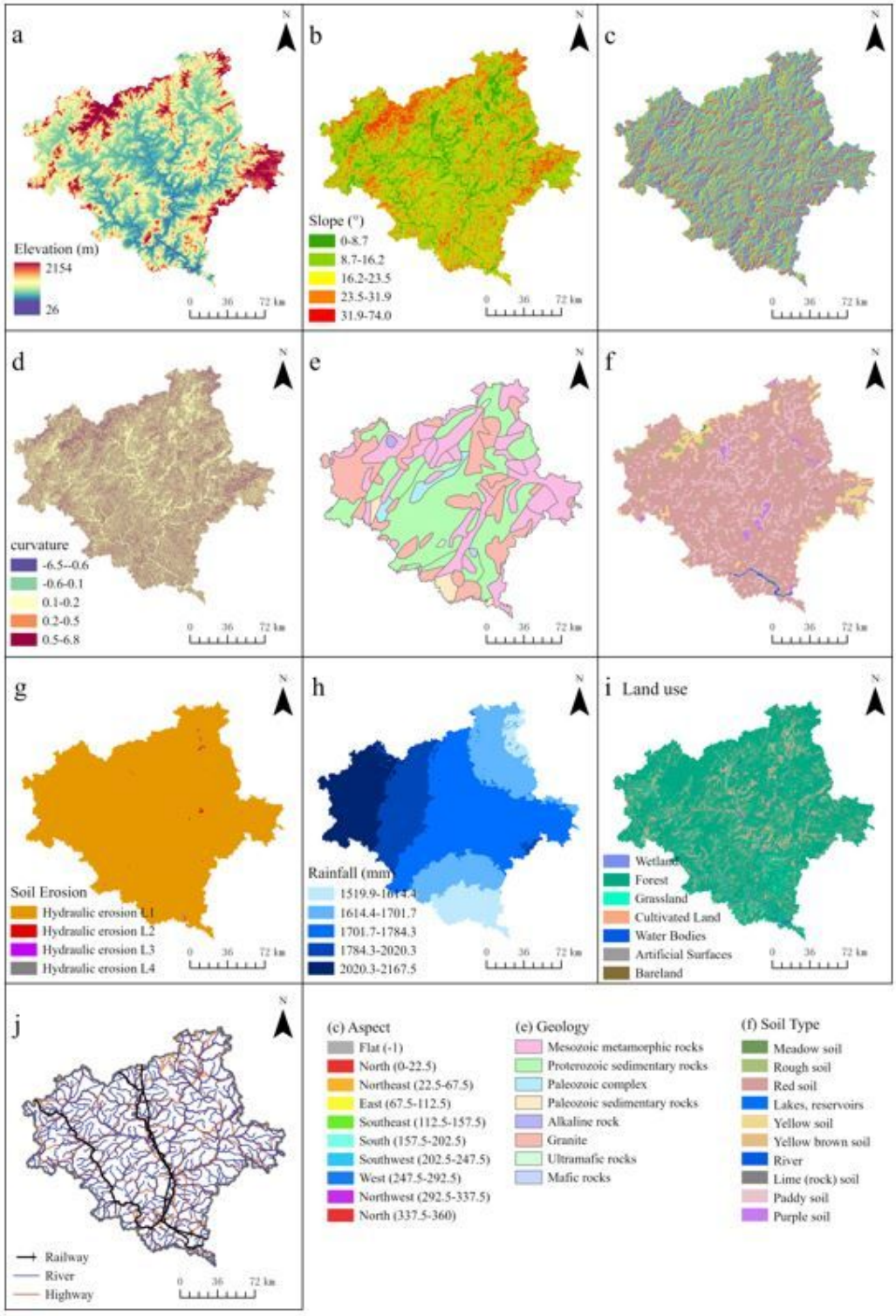

Figure 2

Landslide conditional factors: (a) elevation; (b) slope; (c) aspect; (d) curvature; (e) lithology; (f) soil type; (g) soil erosion; (h) rainfall; (i) land use; (j) railway, river, highway. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its

authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

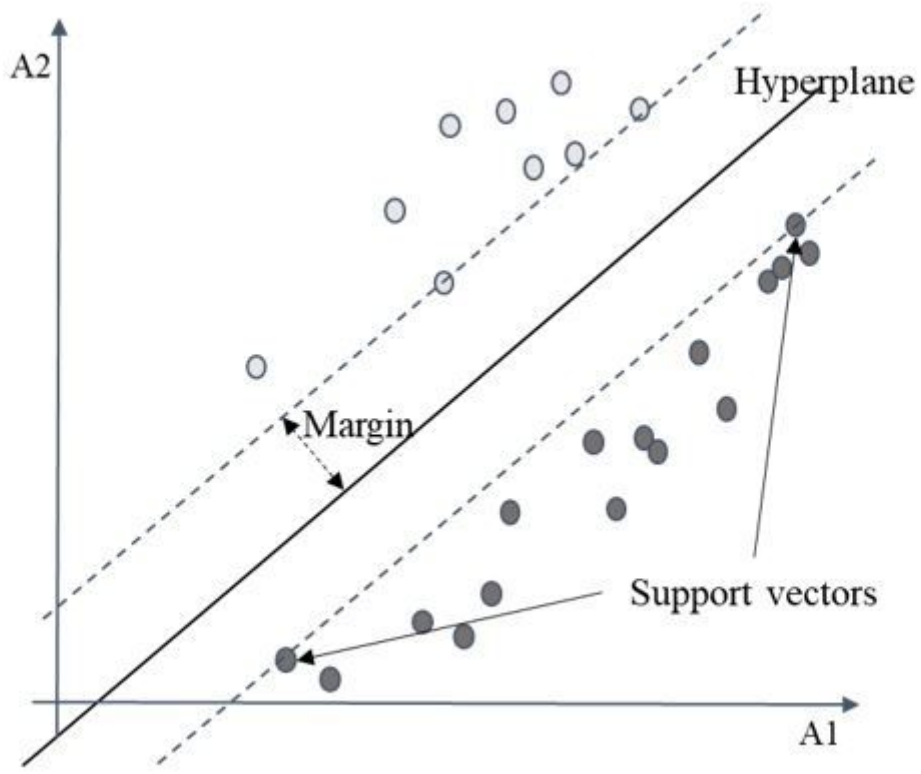

Figure 3

Principle of SVM.

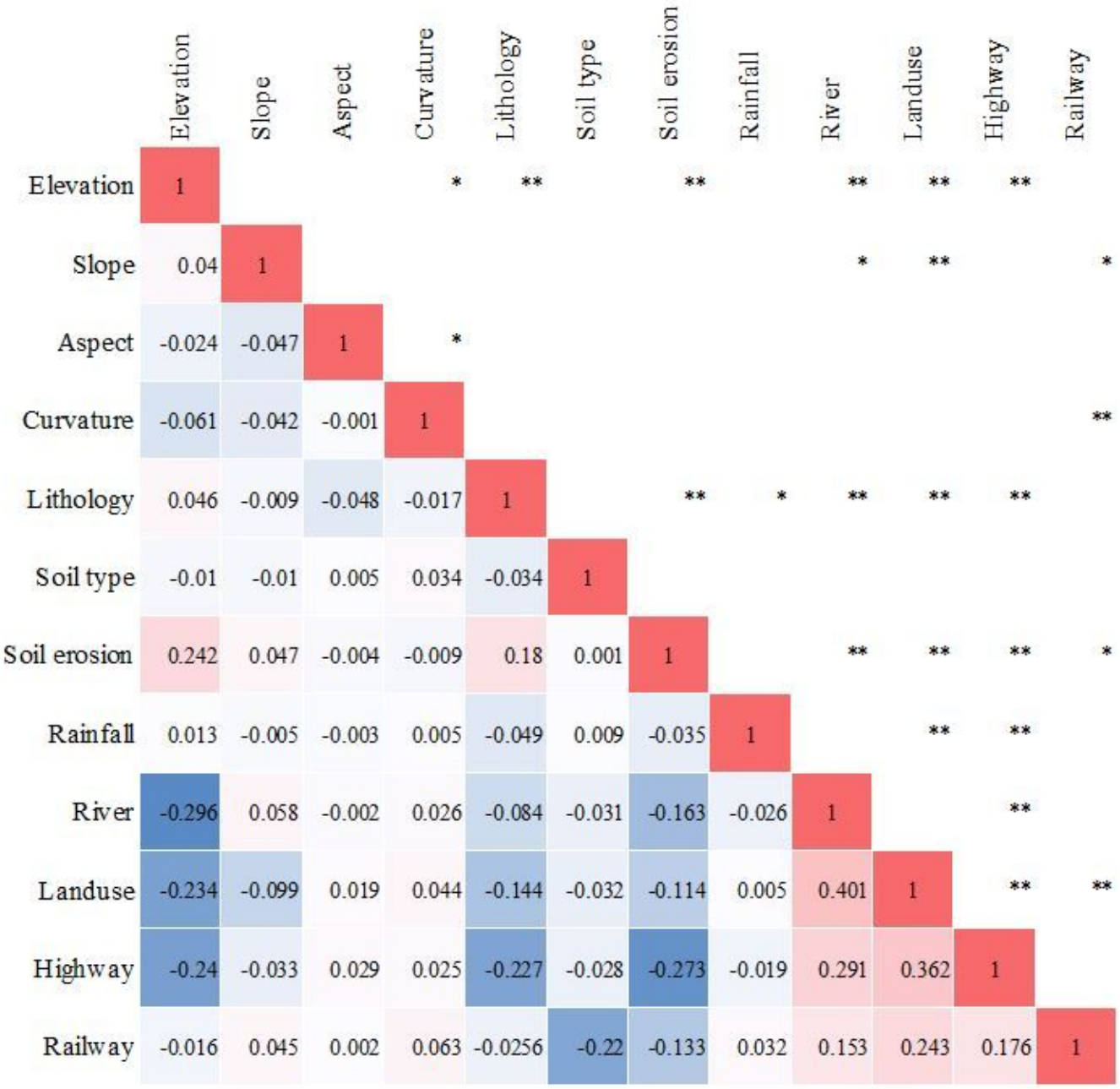

Figure 4

Correlation coefficient of conditional factors. Notes：* means at the significant level α=5%, the correlation is statistically significant, **means at the significant level α=1%, the correlation is statistically significant.

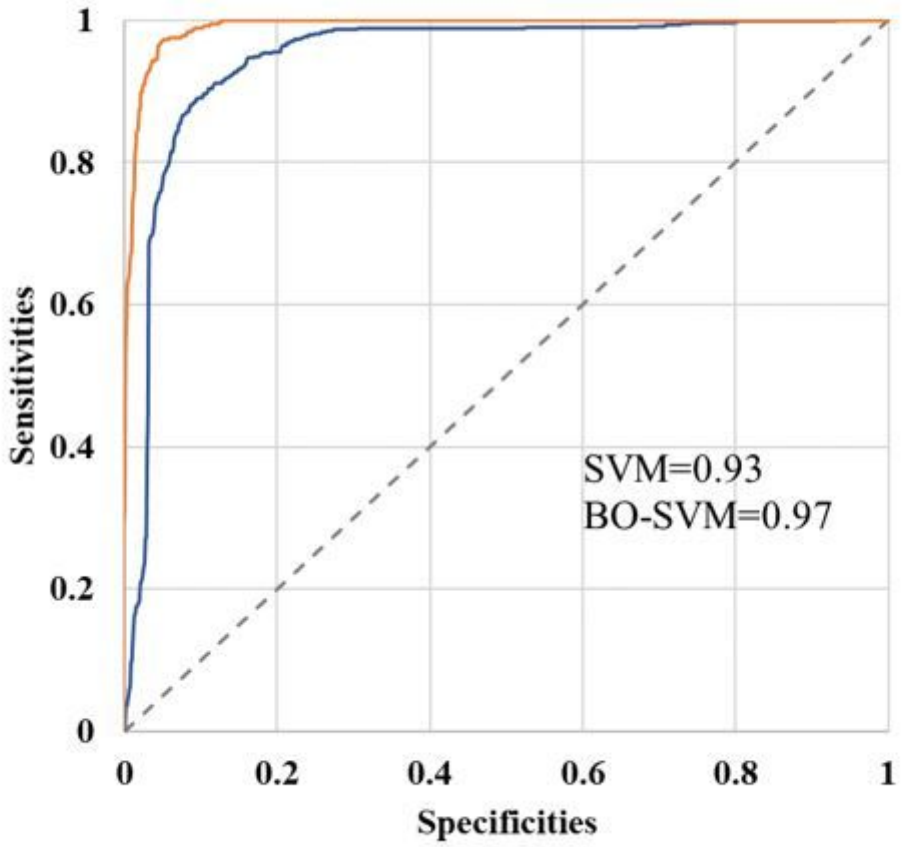

Figure 5

ROC curves of SVM and BO-SVM, AUC is the acronym of area under the ROC curve.

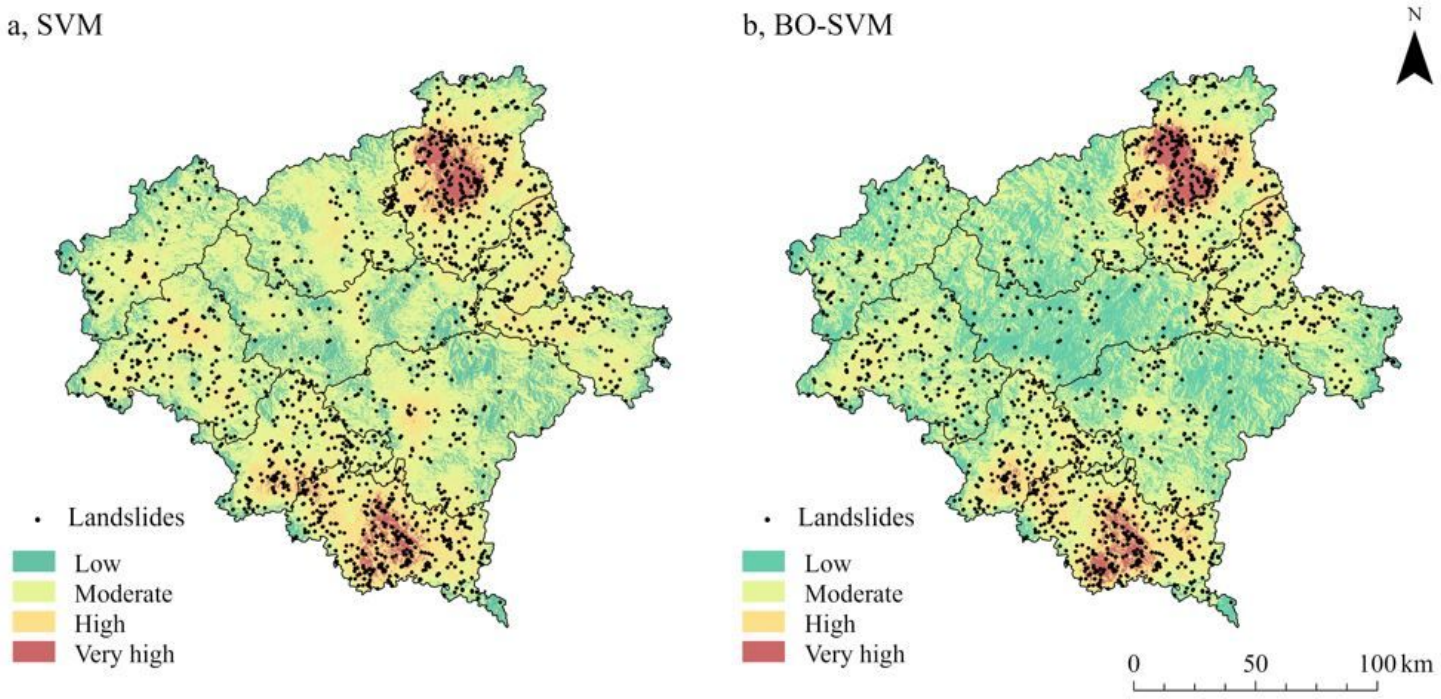

Figure 6

The landslide hazard map of the two models. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.