remote sensing

MDPI

Article

Study on the Uncertainty of Machine Learning Model for Earthquake-Induced Landslide Susceptibility Assessment

Haixia Feng [1], Zelang Miao [2] and Qingwu Hu [1,*]

[1] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China; fenghx@whu.edu.cn
[2] School of Geosciences and Info-Physics, Central South University, Changsha 410017, China; zelang.miao@csu.edu.cn
* Correspondence: huqw@whu.edu.cn

Abstract: The landslide susceptibility assessment based on machine learning can accurately predict the probability of landslides happening in the region. However, there are uncertainties in machine learning applications. In this paper, Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) are used to assess the landslide susceptibility in order to discuss the model uncertainty. The model uncertainty is explained in three ways: landslide susceptibility zoning result, risk area (high and extremely high) statistics, and the area under Receiver Operating Characteristic Curve (ROC). The findings indicate that: (1) Landslides are restricted by influence factors and have the distribution law of relatively concentrated and strip-shaped distribution in space. (2) The percentage of real landslide in risk area is 86%, 87%, 82%, and 61% in SVM, RF, LR, and ANN, respectively. The area under ROC of RF, SVM, LR, and ANN, respectively, is 90.92%, 80.45%, 73.75%, and 71.95%. (3) Compared with the prediction accuracy of the training set and test set from the same earthquake, the accuracy of landslide prediction in the different earthquakes is reduced.

Keywords: earthquake-induced landslide; susceptibility assessment; machine learning; model uncertainty

check for updates

Citation: Feng, H.; Miao, Z.; Hu, Q. Study on the Uncertainty of Machine Learning Model for Earthquake-Induced Landslide Susceptibility Assessment. *Remote Sens.* **2022**, *14*, 2968. https://doi.org/10.3390/rs14132968

Academic Editors: Paraskevas Tsangaratos, Ioanna Ilia, Wei Chen and Haoyuan Hong

Received: 5 May 2022
Accepted: 19 June 2022
Published: 21 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Strong earthquakes can cause landslides, debris flows, and other geological secondary disasters [1]. For example, in China, the Ms8.0 earthquake that struck Wenchuan on 12 May 2008, caused over 200,000 landslides, including around 60,000 landslides with single landslide areas of more than 35,000 m^2. According to data, the death toll from landslides accounted for almost one-third of the overall casualties from the earthquake. Landslides were caused by the Mw6.7 earthquake in Los Angeles in 1994 and the Ms6.8 earthquake in Mid-Niigata, Japan in 2004 [2].

Many researchers have done substantial studies on the susceptibility assessment of landslides during the last several decades. With the advancement of current computer technology, landslide susceptibility assessment is increasingly moving toward a fusion with computer technology. Some machine-learning-based algorithms, such as Artificial Neural (ANN) [3], and Support Vector Machine (SVM) [4], are becoming increasingly popular [5].

Uncertainty is strongly connected to probability, which creates the framework of machine learning (directly or indirectly) [6]. There are two forms of uncertainty in landslide susceptibility assessment using machine learning.

The first is the uncertainty of landslide influence factors, including factor selection, removal, and categorization. Based on the uncertainty of the influence factors in machine learning, scholars combined different methods to carry out a detailed sensitivity analysis of influence factors. It includes the analysis of landslide factors based on the method of determining coefficient [7], the stepwise regression [8], the Bayesian information standard [9], and the cluster analysis [10].

The second is the uncertainty of models, including model selection and input [11]. There are three types of uncertainty in models. The first type is to assess model performance. Ali Sk Ajim et al. discuss the performance of the analytic network process, the Nave Bayes classifier, and the RF classifier [12]. The second type is a multi-model assessment method, which uses the model's advantages to increase the model's accuracy and dependability. Binh Thai Pham et al. find a new classifier based on Composite Hyper-cubes on Iterated Random Projections (CHIRP) for assessing landslide vulnerability in the Uttarakhand Area [13]. The third type is to balance the data in order to improve the model's input data [14].

Landslide susceptibility assessment based on machine learning is a common tool and an effective and economical method to reduce the impact of landslides, which provides technical support for land management and disaster prevention. However, there are uncertainties in machine learning applications. Aiming at the uncertainty of models, this paper takes Lushan earthquake landslide as the research object and explores the uncertainty of machine learning based on the assessment of model performance. The model uncertainty is explained in three ways: landslide susceptibility zoning findings, risk area (high and extremely high) statistics results, and the area under Receiver Operating Characteristic Curve (ROC). Wenchuan earthquake landslides are selected to verify the validation of the proposed models. It is hoped that this study will provide ideas for the practical application of earthquake landslide prediction and help decision-makers make scientific planning and emergency decision more effectively.

2. Materials and Methods

2.1. Study Area

This paper selects the study area with reference to previous achievements [15] of complete landslide inventory, which is located in Sichuan Province's southwest, with coordinates ranging from 102°22′ to 103°19′E and 29°50′ to 30°40′N. The entire area is approximately 5396 km^2. The study area (Figure 1) encompasses three counties of Ya'an city (Lushan, Baoxing, and Tianquan) along the southwest boundary of the Sichuan Basin and its surrounding land. The study area encompasses nearly the whole area badly impacted by the Lushan earthquake [16].

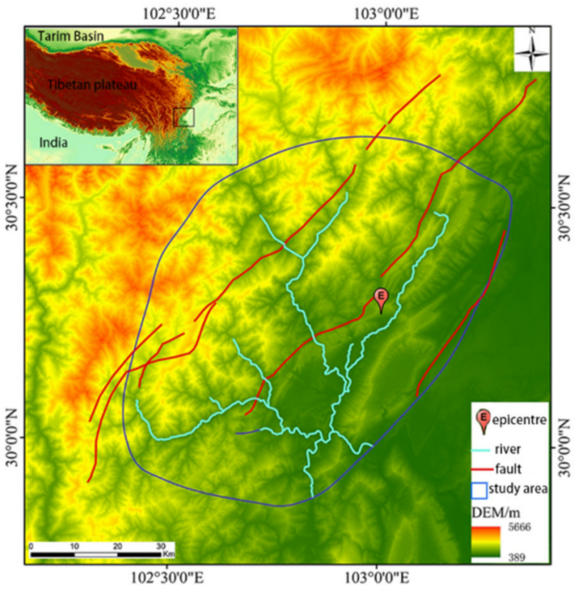


Figure 1. The geographical location of the study area.

Ya'an city is the transitional zone between the Sichuan Basin and the Qinghai-Tibet Plateau, located close to the western Sichuan Plain in the east and the Qinghai-Tibet Plateau in the west. Lushan county is located northeast of the Ya'an area and in the upper reaches

of the Qingyi River [17]. The highly high mountain zone has snow all year in the southwest and northwest marginal areas, and the valley in the region is deep, displaying classic alpine canyon landforms [1].

2.2. Model Uncertainty Assessment Process

The model uncertainty is assessed based on the model training outcomes (Figure 2). The procedure is as follows:

1. Slope unit and effect variables are extracted using a digital elevation model (data resolution of 30 m × 30 m) (slope, aspect, elevation, etc.).
2. Thematic maps are used to generate the categorization results of influence factors.
3. The influence factors that have been chosen are linked to the slope units and exported to Excel. The modeling samples are created in conjunction with the earthquake-induced landslide inventory.
4. Train models and compare the results. Analyze the uncertainty of machine learning from a variety of perspectives.
5. Wenchuan earthquake landslides are selected to verify the validation of the proposed models.

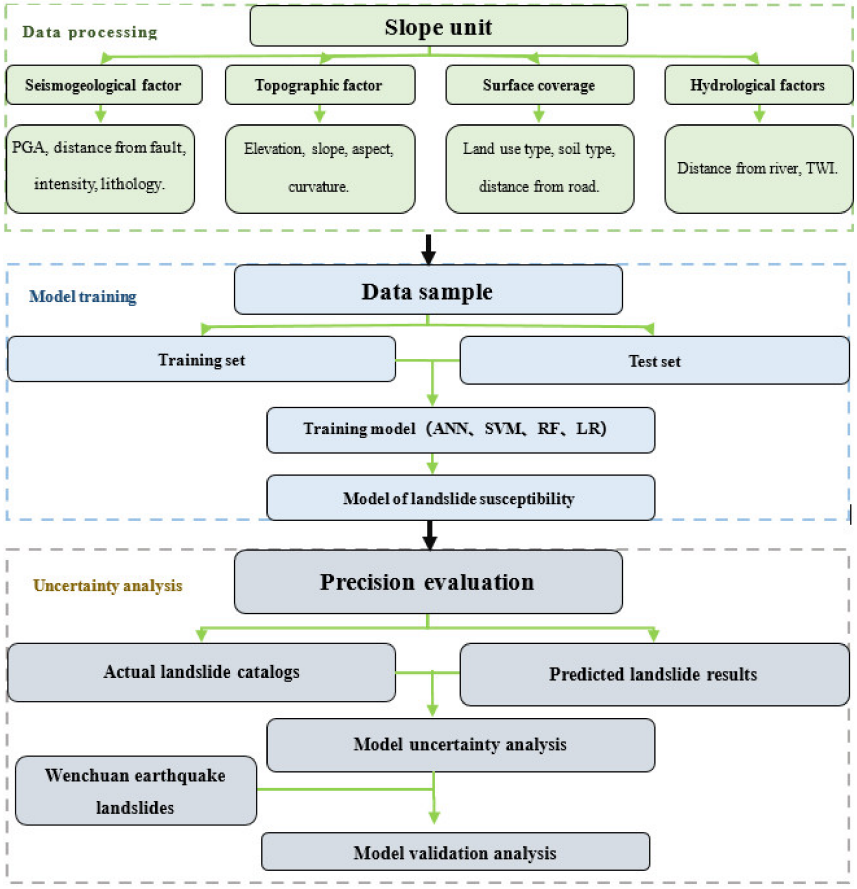


Figure 2. The model uncertainty assessment process.

2.3. Data Processing

Data processing includes the generation of slope units, the classification of influence factors, and the preparation of a training set and a test set.

2.3.1. Slope Units

This work uses GIS as the operational platform to provide slope units with high consistency as the assessment units of Lushan earthquake landslide susceptibility. Slope units are generated by using the hydrological analysis tool to extract valley lines, which

includes key stages like filling depressions, extracting flow direction and accumulation, building river networks, generating catchment basins, and so on. The catchment area is obtained by DEM when combined with hydrology, and its boundary lines are the ridge lines. Then reverse the DEM to create a new catchment area and its boundary lines are the valley lines. Finally, the slope units are obtained by fusing the ridge lines and valley lines and eliminating the patches less than 1 km^2 [18]. 1404 slope units are used to divide the study area (Figure 3).

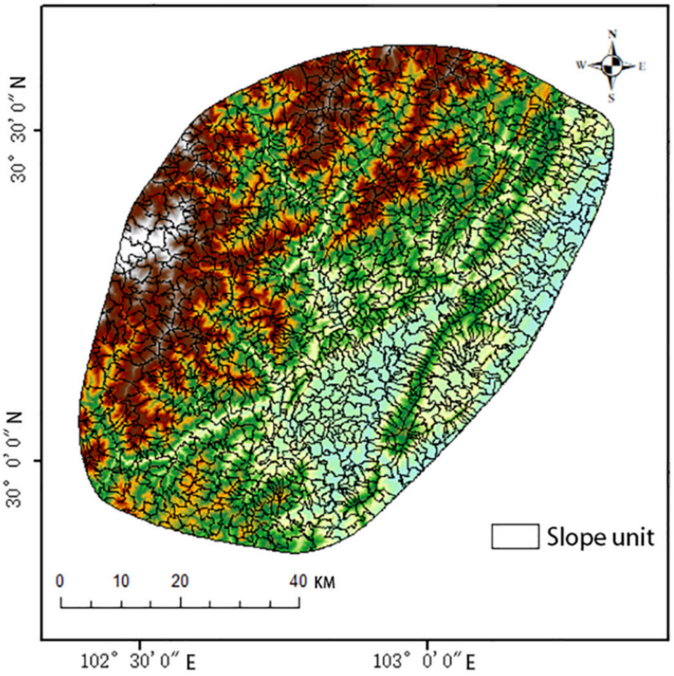


Figure 3. Result of slope unit division in the study area.

2.3.2. Influence Factors

Choosing the appropriate influence factors as an indicator is a daunting challenge. There is no standard to follow in the selection and classification of influence factors and the expert investigation method is often used with reference to previous studies [19]. This article chooses a number of factors to investigate the law of the distribution of landslides, including peak ground acceleration, distance from fault, earthquake intensity, lithology, elevation, slope, aspect, curvature, land use type, soil type, distance from the road, distance from the river, and topographic wetness index. The classification result of influence factors is as follows (Table 1).

2.3.3. Data Sample

The landslide inventory in this paper comes from the research of Xu Chong et al. Xu Chong and others prepared an original, emergency-based co-seismic landslide inventory map based on visual interpretation of available post-earthquake aerial photographs and pre-earthquake satellite images, combined with results from emergency-based field investigations. The original inventory map registered 3884 landslides. Later, 11,761 more landslides were identified by much more careful and time-consuming visual interpretation based on the same aerial photographs and satellite images [15]. This study's sample data set includes 1404 slope units, 683 landslide bearing units, and 721 landslide free units (Figure 4). Each slope unit contains landslide value and influence factor attributes. The landslide value of the landslide bearing unit is 1 and the landslide free unit is 0. Each slope unit has influence factor values associated with it. To establish the link between the influence factor attribute and the landslide attribute, 70% of the sample data is chosen at random as the training sample set [20].

Table 1. The classification of influence factors.

Influence Factor	Number	Standard
peak ground acceleration [g]	5	0.06–0.08; 0.08–0.10; 0.10–0.12; 0.12–0.16; 0.16–0.48
distance from fault [km]	7	0–1.5; 1.5–3; 3–4.5; 4.5–6; 6–7.5; 7.5–9; >9
earthquake intensity	4	ix; viii; vii; vi
lithology	13	before sinian system; ordovician; triassic; permian; granite; jurassic; quaternary; cretaceous; devonian; paleogene; silurian system; tertiary; sinian system
Elevation [km]	9	0.5–1; 1–1.5; 1.5–2; 2–2.5; 2.5–3; 3–3.5; 3.5–4; 4–4.5; 4.5–5
slope	9	0–10; 10–20; 20–30; 30–40; 40–50; 50–60; 60–70; 70–80; >80
aspect	9	flat; north; northeast; east; southeast; south; southwest; west; northwest
curvature	7	0–8; 8–16; 16–24; 24–32; 32–40; 40–48; >48
soil type	5	luvisols; entisols; tierras; ferralsol; rock
distance from the road [km]	7	0–1.5; 1.5–3; 3–4.5; 4.5–6; 6–7.5; 7.5–9; >9
distance from the river [km]	7	0–1.5; 1.5–3; 3–4.5; 4.5–6; 6–7.5; 7.5–9; >9
topographic wetness index	6	0.14–3; 3–6; 6–9; 9–12; 12–15; 15–19
land use type	5	cultivated land; forest land; grass; waters; urban and rural industrial and mining residents

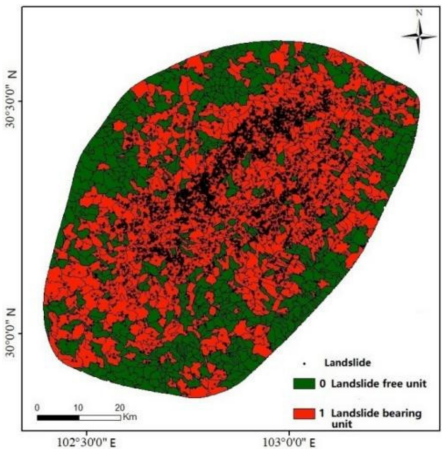


Figure 4. The distribution of landslide free unit and landslide bearing unit.

2.4. Assessment Model

With the advancement of current computer technology, landslide susceptibility assessment is increasingly merging with computer technology. Some machine learning algorithms are becoming increasingly popular. Therefore, four frequently used machine learning algorithms, including ANN, RF, SVM, and LR, are selected in this study to examine the uncertainty of these four models on the landslide problem.

2.4.1. Artificial Neural Network Model

The Artificial Neural Network (ANN) model is a type of training and teaching model. A multi-layer neural network is used in the learning method. Make the neural network attain the given value and receive genuine results by training it [21]. If the input layer is $X = (x_1, x_2, \ldots, x_i, \ldots, x_I)$ and hidden layer $Y = (y_1, y_2, \ldots, y_j, \ldots, y_J)$, then y_j satisfies:

$$y_j = f_1\left(\sum_{i=1}^{I} \omega_{ij} x_i - a_j\right), j = (1, 2, \ldots, j) \quad (1)$$

In the formula, ω_{ij} is the connection weight between the two-layer network, a_{ij} is the threshold between the two-layer networks, and f_1 is the activation function.

2.4.2. Random Forest Model

The Random Forest (RF) model is a predictive learning technique that uses the bagging approach to build multiple categorization regression trees and multiple independent training sets. The primary idea behind the RF is to get results by combining the judgments of several classifiers [22]. The bagging technique selects M samples (about 2/3 of the total data) at random from the original training set.

The estimate is an unbiased estimate of the error derived through cross verification, with the following generalization error bounds:

$$P^* \leq \rho^- \left(1 - s^2\right) / s^2 \tag{2}$$

In the formula, P* is a generalization error. ρ^- is the average correlation between decision trees. s is the average strength of the decision tree. From the expression of generalization error bound.

2.4.3. Support Vector Machine Model

The Support Vector Machine (SVM) is a supervised learning approach for nonlinear covariable transformation based on statistical learning theory. Nonlinear and high-dimensional pattern identification issues are challenging to address using typical statistical approaches because geological risks are very nonlinear [23]. However, using the structural risk minimization principle and the VC dimension theory of statistical theory, SVM may handle these difficulties with a minimal number of samples [24].

The Lagrange multiplier rule is introduced to obtain the extreme value, and the auxiliary function is generated as follows:

$$L = \frac{1}{2} \|\omega\|^2 - \frac{1}{vn} \sum_{i=1}^{n} \xi_i \tag{3}$$

In the formula, the minimization setting of ω is solved by standard procedures and the relaxation factor ξ_i is introduced to adjust the constraints. The $v \in (0, 1]$ is the penalty value of the introduced misclassification.

2.4.4. Logistic Regression Model

The Logistic Regression model (LR) is a regression analysis model that uses binary classification variables on corresponding variables to perform regression analysis [25]. Independent variables are not required to have a normal distribution and can be used to predict the occurrences of binary classification variables. Furthermore, this approach may address the issue of fluctuating dependency [26].

In this model, the relationship between the probability of landslide occurrence and independent variables can be expressed as:

$$Z = B_0 + B_1X_1 + B_2X_2 + \ldots + B_nX_n \tag{4}$$

In order to predict the probability of each grid landslide, the probability calculation formula is:

$$P = 1/\left(1 + e^{-z}\right) \tag{5}$$

In the formula, P is the probability of landslide occurrence. Z is the weighted linear combination of independent variables. $B_i (i = 1, 2, \ldots, n)$ is the weighting coefficient obtained by using sample data, n is the number of independent variables, and $X_i (i = 1, 2, \ldots, n)$ is an independent variable. The maximum likelihood model is used to obtain the relationship between prediction variables and landslide/non-slip samples.

3. Results

This research uses qualitative analysis of the landslide susceptibility zoning results and quantitative analysis of the statistical results of risk area to evaluate and compare the model results of landslides prediction. Finally, the area under ROC is used to assess the model's degree of fit.

3.1. Assessment Model

The entire study area's data set is put into the models. The closer the predicted number is to 0, the less likely the region will experience a landslide disaster, and the closer it is to 1, the more likely the region will experience a landslide disaster. Different models are used to create landslide susceptibility assessment maps over the whole study area. The rivers and faults are placed on the landslide susceptibility zoning findings (Figure 5).

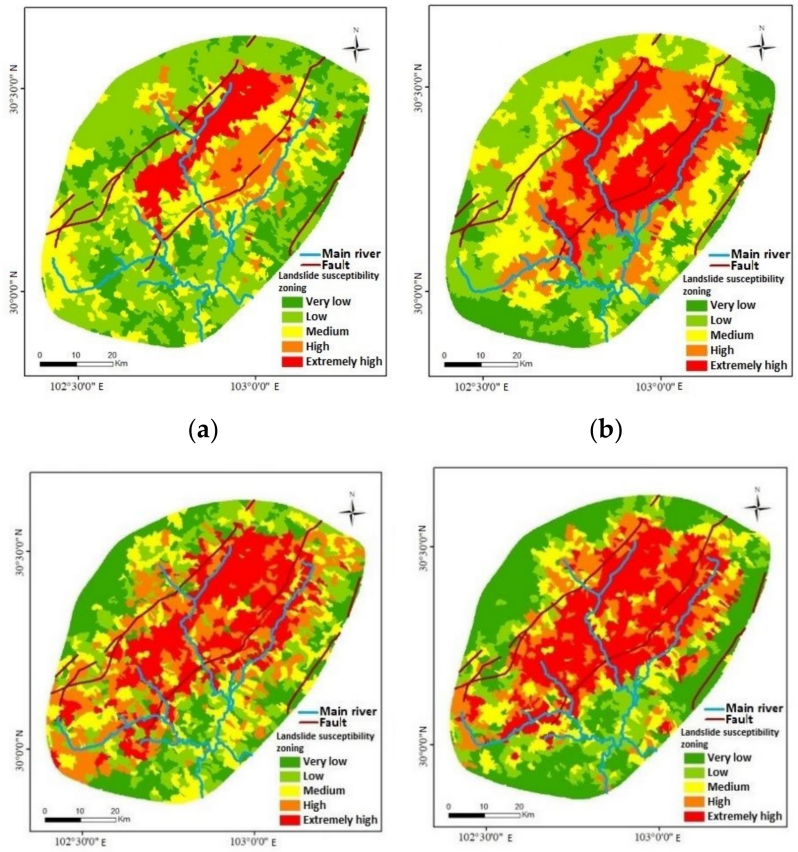


Figure 5. The relationship between the distribution of landslide susceptibility zones, rivers, and faults: (**a**) ANN; (**b**) LR; (**c**) RF; and (**d**) SVM.

The following conclusions can be drawn by qualitatively evaluating the four models from Figure 4:

1. The study area is separated into five portions representing landslide susceptibility ranging, namely, very low, low, medium, high, and extremely high, according to the natural breakpoint technique [27].
2. Landslides are restricted by influence factors and have the distribution law of relatively concentrated and strip-shaped distribution in space. The zones with high landslide susceptibility show a trend distribution from northeast to southwest in the training results of the four models, which is consistent with the trend of faults and rivers. This is in line with Zhao DL's findings on the landslide in Qinghai Province's Hehuang

valley [28]. It can also be seen that the landslide mainly occurs in the area between the two faults.

3. In terms of landslide extraction integrity, the zone with a high probability of landslide occurrence (high and extremely high) is much larger than that with a low probability of landslide occurrence (medium, low, and very low) in the findings of SVM and RF. ANN and LR landslide extractions are concentrated in the study area's center and are unaffected by landslide extraction in the margins.

3.2. Statistical Results of Risk Area

The percentage of the partition area and the percentage of real landslide in the partition are calculated using the zoning statistical tool in ArcGIS10.5 in order to quantitatively compare the susceptibility assessment findings of the four models [29] (Table 2).

Table 2. The statistical results of superimposing the susceptibility partition with the real landslide catalog.

Landslide Susceptibility Partition	Partition Area [km^2]	Percentage of the Partition Area	Area of Real Landslide in the Partition [km^2]	Percentage of Real Landslide in the Partition [%]
		ANN		
very low	293.47	0.19	0.55	0.04
low	607.95	0.40	1.75	0.12
medium	330.06	0.22	3.26	0.23
high	175.70	0.11	2.90	0.21
extremely high	126.10	0.08	5.60	0.40
		LR		
very low	178.44	0.12	0.23	0.02
low	377.20	0.25	0.81	0.06
medium	342.04	0.22	1.47	0.10
high	317.05	0.21	4.14	0.29
extremely high	318.55	0.21	7.42	0.53
		RF		
very low	310.61	0.20	0.21	0.01
low	318.39	0.21	0.45	0.03
medium	259.32	0.17	1.20	0.09
high	313.24	0.20	2.99	0.21
extremely high	331.70	0.22	9.23	0.66
		SVM		
very low	415.99	0.27	0.49	0.04
low	207.65	0.14	0.49	0.04
medium	212.09	0.14	1.05	0.07
high	268.94	0.18	1.78	0.13
extremely high	428.60	0.28	10.26	0.73

In order to more intuitively compare the statistical results of risk area (high and extremely high) in the four models, the "Percentage of the Partition Area" data from the third column of the row in the "high" row and "extremely high" row are selected. Multiply the two values to obtain the percentage of risk area in the total area and draw the bar chart. The "Percentage of Real Landslide in the Partition" data from the fifth column of the row in the "high" row and "extremely high" row are selected. Multiply the two values to obtain the percentage of real landslide in the risk area and draw the line chart. The fraction of the percentage of risk area and the percentage of real landslide in the risk area are depicted in Figure 6.

The statistical results suggest that:

1. The risk area in the ANN results is 302 km^2, or 19% of the whole study area. The real landslide area is 8 km^2, which accounts for 61% of the overall landslide area. Due to the limited risk area division in this model, the real landslide area in the risk

area is only 61%, showing that the prediction model has a problem with landslide leakage extraction.

2. The risk area is 636 km^2, representing 42 percent of the whole study area, according to the LR data. The real landslide area is 12 km^2, accounting for 82% of the entire landslide area. The risk area is 645 km^2 in the RF findings, amounting to about 42% of the overall study area. The real landslide area is 12 km^2, accounting for 87% of the entire landslide area. The majority of the landslides predicted by these two models are focused on the risk area, yielding good outcomes. In RF, the real landslide area accounted for a high proportion of the overall landslide area. As a result, the RF forecast findings are accurate.
3. The risk area in the SVM findings is 697 km^2, accounting for 46% of the whole study area. The real landslide area is 12 km^2, accounting for 86% of the overall landslide area. The proportion of risk area in this model is considerable, suggesting that the prediction model is simple to split the region into risk regions.
4. The results of the above models are higher than the Qiu WR's results in the study of landslide susceptibility in Lingtai county [30]. In Qiu's results, the percentage of the real landslide in the risk area ranges from 50% to 70%, while the results of this study range from 60% to 90%. The possible reason for this is the slope units and the landslides area's center are large in the study, leading to good extraction results for landslides that have a large development area and bigger percentages of the real landslide in the risk area.

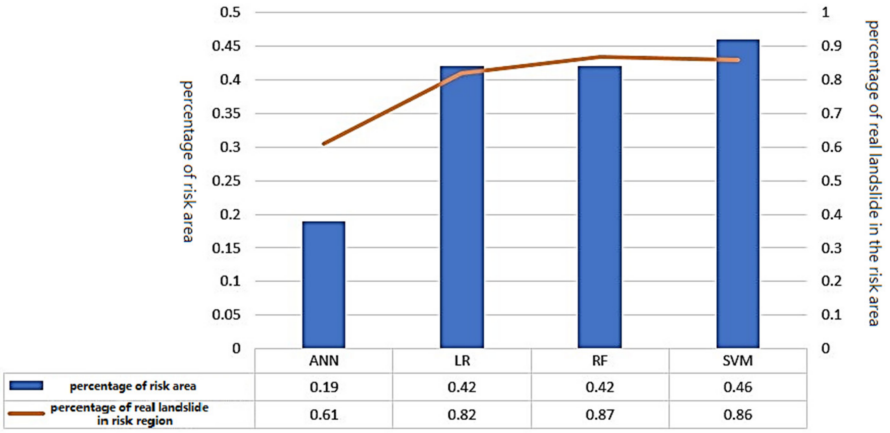


Figure 6. The statistical results of risk area.

3.3. ROC

The Receiver Operating Characteristic Curve (ROC) is a type of integrated indicator that measures the sensitivity and specificity of continuous variables [31]. The following are the assessment criteria: if AUC (Area Under Curve) is less than 0.5, the prediction results are inverse. If the AUC is less than 0.5, the model is random. If the AUC value is between 0.5 and 0.7, the model's accuracy is poor. The model has great accuracy if the AUC is between 0.7 and 0.9. If the AUC is more than 0.9, the model is extremely accurate. The greater the area under the ROC (the steeper the curve), the higher the model accuracy [32]. The ROC (Figure 7) shows that the RF model has the greatest AUC value in the training set, at 90%; the SVM model has an AUC value of 80%; and the LR and ANN models have AUC values of 74% and 72%, respectively.

The accuracy of RF is the highest, which is consistent with the research results of Qiu WR's analysis in Lingtai county and Alusi's analysis of landslide susceptibility in Tonghua County [30,33]. This shows that the RF model has strong applicability in the study of landslide susceptibility. Because it avoids the problem of decision tree overfitting, the RF model has maximum accuracy. It tolerates noise and outliers well, efficiently eliminates overfitting, and is scalable and parallelizable for high-dimensional data classification

applications. In comparison to the other algorithm's capacity to adapt to new samples, the trained network can also provide acceptable output and has great generalization ability.

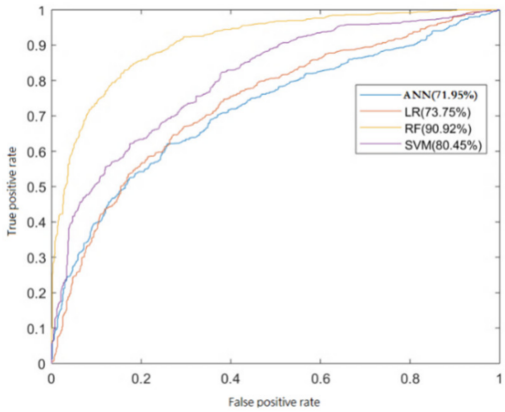
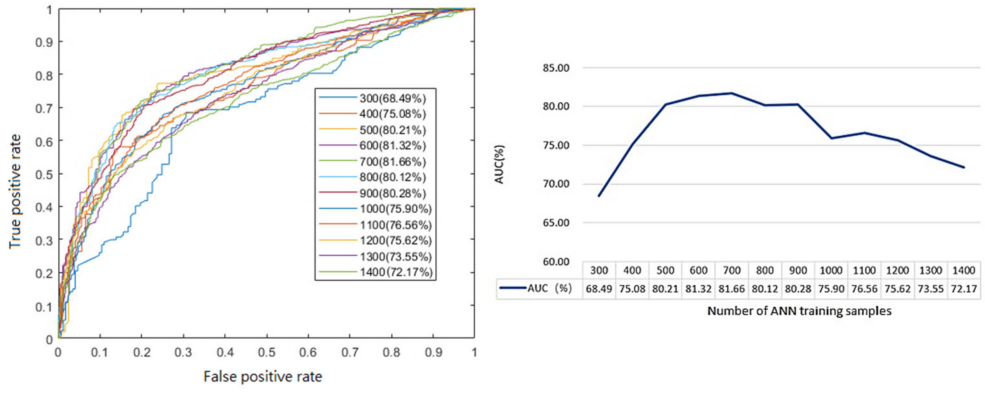


Figure 7. ROC and AUC: ANN, LR, RF, SVM.

The accuracy of LR is consistent with that of qiuhj, however, the accuracy of ANN is lower. The ANN's prediction accuracy is low for the following reasons: there is an inconsistency between forecast and training ability. Predictive ability is also known as generalization ability, whereas training ability is known as approximation ability. This issue occurs when the network has learned too many sample features, and the learned model can no longer reflect the rules contained in the sample. ANN has a sample dependence problem. The network model's prediction and training performance are highly connected to the typicality of the learning samples, and selecting the typical sample variables from the issues to build the training set is a somewhat tough challenge.

As demonstrated in Figure 8, as the number of samples grows, the ANN model's prediction accuracy climbs initially, then falls, and approaches near the limit value when the number of samples exceeds 700. When the number of input values is 300, the model is under fitted and the accuracy is the lowest. This study must adjust for other uncertain elements in order to investigate the model's uncertainty. As a result, the sample size is 1404, resulting in the ANN model's relatively poor accuracy in this study.



(**a**) (**b**)

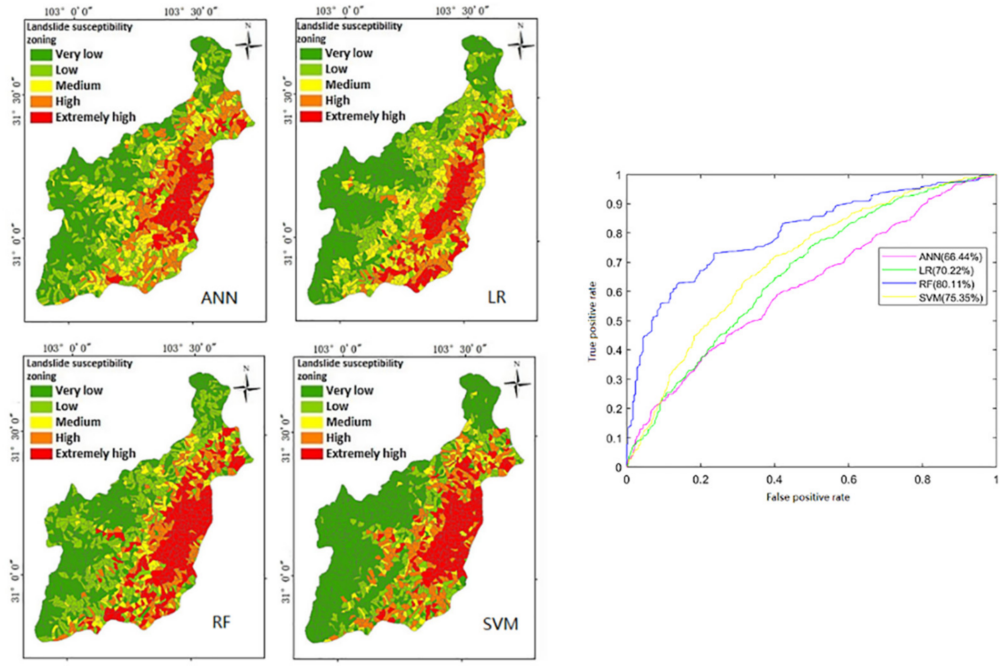
Figure 8. The value of the ANN model using different samples: (**a**) the ROC and AUC of ANN model using different samples; (**b**) the line chart of AUC.

3.4. Validation of the Proposed Models

Wenchuan earthquake landslides are selected to verify the validation of the proposed models. The Wenchuan earthquake and the Lushan earthquake both caused a large number of landslides. The distance between the epicenter of the two earthquakes is only about 80 km away, and both of them are located in the Longmenshan fault [34]. Therefore, the

Wenchuan earthquake and the Lushan earthquake are treated as similar topographic areas in this paper.

Based on the landslide model of the Lushan earthquake, the susceptibility of landslides induced by the Wenchuan earthquake is assessed (Figure 9a). The ROC is used to evaluate the validation of the models (Figure 9b). Then, the validation of proposed models is discussed, which is helpful to understand the uncertainty of machine learning model.



(**a**) (**b**)

Figure 9. Wenchuan earthquake-induced landslide susceptibility assessment and ROC: (**a**) landslide susceptibility assessment (ANN, LR, RF, and SVM); (**b**) ROC.

The results suggest that: The ROC (Figure 9b) shows that the area under ROC of RF, SVM, LR, and ANN, respectively, is 80.11%, 75.35%, 70.22%, and 66.44%. Compared with the prediction accuracy of the training set and test set from the same earthquake, the accuracy of landslide prediction in Wenchuan earthquake is reduced by 5.51%, 3.53%, 10.81%, and 5.1% respectively. The accuracy of most models declines slightly for the prediction with similar regional characteristics. Compared with the prediction accuracy of the training set and test set from the same earthquake, the accuracy of landslide prediction in the different earthquakes is reduced. The probability of earthquake-induced landslide with different regional characteristics can be predicted but the accuracy is reduced.

4. Discussion

From the visual analysis in the Lushan earthquake-induced landslide susceptibility assessment, it can be seen that landslides are restricted by influence factors and have the distribution law of relatively concentrated and strip-shaped distribution in space. The zone with a high probability of landslide occurrence (high and extremely high) is much larger than that with a low probability of landslide occurrence (medium, low and very low) in the findings of SVM and RF. ANN and LR landslide extractions are concentrated in the study. From the qualitative analysis, it can be seen that the result of the percentage of the real landslide in the risk area is high. The possible reason is that the slope units and the landslides area's center are large in the study, leading to good extraction results for landslides that have a large development area and bigger percentages of the real landslide in the risk area. From the ROC analysis, it can be seen that RF has high precision which is fast and reliable for the susceptibility evaluation of landslide. However, compared with the

prediction accuracy of training set and test set from the same earthquake, the accuracy of landslide prediction in different earthquake is reduced. The magnitude of the earthquake, the number of landslides, and the difference in geographical conditions are the possible reasons for the reduction in accuracy. In addition, the difference in interaction between influence factors during different earthquakes may also affect the validation of the models. Adding indicators expressing spatial differences to model training may also increase the validation of the model, which needs to be further proved in future research. At present, the accuracy of most models declines slightly for the prediction with similar regional characteristics. The probability of earthquake-induced landslide with different regional characteristics cannot be accurately predicted. How to improve the generality of the model through a large number of historical earthquake landslide data sets and higher resolution index factors is also a problem to be considered in the future.

5. Conclusions

The uncertainty of machine learning is examined in this study by simulating the landslide susceptibility assessment in a specific location. Using the Lushan earthquake landslide as the research object, modeling of landslide susceptibility is done to examine the uncertainty of the machine learning model by employing thirteen influencing factors, such as peak ground acceleration, distance from the fault, lithology, and slope, among others, via ANN, RF, SVM, and LR. The training model is utilized to assess the study area's landslide susceptibility, and the model's uncertainty is analyzed from three perspectives: qualitative analysis of the landslide susceptibility zoning results, quantitative analysis of the statistical results of risk area (high and extremely high) and ROC, which is used to assess the model's degree of fit.

1. Landslide is restricted by influence factors and has the distribution law of relative concentration and strip distribution in space. The zones with high landslide susceptibility show a trend distribution from northeast to southwest in the training results of the four models. The zone with a high probability of landslide occurrence (high and extremely high) is much larger than that with a low probability of landslide occurrence (medium, low, and very low) in the zoning findings of SVM and RF. ANN and LR landslide extraction is concentrated in the study area's center and is unaffected by landslide extraction in the margins.
2. The proportion of risk area to total area is 46%, 42%, 42%, and 19%, respectively, while the percentage of landslide area in risk area is 86%, 87%, 82%, and 61% in SVM, RF, LR, and ANN, respectively. As a result, while evaluating the safety of landslide treatment, the model's assessment effect is SVM > RF > LR > ANN from high to low.
3. In terms of ROC, the area under the ROC curve for RF, SVM, LR, and ANN, respectively, is 90.92%, 80.45%, 73.75%, and 71.95%. From high to low in terms of prediction accuracy, the sequence is RF > SVM > LR > ANN. In this study, the RF model has good generalization capacity. However, the ANN model has an overfitting problem. Large slope units will lead to a higher percentage of real landslide in the risk area, which is the reason why the percentage of real landslide in the risk area of LR and SVM is higher than ROC.
4. RF has high precision which is fast and reliable for the susceptibility evaluation of landslides. However, compared with the prediction accuracy of the training set and test set from the same earthquake, the accuracy of landslide prediction in the different earthquakes is reduced. How to improve the generality of the model through a large number of historical earthquake landslide data sets and adding the index of spatial differences is also a problem to be considered in the future.

Author Contributions: Conceptualization, H.F. and Z.M.; methodology, H.F. and Z.M.; software, H.F. and Z.M.; validation, H.F. and Z.M.; formal analysis, H.F. and Z.M.; resources, H.F. and Z.M.; data curation, H.F. and Z.M.; writing—original draft preparation, H.F.; writing—Review and editing, H.F., Q.H. and Z.M.; visualization, H.F.; supervision, Q.H. and Z.M.; project administration, Q.H.; funding acquisition, Q.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (Grant No. 2020YFC1521900), and the Science and Technology Planning Project of Guangdong Province, China (Grant No. 2017B020218001).

Data Availability Statement: The data that support the findings of this study are available from the author upon reasonable request.

Acknowledgments: We would like to acknowledge the suggestions given by the reviewers and editor.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ma, S.Y. Study on Seismic Landslide Risk Assessment Based on Newmark Model: A Case Study of Lushan Earthquake and Wenchuan Earthquake. Master's Thesis, Institute of Geology, China Earthquake Administration, Beijing, China, 2018.
2. Lanxin, D.; Qiang, X.; Xuanmei, F.; Ming, C.; Qin, Y.; Fan, Y.; Jing, R. A preliminary study on spatial distribution and susceptibility assessment of geological hazards induced by the August 8, 2017 Jiuzhaigou earthquake. *J. Eng. Geol.* **2017**, *4*, 1151–1164.
3. Zhao, T. Study on landslide risk assessment method based on artificial neural network. *Urban Constr.* **2013**, *20*, 306–307.
4. Dai, F.C.; Yao, X.; Tan, G.H. Support vector machine model for Spatial Prediction of landslide Disaster and its Application. *Earth Sci. Front.* **2007**, *6*, 153–159.
5. Liu, J.; Li, S.L.; Chen, T. Assessment of landslide susceptibility based on optimization random forest model. *Inf. Sci. Technol. Wuhan Univ.* **2018**, *43*, 1085–1091.
6. Hammer, B.; Villmann, T. How to process uncertainty in machine learning. In Proceedings of the ESANN 2007, 15th European Symposium on Artificial Neural Networks, Bruges, Belgium, 25–27 April 2007.
7. Liu, L.N.; Xu, C.; Chen, J. Sensitivity analysis of 2013 Lushan earthquake landslide factors based on CF method supported by GIS. *J. Eng. Geol.* **2014**, *22*, 1176–1186.
8. Zhang, J.Y.; Ni, W.D.; Shi, Y. The application of stepwise regression analysis in the sensitivity analysis of landslide factors. *Sci. Technol. Econ. Mark.* **2008**, *6*, 3.
9. Li, X.P.; Tang, H.M. Application of Bayesian information standard in sensitivity Analysis of landslide factors. *Rock Soil Mech.* **2006**, *8*, 1393–1397.
10. Basu, T.; Pal, S. A GIS-based factor clustering and landslide susceptibility analysis using AHP for Gish River Basin, India. *Environ. Dev. Sustain.* **2020**, *22*, 4787–4819. [CrossRef]
11. Uncertainty in Position Estimation Using Machine Learning. Available online: http://arxiv.org/abs/2106.02370 (accessed on 5 May 2022).
12. Ali, S.A.; Parvin, F.; Vojteková, J.; Costache, R.; Linh, N.T.T.; Pham, Q.B.; Vojtek, M.G.L.; Ahmad, A.; Ghorbani, M.A. GIS-Based Landslide Susceptibility Modeling: A Comparison between Fuzzy Multi-Criteria and Machine Learning Algorithms. *Geosci. Front.* **2021**, *12*, 857–876. [CrossRef]
13. Binh, T.P. A Novel Classifier Based on Composite Hyper-cubes on Iterated Random Projections for Assessment of Landslide Susceptibility. *J. Geol. Soc. India* **2018**, *91*, 355–362.
14. Liu, X.L. Landslide susceptibility assessment based on Sampling technique and Bayesian spatial statistical model. Master's Thesis, Southwest University of Science and Technology, Mianyang, China, 2021.
15. Xu, C.; Xu, X.; Shyu, J.B.H. Database and spatial distribution of landslides triggered by the Lushan, China Mw 6.6 earthquake of 20 April 2013. *Geomorphology* **2015**, *248*, 77–92. [CrossRef]
16. Shi, L.L. Temporal and Spatial Variation of Vegetation Coverage in Lushan County of Earthquake Region. Master's Thesis, Sichuan Normal University, Chengdu, China, 2014.
17. Liu, L.N. Assessment of Debris Flow Susceptibility in Lushan Earthquake Area. Master's Thesis, China University of Geosciences, Beijing, China, 2015.
18. Jia, J.; Guo, M.Z.; Yao, K.; Liu, H. Geological hazard assessment based on information model supported by slope unit. *Henan Sci.* **2017**, *35*, 787–792.
19. Wang, L.; Sawada, K.; Moriguchi, S. Landslide susceptibility analysis with logistic regression model based on FCM sampling strategy. *Comput. Geosci.* **2013**, *57*, 81–92. [CrossRef]
20. Liu, T.D.; Du, T.C. Research on welding tube factory inventory prediction based on BP neural network. *J. Beijing Inst. Petrochem. Technol.* **2017**, *25*, 53–57.
21. Ke, F.Y.; Li, Y.Y. Landslide geological disaster prediction method based on BP neural network. *Eng. Investig.* **2014**, *8*, 55–60.
22. Li, L. Study on Landslide Disaster in Chongqing Based on Random Forest Model. Master's Thesis, Chongqing Normal University, Chongqing, China, 2015.
23. Gupta, S.K.; Dericks, P.S. Data Imbalance in Landslide Susceptibility Zonation: A Case Study of Mandakini River Basin, Uttarakhand, India. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, Geoscience and Remote Sensing Symposium, IGARSS 2020-2020 IEEE International, Waikoloa, HI, USA, 26 September 2020.

24. Ma, Z.J.; Chen, H.L.; Yang, S.F. Landslide disaster prediction based on support vector machine theory—A case study of Qingyuan district, Zhejiang province. *J. Zhejiang Univ. Sci.* **2003**, *5*, 592–596.
25. Zhang, Y.; Wen, H.; Xie, P.; Hu, D.; Zhang, J.; Zhang, W. Hybrid-optimized logistic regression model of landslide susceptibility along mountain highway. *Bull. Eng. Geol. Environ. Off. J. IAEG* **2021**, *80*, 7385–7401. [CrossRef]
26. Xu, C.; Xu, X.; Dai, F.; Wu, Z.; He, H.; Shi, F.; Wu, X.; Xu, S. Risk assessment and test of landslide in Wenchuan earthquake based on logistic regression model. *Hydrogeol. Eng. Geol.* **2013**, *3*, 98.
27. Xu, C.; Dai, F.C.; Yao, X.; Zhao, Z.; Xiao, J.A. Assessment of Landslide Susceptibility in Wenchuan Earthquake Based on GIS and Deterministic Coefficient Analysis Method. In Proceedings of the Chinese Academy of Sciences (2010) Academic Annual Conference Proceedings (MIDDLE), Beijing, China, 13 January 2011.
28. Zhao, D.L.; Lan, C.Z.; Hou, G.L.; Xu, C.J.; Li, W.Z. Evaluation of geological hazard susceptibility in Hehuang valley of Qinghai Province. *J. Geomech.* **2021**, *27*, 83–95.
29. Wu, W.Y. Landslide Susceptibility Analysis of Ludian Earthquake in Yunnan Province. Master's Thesis, Institute of Earthquake Prediction, China Earthquake Administration, Beijing, China, 2018.
30. Qiu, W.R.; Wu, B.Y.; Pan, X.S. Application of several clustering optimization machine learning methods in landslide susceptibility evaluation in Lingtai County. *Northwest Geol.* **2020**, *53*, 222–233.
31. Liu, Y.; Wang, N.T.; Zhou, C.; Xie, J.L.; Li, Y.Y. Assessment of landslide susceptibility in Fengjie County of the Three Gorges Reservoir Area based on ROC curve and deterministic coefficient method integrated model. *Saf. Environ. Eng.* **2020**, *27*, 61–70.
32. Li, Z.Y. Big data under the background of ROC curve is introduced and application. *Sci. Trib.* **2021**, *14*, 81–84.
33. Alu, S.; Zhang, J.Q.; Tong, Z.J. Study on the Influence of Natural and Human Factors on the Evaluation Results in Landslide Hazard Susceptibility Evaluation. In Proceedings of the 8th Annual Meeting of Risk Analysis Professional Committee of China Disaster Prevention Association, Xi'an, China, 20–21 October 2018; pp. 26–31.
34. Chen, C.; Hu, K.H. Comparative study on the distribution law of landslides in Wenchuan, Lushan and Ludian earthquakes. *J. Eng. Geol.* **2017**, *25*, 806–814.