



Article

AI-Based Susceptibility Analysis of Shallow Landslides Induced by Heavy Rainfall in Tianshui, China

Tianjun Qi ¹, Yan Zhao ² , Xingmin Meng ^{1,2,*}, Guan Chen ² and Tom Dijkstra ³

¹ MOE Key Laboratory of Western China's Environmental Systems, College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China; qitj16@lzu.edu.cn

² School of Earth Sciences, Lanzhou University, Lanzhou 730000, China; zhaoyandz@lzu.edu.cn (Y.Z.); gchen@lzu.edu.cn (G.C.)

³ School of Architecture, Building and Civil Engineering, Loughborough University, Loughborough LE11 3TU, UK; T.A.Dijkstra@lboro.ac.uk

* Correspondence: xmmeng@lzu.edu.cn

Abstract: Groups of landslides induced by heavy rainfall are widely distributed on a global basis and they usually result in major losses of human life and economic damage. However, compared with landslides induced by earthquakes, inventories of landslides induced by heavy rainfall are much less common. In this study we used high-precision remote sensing images before and after continuous heavy rainfall in southern Tianshui, China, from 20 June to 25 July 2013, to produce an inventory of 14,397 shallow landslides. Based on the results of landslide inventory, we utilized machine learning and the geographic information system (GIS) to map landslide susceptibility in this area and evaluated the relative weight of various factors affecting landslide development. First, 18 variables related to geomorphic conditions, slope material, geological conditions, and human activities were selected through collinearity analysis; second, 21 selected machine learning models were trained and optimized in the Python environment to evaluate the susceptibility of landslides. The results showed that the ExtraTrees model was the most effective for landslide susceptibility assessment, with an accuracy of 0.91. This predictive ability means that our landslide susceptibility results can be used in the implementation of landslide prevention and mitigation measures in the region. Analysis of the importance of the factors showed that the contribution of slope aspect (SA) was significantly higher than that of the other factors, followed by planar curvature (PLC), distance to river (DR), distance to fault (DTF), normalized difference vegetation index (NDVI), distance to road (DTR), and other factors. We conclude that factors related to geomorphic conditions are principally responsible for controlling landslide susceptibility in the study area.

Keywords: heavy rainfall; landslide inventory; machine learning; susceptibility mapping



Citation: Qi, T.; Zhao, Y.; Meng, X.; Chen, G.; Dijkstra, T. AI-Based Susceptibility Analysis of Shallow Landslides Induced by Heavy Rainfall in Tianshui, China. *Remote Sens.* **2021**, *13*, 1819. <https://doi.org/10.3390/rs13091819>

Academic Editors: Lefei Zhang and Cristiano Tolomei

Received: 16 March 2021

Accepted: 2 May 2021

Published: 7 May 2021

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



Copyright: © 2021 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

Extreme rainfall events and earthquakes are the two main factors inducing regional landslides [1–6]. Compared with the spatial distribution of earthquakes, which are concentrated in plate margins and intracontinental orogenic belts [1,7], landslides induced by heavy rainfall are more widely distributed on a global basis [8–12]. As the landslides induced by heavy rainfall are characterized by wide distribution, high density, and long travel distance, such landslide events often cause many casualties and major property losses and ecological damage [13,14]. Landslide inventories are an essential basis for studying the formation, distribution, landscape evolution, susceptibility, and risk assessment of regional landslides [5,15–17]. An event inventory shows landslides caused by a single trigger, such as an earthquake, rainfall event, or snowmelt event [18]. Compared with the widespread concerns related to landslide disasters caused by earthquakes, inventories of landslides induced by heavy rainfall are much fewer in number [19]. Nevertheless, in recent years, the rapid development of space radar, satellite remote sensing, small unmanned aerial

vehicles (UAV), and other technologies have provided high-precision images for landslide interpretation induced by heavy rainfall events, which greatly facilitate the production of corresponding landslide inventories [16,20–22].

Landslides are one of the most important natural hazards in China as they are widespread and cause substantial damage and fatalities every year [23]. Consequently, the development of methods to reduce the threat of landslides has long been an important component of landslide research in China and elsewhere. As a method of assessing areas with high landslide susceptibility, landslide susceptibility analysis is important for disaster prevention and mitigation [24–28]. Regional evaluations of landslide susceptibility based on physical models and GIS technology are widely used, and they are important for assessing the future risks of landslides and debris flows, and they can make a major contribution to disaster prevention and control planning [24,29,30]. However, these two traditional methods are limited by problems of efficiency and cost and by their limited ability to obtain useful information from complex datasets, together with their high dependence on human subjectivity [31,32]. In recent years, with the rapid development of Artificial Intelligence (AI) technology, machine learning provides possibilities for improving the accuracy and efficiency of geological hazard susceptibility evaluation [31,33,34]. Machine learning has achieved outstanding results in landslide and debris flow hazard analysis in several regions [35–37]; however, as a new technology, the effect of machine learning in different environments needs to be further examined.

From 20 June to 25 July 2013, continuous heavy rainfall induced a large number of shallow landslides and debris flows in southern Tianshui, China. They resulted in 24 deaths and one missing person; in addition, 2386 houses collapsed and 6666 were damaged. The direct economic losses were USD 1.24 billion, in addition to the losses of life and property and the trauma caused to the local inhabitants [38]. The study had three main components: (i) Comparison of high-precision image data before and after rainfall events was used to interpret and catalog landslides induced by heavy rainfall. (ii) The effect of machine learning on the susceptibility evaluation of shallow landslides induced by heavy rainfall in an area of high vegetation cover was examined. (iii) Evaluating the relative contribution of various factors affecting landslide formation in a high vegetation coverage area.

2. Study Area

The study area is located in southern Tianshui, China (Figure 1). The geological structure of the region is complex, influenced mainly by the Qinling Mountains latitudinal structural belt, the Qilv-Holland Arc structural belt, the West Qinling Mountains northeast structural belt, and the Longxi spiral structural belt. The main lithological unit in the area is the Devonian Shujiaba formation, mainly composed of marl, slate with thin layers of limestone and metasandstone. Yanshanian biotite granite and medium-coarse grained granite are exposed in several areas in the north and east. Neogene strata are dominant in the west and they have an unconformable relationship with the other strata. The lithology is mainly gray-white and gray-green clay and red mudstone with a sandy conglomerate. The thickness of the formation exceeds 1000 m. In addition, carboniferous glutenite and late Devonian slate and sandstone are sporadically exposed in the north and south parts of the study area.

The geomorphology of the study area is dominated by the intermediate- and low-elevation mountains, with altitudes ranging between 1239 m and 2249 m. Because it is located on the southern edge of the Chinese Loess Plateau, the area has a cover of quaternary loess, forming a dual-stratum structure of bedrock and overlying loess. The development of a large pore space and vertical joints in the loess makes it highly permeable, and in addition, the bedrock has a low permeability; therefore, the excess pore water pressure caused by heavy rainfall, combined with the seepage force at the stratum interface, is likely the main reason for the extensive occurrence of shallow landslides in the area [39–41].

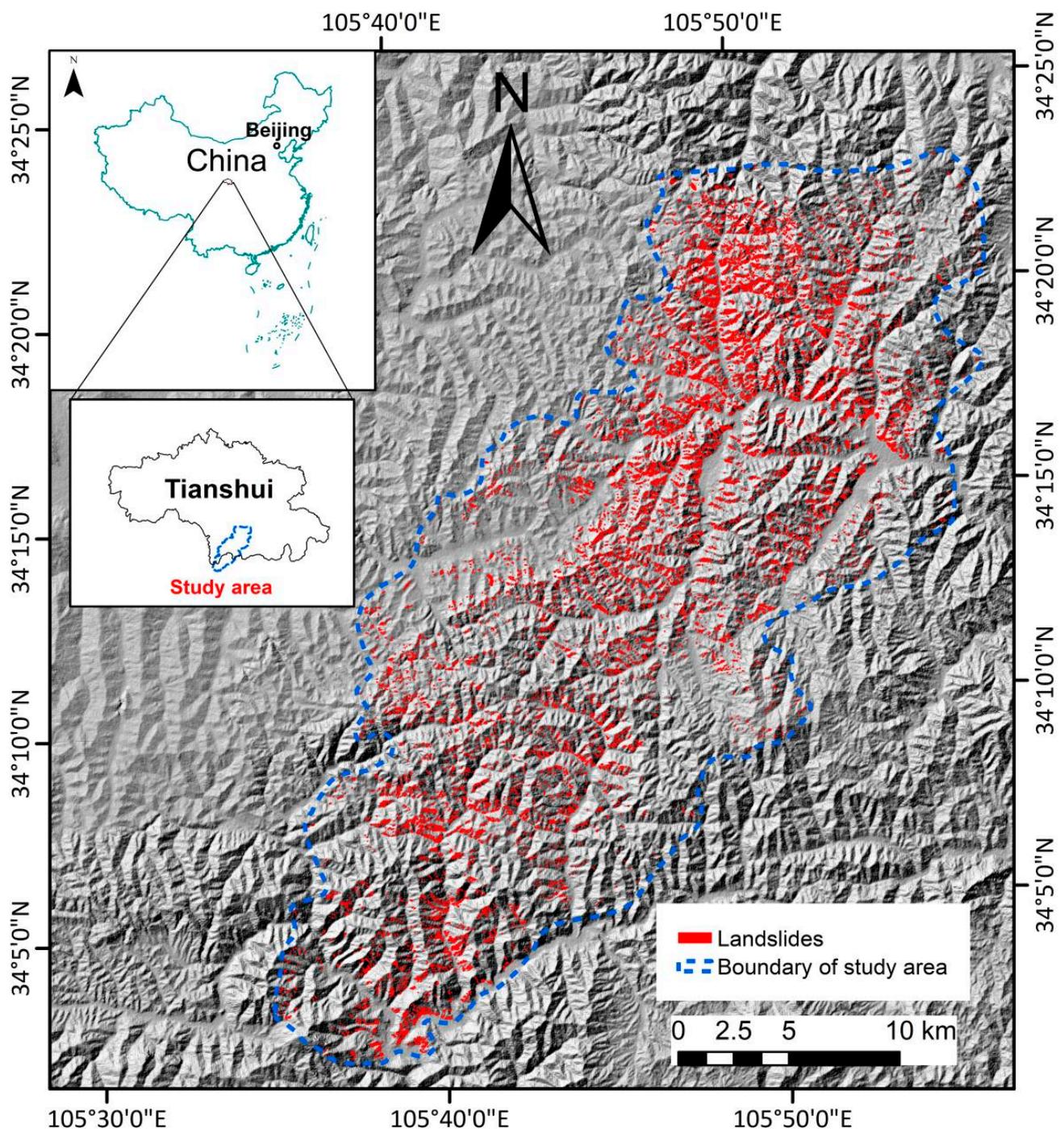


Figure 1. Location of the study area and landslide inventory map.

The study area is located in the transitional region between a semi-humid and semi-arid climate. The climate type is a warm temperate continental climate. The annual average temperature is ~6–11 centigrade, with the highest temperatures generally occurring in July, with a relative humidity of 66%. The average annual rainfall is 800–900 mm, and the seasonal distribution of rainfall is very uneven; most of the rainfall occurs between July to September, which comprises 68% of the annual total. The vegetation coverage in the area is high (generally > 70%) and species-rich.

3. Methods

3.1. Landslide Inventory and Mapping

Guzzetti et al. [18] comprehensively summarized landslide inventory methods and divided them into two categories: traditional methods based mainly on geomorphological field mapping and visual interpretation of aerial photos; and new methods based on high-precision satellite image interpretation, analysis of surface morphology using Airborne LiDAR (Light Detection and Ranging), and the automated and semi-automated recognition of landslides. The former is expensive in terms of time and cost, and for these reasons, traditional methods are gradually being superseded by high-precision optical image interpretation and semi-automatic and automated recognition. However, due to the limitations of automated and semi-automated methods in terms of the accuracy of recognition, they cannot provide a truly comprehensive landslide inventory [42–44]. Therefore, high-precision image interpretation has become the most commonly used method for landslide inventory development in the case of recent earthquakes and rainfall and other events [5,45]. In areas with high vegetation coverage, landslide scars induced by heavy rainfall are clearly resolved in optical images. Therefore, in this study, we downloaded 2 m × 2 m Google earth images before and after an interval of continuous rainfall and used them for comparative analysis (Figure 2), in order to provide a detailed recognition of rainfall-induced landslides. Landslide interpretation, inventorying, and mapping was conducted using ArcGIS 10.2 software (The company is located in Redlands, CA, USA).

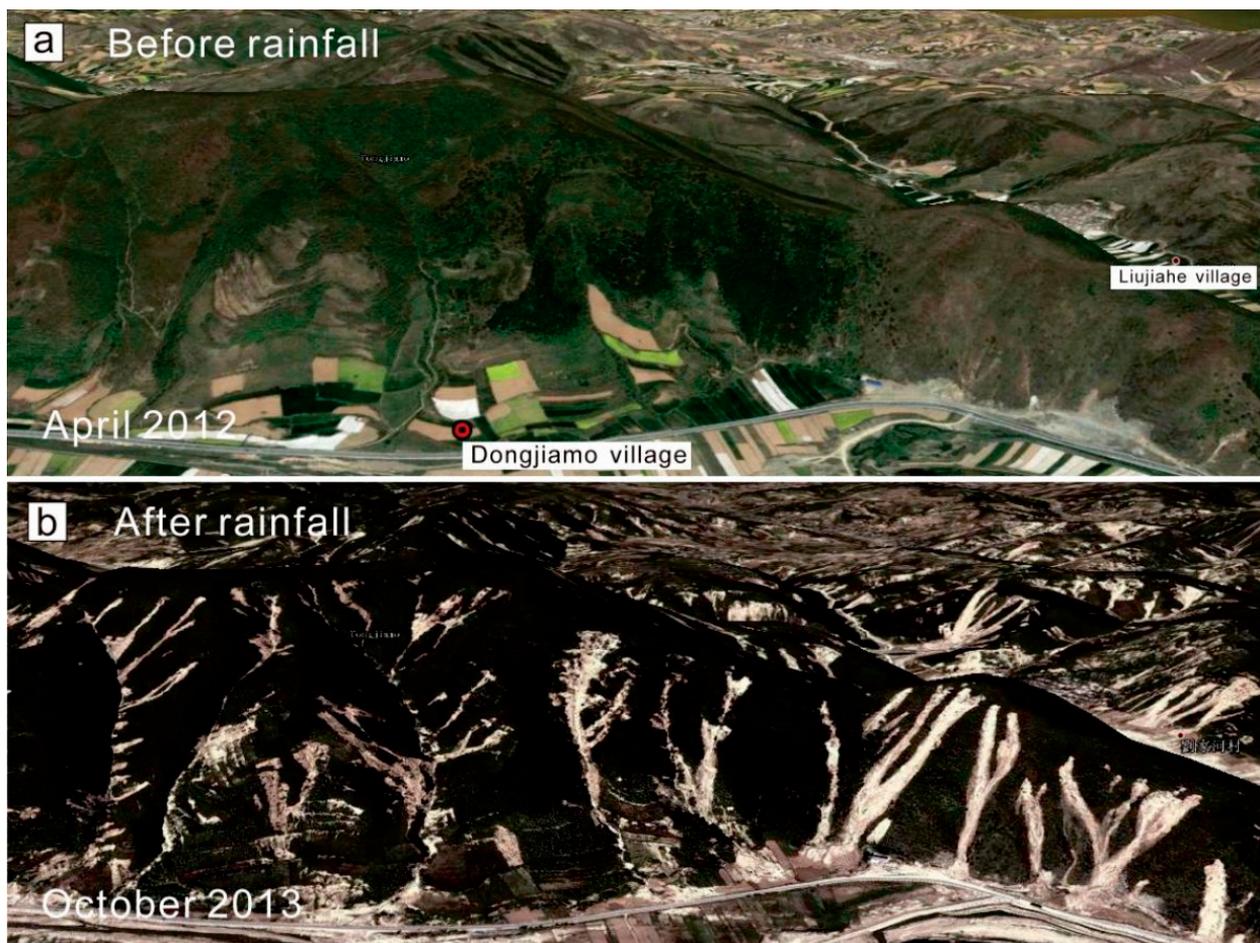


Figure 2. An example of the comparison of Google earth images before and after rainfall in the same area.

3.2. Landslide Susceptibility Evaluation Based on Machine Learning

The process of modeling with machine learning includes the selection and preparation of parameters, data acquisition and processing, and model selection, fitting, and evaluation. A flow chart of the process is shown in Figure 3. The selection of a suitable terrain mapping unit is the basis of landslide sensitivity analysis [25]. At present, the grid cell is still the most commonly used terrain element in most of the literature [25,26]. In this study, in order to balance the amount of information of grid acquisition, data volume, and calculation efficiency, a 100 m × 100 m grid was selected as the evaluation unit for landslide susceptibility analysis. A total of 65,472 grid cells were defined in the study area, of which 13,859 grid cells corresponded to landslides. In this study, the extraction of geomorphic factors was based on 12.5 m × 12.5 m Digital Elevation Mode (DEM) data from ALOS Satellite (Figure 4A), and lithology and faults were derived from 1:50,000 geological mapping data (Figure 4B) (source from China Geological Survey).

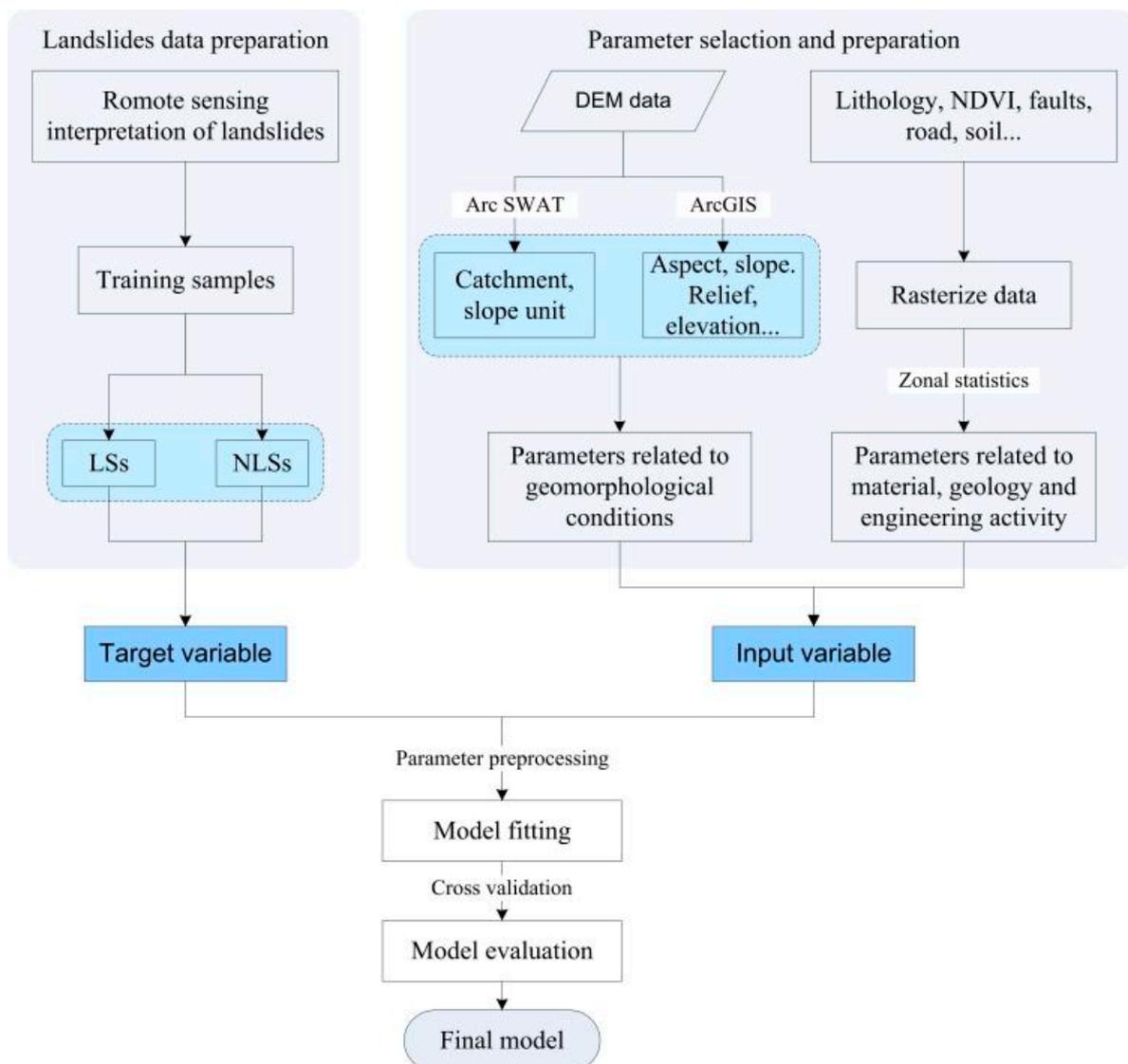


Figure 3. Modeling flow chart (Note: LSs indicate landslides, and NLSs indicate non-landslides).

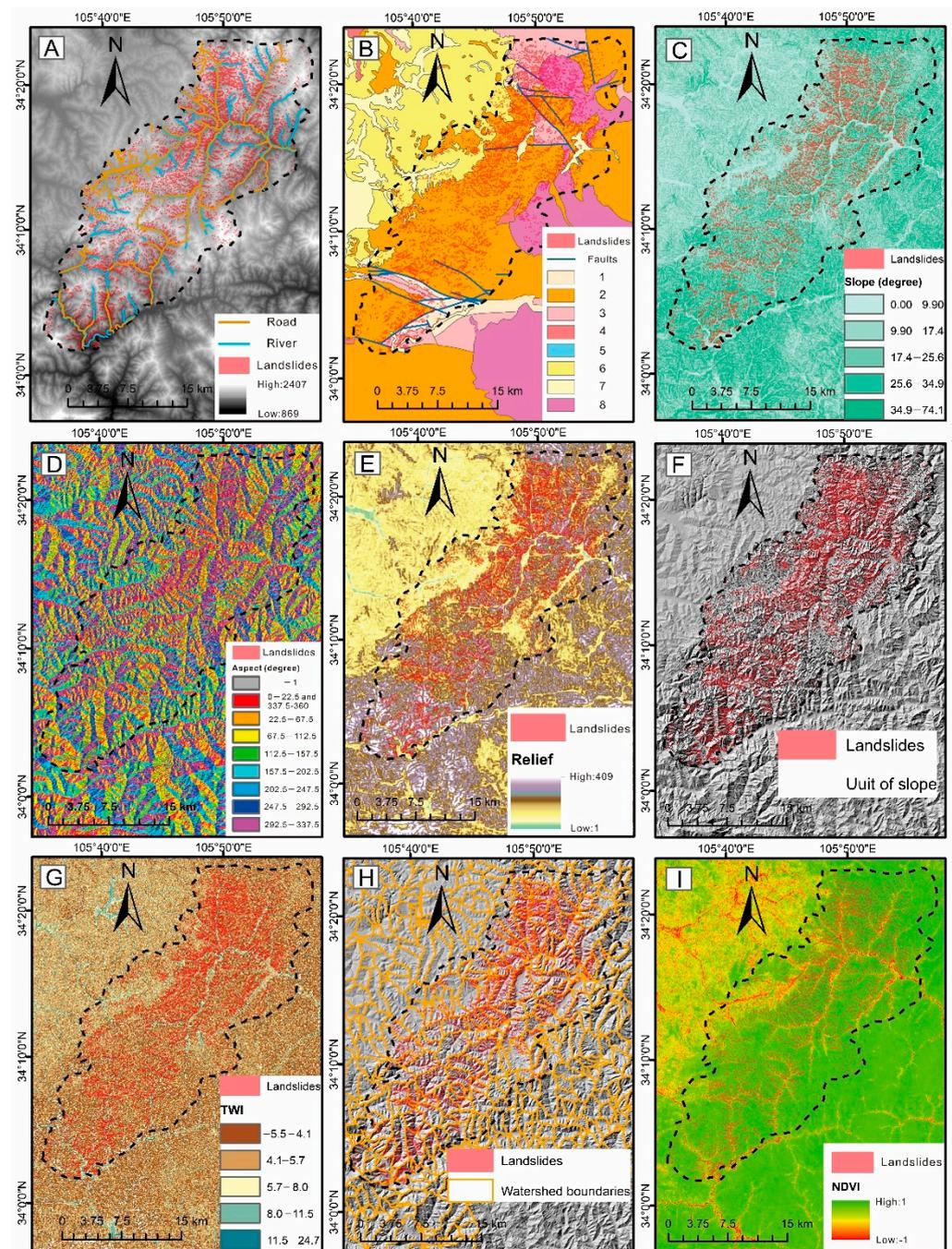


Figure 4. Shows part of the landslide-influencing factors used in this study. (A) Digital Elevation Mode (DEM); (B) Lithology and fault, 1. Carboniferous glutenite, 2. Middle Devonian marl, slate, 3. Late Devonian slate and sandstone, 4. Paleogene sandstone, 5. Jurassic sandstone and limestone, 6. Neogene mudstone, 7. Quaternary sediments, 8. Yanshanian granite; (C) Slope; (D) Slope aspect; (E) Local relief; (F) Unit of slope; (G) TWI; (H) Watershed; (I) NDVI.

3.2.1. Selection of Factors Influencing Landslides

As a complex process of material transport and energy transfer on the earth's surface, the formation and distribution of landslides are determined by the effects of climate, hydrology, geology, landforms, human activity, and other factors [46]. To a large extent, the formation and distribution of landslides is related to specific local factors; for example, in orogenic belts, slope, lithology, and structure are important factors [47,48], while in mountains with low and intermediate altitudes, rainfall, soil properties, and engineering activity are important [49,50]. Therefore, there is no consensus regarding which factors

should be used in landslide susceptibility evaluation. In this study, based on an evaluation of previous studies [24–28] and with regard to the specific environment of the Tianshui area, we selected geomorphic factors, landslide material and geological conditions, and human activity as the three categories of factors for landslide susceptibility analysis (Table 1).

Table 1. Fields and characteristics of the spatial database.

	Field	Parameter	Units
1	ID	Identification field	/
2	LO	Landslide occurrence or not	/
3	AS	Average slope	°
4	SA	Slope aspect	°
5	LR	Local relief	km
6	PRC	Profile curvature	/
7	PLC	Planar curvature	/
8	SUA	Slope unit area	km ²
9	E	Elevation	km
10	TWI	Topographic wetness index	/
11	WA	Watershed area	km ²
12	NDVI	Normalized Difference Vegetation Index	/
13	FLI	Formation lithological index	/
14	DTF	Distance to fault	km
15	ST	Soil type	/
16	SC	Sand content	%
17	SG	Gravel content	%
18	SIC	Silt content	%
19	CC	Clay content	%
20	SB	Soil bulk	N/m ³
21	DR	Distance to river	km
22	DTR	Distance to road	km

Continuous heavy rainfall was the primary cause of the groups of shallow landslide events. The shallow landslides induced by heavy rainfall are widely distributed all over the world. Critical rainfalls that induce shallow landslides are an important factor in the study of landslide triggering thresholds. The records of six rainfall stations in the study area show that the accumulated rainfall from 20 June to 25 July 2013 is more than 230 mm, which reaches the threshold of shallow landslide in many studies [51,52]. In other words, in the event, the rainfall intensity meets the critical threshold in the whole area. In addition, due to the study area being small, the error and resolution of the existing rainfall data cannot meet the factor requirements of machine learning susceptibility evaluation. Therefore, we chose to carry out unified value processing for the rainfall conditions of this evaluation, focusing on the influence of geomorphic conditions, geological structure, and material composition on the susceptibility of shallow landslide. The specific parameters adopted and the rationale for their use are described below.

Parameters Related to Geomorphological Conditions

Average slope (AS) (Figure 4C). Slope is one of the most important factors influencing stability. Different slope angles can affect the magnitude of normal stress and shear stress on the potential failure surface.

Slope aspect (SA) (Figure 4D). Slope aspect strongly affects hydrological processes via evapotranspiration and weathering processes in a given microclimatic environment [53].

Local relief (LR) (Figure 4E). The potential energy of the slope is determined by its relief. Statistical analysis of landslides shows that topographic relief is an important factor affecting the spatial distribution of a landslide [22,54].

Profile curvature (PRC) and planar curvature (PLC). Profile and planar curvature are important parameters reflecting the morphological characteristics of slopes. Curvature is defined as the rate of change of slope gradient or aspect, usually in a specific direction [55].

Slope unit area (SUA) (Figure 4F). The slope unit is the fundamental spatial domain used in quantitative geomorphological analyses. A slope unit can be used for terrain zonation, using methods such as sensitivity modeling and hydrological and erosion modeling; and those based on the geographical environment, including ecology, agriculture, forestry, land use, and other aspects [56]. The slope unit used in this study was extracted by the ArcSWAT module of ArcGIS software. Specifically, it was obtained from the ridgeline and river network extracted under the condition of a 100-hectare flow accumulation.

Elevation (E). Temperature, rainfall, vegetation type, and microorganisms are dependent on elevation. These factors can affect soil layer thickness: the lower the altitude, the thicker the soil layer, while high mountain areas are mainly bare hard rock. In some cases, precipitation and the incidence of landslides increase with increasing altitude [57].

Topographic wetness index (TWI) (Figure 4G). TWI reflects the distribution of soil moisture, and soil moisture content in turn strongly affects the cohesion and internal friction angle of slope materials.

Watershed area (WA) (Figure 4H). Záruba and Mencl [58] observed a relationship between the occurrence of landslides and watershed area. The larger the watershed area, the greater the amount of water seeping into the ground, which increases slope instability [59,60].

Parameters Related to Landslide Materials and Geological Conditions

Normalized Difference Vegetation Index (NDVI) (Figure 4I). The NDVI effectively reflects vegetation coverage, which has important effects on slope stability by reducing the rainfall infiltration rate. The vertical and horizontal growth of plant roots also increases slope stability [61]. NDVI was derived from Landsat-8 images (June 2016) with a resolution of 30-m (Landsat-8 image courtesy of the US Geological Survey).

Formation lithological index (FLI) (Figure 4B). Lithology affects the spatial distribution of landslides. The structural characteristics of the bedrock promote landslide initiation in several ways: (1) by producing weak surfaces that are prone to sliding; (2) by facilitating the introduction of groundwater into the overlying soil mantle; and (3) by destabilizing the regolith because of weathering [46].

Distance to fault (DTF). The two principal effects of faults on landslides are (1) a fault plane can act as the dominant structural plane in the formation of a sliding surface, and (2) rock mass damage caused by fault activity may lead to slope instability.

Soil type (ST). Like lithology, soil is also the material basis of landslide formation. There are substantial differences in soil microstructure, water permeability, and vegetation growth between soil types.

Contents of sand (SC), gravel (SG), silt (SIC), and clay (CC). The grain-size composition of the soil determines the cohesion, shear strength, and hydraulic conductivity of the slope, and thus its stability.

Distance to river (DR). In many areas, landslides are clustered along rivers, and landslide density decreases with increasing distance from rivers [17,62]. Fluvial incision provides the potential energy for the development of a landslide, while the lateral erosion of a river can destroy the slope toe, causing slope instability.

3.2.2. Parameter Preprocessing

Parameters Related to Landslide Materials and Geological Conditions

In order to eliminate collinearity among the selected parameters, a heat map of a parameter correlation matrix was calculated using the Seaborn Python visualization package (<https://seaborn.pydata.org/generated/seaborn.heatmap.html#seaborn.heatmap>, accessed date: 10 November 2020) (Figure 5). Strongly correlated parameters have a certain degree of redundancy and they also affect the stability of the model operation. Through the heat map of the parameter correlation matrix, it was found that several parameters selected in the study have a strong correlation, for example, the following correlations coefficients were obtained: SB vs. SC (0.86); SIC vs. SC (0.93); SB vs. CC (0.97), in other words, they have an almost consistent influence on landslide development. Therefore, SB and SC were excluded from this study.

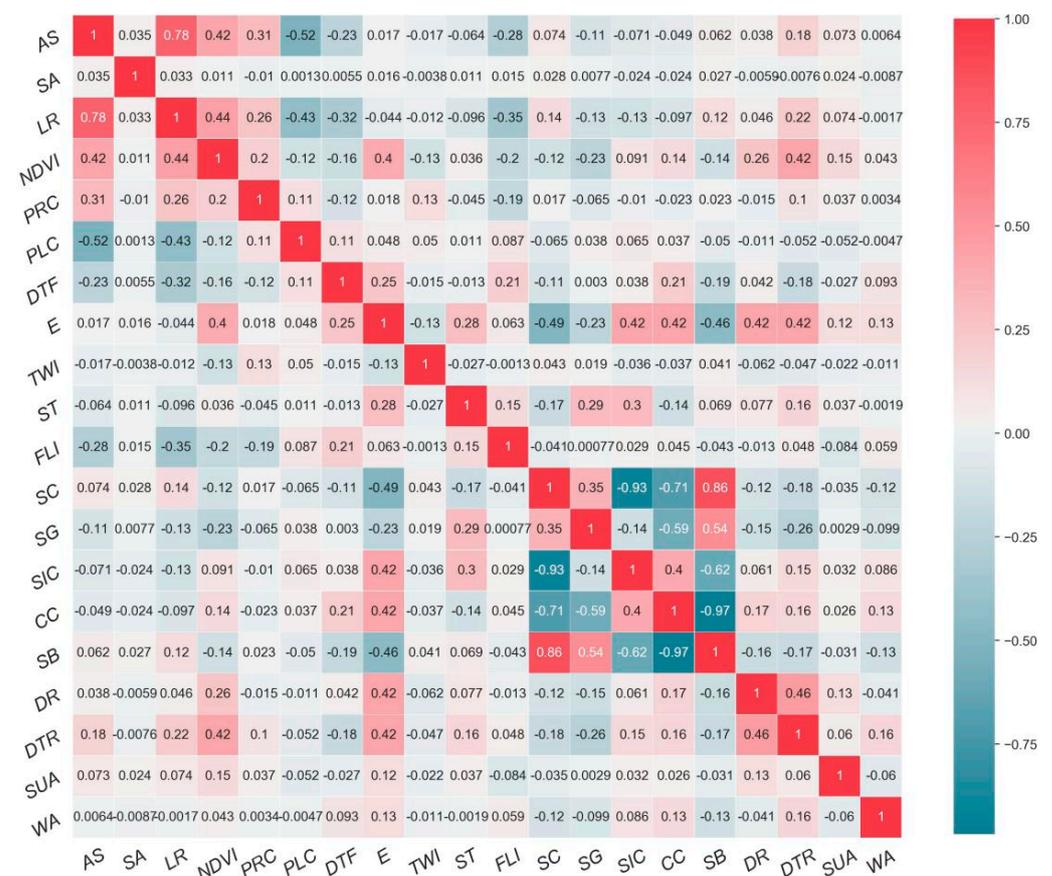


Figure 5. Heat map of the parameter correlation matrix.

Resampling

In a machine learning algorithm, if the ratio of non-landslides (NLSs) to landslides (LSs) is 1:1, machine learning may focus on the classification of LSs rather than of NLSs. However, in the present study, the ratio of NLSs to LSs is close to 4:1 (Figure 6). In order to maintain a balance between the two types of samples, SMOTE (synthetic mineral oversampling technology) was used to increase the number of LS samples [63]. This method randomly selects a nearest neighbor sample B from A (a sample in NDFs), and then, randomly selects a point C from the relationship between A and B, as a new minority sample. After resampling, the ratio of NLS sample size to LS sample size is 1:1, which provides a balance between the data samples.

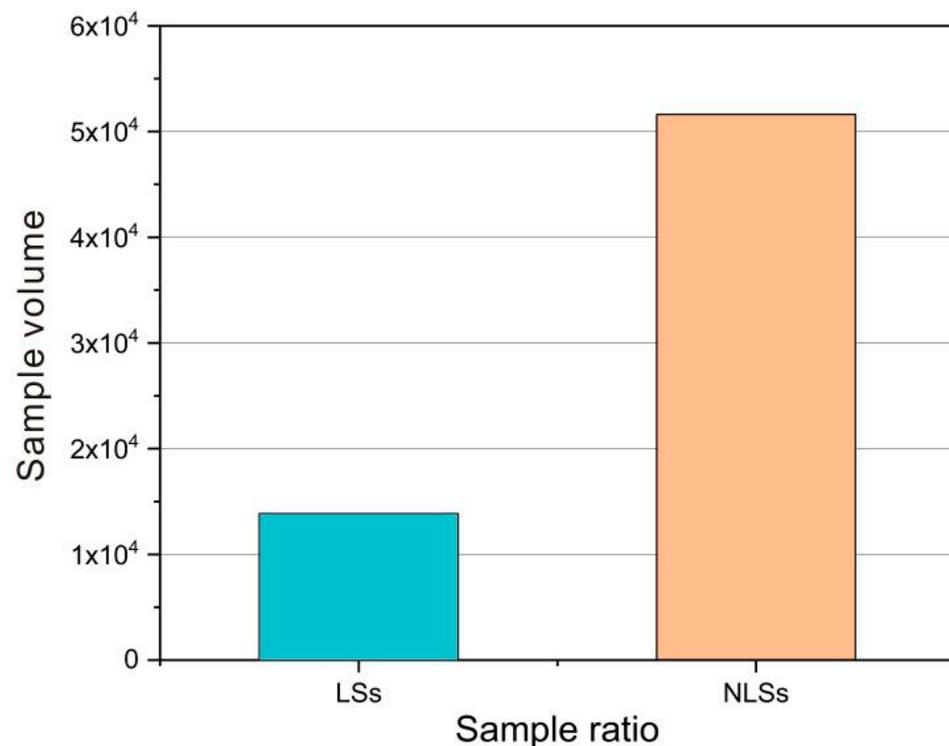


Figure 6. Sample ratio for LSs (landslides) and NLSs (non-landslides).

Data Standardization

The data were standardized in order to improve the accuracy of the model algorithm and to speed up the convergence of the model. In addition, several machine learning algorithms are very sensitive to feature scales. Therefore, we used a standard scalar algorithm (from Scikit-learn, <https://scikit-learn.org>, accessed date: 10 November 2020) to normalize the factors by removing the mean and scaling according to the variance. Scikit-learn is a Python library that provides a standard interface for implementing machine learning algorithms [64].

3.2.3. Candidate Machine Selection

We chose 21 types of model algorithms that are widely used in machine learning [35]. Via inspection and testing, we chose the most suitable model algorithm for landslide susceptibility evaluation in an area of dense vegetation.

Ensemble Methods

The principle of the ensemble method is to combine several classifiers (or different parameters of an algorithm) to improve the effectiveness of each single classifier. The classifiers can be divided into average methods and boosting methods. AdaBoost, Gradient Tree Boosting (GDBT), Bagging, Random Forest, and Extra Trees were selected in this study.

Generalized Linear Models (GLMs)

The generalized linear model is an extension of the linear model. The relationship between the mathematical expectation of the response variable and the prediction variable of the linear combination is established by the relationship function. Logistic Regression (LR), Passive Aggressive, Ridge, Stochastic Gradient Descent (SGD), and Perceptron were used in this study.

Naïve Bayes (NB)

Naive Bayes classification is based on Bayesian probability. Assuming that the attributes are independent of each other, the probability of each feature is obtained and the larger one is taken as the prediction result. Gaussian Naive Bayes and Bernoulli Naive Bayes were selected.

Nearest Neighbors

The principle of the nearest neighbor method is to find a specified number of nearest sample points and then use them to predict new points.

Support Vector Machines (SVM)

The principle of SVM is to solve the separation hyperplane, which can correctly divide the training dataset and provide the largest geometric interval. Support Vector Classification (SVC), Linear SVC, and Nu-SVC were selected.

Trees

The tree classifier is a tree structure in which each internal node represents a judgment of an attribute, and each branch represents an output of the judgment result. Finally, each leaf node represents a classification result. Decision Tree and Extra Tree were selected.

Discriminant Analysis

Discriminant analysis is a method of multivariate statistical analysis that classifies the studied objects according to several observed indexes. Linear discriminant and quadratic discriminant analyses were selected.

eXtreme Gradient Boosting (XGBoost)

XGBoost is a boosting algorithm and a type of lifting tree model. It implements the GBDT algorithm efficiently and makes many improvements to the algorithm, integrating numerous tree models to produce a strong classifier.

3.2.4. Model Fitting and Tuning

The initial model is trained by the training data in a cross-validation dataset. The models are then sorted according to the average accuracy score (ACC) of the test data in the cross-validation data set. ACC represents the correct allocation rate of all samples involved in the modeling. It can be seen in Figure 7 that the overall fitting effect of the comprehensive model is better than that of other models, and the highest score was achieved by ExtraTrees, followed by RandomForest, Bagging, and KNeighbors. ACC is calculated as follows:

$$ACC = (TP + TN)/(TP + FN + FP + TN) \quad (1)$$

True positive (TP): the predicted class is positive, and the prediction agrees with the actual class;

False positive (FP): the predicted class is positive, and the prediction disagrees with the actual class;

True negative (TN): the predicted class is negative, and the prediction agrees with the actual class;

False negative (FN): the predicted class is negative, and the prediction disagrees with the actual class.

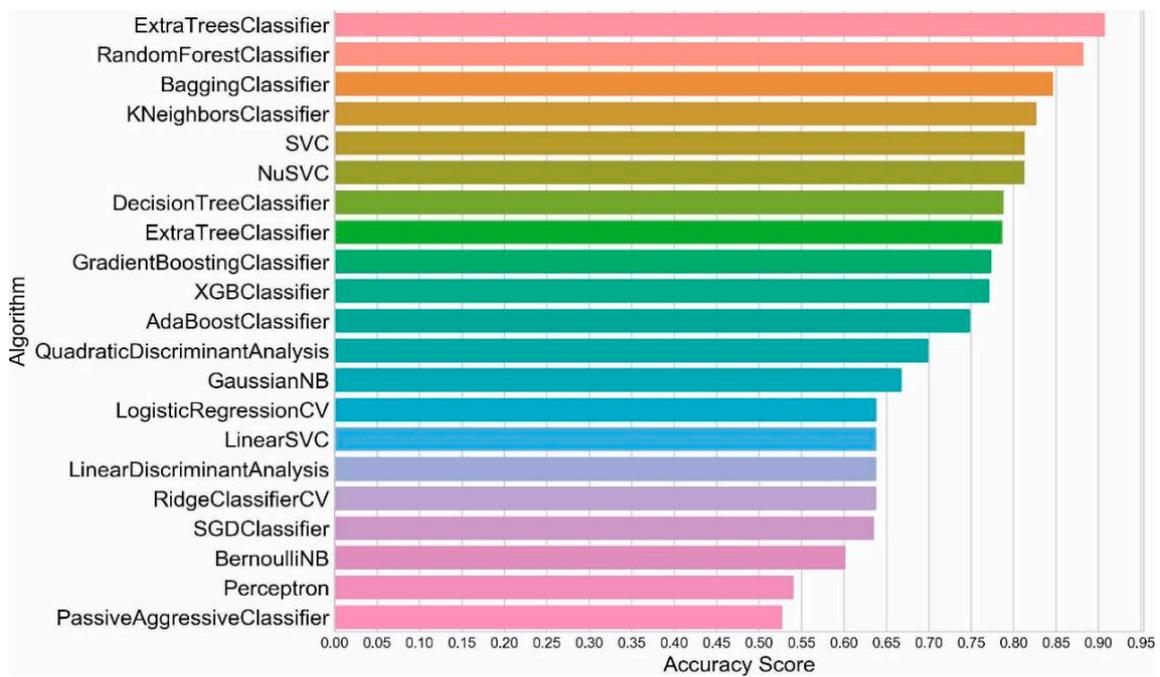


Figure 7. Ranking of model accuracy scores.

The terms are listed in Table 2 and are defined as follows:

Table 2. Confusion matrix results.

		Predicted Label	
		Positive	Negative
True label	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

We selected the first four models for optimization (Figure 8). The model was fitted using a parametric grid method and the grid search cross-validation method, and the best super parameters were found by AUC (area under the receiver operating characteristic curve) scoring method. According to the optimal super parameters of each model given in Table 3, the training set of the model was cross-validated 10 times, and the models were reordered according to the average accuracy score of the test data. After optimization, the performance of the four models was seen to have improved. ExtraTrees remained the optimal model, with a test data ACC of 0.91, and an average AUC of 0.97 after 10-time cross validation (Figure 9). AUC represents a trade-off between sensitivity and specificity. After optimization, the accuracy of the Bagging model was significantly improved.

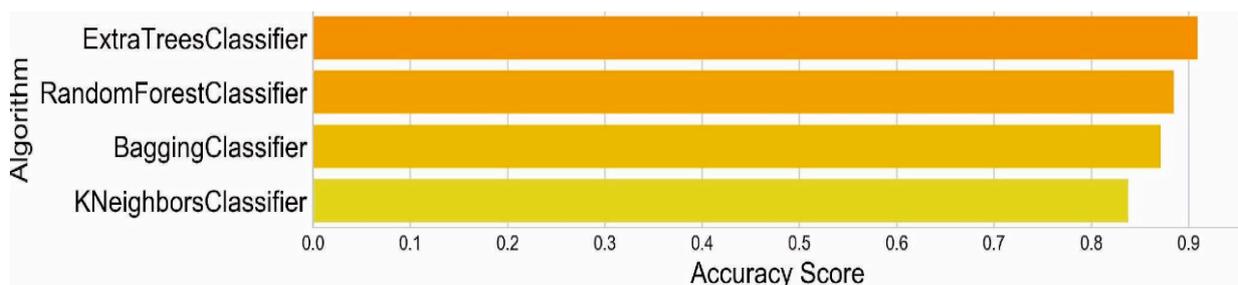


Figure 8. Ranking of ACC scores after model optimization.

Table 3. Model optimal super parameter results and the time consumed. (Note: Please refer to the Scikit-learn website for the explanation of each parameter and its role in model adjustment: <https://scikit-learn.org>, accessed date: 10 November 2020).

	Classifier Algorithm	Best Parameter	Runtime (s)
1	ExtraTreesClassifier	'n_estimators' = 500 'random_state' = 0 'criterioin' = gini	50,615.69
2	RandomForestClassifier	'criterioin' = 'entropy' 'max_depth' = 54 'n_estimators'=500 'oob_score' = True	329,402.72
3	BaggingClassifier	'max_samples' = 1.0 'n_estimators' = 500	35,085.42
4	KNeighborsClassifier	'algorithm' = auto 'n_neighbors' = 8 'weithts' = 'distance'	41,816.73

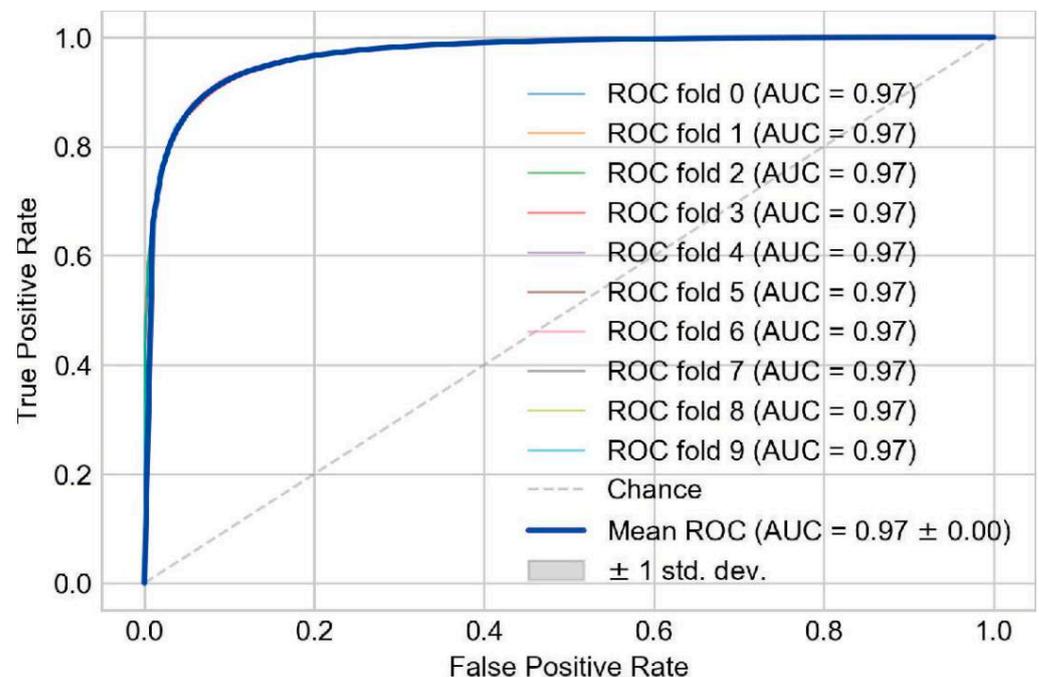


Figure 9. Receiver Operating Characteristic Curve (ROC) and AUC of ExtraTrees using 10-time cross validation.

4. Results

4.1. Landslide Inventory and Mapping

Comparison of the remote sensing images before and after rainfall event enabled us to identify 14,397 landslides in an area of 655 km². The interpretation results are shown in Figure 1. The landslide density reached 22/km², with the largest landslide area being 39,637 m². The average landslide area is 907 m². The total area of all landslides in the study region is 13.06 km², accounting for 2% of the total. In the landslide inventory, the largest 10 landslides account for 1.8% of the total landslide area, while the top 10% of large landslides account for 10.9%. Compared with the results of landslide inventories in other areas, the proportion of large-scale landslides of the total landslide area is relatively small [53,65].

Kernel density spatial analysis, with a default radius of 1 km (unweighted) and with area weighting, was carried out using the ArcGIS 10.2 toolbox (Figure 10). The results show that the spatial distribution of landslides induced by heavy rainfall is not completely uniform. In the case of a non-weighted distribution, the landslides are clustered in the north and south, and the density and area of the cluster in the north are higher than those of the south (Figure 10A). From the area-weighted distribution (Figure 10B), it was found that large landslides are mainly concentrated in the northern region and compared to the non-weighted distribution, the distribution range of large landslides in the northern region is larger and more dispersed.

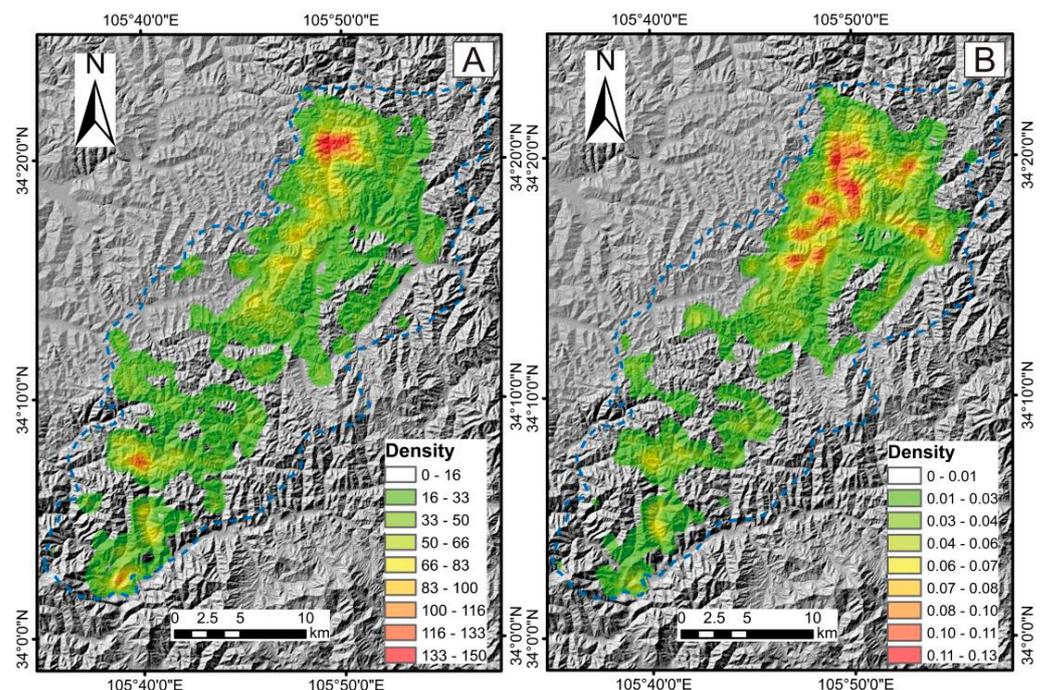


Figure 10. Kernel density analysis (search radius 1 km). (A) Non-weighted landslide density. (B) Area-weighted landslide density.

4.2. Landslide Susceptibility Mapping

Although the ExtraTrees model achieved the highest score after optimization (Figure 8), it can only output the classification result (i.e., 0/1) and cannot generate a probability value. However, probability values are needed to produce a landslide susceptibility map. Therefore, the RandomForest (Super parameter: ‘criterion’ = ‘entropy’, ‘max_depth’ = 54, ‘n_estimators’ = 300, ‘obb_score’ = True) was used as the final model to produce a landslide susceptibility map for the study area. The methods provided by Scikit-learn were used to construct the probability set. In the binary case, the probabilities are calibrated using Platt scaling (Platt “Probabilistic outputs for RandomForests and comparisons to regularized likelihood methods”): logistic regression of the RandomForest scores and fitting by additional cross-validation of the training data [66].

The natural discontinuity method was used to divide the probability values of landslide susceptibility into five categories [67]: very low, low, moderate, high, and very high (Figure 11), and the corresponding proportions were 41.2%, 24.5%, 13.1%, 6.2%, and 15.0% respectively. It can be seen that the proportion of the landslide susceptibility area does not decrease with increasing sensitivity. The proportion of the extremely low susceptibility area is the highest, followed by the low susceptibility area, and the proportion of the high susceptibility area is the lowest. The spatial distribution of the very high susceptibility area is not completely uniform: the northern region has a more dense and larger range of distribution characteristics than the central and southern regions, which is consistent with

the actual distribution of landslides. The area surrounding the study area has mainly a very low susceptibility, which may be related to rainfall intensity.

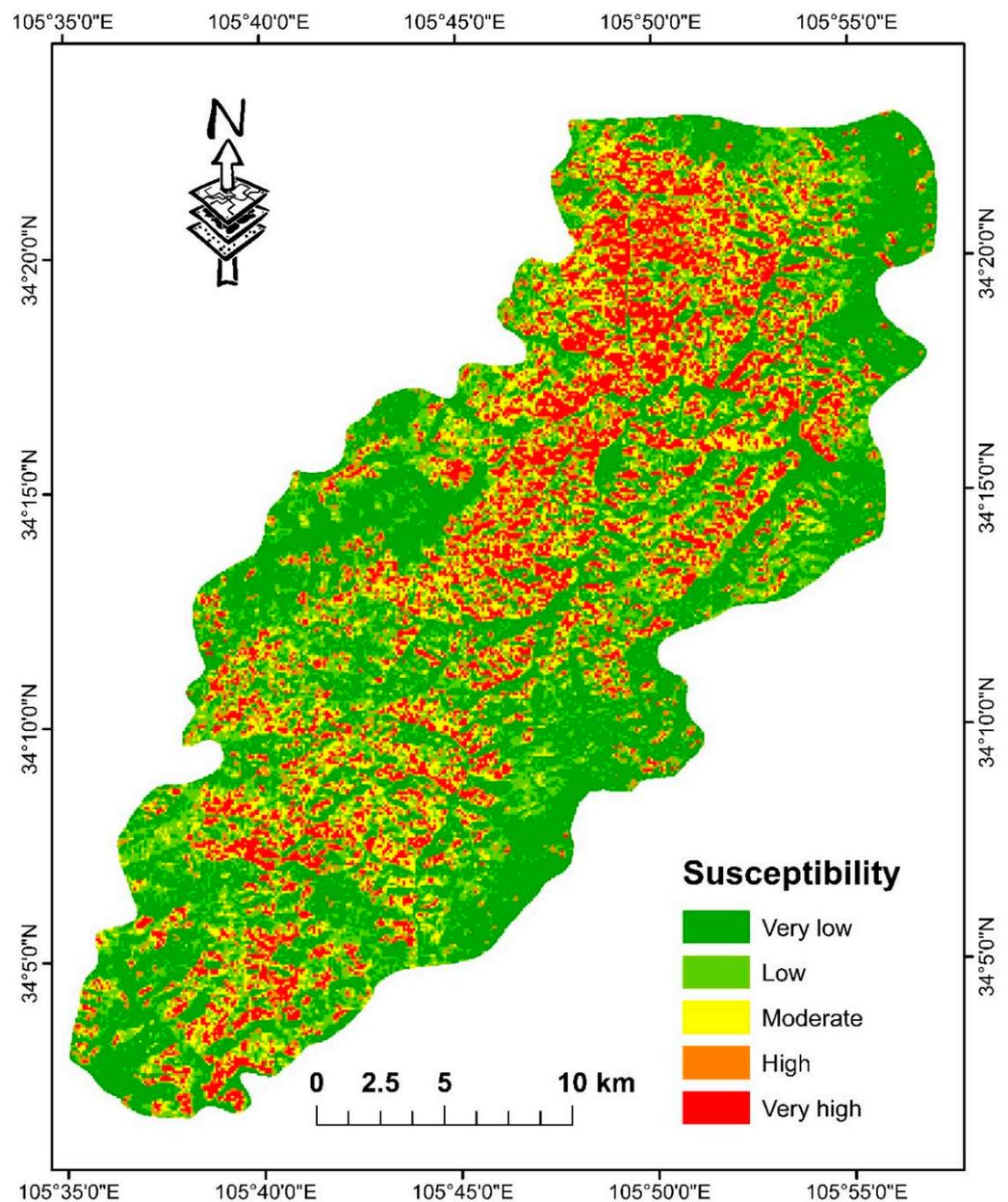


Figure 11. Landslide susceptibility map of the Tianshui area calculated using the RandomForest model.

5. Discussion

The interpretability of the model helps to determine the potential relationship between different influencing factors and landslide susceptibility. This in turn enables the landslide susceptibility assessment results to be applied outside the study area and to increase our understanding of the contribution of the various factors influencing landslide formation under similar environmental conditions.

The calculated weight of each factor is shown in Figure 12, from which it can be seen that although all of the factors contribute to the landslide development, there are differences in the size of their contributions. Among the factors, the contribution of SA (14%) is significantly higher than those of the other factors, followed by PLC (8%). DR, DTF, NDVI, and DTR have contributions of 7%, and the other factors contribute less than 6%. Geomorphic factors can be seen to be the most important controls for

landslide susceptibility, while factors related to the landslide material and geological conditions play a secondary role, and the impact of engineering activity is relatively small. In the present study, four soil-related factors (SIC, ST, SG, and CC) made only small contributions to landslide development, which may be related to inaccuracies in the soil data. Specifically, from the distribution of SA (Figure 13A), LSs accounted for the highest proportion when $94^\circ < SA < 246^\circ$, indicating that sunlit slopes are more prone to landslides than shaded slopes. The influence of this aspect on the spatial distribution of landslides has been concerning for a long time. Many research results show that the influence of the slope aspect on landslide formation is mainly manifested in three categories: (i) The microclimate of slopes with different orientations shows regular differences. Compared with the shaded slope, the sunlit slope has higher temperature and precipitation, the physical and chemical weathering rate is, therefore, faster, forming a thicker soil layer, ensuring the material source of landslide formation [51,68]. (ii) Comparing with the shaded slope, the vegetation coverage of the sunlit slope is low, and most of it consists of shrubs and herbs [69]. The influence of shallow roots on the stability of the landslide on the sunlit slope is significantly weaker than that of the vertical roots of trees on the shaded slope. The influence of vegetation on slope stability is bidirectional, and it has adverse effects on the development of deep landslides, while vegetation can restrain the development of shallow landslides [70]. (iii) The continuous alternation of wet and dry on a sunlit slope can easily form the macropore system in the unsaturated zones of the slopes, which is conducive to the rapid infiltration of precipitation, thus is unfavorable to slope stability [71,72].

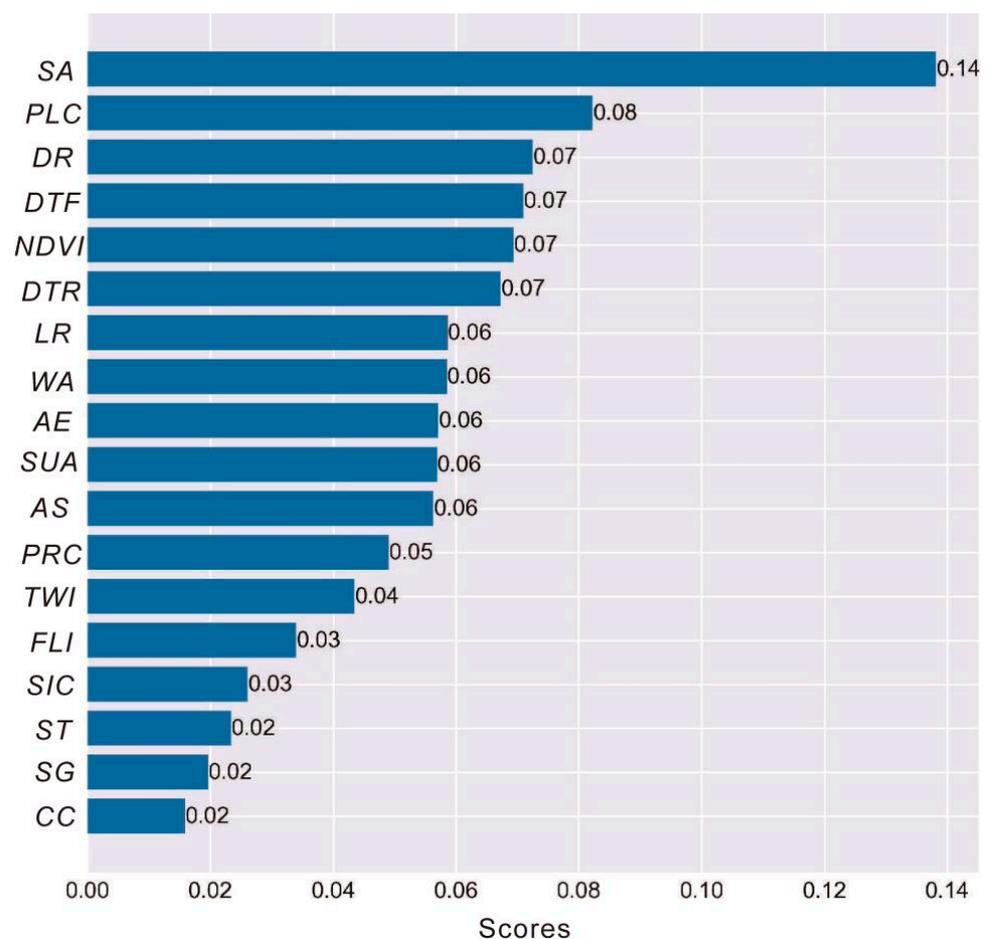


Figure 12. Calculated importance of the parameters.

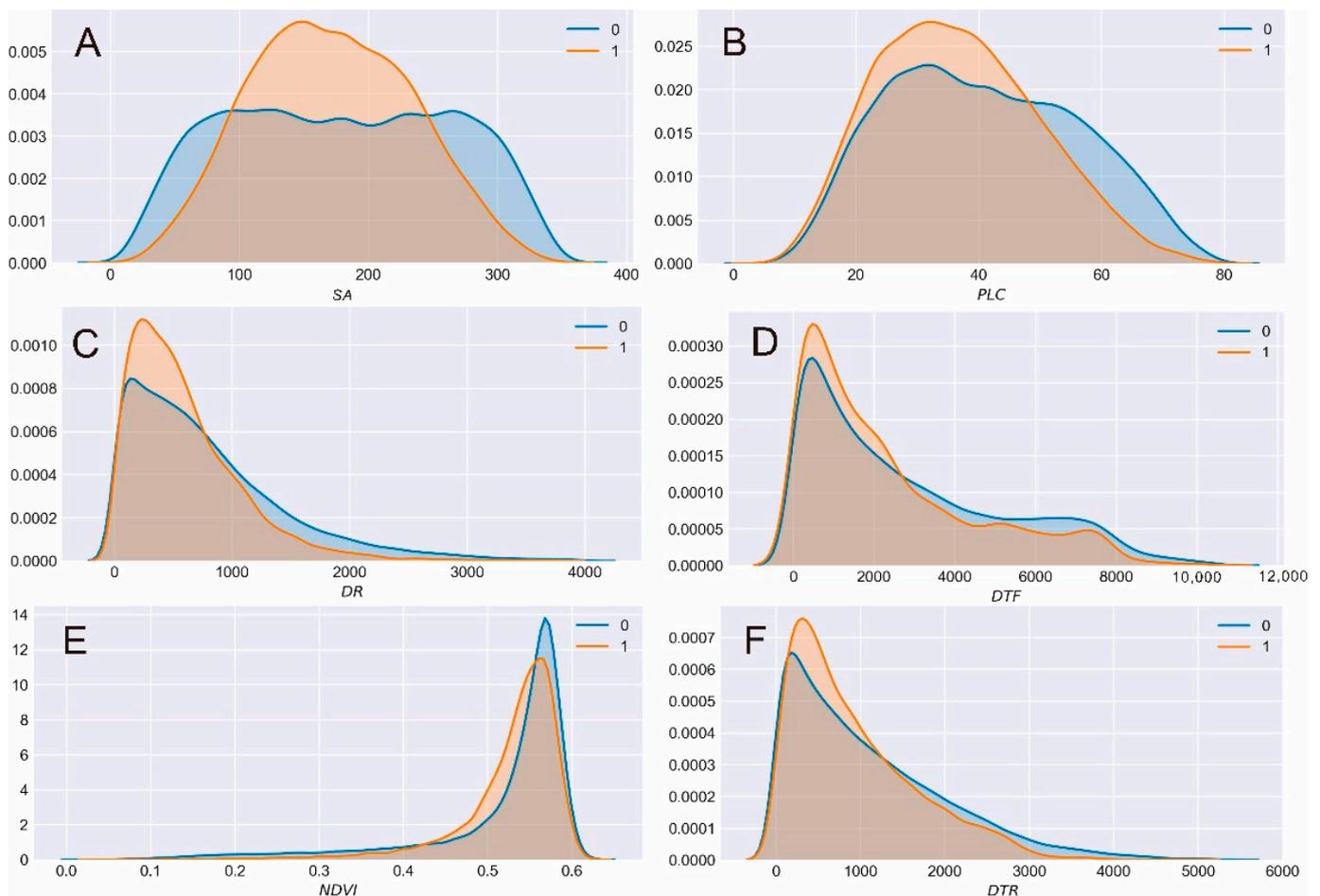


Figure 13. Distribution of (A) Slope aspect (SA), (B) Planar curvature (PLC), (C) Distance to river (DR), (D) Distance to fault (DTF), (E) Normalized Difference Vegetation Index (NDVI), and (F) Distance to road (DTR). '1' (0) indicates grid cells prone to (not prone to) landslides.

For the ranges of $5 < PLC < 48$, $84 < DR < 760$, $0 < DTF < 2677$, $0.03 < NDVI < 0.56$, $150 < DTR < 1270$ (Figure 13), the proportion of LSs is higher, which indicates that landslides are more likely to occur within these ranges. For this landslide event, the shape of the slope was the second most important factor after slope aspect because there are significant differences in the ponding capacity and degree of surface differentiation of different types of slopes, such as concave, convex, and flat slopes [73]. This difference may be amplified under the effect of heavy rainfall, thus strongly influencing the development of landslides. DR, DTF, and DTR have similar distribution characteristics. The intensity of fluvial erosion, the damage of fault to rock mass strength, and the level of engineering activity all decrease with increasing distance from these elements [17,74]. Therefore, the smaller the distance from these features, the higher the proportion of LSs, and with the increase of distance, NLSs gradually occupy a higher proportion. The influence of vegetation on landslide development is reflected by the fact that areas of low vegetation coverage are more prone to landslides [61]. This is mainly because vegetation delays rainfall infiltration, increases evaporation, and well-developed root systems significantly increase slope stability. Similar to the results of some studies, slope aspect is the most important factor in the susceptibility evaluation [73]. However, due to the differences in geographical location, climate environment, topography, and vegetation types, the role of slope aspect in many regions may be very different. For example, the weight of landslide susceptibility factors in an orogenic belt and in a mountainous hilly region may be completely different. The location and climatic conditions of the study area determine that the microclimate of different slope aspects is very important to the development of landslides, and the microclimate signif-

icantly affects the vegetation type and coverage, water system development, humidity index, soil thickness, and other factors, thus profoundly affecting the spatial distribution of landslides [53].

The LSs based on machine learning are relatively flexible and practical, and they are readily applicable in disaster prevention and land management. The research results can be used as a reference for decision-makers and planners in the study area. With the increasing population pressure in western China, there is a trend towards an increased settlement on steep hillsides. Therefore, in order to protect human life and property, landslide susceptibility maps can be used as a basic tool for land management and planning in future construction projects in such areas.

6. Conclusions

We have compared high-precision remote sensing images in southern Tianshui before and after an interval of continuous high rainfall (from 20 June to 25 July 2013), with the aim of identifying rainfall-induced landslides. According to the inventory map of landslides, various machine learning methods were applied to landslide susceptibility evaluation, and we selected the optimal model for landslide susceptibility evaluation in areas of low and medium elevation mountains with a high vegetation coverage and produced a landslide susceptibility map. Finally, in order to better understand the factors controlling landslide susceptibility, we analyzed the role and weight of each influencing factor in the training process. The main conclusions are as follows:

- (1) The 21 initial models were trained with the training data in the cross-validation dataset, and the models were then sorted according to the average accuracy score (ACC). The results showed that the overall fitting effect of the comprehensive model was better than for the other models. The ExtraTrees model had the highest score, with an average test data accuracy of 0.91, and the average AUC after 10-times cross validation was 0.97. This model can be effectively used for the susceptibility evaluation of shallow landslides.
- (2) Among all of the selected evaluation factors, slope aspect made a larger contribution to landslide development than the other factors. For $94^\circ < SA < 246^\circ$, LSs accounted for the highest proportion, which indicates that sunlit slopes are significantly more prone to landslides than shaded slopes, followed by PLC, DR, DTF, NDVI, and DTR. Geomorphic conditions are the most important factors in triggering landslides induced by heavy rainfall, followed by fluvial erosion and fault distribution, while human activities have only a small influence.
- (3) In the evaluation of landslide susceptibility based on machine learning, the prediction performance of various models is significantly different. Extensive comparative prediction in different environments, closely linking the model evaluation with the goals of the study and increasing the understanding of the ability and limitations of the model are the key to model selection in the future, so as to strengthen the application of artificial intelligence technology in the field of geological disaster prevention and improve the prediction accuracy and efficiency.

Author Contributions: T.Q. and Y.Z. designed this study, performed the main analysis, and wrote the paper. X.M. and T.D. were mainly involved in supervision and discussion. G.C. contributed to the revising of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Key R&D Program of China (Grant No. 2018YFC1504704), Science and Technology Major Project of Gansu Province (Grant No. 19ZD2FA002), Program for International S&T Cooperation Projects of Gansu Province (Grant No. 2018-0204-GJC-0043), Fundamental Research Funds for the Central Universities (lzujbky-2018-46), and the Key Research and Development Program of Gansu Province (Grant No. 18YF1WA114).

Acknowledgments: The Digital Elevation Model data were provided by the Japan Aerospace Exploration Agency (JAXA). Soil data were supported by the Chinese Soil Science Database. The authors would like to acknowledge Jan Bloemendal for his comments which improved this paper and for his English language corrections.

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

References

1. Keefer, D.K. Landslides caused by earthquakes. *Geol. Soc. Am. Bull.* **1984**, *95*, 406–421. [[CrossRef](#)]
2. Guzzetti, F.; Cardinali, M.; Reichenbach, P.; Cipolla, F.; Sebastiani, C.; Galli, M.; Salvati, P. Landslides triggered by the 23 November 2000 rainfall event in the Imperia Province, Western Liguria, Italy. *Eng. Geol.* **2004**, *73*, 229–245. [[CrossRef](#)]
3. Yamagishi, H.; Iwahashi, J. Comparison between the two triggered landslides in mid-Niigata, Japan by July 13 heavy rainfall and October 23 intensive earthquakes in 2004. *Landslides* **2007**, *4*, 389–397. [[CrossRef](#)]
4. Owen, L.A.; Kamp, U.; Khattak, G.A.; Harp, E.L.; Keefer, D.K.; Bauer, M.A. Landslides triggered by the 8 October 2005 Kashmir Earthquake. *Geomorphology* **2008**, *94*, 1–9. [[CrossRef](#)]
5. Sato, H.P.; Harp, E.L. Interpretation of earthquake-induced landslides triggered by the 12 May 2008, M7.9 Wenchuan earthquake in the Beichuan area, Sichuan Province, China using satellite imagery and Google Earth. *Landslides* **2009**, *6*, 153–159. [[CrossRef](#)]
6. Pánek, T.; Brázdil, R.; Klime, J.; Smolková, V.; Hradeck, J.; Zahradník, P. Rainfall-induced landslide event of May 2010 in the eastern part of the Czech Republic. *Landslides* **2011**, *8*, 507–516. [[CrossRef](#)]
7. Romanowicz, B. Spatiotemporal patterns in the energy release of great earthquakes. *Science* **1993**, *260*, 1923–1926. [[CrossRef](#)]
8. Kirschbaum, D. Global Catalog of Rainfall-Triggered Landslides for Spatial and Temporal Hazard Characterization. In *Landslide Science for a Safer Geoenvironment*; Springer: Cham, Switzerland, 2014; pp. 809–814.
9. Guzzetti, F.; Gariano, S.L. Landslides in a changing climate. *Earth Sci. Rev.* **2016**, *162*, 227–252.
10. Hong, Y.; Adler, R.; Huffman, G. Evaluation of the potential of NASA multi-satellite precipitation analysis in global landslide hazard assessment. *Geophys. Res. Lett.* **2018**, *33*, 1–5. [[CrossRef](#)]
11. Marc, O.; Stumpf, A.; Malet, J.P.; Gosset, M.; Uchida, T.; Chiang, S.H. Initial insights from a global database of rainfall-induced landslide inventories: The weak influence of slope and strong influence of total storm rainfall. *Earth Surf. Dyn.* **2018**, *6*, 903–922. [[CrossRef](#)]
12. Fustos, I.; Abarca, D.R.R.; Yaeger, P.M.; Valenzuela, M.S. Rainfall-induced landslides forecast using local precipitation and global climate indexes. *Nat. Hazards* **2020**, *102*, 115–131. [[CrossRef](#)]
13. Dai, F.C.; Lee, C.F.; Wang, S.J. Characterization of rainfall-induced landslides. *Int. J. Remote Sens.* **2003**, *24*, 4817–4834. [[CrossRef](#)]
14. Chen, H.; Dadson, S.; Chi, Y.G. Recent rainfall-induced landslides and debris flow in northern Taiwan. *Geomorphology* **2006**, *77*, 112–125. [[CrossRef](#)]
15. Guzzetti, F.; Reichenbach, P.; Cardinali, M.; Galli, M.; Ardizzone, F. Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* **2005**, *72*, 272–299. [[CrossRef](#)]
16. Haneberg, W.C.; Cole, W.F.; Kasali, G. High-resolution lidar-based landslide hazard mapping and modeling, UCSF Parnassus Campus, San Francisco, USA. *Bull. Eng. Geol. Environ.* **2009**, *68*, 263–276. [[CrossRef](#)]
17. Larsen, I.J.; Montgomery, D.R. Landslide erosion coupled to tectonics and river incision. *Nat. Geosci.* **2012**, *5*, 468–473. [[CrossRef](#)]
18. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.T. Landslide inventory maps: New tools for an old problem. *Earth Sci. Rev.* **2012**, *112*, 42–66. [[CrossRef](#)]
19. Tanyaş, H.; Westen, C.J.; Allstadt, K.E.; Jibson, R.W. Factors controlling landslide frequency-area distributions. *Earth Surf. Process. Landf.* **2018**, *44*, 900–917. [[CrossRef](#)]
20. Glenn, N.F.; Streutker, D.R.; Chadwick, D.J.; Thackray, G.D.; Dorsch, S.J. Analysis of LiDAR-derived topographic information for characterizing and differentiating landslide morphology and activity. *Geomorphology* **2006**, *73*, 131–148. [[CrossRef](#)]
21. Santangelo, M.; Cardinali, M.; Rossi, M.; Mondini, A.C.; Guzzetti, F. Remote landslide mapping using a laser rangefinder binocular and GPS. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 2539–2546. [[CrossRef](#)]
22. Pánek, T.; Břežný, M.; Kapustová, V.; Lenart, J.; Chalupa, V. Large landslides and deep-seated gravitational slope deformations in the Czech Flysch Carpathians: New LiDAR-based inventory. *Geomorphology* **2019**, *346*, 1–18. [[CrossRef](#)]
23. Huang, R. Large-scale landslide and their sliding mechanisms in China since the 20th Century. *Chin. J. Rock Mech. Eng.* **2007**, *26*, 433–454.
24. Ayalew, L.; Yamagishi, H.; Ugawa, N. Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. *Landslides* **2004**, *1*, 73–81. [[CrossRef](#)]
25. Nefeslioglu, H.A.; Duman, T.Y.; Durmaz, S. Landslide susceptibility mapping for a part of tectonic Kelkit Valley (Eastern Black Sea region of Turkey). *Geomorphology* **2008**, *94*, 401–418. [[CrossRef](#)]
26. Baeza, C.; Corominas, J. Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. *Earth Surf. Process. Landf.* **2010**, *26*, 1251–1263. [[CrossRef](#)]
27. Pradhan, B.; Lee, S. Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. *Landslides* **2010**, *7*, 13–30. [[CrossRef](#)]
28. Demir, G. GIS-based landslide susceptibility mapping for a part of the North Anatolian fault Zone between Reşadiye and Koyulhisar (Turkey). *Catena* **2019**, *183*, 1–12. [[CrossRef](#)]

29. Saha, A.K.; Gupta, R.P.; Sarkar, I.; Arora, M.K.; Csaplovics, E. An approach for GIS-based statistical landslide susceptibility zonation—with a case study in the Himalayas. *Landslides* **2005**, *2*, 61–69. [[CrossRef](#)]
30. Yesilnacar, E.; Topal, T. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek Region (Turkey). *Eng. Geol.* **2005**, *79*, 251–266. [[CrossRef](#)]
31. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [[CrossRef](#)]
32. Aditian, A.; Kubota, T.; Shinohara, Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology* **2018**, *318*, 101–111. [[CrossRef](#)]
33. Marjanović, M.; Kovačević, M.; Bajat, B.; Voženilek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **2011**, *123*, 225–234. [[CrossRef](#)]
34. Pham, B.T.; Bui, T.D.; Prakash, I.; Dholakia, M.B. Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan Area (India) using GIS. *Catena* **2017**, *149*, 52–63. [[CrossRef](#)]
35. Feng, Q.; Zhao, Y.; Meng, X.; Su, X.; Qi, T.; Yue, D. Application of machine learning to debris flow susceptibility mapping along the China-Pakistan Karakoram highway. *Remote Sens.* **2020**, *12*, 2933.
36. Napoli, M.D.; Carotenuto, F.; Cevasco, A.; Confuorto, P.; Calcaterra, D. Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. *Landslides* **2020**, *17*, 1897–1914. [[CrossRef](#)]
37. Zhao, Y.; Meng, X.; Qi, T.; Qing, F.; Chen, G. AI-based identification of low-frequency debris flow catchments in the Bailong River Basin, China. *Geomorphology* **2020**, *359*, 107–125. [[CrossRef](#)]
38. Guo, F.; Meng, X.; Li, Z.; Xie, Z.; Chen, G.; He, Y. Characteristics and causes of assembled geo-hazards induced by the rainstorm on 25th July 2013 in Tianshui City, Gansu, China. *Mt. Res.* **2015**, *1*, 100–107.
39. Dijkstra, T.A.; Rogers, C.D.F.; Smalley, I.J.; Derbyshire, E.; Li, Y.J.; Meng, X.M. The loess of north-Central China: Geotechnical properties and their relation to slope stability. *Eng. Geol.* **1994**, *36*, 153–171. [[CrossRef](#)]
40. Wang, Z.; Kang, G.; Ma, C.; Miao, T. A study on the generating mechanism of vertical joints in loess. *Int. J. Rock Mech. Min. Sci. Geomech.* **1994**, *31*, 259–260.
41. Sun, P.; Wang, G.; Wu, L.Z.; Lgwe, O.; Zhu, E. Physical model experiments for shallow failure in rainfall-triggered loess slope, Northwest China. *Bull. Eng. Geol. Environ.* **2019**, *78*, 43–63. [[CrossRef](#)]
42. Tarolli, P.; Tarboton, D.G. A new method for determination of most likely landslide initiation points and the evaluation of digital terrain model scale in terrain stability mapping. *Hydrol. Earth Syst. Sci.* **2006**, *10*, 663–677. [[CrossRef](#)]
43. Passalacqua, P.; Tarolli, P.; Foufoula, G.E. Testing space-scale methodologies for automatic geomorphic feature extraction from lidar in a complex mountainous landscape. *Water Resour. Res.* **2010**, *46*, 1–17. [[CrossRef](#)]
44. Tarolli, P.; Sofia, G.; Fontana, G.D. Geomorphic features extraction from high-resolution topography: Landslide crowns and bank erosion. *Nat. Hazards* **2012**, *61*, 65–83. [[CrossRef](#)]
45. Pandey, P. Inventory of rock glaciers in Himachal Himalaya, India using high-resolution Google Earth imagery. *Geomorphology* **2019**, *340*, 103–115. [[CrossRef](#)]
46. Clague, J.J.; Stead, D. *Landslides: Types, Mechanisms and Modeling*; Cambridge University Press: Cambridge, UK, 2012; pp. 159–171.
47. Carlini, M.; Chelli, A.; Francese, R.; Giacomelli, S.; Giorgi, M.; Quagliarini, A.; Carpena, A.; Tellini, C. Landslides types controlled by tectonics-induced evolution of valley slopes (northern Apennines, Italy). *Landslides* **2018**, *15*, 283–296. [[CrossRef](#)]
48. Mishra, B.K.; Bhattacharjee, D.; Chattopadhyay, A.; Prusty, G. Tectonic and lithologic control over landslide activity within the Larji-Kullu tectonic window in the higher Himalayas of India. *Nat. Hazards* **2018**, *92*, 673–697. [[CrossRef](#)]
49. Wang, J.; Liang, Y.; Zhang, H.; Wu, Y.; Lin, X. A loess landslide induced by excavation and rainfall. *Landslides* **2014**, *11*, 141–152. [[CrossRef](#)]
50. Peng, J.; Fan, Z.; Wu, D.; Zhuang, J.; Dai, F.; Chen, W.; Zhao, C. Heavy rainfall triggered loess–mudstone landslide and subsequent debris flow in Tianshui, China. *Eng. Geol.* **2015**, *186*, 79–90. [[CrossRef](#)]
51. Palladino, M.R.; Viero, A.; Turconi, L.; Brunetti, M.T.; Peruccacci, S.; Melillo, M.; Luino, F.; Deganutti, A.M.; Guzzetti, F. Rainfall thresholds for the activation of shallow landslides in the Italian Alps: The role of environmental conditioning factors. *Geomorphology* **2018**, *303*, 53–67. [[CrossRef](#)]
52. Kanungo, D.P.; Sharma, S. Rainfall thresholds for prediction of shallow landslides around Chamoli-Joshimath region, Garhwal Himalayas, India. *Landslides* **2014**, *11*, 629–638. [[CrossRef](#)]
53. Burnett, B.N.; Meyer, G.A.; Mcfadden, L.D. Aspect-related microclimatic influences on slope forms and processes, northeastern Arizona. *J. Geophys. Res. Earth Surf.* **2008**, *113*, 1–18. [[CrossRef](#)]
54. Břežný, M.; Pánek, T. Deep-seated landslides affecting monoclinial flysch morphostructure: Evaluation of LiDAR-derived topography of the highest range of the Czech Carpathians. *Geomorphology* **2017**, *285*, 44–57. [[CrossRef](#)]
55. Wilson, J.P.; Gallant, J.C. *Terrain Analysis Principles and Applications*; Wiley: Toronto, ON, Canada, 2000; p. 479.
56. Alvioli, M.; Guzzetti, F.; Marchesini, I. Parameter-free delineation of slope units and terrain subdivision of Italy. *Geomorphology* **2020**, *358*, 107–124. [[CrossRef](#)]
57. Gallart, F.; Clotet, P.N. Some aspects of the geomorphic processes triggered by an extreme rainfall event: The November 1982 flood in the eastern Pyrenees. *Catena Suppl.* **1988**, *13*, 79–95.
58. Záruba, Q.; Mencl, V. *Landslides and Their Control*; Elsevier: Amsterdam, The Netherlands, 1969; p. 236.

59. Okimura, T. A prediction system for the site of probable surface failure of mountain-slope by topographical factors. *Proc. Jpn. Soc. Civil Eng.* **1983**, *338*, 131–138. [[CrossRef](#)]
60. Oyagi, N. Landslides in Weathered Rocks and Residual Soils in Japan and Surrounding Areas: A State of the Art Report. In Proceedings of the 4th International Symposium on Landslides, Toronto, ON, Canada, 16–21 September 1984; pp. 1–31.
61. Stokes, A.; Atger, C.; Bengough, A.G.; Fourcaud, T.; Sidle, R.C. Desirable plant root traits for protecting natural and engineered slopes against landslides. *Plant Soil* **2009**, *324*, 1–30. [[CrossRef](#)]
62. Tsou, C.; Chigira, M.; Matsushi, Y.; Chen, S. Deep-seated gravitational deformation of mountain slopes caused by river incision in the Central Range, Taiwan: Spatial distribution and geological characteristics. *Eng. Geol.* **2015**, *196*, 126–138. [[CrossRef](#)]
63. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
64. Bisong, E. Introduction to Scikit-learn. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*; Apress: Berkeley, CA, USA, 2019; pp. 215–229.
65. Korup, O. Distribution of landslides in southwest New Zealand. *Landslides* **2005**, *2*, 43–51. [[CrossRef](#)]
66. Wu, T.F.; Lin, C.J.; Weng, R.C. Probability estimates for multi-class classification by pairwise coupling. *J. Mach. Learn. Res.* **2004**, *5*, 975–1005.
67. Roy, J.; Saha, S.; Arabameri, A.; Blaschke, T.; Bui, D.T. A novel ensemble approach for landslide susceptibility mapping (ISM) in Darjeeling and Kalimpong districts, West Bengal, India. *Remote Sens.* **2019**, *11*, 2866. [[CrossRef](#)]
68. Cooper, A.W. An example of the role of microclimate in soil genesis. *Soil Sci.* **1960**, *90*, 109–120. [[CrossRef](#)]
69. Nevo, E.; Fragman, O.; Dafni, A.; Beiles, A. Biodiversity and interslope divergence of vascular plants caused by microclimate differences at “Evolution Canyon” lower nahal Oren, Mount Carmel, Israel. *Isr. J. Plant Sci.* **1999**, *47*, 49–59. [[CrossRef](#)]
70. Xue, W.P.; Zhao, Z.; Li, P.; Cao, Y. Researches on root distribution characteristics of Robinia Pseudoacacia stand in Wangdonggou on different site conditions. *J. Agric. Sci. Technol.* **2003**, *31*, 27–32.
71. Hussein, M.A. Changes in microstructure, voids and b-fabric of surface samples of a Vertisol caused by wet/dry cycles. *Geoderma* **1998**, *85*, 63–82. [[CrossRef](#)]
72. Luiz, F.; Osny, O.S.; Bacchi, K.P. Gammaray computed tomography to evaluate wetting/drying soil structure changes. *Nucl. Instrum. Methods Phys. Res.* **2005**, *229*, 443–456.
73. Begueria, S. Changes in land cover and shallow landslide activity: A case study in the Spanish Pyrenees. *Geomorphology* **2006**, *74*, 196–206. [[CrossRef](#)]
74. Reid, M.; Iverson, R. Gravity-driven groundwater flow and slope failure potential 2. Effects of slope morphology, material properties, and hydraulic heterogeneity. *Water Resour. Res.* **1992**, *3*, 935–950. [[CrossRef](#)]