

Article

# Application of Machine Learning to Debris Flow Susceptibility Mapping along the China–Pakistan Karakoram Highway

Feng Qing<sup>1,2</sup>, Yan Zhao<sup>3,\*</sup> , Xingmin Meng<sup>1,3,4</sup>, Xiaojun Su<sup>1</sup> , Tianjun Qi<sup>1</sup> and Dongxia Yue<sup>1</sup>

<sup>1</sup> College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China; qingf16@lzu.edu.cn (F.Q.); xmmeng@lzu.edu.cn (X.M.); suxj17@lzu.edu.cn (X.S.); qitj16@lzu.edu.cn (T.Q.); dxyue@lzu.edu.cn (D.Y.)

<sup>2</sup> Department of Emergency Management of Gansu Province, Lanzhou 730000, China

<sup>3</sup> School of Earth Sciences, Lanzhou University, Lanzhou 730000, China

<sup>4</sup> Gansu Tech Innovation Centre for Environmental Geology and Geohazard Prevention, Lanzhou 730000, China

\* Correspondence: zhaoyandz@lzu.edu.cn

Received: 23 July 2020; Accepted: 7 September 2020; Published: 10 September 2020



**Abstract:** The China–Pakistan Karakoram Highway is an important land route from China to South Asia and the Middle East via Pakistan. Due to the extremely hazardous geological environment around the highway, landslides, debris flows, collapses, and subsidence are frequent. Among them, debris flows are one of the most serious geological hazards on the Karakoram Highway, and they often cause interruptions to traffic and casualties. Therefore, the development of debris flow susceptibility mapping along the highway can potentially facilitate its safe operation. In this study, we used remote sensing, GIS, and machine learning techniques to map debris flow susceptibility along the Karakoram Highway in areas where observation data are scarce and difficult to obtain by field survey. First, the distribution of 544 catchments which are prone to debris flow were identified through visual interpretation of remote sensing images. The factors influencing debris flow susceptibility were then analyzed, and a total of 17 parameters related to geomorphology, soil materials, and triggering conditions were selected. Model training was based on multiple common machine learning methods, including Ensemble Methods, Gaussian Processes, Generalized Linear models, Navies Bayes, Nearest Neighbors, Support Vector Machines, Trees, Discriminant Analysis, and eXtreme Gradient Boosting. Support Vector Classification (SVC) was chosen as the final model after evaluation; its accuracy (ACC) was 0.91, and the area under the ROC curve (AUC) was 0.96. Among the factors involved in SVC, the Melton Ratio (*MR*) was the most important, followed by drainage density (*DD*), Hypsometric Integral (*HI*), and average slope (*AS*), indicating that geomorphic conditions play an important role in predicting debris flow susceptibility in the study area. SVC was used to map debris flow susceptibility in the study area, and the results will potentially facilitate the safe operation of the highway.

**Keywords:** debris flow; machine learning; susceptibility mapping; Karakoram Highway

## 1. Introduction

The China–Pakistan Karakoram Highway (KKH) is an important land route from China to South Asia and the Middle East via Pakistan [1,2]. Debris flows are one of the most serious geological hazards affecting the KKH [3]. A total of 150 debris flows occurred along the KKH from 2008 to 2011 and were investigated by Liao et al. [4]. They are rapid, surging flows of water-charged clastic sediments flowing in a steep channel [5]. For example, on 2 March 2015, a debris flow occurred in section 1607 of the KKH, on the border between Aktau county and Shufu county in Kashi Prefecture,

Xinjiang, trapping more than 300 vehicles and more than 600 people (according to People's Daily: <http://xj.people.com.cn/n/2015/0305/c188514-24068832.html>). On 6 July 2016, a debris flow occurred in Yecheng County in Xinjiang, killing 35 people and seriously damaging the road (according to People's Daily: <http://en.people.cn/n3/2016/0708/c90882-9083370.html>). On 11 July 2017, a mud rock flow occurred in Bulunkou Township, Aketao County, Xinjiang Autonomous Prefecture, burying several roads and more than 200 vehicles, and trapping more than 1000 persons (according to China Network: [http://www.china.com.cn/news/2017-07/12/content\\_41202196.htm](http://www.china.com.cn/news/2017-07/12/content_41202196.htm)).

Although debris flow susceptibility mapping can provide rapid support for regional debris flow prevention, few studies of debris flows have been conducted in the area. Yang et al. [3,6] evaluated the degree of activity of a glacial debris flow and the formation conditions. The associated risk of glacial debris flow disasters along the KKH was evaluated based on factors such as occurrence frequency, catchment area, volume of the alluvial fan, estimated outflow, vegetation coverage, slope, and altitude. Ali et al. [7] produced a landslide susceptibility map along the KKH using different conditioning and triggering factors for landslide occurrences, including lithology, seismicity, rainfall intensity, faults, elevation, slope angle, aspect, curvature, land cover, and hydrology. The results indicated that active faults, seismicity, and slope angle were the main controls on the spatial distribution of landslides. Overall, however, there is a lack of research on debris flow susceptibility mapping in the region.

Susceptibility mapping is an important method for disaster reduction of landslides and debris flows [8–10]. Debris flow susceptibility is the likelihood of a debris flow occurring in an area on the basis of local conditions [11]. Relevant models are divided into physical and statistical models. At present, methods based on physical models and GIS have been applied to evaluate the susceptibility of regional debris flow. For example, Carrara et al. [12] developed five models to predict location of debris flow source areas from a large set of environmental characteristics and a detailed inventory of debris flows; Bregoli et al. [13] introduced two physical models, which simulate the generation of debris flow through shallow landslide; Si et al. [14] assessed debris flow susceptibility using the Integrated Random Forest-Based Steady-State Infinite Slope method. Calista et al. [15] assessed the rockfall and debris flow hazard in the SW escarpment of the Montagna del Morrone Ridge. Carabella et al. [16] assessed the post wildfire landslide hazard of the 2017 Montagna Del Morrone fire.

Many statistical models are used in debris flow susceptibility assessment, such as analytic hierarchy processes, fuzzy logic, analytic hierarchy, network analysis, and information value [17–26]. Statistical models clearly have an important role in debris flow evaluation, but their application is limited because of the limited ability to obtain useful information from complex data and the high degree of human subjectivity [27–29].

As a result of the rapid development of AI technology, machine learning is increasingly being used in the evaluation of debris flow susceptibility, with good results, and this compensates for some of the shortcomings of traditional statistical models. Liang et al. (2012) [30] compared the effects of different machine learning models on debris flow hazard assessment. Chevalier et al. (2013) [31] assessed debris flow susceptibility in the central–eastern Pyrenees, using fluvio-morphological parameters and data mining. Addison et al. (2019) [32] assessed post wildfire debris flow occurrence using a classifier tree. Zhang et al. (2019) [28] mapped debris flow susceptibility based on different machine learning methods, using topography, vegetation, human activity, and soil factors. Dou et al. (2019) [33] proposed a method for the improved determination of controlling factors for debris flow susceptibility mapping, based on information fusion. Di et al. (2019) [29] assessed debris flow susceptibility in southwest China using a Gradient Boosting machine learning method. Xiong et al. (2020) [34] compared the application of different machine learning methods in debris flow susceptibility mapping using topographical, geological, edaphic, meteorological, land-cover, and sociometric factors. With the rapid increase in the availability of geographic data, machine learning technology can quickly learn and model a large amount of data, determine relatively important factors, and solve practical problems [30,35].

Therefore, machine learning methods play an increasingly important role in debris flow susceptibility assessment. However, it is necessary to explore whether a machine learning method is

suitable for the study area. Furthermore, the impact of different factors on the susceptibility of debris flow through a machine learning model needs to be analyzed, and the machine learning method needs lower capital and calculation costs compared with a physical model. Therefore, this paper chooses a machine learning method to map debris flow susceptibility along KKH.

The purpose of the present study was to develop a methodology for rapidly mapping debris flow susceptibility along the KKH, using remote sensing, GIS, and machine learning technology. First, the inventory of catchments which are prone to debris flows was obtained by visual interpretation of remote sensing images and field survey, and then, the factors influencing debris flows were collected and processed. Finally, debris flow susceptibility mapping was performed based on multiple machine learning methods.

## 2. Study Area

The study area is located in the northwest part of the Himalayan orogenic belt in northern Pakistan and northwest China, passing through the Karakoram, Himalayan, and Hindu Kush mountain ranges (Figure 1). The KKH starts from Kashgar in Xinjiang, China, passing through Khunjerab (at the border between China and Pakistan), the Himalayas, west of the Karakoram Mountains, and the northern edge of the Pamir Plateau, reaching Thakot in northern Pakistan. The total length exceeds 1000 km. From Khunjerab to the northern city of Thakot in Pakistan, the decrease in altitude is up to 4500 m, and the landform characteristics are unique, mainly mountains, valleys, glaciers, and ice margins.

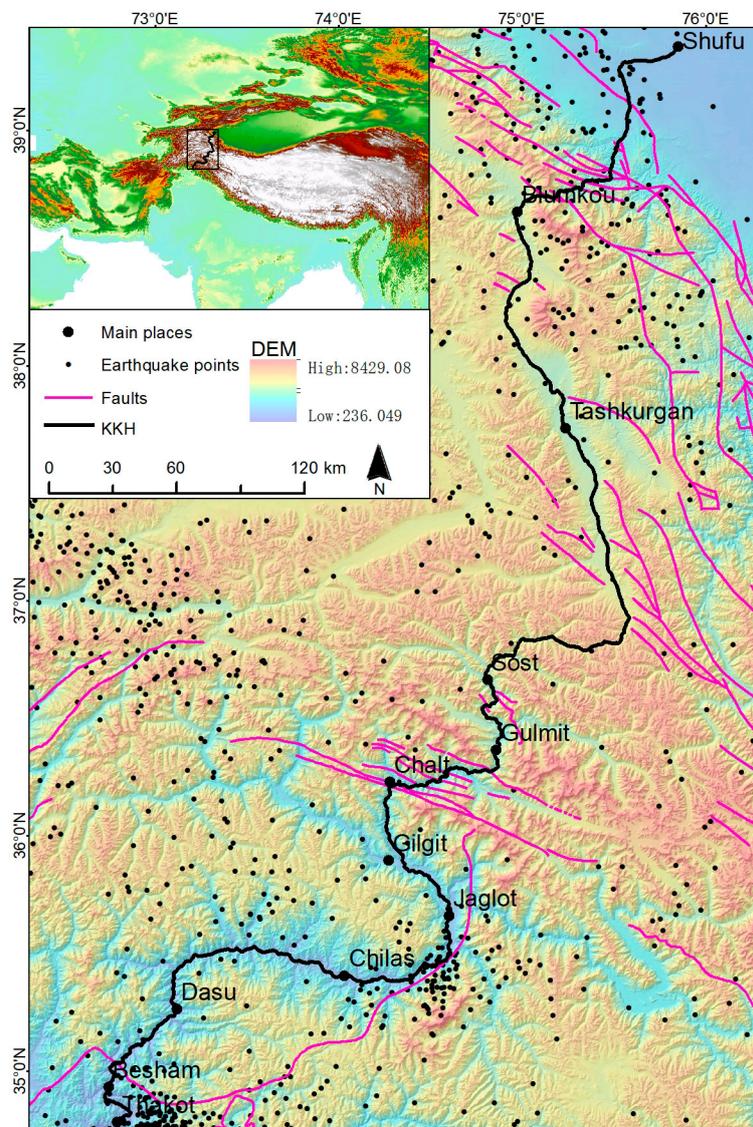
The KKH spans a wide area; the geological environment is complex. The sedimentary and metamorphic rocks exposed along the road mainly comprise Proterozoic, Paleozoic Carboniferous, Permian, Mesozoic Triassic, and Cretaceous strata. The lithologies are mainly mudstone, limestone, and slate, and there are large variations in lithology. Differential weathering and erosion lead to steep topography and widely distributed loose deposits [36], providing favorable terrain conditions and abundant materials for landslides.

The study region is located in the transitional zone between the monsoon zone of Central Asia and South Asia. In the northern part of the area, there is a high-elevation inland plateau climate, with low precipitation, strong solar radiation, low temperature, extensive glacier coverage, and pronounced freeze–thaw weathering. The southern foothills of the mountainous area are influenced by water vapor derived from the Indian Ocean, and the rainfall is greater, with the annual precipitation reaching 600–1000 mm [7]. The rainfall is concentrated in summer, and the high elevation mountainous terrain leads to pronounced small-scale climatic variability and extreme rainfall events. The climate along the highway has a pronounced vertical differentiation, with hot, dry valley, alpine, and glacial climates, with strong contrasts. The average annual rainfall in valley areas is less than 150 mm and the average annual rainfall of the high-altitude mountain areas, where the majority of the rainstorms occur, reaches 2000 mm.

This area is located in the active collision zone of the Indian and Asian plates [37]. The significant tectonic features are the main Karakoram thrust and Karakoram fault [38]. The main Karakoram thrust is the collision zone on the southern margin of the Eurasian plate and extends to Baltistan through the Hashupa, Shigar, and Shyok valleys, and is an active thrust fault with a large angle, along which many earthquakes occurred [39] (Figure 1).

As the geological environment around the highway is very hazardous, geological and geomorphological processes such as landslides, mudslides, collapses, and subsidence along the highway are frequent; they are major obstacles to the smooth operation of the highway, often blocking traffic and seriously affecting its normal operation [7]. Debris flows on both sides of the China–Pakistan highway are widely developed, and they can be divided into snowmelt debris flows and flood debris flows. The snowmelt debris flows are concentrated in the glacial areas of Kungai Mountain and Gongge Mountain, on both sides of the highway, with an altitude of more than 4000 m. The water source is mainly snowmelt and the loose material mainly consists of moraine and old debris flow deposits [6]. Flood type debris flows are widely distributed in the study area. Generally, the formation area is large,

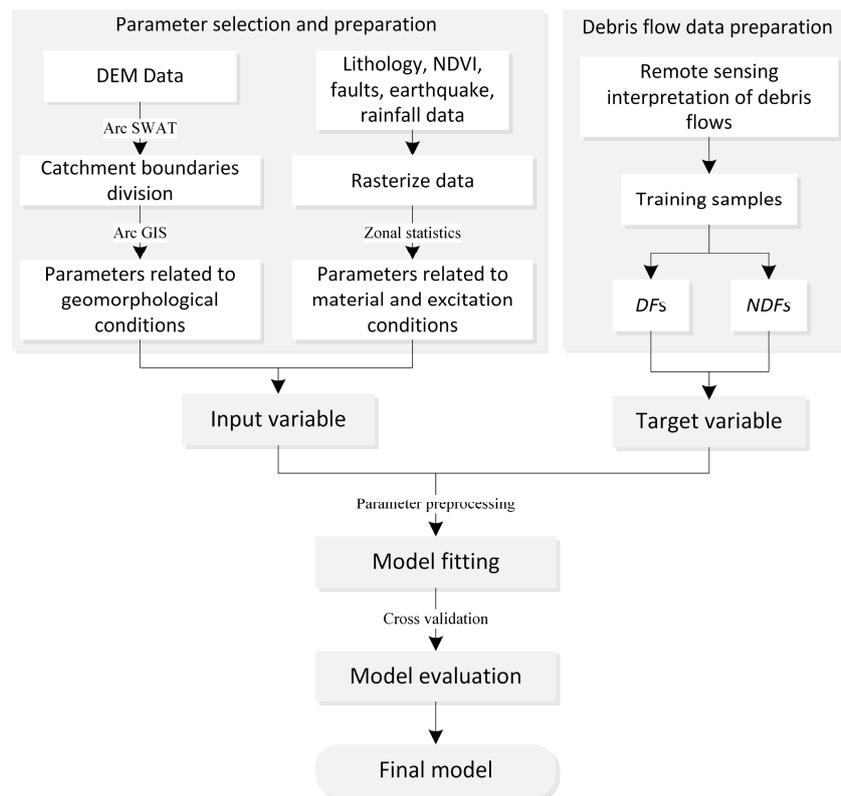
the slopes are steep, and there is a high density of gullies. These factors, combined with the rainstorms, result in the widespread occurrence of debris flows.



**Figure 1.** Location of the study area and distributions of faults, the KKH, earthquake sites, and the major settlements. (The Digital Elevation Model (DEM) data were provided by the International Scientific and Technical Data Mirror Site, Computer Network Information Center, Chinese Academy of Sciences, <http://www.gscloud.cn>; faults were provided by China Geological Survey; the earthquake sites were from earthquake database supplied by United States Geological Survey (USGS), and the sites with magnitude above 4 on the Richter scale since 2000 were selected, <https://earthquake.usgs.gov/earthquakes>).

### 3. Methods

The modeling process included several steps: parameter selection and preparation, data collection and processing, debris flow catchment preparation, model selection, model fitting, and model evaluation. A flow chart of the process is illustrated in Figure 2.



**Figure 2.** Modeling flow chart. (Note: DFs indicate debris flow catchments, and NDFs indicate non-debris flow catchments).

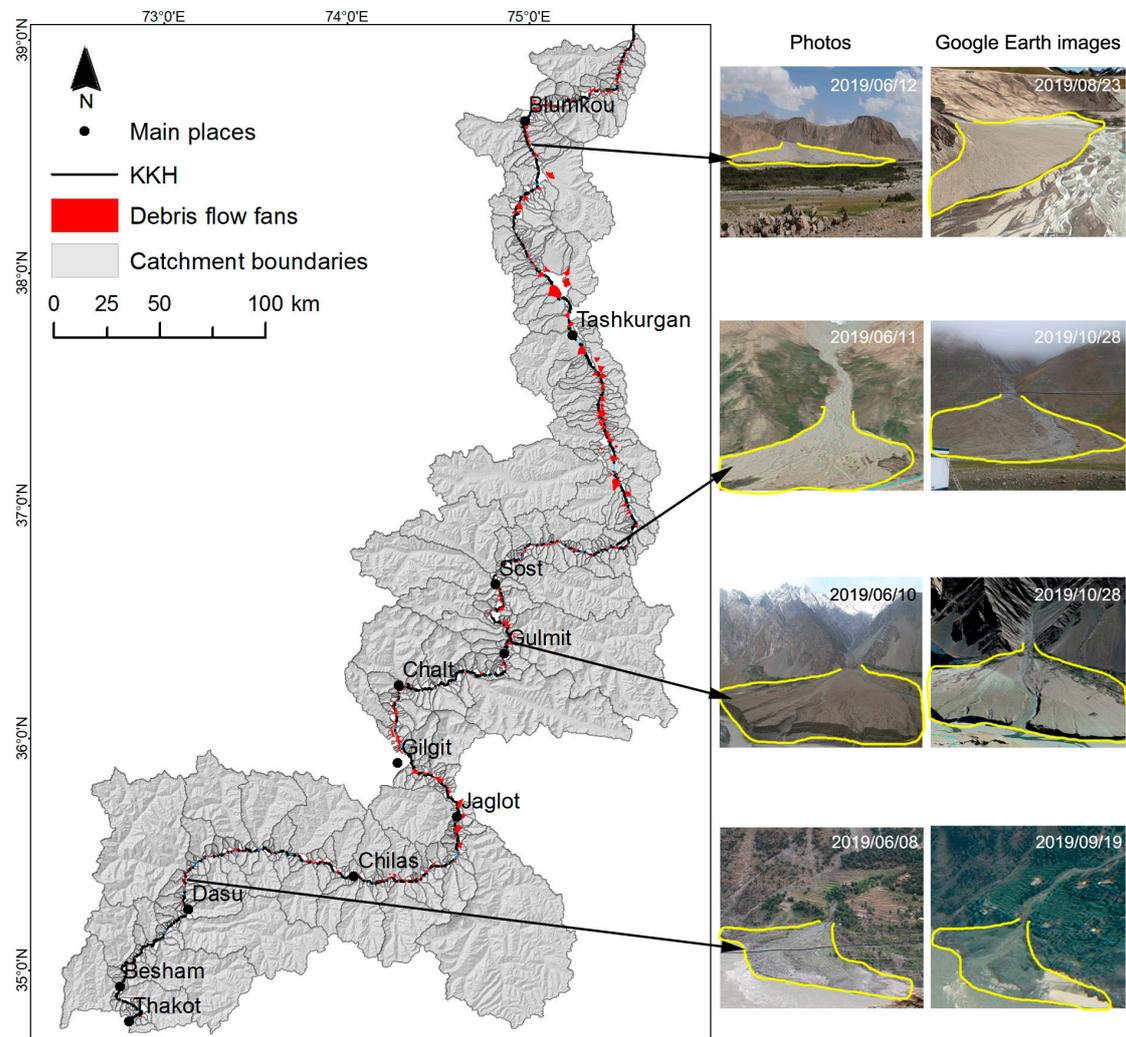
### 3.1. Catchment Boundaries Division

Catchments were selected as mapping units because of their significant conceptual and operational advantages [40,41]. Based on GIS and Soil & Water Assessment Tool (SWAT) tools, a DEM of the study area with a 60-m resolution (the DEM data were provided by the International Scientific and Technical Data Mirror Site, Computer Network Information Center, Chinese Academy of Sciences, <http://www.gscloud.cn>) was used for hydrological analysis and watershed boundary division. Taking the primary tributaries of the water system along the China–Pakistan highway as the division standard of the catchment boundary, 653 catchments were divided after being sorted manually, with the drainage area ranging from 0.74 to 3990 km<sup>2</sup> (Figure 3).

### 3.2. Inventory of Debris Flows

From 6 to 12 June 2019, our team carried out a survey on the disasters along the highway; a rough survey of the debris flow was carried out for safety reasons. However, field investigation can only identify the obvious debris flow fans, which may not be able to effectively identify the debris flow accumulation fan that has been damaged. Under the conditions of a large study area, few historical references, and difficult field conditions, remote sensing interpretation is a feasible and attractive approach [42]. Therefore, Google Earth images were used to avoid the time limit of field investigation. In order to minimize the problem of debris flow accumulation fans in some catchments being damaged and unrecognized, historical aerial images visible in Google Earth (the time is from 2005 to 2020) were interpreted; the ancient debris flow accumulation fans were not interpreted, and only the debris flow catchments that were still active in recent years were considered. According to Chevalier [31], criteria such as vegetation's change in the landscape, landslide scar(s), clear visibility of a torrent/gully/stream where roughness could be assessed, or the presence of potential deposition fans were considered.

The identification results are shown in Figure 3. A total of 544 catchments prone to debris flow (DFs) were identified in the study area, and the remaining 109 catchments were not prone to debris flow (NDFs).



**Figure 3.** Results of the identification results of debris flow accumulation fans along the KKH, and comparison of Google Earth images and field photos of typical debris flow fans. (The photo was taken by Xiaojun Su from the field survey).

### 3.3. Selection of Factors Influencing Debris Flows

The selection and processing of prediction factors have an important influence on the accuracy of debris flow susceptibility mapping [34]. The factors influencing debris flow susceptibility were divided into three categories: geomorphological, materials, and triggering. The specific selection parameters and their bases are as follows.

#### 3.3.1. Parameters Related to Geomorphological Conditions

The catchment area (CA) can reflect the sediment yield and concentration within the basin [17]. Channel length (CL) [17,43] reflects the potential length of debris flow movement and the mass of loose materials along its path. Catchment perimeter (CP) reflects the morphological characteristics of the catchment.

Average slope (AS) is the average value of the slope within the boundary of each catchment, which can reflect the overall slope of the catchment and slope stability. Catchment relief (CR) is the

height difference between the top and the outlet of the catchment [44]. This can reflect the overall potential energy conditions of the catchment. The relief ratio (*RR*) reflects the overall steepness of the channel, obtained by dividing *CR* by the *CL* [45]. *AS*, *CR*, and *RR* are important factors influencing debris flows, and they can provide sufficient potential energy conditions for the initiation and process of debris flow [44].

Cut density (*CD*) is the ratio of the channel length to *CA*. This reflects the geological and geomorphic conditions in the catchment, such as the weathering resistance and geomorphic development of rocks [46]. Drainage density (*DD*) is the ratio of *RR* to *CP*. Circularity ratio (*CR2*) is the ratio of *CP* to *CL* [47], reflecting the morphological characteristics of the catchment.

Hypsometric Integral (*HI*) reflects the slope distribution in the catchment [45]; it was calculated using the relief ratio method proposed by Pike and Wilson (1971) [48].

$$HI = (H_{\text{mean}} - H_{\text{min}})/(H_{\text{max}} - H_{\text{min}}) \quad (1)$$

where  $H_{\text{mean}}$ ,  $H_{\text{max}}$ , and  $H_{\text{min}}$  are the average, maximum, and minimum elevations in the catchment, respectively.

The Melton Ratio (*MR*) is the ratio of the *CR* to the square root of the *CA*. This reflects the dynamics and the susceptibility of the catchment to debris flows, and is widely used in the study of debris flows [16,43,49–51].

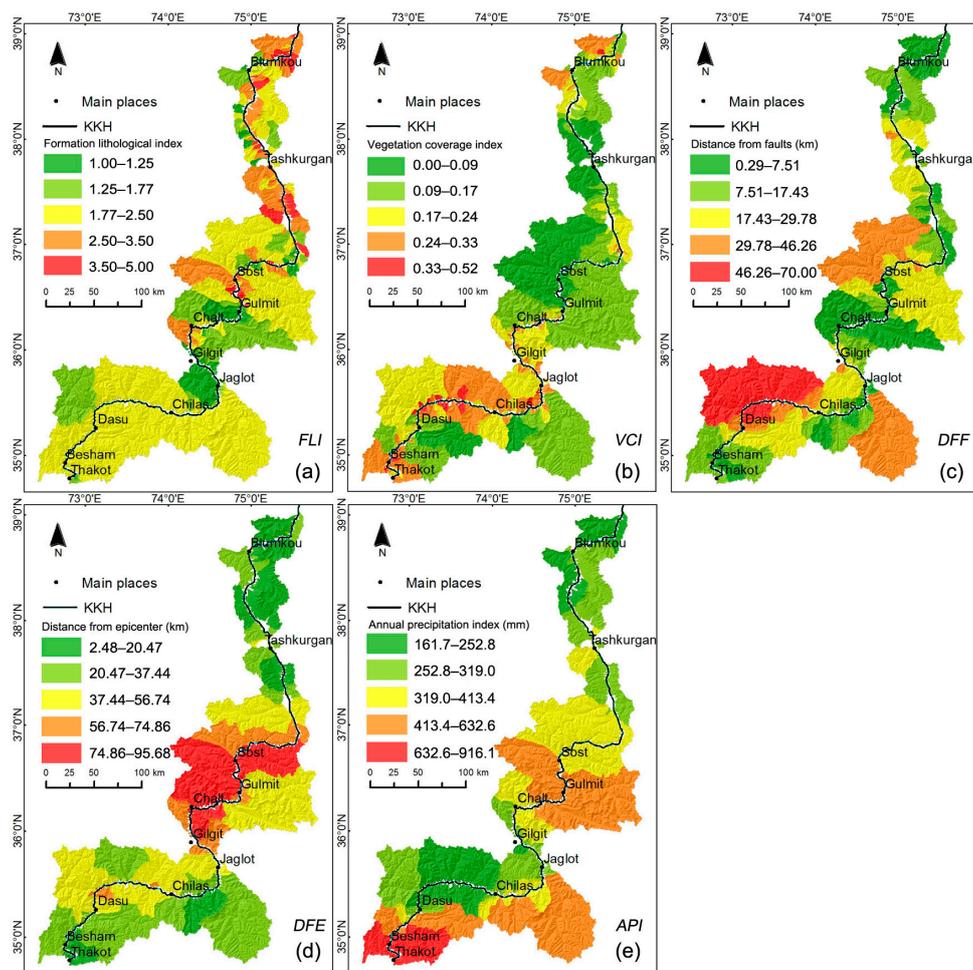
$$MR = CR / \sqrt{CA} \quad (2)$$

### 3.3.2. Parameters Related to Soil Material and Geology Conditions

Catchment lithology determines the material background conditions for the formation of debris flows. As lithological data cannot be assigned to the catchment unit, in order to simplify the calculation, we used the solution proposed by Zhao et al. [35], which involves dividing the stratigraphic lithological data (vector maps published in December 2002, and supplied by China Geological Survey, 1:500,000) into five levels: very hard, hard, medium, soft, and very soft, according to the rock strength, and assigning values of 1–5, respectively (Table 1). The vector data of strata were then transformed into grid data using a resolution of  $30 \times 30$  m, so that each grid has an independent lithological strength value. Through the statistical analysis of each catchment, the average lithological strength value of each catchment is obtained as the formation lithology index (*FLI*) of that catchment (Figure 4a). The higher the *FLI* value, the softer the overall lithology of the catchment, and the greater the quantity of loose materials that can be provided.

**Table 1.** Lithology composition and classification of the study area.

Major Lithology	Relative Strength	Rock Strength
Granite, diorite, granodiorite, quartz diorite, gabbro	Very hard	1
Schist, gneiss, marble, quartzite, slate	Hard	2
Mudstone, sandstone, sandy mudstone, argillaceous siltstone, clastic rock with siltstone	Medium	3
Limestone, phyllite, shale	Soft	4
Alluvial, proluvial, lacustrine, marine, fluvial and fluvial sediments, glaciers deposits	Very soft	5



**Figure 4.** Distribution of parameters of FLI (a), VCI (b), DFF (c), DFE (d) and API (e).

Vegetation coverage is one of the most important parameters for evaluating DFs [28]. NDVI data cannot be assigned according to the catchment unit, and the average NDVI value of each catchment was used as the vegetation coverage index (VCI) of that catchment (Figure 4b). The higher the VCI, the greater the vegetation coverage. NDVI was derived from Landsat-8 images (June 2017) with a resolution of 30-m (Landsat-8 image courtesy of the U.S. Geological Survey).

Faults trigger earthquakes, producing discontinuities in rocks, and as a consequence, they provide rock debris that can be mobilized [24]. Therefore, the distance (km) from fault (DFF) is an important parameter. Based on fault distribution data (vector maps published in December 2002, and supplied by China Geological Survey), the average distance between each catchment and fault was calculated (Figure 4c). The lower the DFFs, the greater the influence of the fault and the higher the degree of fracture of the adjacent rock mass.

Chen et al. [52] found that debris flow occurrence was closely related to the impact of earthquakes and droughts, because they can increase the amount of loose materials. Wu et al. [53] evaluated the debris flow susceptibility of the Longxi River in the Wenchuan earthquake area. They summed the volumes of the source debris accumulations to reflect the impact of earthquakes on the debris flow susceptibility. Therefore, the average distance (km) from previous earthquake epicenters (DFE) is an important parameter. Based on the earthquake distribution data (the earthquake database was provided by USGS, and the sites with magnitude above 4 on the Richter scale since 2000 were selected, <https://earthquake.usgs.gov/earthquakes>), the average distance from the epicenter to each catchment was calculated (Figure 4d). The lower the DFE, the greater the earthquake influence, and the greater the slope instability.

### 3.3.3. Parameters Related to Triggering Conditions

Rainfall is the major factor triggering debris flow hazards [30]. In this study, the average annual rainfall of each catchment was calculated as the annual precipitation index (*API*) (Figure 4e). Annual precipitation was derived from TRMM\_3B43 images with a resolution of 0.25° (Tropical Rainfall Measuring Mission (TRMM) (2011), TRMM (TMPA/3B43) Rainfall Estimate L3 1 month 0.25° × 0.25° V7, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: 6 April 2018, <https://doi.org/10.5067/TRMM/TMPA/MONTH/7>). It provides monthly precipitation data in mm/hr. The data product represents the average level of monthly precipitation. The monthly precipitation in 2017 is calculated and accumulated to obtain the annual precipitation in the study area (mm).

In summary, a spatial database of catchments in the study area was established (Table 2).

**Table 2.** Fields and characteristics of the spatial database.

		Fields	Parameters	Unit	Type	
1		ID	Identification field	/	Short	
2		DF	Debris flow catchment or not	/	1/0	
3	Parameters related to geomorphological conditions	CA	Catchment area	km <sup>2</sup>	Double	
4		CL	Channel length	km	Double	
5		CR	Catchment relief	km	Double	
6		RR	Relief ratio	‰	Double	
7		AS	Average slope	°	Double	
8		CP	Catchment perimeter	km	Double	
9		CD	Cut density	/	Double	
10		DD	Drainage density	/	Double	
11		CR2	Circularity ratio	/	Double	
12		HI	Hypsometric Integral	/	Double	
13		MR	Melton ratio	/	Double	
14		Parameters related to material conditions	VCI	Vegetation coverage index	/	Double
15			FLI	Formation lithological index	/	Double
16	DFD		Distance from faults	km	Double	
17	DFE		Distance from epicenter	km	Double	
18	Parameters related to triggering conditions	API	Annual precipitation index	mm/year	Double	

### 3.4. Parameter Preprocessing

#### 3.4.1. Collinearity Processing

For some models, strongly correlated parameters contain some degree of redundancy, and substantial collinearity will lead to model instability. In this study, a parameter correlation matrix heat map (Figure 5) was calculated based on the Seaborn Python visualization package. It was found that there were several pairs of variables which were highly correlated; for example, the following correlation coefficients were obtained: 0.92 for *CA* vs. *CP*; 0.88 for *CL* vs. *CA*; 0.99 for *CL* vs. *CP*; 0.92 for *RR* vs. *DD*; and 0.91 for *RR* and *MR*. Therefore, *CL*, *CP*, and *RR* were eliminated.

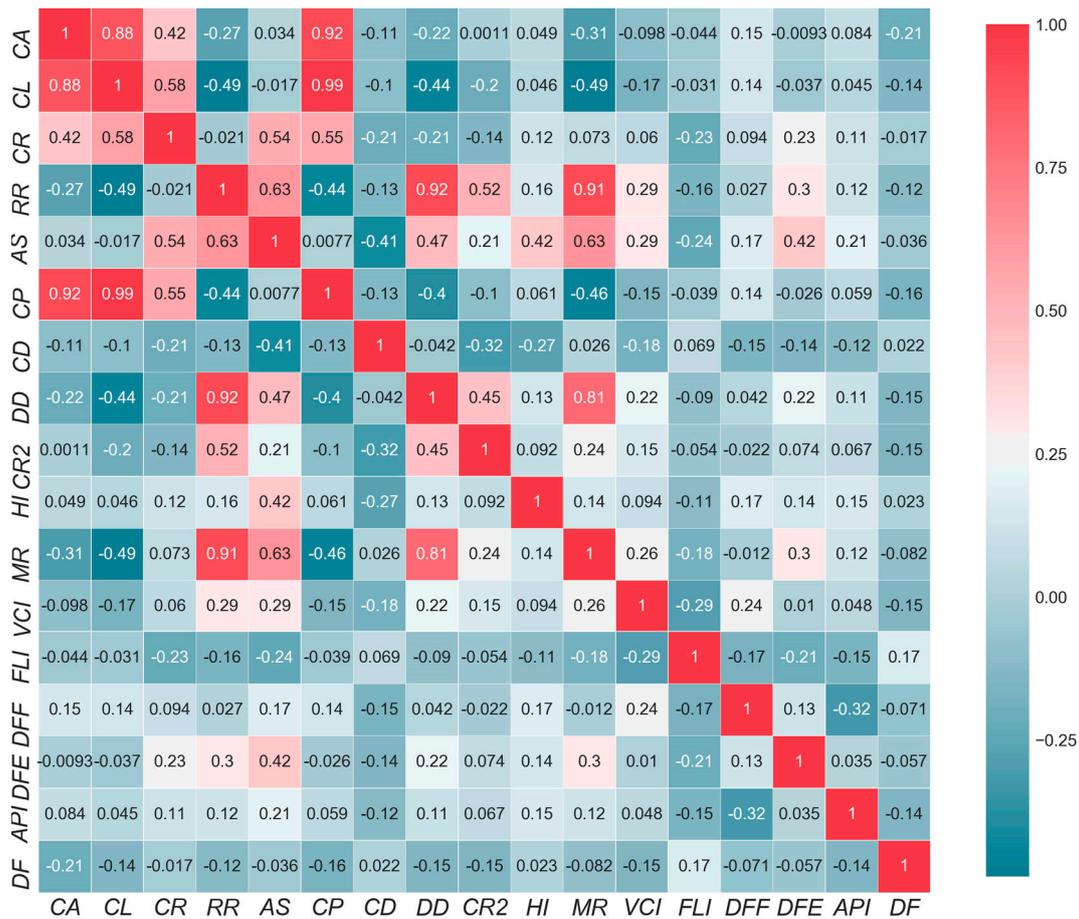


Figure 5. Heat map of the parameter correlation matrix.

### 3.4.2. Resampling

For the sample data collected in this study, the ratio of DFs to NDFs is 5:1 (Figure 6). This data imbalance will cause the learning machine to pay more attention to the classifications with DFs and hence, obtain an inaccurate model. In this study, SMOTE (the synthetic mineral oversampling technique) [54] was used to increase the number of NDF samples. 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 sample size of NDFs was increased to 544, and the ratio of NDFs to DFs was 1:1.

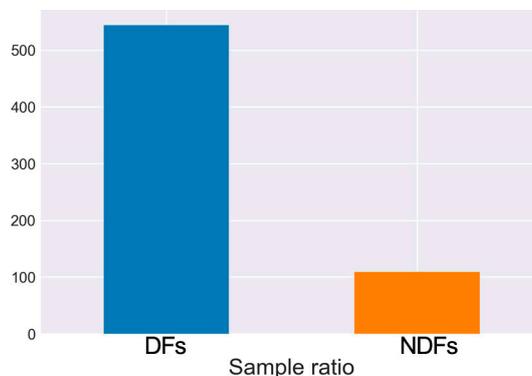


Figure 6. Sample ratio for DFs and NDFs.

### 3.4.3. Data Standardization

Data standardization can speed up model convergence and improve model accuracy. Moreover, some machines are very sensitive to feature scales; therefore, in this study, we used a standard scaler algorithm (from Scikit-learn, <https://scikit-learn.org>) to standardize the features by removing the mean and scaling to unit variance. Scikit-learn is a Python library that provides a standard interface for implementing machine learning algorithms [55].

### 3.4.4. Generating a Cross-Validation Dataset

Using the cross-validation algorithm in Scikit-learn, 70% of the training data were randomly selected for model training, and the remaining 30% were used as test data to evaluate the model; this was repeated 10 times (Figure 7). In this way, the algorithm can use different training data subsets to build the model, and test data are used to evaluate model performance of the model and prevent the model from being over fitted.

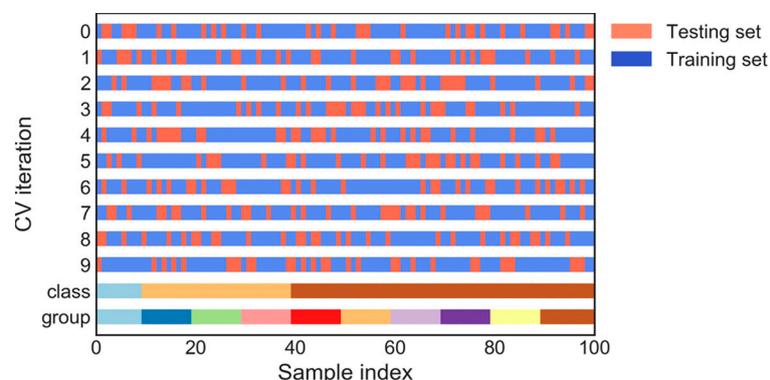


Figure 7. Generation of the cross-validation dataset.

## 3.5. Candidate Machine Selection

In order to find the most appropriate model to solve the research problem, we selected a variety of widely used machine learning methods. Typical algorithms were selected of each type and 22 model methods were finally selected.

### 3.5.1. Ensemble Methods

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

### 3.5.2. Gaussian Processes (GP)

Gaussian Processes are a generic supervised learning method designed to solve regression and probabilistic classification problems.

### 3.5.3. Generalized Linear Models (GLM)

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 connection function. Logistic Regression (LR), Passive Aggressive, Ridge, Stochastic Gradient Descent (SGD), and Perceptron were selected in this study.

#### 3.5.4. Naive Bayes (NB)

Naive Bayes classification is based on the concept of Bayes 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.

#### 3.5.5. 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.

#### 3.5.6. Support Vector Machines (SVM)

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

#### 3.5.7. Trees

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

#### 3.5.8. Discriminant Analysis

Discriminant Analysis is a method of multivariate statistical analysis which classifies the studied objects according to several observed indexes. Linear Discriminant and Quadratic Discriminant were selected.

#### 3.5.9. eXtreme Gradient Boosting (XGBoost)

XGBoost is a boosting algorithm, and it is 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.

### 3.6. Model Fitting and Tuning

The training data in the cross-validation dataset were used to train the initial model. The models were then sorted according to the average accuracy score (ACC) of the test data in the cross-validation dataset (Figure 8). The ACC (Formula 3) indicates the rate of correct assignment of all samples participating in the modeling (Table 3). It is evident from Figure 8 that the overall fitting effect of the integrated model is better than that of the other models. The highest score was achieved by Extra Trees, followed by Random Forest, Gaussian Process, Gradient Boosting, and XGBoost.

$$ACC = (TP + TN)/(TP + FN + FP + TN) \quad (3)$$

Among them (Table 3):

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

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

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

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

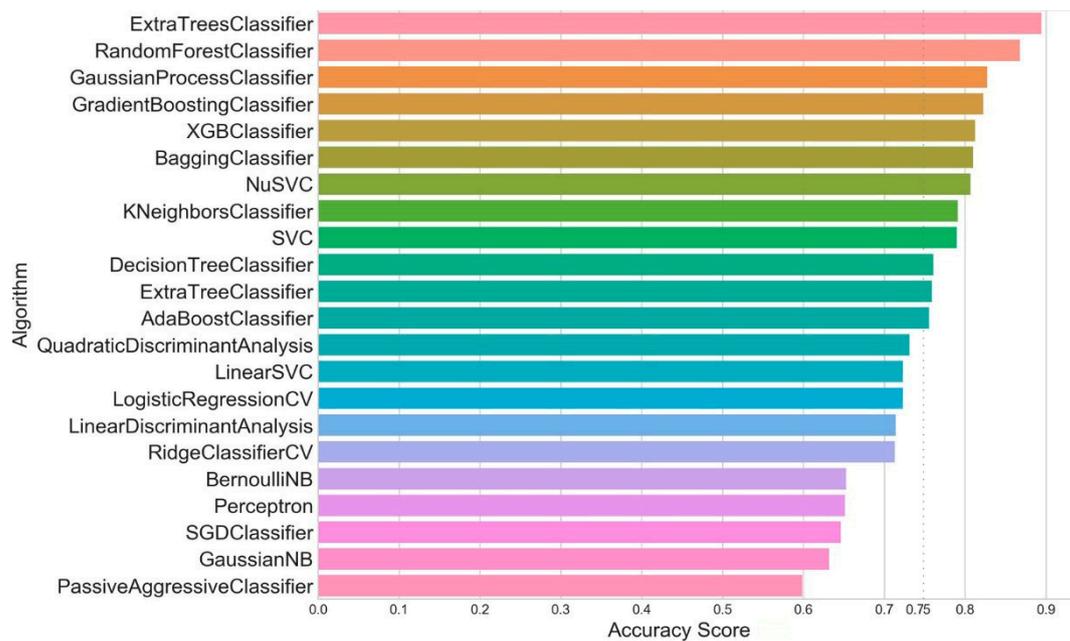


Figure 8. Ranking of model accuracy scores.

Table 3. Confusion matrix.

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

The top 12 models (ACC > 0.75) were selected for model optimization. Parameter grid and grid search cross-validation were used to fit the model, and the AUC (area under the receiver operating characteristic curve) score was used to search for the best super parameter. According to the optimal super parameters of each model given in Table 4, 10-fold cross-validation was again performed on the model training set, and the models were reordered according to the average accuracy score of the test data (Figure 9). It can be seen that the overall accuracy of the models had been improved. After model optimization, the SVC model became the optimal model; its test data ACC is 0.91, and the average AUC of 10-fold cross-validation is 0.96 (Figure 10). The AUC indicates the tradeoff between sensitivity and specificity [56]. From the perspective of the time needed to find the optimal parameters, the most time consuming model was Gradient Boosting (112,533.93 s), the least time consuming was Extra Tree (1.53 s), and the time consumption of SVC was 57.79 s. Therefore, SVC provided high efficiency combined with the highest accuracy.

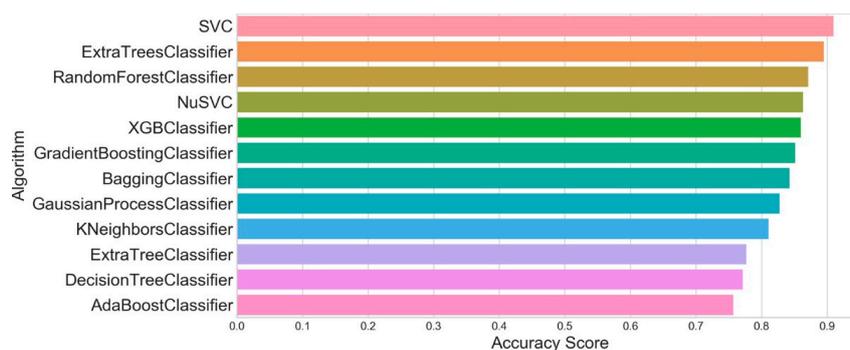
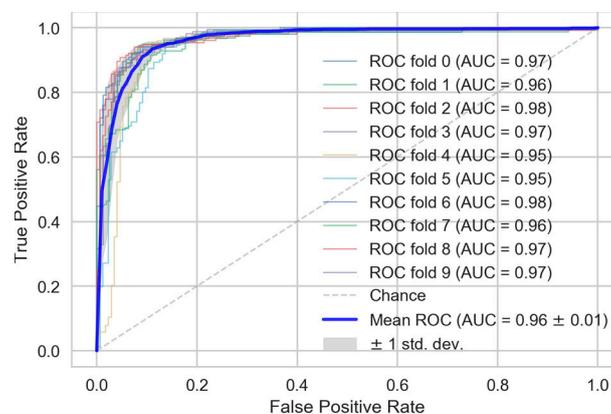


Figure 9. Ranking of ACC scores after model optimization.

**Table 4.** 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>).

	Classifier Algorithm	Best Parameter	Runtime (s)
1	AdaBoostClassifier	'algorithm' = 'SAMME.R' 'learning_rate' = 0.25 'n_estimators' = 300	189.67
2	BaggingClassifier	'max_samples' = 1.0 'n_estimators' = 300	131.67
3	ExtraTreesClassifier	'criterion' = 'entropy' 'max_depth' = 30 'n_estimators' = 150	603.41
4	GradientBoostingClassifier	'criterion' = 'friedman_mse' 'learning_rate' = 0.1 'max_depth' = 6 'n_estimators' = 300	112,533.93
5	RandomForestClassifier	'criterion' = 'gini' 'max_depth' = 18 'n_estimators' = 250 'oob_score' = True	1953.35
6	GaussianProcessClassifier	'max_iter_predict' = 10	23.94
7	KNeighborsClassifier	'algorithm' = 'auto' 'n_neighbors' = 5 'weights' = 'distance'	6.90
8	SVC	'C' = 4 'decision_function_shape' = 'ovo' 'gamma' = 0.5 'probability' = True	57.79
9	NuSVC	'nu': 0.3	3.69
10	DecisionTreeClassifier	'criterion' = 'entropy' 'max_depth' = 10	4.54
11	ExtraTreeClassifier	'criterion' = 'gini' 'max_depth' = 10	1.53
12	XGBClassifier	'learning_rate' = 0.1 'max_depth' = 6 'n_estimators' = 300	368.50

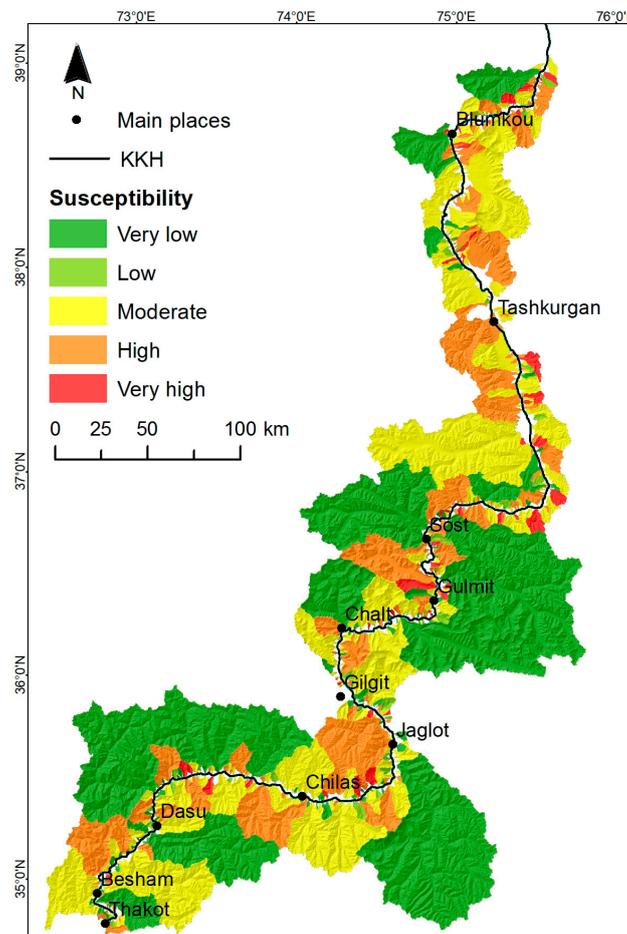


**Figure 10.** Receiver Operating Characteristic Curve (ROC) and AUC of SVC using 10-fold cross-validation.

#### 4. Results

The optimal model SVC (Super parameter: 'C' = 4, 'decision\_function\_shape' = 'ovo', 'gamma' = 0.5, 'probability' = True) was selected as the final model to map the landslide susceptibility of the study area. The SVC can only output classification results (i.e., 0/1), so the model cannot directly output the probability; however, the probability value is needed to map debris flow susceptibility. We used the method provided by Scikit-learn to set the probability to true when constructing the SVC function option, so that the model could output the classification probability (0–1). In the binary case, the probabilities are calibrated using Platt scaling (Platt "Probabilistic outputs for SVMs and comparisons to regularized likelihood methods"): logistic regression of the SVM's scores and fitting by an additional cross-validation of the training data [57].

The probability value of debris flow susceptibility was divided into five categories using the natural fracture method [58]: very low, low, moderate, high, and very high. The susceptibility mapping results (Figure 11) show that debris flow susceptibility in large catchments (>100 km<sup>2</sup>) is generally low, which may be because the fluvial processes take over the debris flow processes in these catchments. The areas of high and very high debris flow susceptibility are evenly distributed in the study area, and there is no obvious spatial distribution pattern. The proportion of debris flow susceptibility ranges from low to high, with values of 16.7%, 7.2%, 41.5%, 26.5%, and 8.1%, respectively. The areas with moderate susceptibility account for the highest proportion, while the areas of low susceptibility and very high susceptibility account for the lowest proportion. In order to help in understanding the magnitude of the problem as well as in early planning of the tasks, the lengths of the part of highway under low to high susceptibility were counted according to the fans interpreted in Google Earth, with lengths of 27.97, 22.51, 135.46, 58.78, and 22.85 km, respectively.



**Figure 11.** Susceptibility mapping results of the KKH area.

## 5. Discussion

The interpretability of the model can help determine the potential relationship between the influencing factors and debris flow susceptibility from the data, which is very important for understanding the factors influencing debris flow susceptibility in the study area. The SVC model is a “black box” process which cannot directly indicate the importance of the variables employed. Here, the importance of the variables was calculated using the Permutation Importance algorithm provided by the extension module of ELI5 (ELI5 is a Python library, which allows the visualization and debugging of various Machine Learning models using unified API. It has built-in support for several ML frameworks and provides a way to explain black-box models, <https://eli5.readthedocs.io/en/latest/index.html>). This determines the importance of parameters by measuring how to reduce the score when a feature is not available. This method is also known as “rank importance” or “mean decreased accuracy (MDA)” [59].

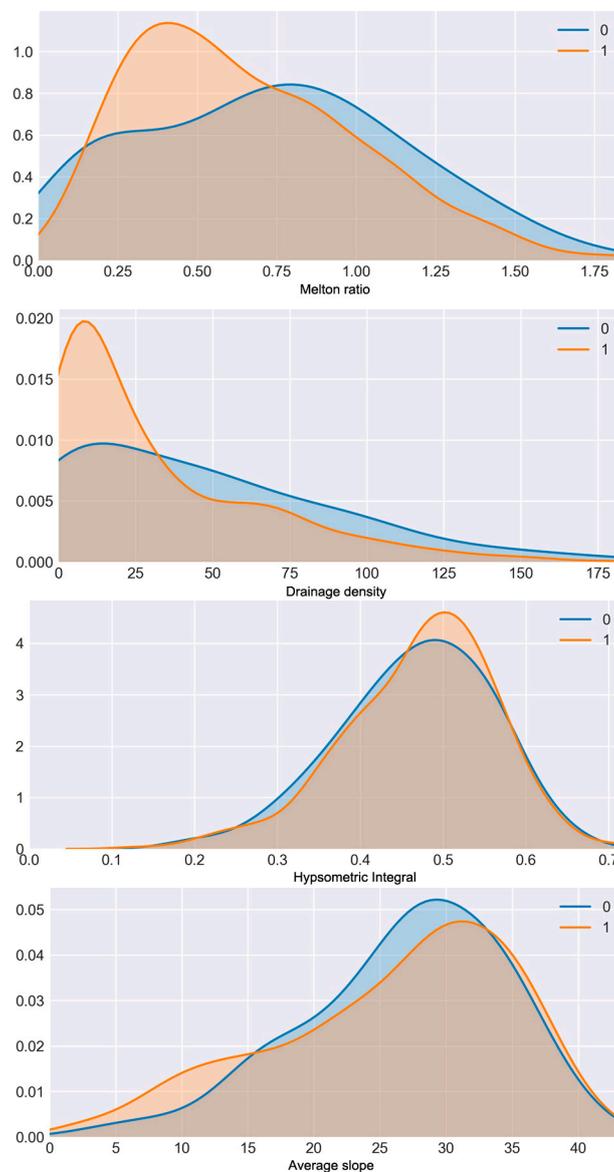
From the importance of the calculation results (Table 5), it can be seen that all of the parameters involved in the SVC have some degree of importance. Among them, *MR* is the most important. *MR* is a parameter proposed by Melton (1965) [60] to reflect the dynamic characteristics of the catchment and the debris flow susceptibility; it is widely used in the evaluation and classification of debris flows [43,50,51,61]. This is followed by *DD*, *HI*, and *AS*, which are related to the geomorphic conditions. This shows that the parameters related to these factors play an important role in the prediction of debris flow susceptibility in the study area. In the study of Xiong et al. [34], several geomorphological parameters (mean altitude, altitude difference, and groove gradient) were also the most important factors for the prediction of debris flow susceptibility. In the study of Zhang et al. [28], aspect was the most important factor, followed by rainfall, elevation, and slope curvature. In addition, there are three parameters related to material conditions: *DFE*, *DFE*, and *VCI*, followed by other factors. Among the factors related to soil material and geology conditions, the most important factor is the *DFE*, followed by *DFE* and *VCI*. Due to the lack of detailed data of snow cover and glacier melting, the annual rainfall was assumed to be the main trigger factor of debris flow susceptibility. From the parameter importance ranking (Table 5), we can see that the importance of *API* is not as high as we assumed. This may be due to the lack of data of glaciers, snow, and frozen soil. This is a defect of this paper that we hope to improve in future work, in order to obtain a better debris flow susceptibility map.

**Table 5.** Calculated importance of the features.

Features	Weight
<i>MR</i>	0.1483 ± 0.0156
<i>DD</i>	0.1414 ± 0.0194
<i>HI</i>	0.1256 ± 0.0210
<i>AS</i>	0.1239 ± 0.0196
<i>DFE</i>	0.1131 ± 0.0223
<i>DFE</i>	0.1123 ± 0.0196
<i>VCI</i>	0.1121 ± 0.0272
<i>CR2</i>	0.1110 ± 0.0152
<i>CR</i>	0.1077 ± 0.0150
<i>API</i>	0.1015 ± 0.0133
<i>FLI</i>	0.0901 ± 0.0213
<i>CD</i>	0.0866 ± 0.0054
<i>CA</i>	0.0748 ± 0.0080

The distributions of *MR*, *DD*, *HI*, and *AS* are shown in Figure 12. It can be seen that the proportion of DFs is higher when  $0.15 < MR < 0.74$ ,  $DD < 32$ ,  $0.45 < HI < 0.58$ , and  $AS < 15$ ,  $AS > 33$ , which indicates that a debris flow is more likely to occur in these areas. Among these four parameters, *MR* has the strongest ability to distinguish DF and NDF. It can be seen from the figure that the differences in the peak distribution of DF and NDF in *MR* are obvious, but that this is not the case for the difference

in the other three parameters. This indicates that *MR* is an important factor affecting debris flow susceptibility in the study area.



**Figure 12.** Distribution of *MR*, *DD*, *HI*, and *AS*. “1” indicates catchments prone to debris flow, and “0” indicates catchments not prone to debris flow.

## 6. Conclusions

We have applied a variety of machine learning methods to the mapping of debris flow susceptibility of the Karakoram Highway. Good prediction results were obtained. Comparison of super parameter optimization and model prediction accuracy indicated that the Support Vector Classification (SVC) model had the best performance, with an average test data accuracy of 0.91 and an average AUC of 10-fold cross-validation of 0.96. The search time for the optimal parameters of the model was 57.79 s, which shows that it combined high efficiency with the highest prediction accuracy. Hence, it provides a valuable tool for mapping debris flow susceptibility. *MR* was the most important predictor in the final model, followed by *DD*, *HI*, *AS*, *DFE*, *DFI*, *VCI*, and other factors. This shows that geomorphic conditions play an important role in the prediction of debris flow susceptibility. Among the factors related to soil material and geology conditions, the most important factor is the *DFE*, followed by *DFI*

and *VCI*. The importance of *API* is not as high as we assumed, which may be due to the lack of data of glaciers, snow, and frozen soil.

The susceptibility mapping results show that debris flow susceptibility in large catchments is generally low, which may be because the fluvial processes take over the debris flow processes in these catchments. The areas of high and very high debris flow susceptibility are evenly distributed within the study area, and there is no obvious spatial distribution pattern. The area of moderate debris flow susceptibility accounts for the highest proportion, while the areas of low and very high susceptibility account for the lowest proportion. Overall, we suggest that our approach can provide valuable inputs to policies, promoting the safe operation of highways.

**Author Contributions:** F.Q. and Y.Z. designed this study, performed the main analysis, and wrote the paper. X.M. directed and revised the manuscript. X.S. and T.Q. contributed to the data preparation. D.Y. revised and polished the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Key Research and Development Program of China (Grant Nos. 2017YFC1501005, 2018YFC1504704); Major Scientific and Technological Projects of Gansu Province (No. 19ZD2FA002); National Natural Science Foundation of China (No. 41661144046); Program for International S&T Cooperation Projects of Gansu Province (No. 2018-0204-GJC-0043); and the Fundamental Research Funds for the Central Universities (Nos. lzujbky-2018-k14, lzujbky-2017-it92, lzujbky-2020-sp03).

**Acknowledgments:** The DEM data were provided by the International Scientific and Technical Data Mirror Site, Computer Network Information Center, Chinese Academy of Sciences.

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

## Abbreviations

Names	Abbreviations
China–Pakistan Karakoram Highway	KKH
Catchments prone to debris flow	DFs
Catchments not prone to debris flow	NDFs
Catchment area	CA
Channel length	CL
Catchment perimeter	CP
Average slope	AS
Catchment relief	CR
Relief ratio	RR
Cut density	CD
Drainage density	DD
Circularity ratio	CR2
Hypsometric Integral	HI
Melton ratio	MR
Vegetation coverage index	VCI
Formation lithological index	FLI
Distance from faults	DF
Distance from epicenter	DFE
Annual precipitation index	API
Gradient Tree Boosting	GDBT
Gaussian Processes	GP
Generalized Linear models	GLM
Logistic Regression	LR
Stochastic Gradient Descent	SGD
Naive Bayes	NB
Support Vector Machines	SVM
Support Vector Classification	SVC

Names	Abbreviations
eXtreme Gradient Boosting	XGBoost
Average accuracy score	ACC
True positive	TP
False positive	FP
True negative	TN
False negative	FN
Area under the receiver operating characteristic curve	AUC
Receiver Operating Characteristic Curve	ROC
Mean decreased accuracy	MDA

## References

- Zhao, F.; Meng, X.; Zhang, Y.; Chen, G.; Su, X.; Yue, D. Landslide susceptibility mapping of karakorum highway combined with the application of SBAS-InSAR technology. *Sensors* **2019**, *19*, 2685. [[CrossRef](#)] [[PubMed](#)]
- Rehman, M.U.; Zhang, Y.; Meng, X.; Su, X.; Catani, F.; Rehman, G.; Yue, D.; Khalid, Z.; Ahmad, S.; Ahmad, I. Analysis of landslide movements using interferometric synthetic aperture radar: A case study in Hunza-Nagar Valley, Pakistan. *Remote Sens.* **2020**, *12*, 2054. [[CrossRef](#)]
- Yang, Z.; Zhu, Y.; Zou, D.H.S.; Liao, L. Activity degree evaluation of glacial debris flow along international Karakorum Highway (KKH) based on fuzzy theory. *Adv. Mater. Res.* **2011**, *261–263*, 1167–1171. [[CrossRef](#)]
- Liao, L.; Zhu, Y.; Steve Zou, D.H.; Yang, Z.; Muhammad, W.; Chen, J.; Wang, Y.; Ye, C. Key point of bridge damage caused by glacial debris flows along international karakorum highway, Pakistan. *Appl. Mech. Mater.* **2013**, *256–259*, 2713–2723. [[CrossRef](#)]
- Jomelli, V.; Pavlova, I.; Eckert, N.; Grancher, D.; Brunstein, D. A new hierarchical Bayesian approach to analyse environmental and climatic influences on debris flow occurrence. *Geomorphology* **2015**, *250*, 407–421. [[CrossRef](#)]
- Yang, Z.; Zhu, Y.; Zou, D.H.S. Formation conditions and risk evaluation of glacial debris flow disasters along International Karakorum Highway (KKH). In Proceedings of the 5th International Conference on Debris-Flow Hazards Mitigation: Mechanics, Prediction and Assessment, Padua, Italy, 14–17 June 2011; pp. 1031–1037. [[CrossRef](#)]
- Ali, S.; Biermanns, P.; Haider, R.; Reicherter, K. Landslide susceptibility mapping by using a geographic information system (GIS) along the China-Pakistan Economic Corridor (Karakoram Highway), Pakistan. *Nat. Hazards Earth Syst. Sci.* **2019**, *19*, 49–71. [[CrossRef](#)]
- Marsala, V.; Galli, A.; Paglia, G.; Miccadei, E. Landslide susceptibility assessment of Mauritius Island (Indian ocean). *Geosciences* **2019**, *9*, 493. [[CrossRef](#)]
- Fell, R.; Corominas, J.; Bonnard, C.; Cascini, L.; Leroi, E.; Savage, W.Z. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Eng. Geol.* **2008**, *102*, 85–98. [[CrossRef](#)]
- Rahman, M.S.; Ahmed, B.; Di, L. Landslide initiation and runoff susceptibility modeling in the context of hill cutting and rapid urbanization: A combined approach of weights of evidence and spatial multi-criteria. *J. Mt. Sci.* **2017**, *14*, 1919–1937. [[CrossRef](#)]
- Bertrand, M.; Liébault, F.; Piégay, H. Debris-flow susceptibility of upland catchments. *Nat. Hazards* **2013**, *67*, 497–511. [[CrossRef](#)]
- Carrara, A.; Crosta, G.; Frattini, P. Comparing models of debris-flow susceptibility in the alpine environment. *Geomorphology* **2008**, *94*, 353–378. [[CrossRef](#)]
- Bregoli, F.; Medina, V.; Chevalier, G.; Hürlimann, M.; Bateman, A. Debris-flow susceptibility assessment at regional scale: Validation on an alpine environment. *Landslides* **2015**, *12*, 437–454. [[CrossRef](#)]
- Si, A.; Zhang, J.; Zhang, Y.; Kazuva, E.; Dong, Z.; Bao, Y.; Rong, G. Debris flow susceptibility assessment using the integrated random forest based steady-state infinite slope method: A case study in Changbai Mountain, China. *Water* **2020**, *12*, 2057. [[CrossRef](#)]
- Calista, M.; Menna, V.; Mancinelli, V.; Sciarra, N.; Miccadei, E. Rockfall and debris flow hazard assessment in the SW escarpment of montagna del morrone ridge (Abruzzo, Central Italy). *Water* **2020**, *12*, 1206. [[CrossRef](#)]

16. Carabella, C.; Miccadei, E.; Paglia, G.; Sciarra, N. Post-wildfire landslide hazard assessment: The case of the 2017 montagna del morrone fire (central apennines, Italy). *Geosciences* **2019**, *9*, 175. [[CrossRef](#)]
17. Chang, T.C.; Chao, R.J. Application of back-propagation networks in debris flow prediction. *Eng. Geol.* **2006**, *85*, 270–280. [[CrossRef](#)]
18. Greco, R.; Sorriso-Valvo, M.; Catalano, E. Logistic Regression analysis in the evaluation of mass movements susceptibility: The Aspromonte case study, Calabria, Italy. *Eng. Geol.* **2007**, *89*, 47–66. [[CrossRef](#)]
19. Yalcin, A. GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations. *Catena* **2008**, *72*, 1–12. [[CrossRef](#)]
20. Kappes, M.S.; Malet, J.P.; Remaître, A.; Horton, P.; Jaboyedoff, M.; Bell, R. Assessment of debris-flow susceptibility at medium-scale in the Barcelonnette Basin, France. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 627–641. [[CrossRef](#)]
21. Horton, P.; Jaboyedoff, M.; Rudaz, B.; Zimmermann, M. Flow-R, a model for susceptibility mapping of debris flows and other gravitational hazards at a regional scale. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 869–885. [[CrossRef](#)]
22. Chang, M.; Tang, C.; Zhang, D.D.; Ma, G.C. Debris flow susceptibility assessment using a probabilistic approach: A case study in the Longchi area, Sichuan province, China. *J. Mt. Sci.* **2014**, *4*, 1001–1014. [[CrossRef](#)]
23. De Carvalho Faria Lima Lopes, L.; de Almeida Prado Bacellar, L.; de Amorim Castro, P.T. Assessment of the debris-flow susceptibility in tropical mountains using clast distribution patterns. *Geomorphology* **2016**, *275*, 16–25. [[CrossRef](#)]
24. Li, Y.; Wang, H.; Chen, J.; Shang, Y. Debris flow susceptibility assessment in the Wudongde dam area, China based on rock engineering system and fuzzy C-means algorithm. *Water* **2017**, *9*, 669. [[CrossRef](#)]
25. Kang, S.; Lee, S.R. Debris flow susceptibility assessment based on an empirical approach in the central region of South Korea. *Geomorphology* **2018**, *308*, 1–12. [[CrossRef](#)]
26. Qin, S.; Lv, J.; Cao, C.; Ma, Z.; Hu, X.; Liu, F.; Qiao, S.; Dou, Q. Mapping debris flow susceptibility based on watershed unit and grid cell unit: A comparison study. *Geomat. Nat. Hazards Risk* **2019**, *10*, 1648–1666. [[CrossRef](#)]
27. 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**, *308*, 101–111. [[CrossRef](#)]
28. Zhang, Y.; Ge, T.; Tian, W.; Liou, Y.A. Debris flow susceptibility mapping using machine-learning techniques in Shigatse area, China. *Remote Sens.* **2019**, *11*, 2801. [[CrossRef](#)]
29. Di, B.; Zhang, H.; Liu, Y.; Li, J.; Chen, N.; Stamatopoulos, C.A.; Luo, Y.; Zhan, Y. Assessing susceptibility of debris flow in southwest China using gradient boosting machine. *Sci. Rep.* **2019**, *9*, 1–12. [[CrossRef](#)]
30. Liang, W.J.; Zhuang, D.F.; Jiang, D.; Pan, J.J.; Ren, H.Y. Assessment of debris flow hazards using a Bayesian Network. *Geomorphology* **2012**, *171–172*, 94–100. [[CrossRef](#)]
31. Chevalier, G.G.; Medina, V.; Hürlimann, M.; Bateman, A. Debris-flow susceptibility analysis using fluvio-morphological parameters and data mining: Application to the Central-Eastern Pyrenees. *Nat. Hazards* **2013**, *67*, 213–238. [[CrossRef](#)]
32. Addison, P.; Oommen, T.; Sha, Q. Assessment of post-wildfire debris flow occurrence using classifier tree. *Geomat. Nat. Hazards Risk* **2019**, *10*, 505–518. [[CrossRef](#)]
33. Dou, Q.; Qin, S.; Zhang, Y.; Ma, Z.; Chen, J.; Qiao, S.; Hu, X.; Liu, F. A Method for improving controlling factors based on information fusion for debris flow susceptibility mapping: A case study in Jilin Province, China. *Entropy* **2019**, *21*, 695. [[CrossRef](#)]
34. Xiong, K.; Adhikari, B.R.; Stamatopoulos, C.A.; Zhan, Y.; Wu, S.; Dong, Z.; Di, B. Comparison of different machine learning methods for debris flow susceptibility mapping: A case study in the Sichuan Province, China. *Remote Sens.* **2020**, *12*, 295. [[CrossRef](#)]
35. Zhao, Y.; Meng, X.; Qi, T.; Qing, F.; Xiong, M.; Li, Y.; Guo, P.; Chen, G. AI-based identification of low-frequency debris flow catchments in the Bailong River basin, China. *Geomorphology* **2020**, *359*, 107125. [[CrossRef](#)]
36. Derbyshire, E.; Fort, M.; Owen, L.A. Geomorphological hazards along the Karakoram Highway: Khunjerab pass to the Gilgit River, Northernmost Pakistan. *Erdkunde* **2001**, *55*, 49–71. [[CrossRef](#)]

37. Searle, M.P.; Khan, M.A.; Fraser, J.E.; Gough, S.J.; Jan, M.Q. The tectonic evolution of the Kohistan-Karakoram collision belt along the Karakoram Highway transect, north Pakistan. *Tectonics* **1999**, *18*, 929–949. [[CrossRef](#)]
38. Goudie, A.S.; Kalvoda, J. Recent geomorphological processes in the Nagar region, Hunza Karakoram. *Acta Univ. Carol. Geogr.* **2004**, *39*, 135–148.
39. Verma, R.K.; Chandra Sekhar, C. Focal mechanism solutions and nature of plate movements in Pakistan. *J. Geodyn.* **1986**, *5*, 331–351. [[CrossRef](#)]
40. Guzzetti, F.; Carrara, A.; Cardinali, M.; Reichenbach, P. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* **1999**, *31*, 181–216. [[CrossRef](#)]
41. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Science Rev.* **2018**, *180*, 60–91. [[CrossRef](#)]
42. Zhang, H.; Chi, T.; Fan, J.; Liu, T.; Wang, W.; Yang, L.; Zhao, Y.; Shao, J.; Yao, X. Debris-flows scale predictions based on basin spatial parameters calculated from Remote Sensing images in Wenchuan earthquake area. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Beijing, China, 22–26 April 2013; p. 012091.
43. Wilford, D.J.; Sakals, M.E.; Innes, J.L.; Sidle, R.C.; Bergerud, W.A. Recognition of debris flow, debris flood and flood hazard through watershed morphometrics. *Landslides* **2004**, *1*, 61–66. [[CrossRef](#)]
44. Zhou, W.; Tang, C.; Van Asch, T.W.J.; Chang, M. A rapid method to identify the potential of debris flow development induced by rainfall in the catchments of the Wenchuan earthquake area. *Landslides* **2016**, *13*, 1243–1259. [[CrossRef](#)]
45. Johnson, P.A.; McCuen, R.H.; Hromadka, T.V. Magnitude and frequency of debris flows. *J. Hydrol.* **1991**, *123*, 69–82. [[CrossRef](#)]
46. Zhang, W.; Chen, J.P.; Wang, Q.; An, Y.; Qian, X.; Xiang, L.; He, L. Susceptibility analysis of large-scale debris flows based on combination weighting and extension methods. *Nat. Hazards* **2013**, *66*, 1073–1100. [[CrossRef](#)]
47. Miller, V.C. A quantitative geomorphic study of drainage basin characteristics in the Clinch Mountain area, Virginia and Tennessee. *Dep. Geol. Columbia Univ. N. Y.* **1953**, *65*, 1–30.
48. Pike, R.J.; Wilson, S.E. Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. *Bull. Geol. Soc. Am.* **1971**, *82*, 3087–3110. [[CrossRef](#)]
49. Melton, M.A. *An Analysis of the Relations among Elements of Climate, Surface Properties, and Geomorphology*; Columbia University: New York, USA, 1957; p. 99.
50. Jackson, L.E.; Kostaschuk, R.A.; MacDonald, G.M. Identification of debris flow hazard on alluvial fans in the Canadian Rocky Mountains. *GSA Rev. Eng. Geol.* **1987**, *7*, 115–124. [[CrossRef](#)]
51. Bovis, M.J.; Jakob, M. The role of debris supply conditions in predicting debris flow activity. *Earth Surf. Process. Landforms* **1999**, *24*, 1039–1054. [[CrossRef](#)]
52. Chen, N.S.; Zhang, Y.; Tian, S.F.; Deng, M.F.; Wang, T.; Liu, L.H.; Liu, M.; Hu, G. Effectiveness analysis of the prediction of regional debris flow susceptibility in post-earthquake and drought site. *J. Mt. Sci.* **2020**, *17*, 329–339. [[CrossRef](#)]
53. Wu, S.; Chen, J.; Xu, C.; Zhou, W.; Yao, L.; Yue, W.; Cui, Z. Susceptibility assessments and validations of debris-flow events in meizoseismal areas: Case study in China's Longxi River watershed. *Nat. Hazards Rev.* **2020**, *21*, 05019005. [[CrossRef](#)]
54. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
55. 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.
56. Kern, A.N.; Addison, P.; Oommen, T.; Salazar, S.E.; Coffman, R.A. Machine Learning Based Predictive Modeling of Debris Flow Probability Following Wildfire in the Intermountain Western United States. *Math. Geosci.* **2017**, *49*, 717–735. [[CrossRef](#)]
57. 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.
58. Roy, J.; Saha, S.; Arabameri, A.; Blaschke, T.; Bui, D.T. A novel ensemble approach for landslide susceptibility mapping (LSM) in Darjeeling and Kalimpong districts, West Bengal, India. *Remote Sens.* **2019**, *11*, 2866. [[CrossRef](#)]

59. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
60. Melton, M.A. The geomorphic and paleoclimatic significance of alluvial deposits in southern arizona. *J. Geol.* **1965**, *73*, 1–38. [[CrossRef](#)]
61. Wilkerson, F. Landslide recognition: Identification, movement, and causes. *Geomorphology* **1997**, *2*, 171–173. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).