

Article

An RVM-Based Model for Assessing the Failure Probability of Slopes along the Jinsha River, Close to the Wudongde Dam Site, China

Yanyan Li ^{1,*}, Jianping Chen ² and Yanjun Shang ¹

¹ Key Laboratory of Shale Gas and Geoenvironment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China; jun94@mail.iggcas.ac.cn

² College of Construction Engineering, Jilin University, Changchun 130026, China; chenjpwq@126.com

* Correspondence: lee_xandy@126.com; Tel.: +86-10-8299-8632

Academic Editor: Vincenzo Torretta

Received: 27 October 2016; Accepted: 21 December 2016; Published: 27 December 2016

Abstract: Assessing the failure potential of slopes is of great significance for land use and management. The objective of this paper is to develop a novel model for evaluating the failure probability of slopes based on a relevance vector machine (RVM), with a special attention to the characteristics of failed slopes along the lower reaches of the Jinsha River, close to the Wudongde dam site. Seven parameters that influence the occurrence of landslides were selected as environmental factors; namely lithology, slope angle, slope height, slope aspect, slope structure, distance from faults, and land use. A total of 55 landslides mapped in the study area were used to train and test the RVM model. The results suggest that the accuracy of the model in predicting the failure probability of slopes, using both training and testing data sets, is very high and deemed satisfactory. To validate the model performance, it was applied to 28 landslide cases identified in the upper reaches of the Jinsha River, where environmental and geological conditions are similar to those of the study area. An accuracy of approximately 92.9% was obtained, which demonstrates that the RVM model has a good generalization performance.

Keywords: landslide; relevance vector machine; influencing factor; failure probability

1. Introduction

Landslides are among the most notable geological processes that frequently occur in mountainous regions, resulting in financial losses measured in hundreds of billions of euros annually, injuries, and fatalities [1–3]. These disasters have caught unprecedented attention in the world; lots of researchers are continuously carrying out relevant research, mainly focusing on the prediction of landslides [4,5].

Various approaches have been developed for landslide hazard assessment, and can be grouped into three broad categories; analytic, statistical, and soft computing methods [6]. Analytic approaches consider the failure mechanisms of slopes and can provide an accurate prediction for the instability of slopes. However, when the study area is large, application of these approaches may be difficult. Statistical analyses such as discriminant analysis [4,7], logistic regression [8,9], and Bayes learning [10,11], are deemed to be more suitable for geological hazard assessment in large and complex areas [12,13]. With the development of science and technology, soft computing techniques, such as data mining and artificial intelligence, have also been widely used in geological hazard assessment. Examples of these approaches include artificial neural networks [14–17], genetic algorithms [18,19], decision trees [20,21], and support vector machines [22–25]. However predictions provided by these methods are not probabilistic.

Hybrid methods, which are established by combining statistical approaches and artificial intelligence, have also been adopted for assessing geological hazards; these include artificial neural network (ANN)-Bayes analysis [14], ANN-fuzzy logic [26], and neuro-fuzzy inference systems [27,28]. However the ANN-based approaches cannot provide objective and steady assessment results because their outcomes are operator dependent [13,15].

The Bayes learning algorithm is considered to be an effective tool for knowledge representation and reasoning under the influence of uncertainty [10]. Based on this algorithm, a recently developed machine learning technique, relevance vector machine (RVM), was originally introduced by Tipping [29]. As a Bayesian treatment of the sparse learning problem, the RVM can yield a probabilistic output [30].

In this study, a novel empirical model for slope failure analyses based on RVM is presented. We selected the lower reaches of the Jinsha River close to the Wudongde dam site as the study area; 55 landslides mapped in the region were utilized to train and test the RVM model. To evaluate the validity of the model, it was applied to another landslide site where the environmental conditions are similar to those of the study area.

2. Study Area

The study area (Figure 1) lies along the lower reaches of the Jinsha River and is the reservoir region of the Wudongde hydropower station, which is located in the mountains separating the Sichuan and Yunnan provinces. The occurrence of landslides not only poses a threat to human lives and properties, but also affects the stability of the Wudongde dam. Thus assessing the failure potential of slopes in this area is of great significance.

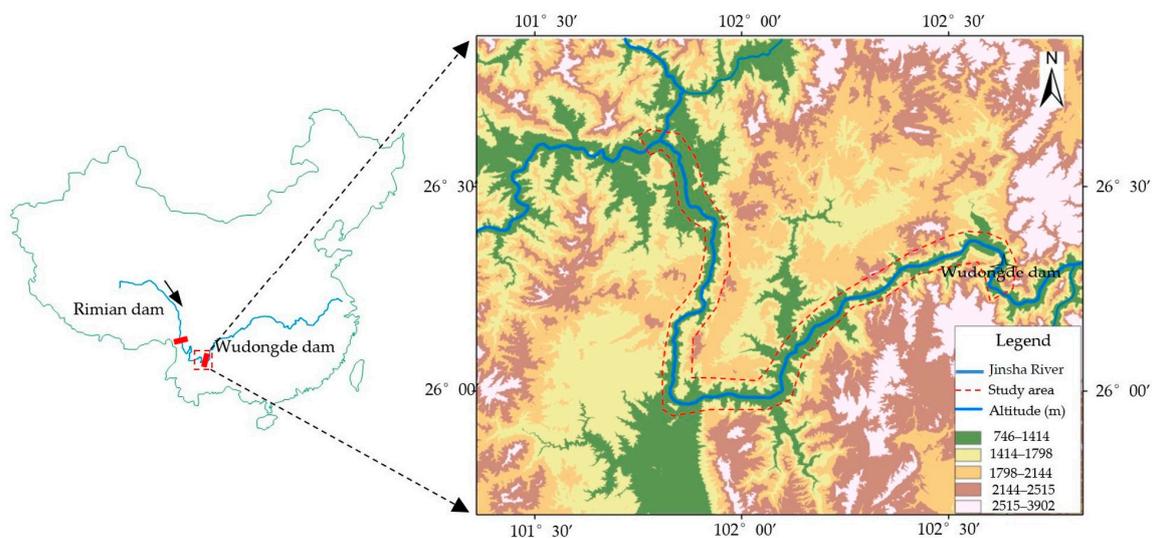


Figure 1. Location map of the study area.

The elevation ranges from 740 to 3900 m. This area experiences a low-latitude plateau subtropical monsoon climate characterized by plenty of sun and large evaporation capacity throughout the year [31]. The mean annual temperature is 20.9 °C. It receives a mean annual rainfall varying from 600 to 800 mm, while the mean annual evaporation is 698 mm.

2.1. Geological and Geomorphological Settings

From the tectonic standpoint, the study area is located in the eastern section of the Tethys-Himalaya tectonic domain, which is one of the tectonic zones of the Himalaya characterized by intense compressing and folding. The predominant regional structures are large-scale faults constituting the famous Chuan-Dian N-S tectonic belt [32]. A total of 13 regional faults are

situated in this region and dominantly trend approximately N-S (Figure 2). Several strong earthquakes have been triggered by these faults since 1955 [33], such as the Lazha earthquake (magnitude scale 6.7, 1955) which was triggered by the Tanglang-Yimeng fault, and the Panzhihua earthquake (magnitude scale 6.1, 2008), triggered by the Mopanshan-Lvzhijiang fault.

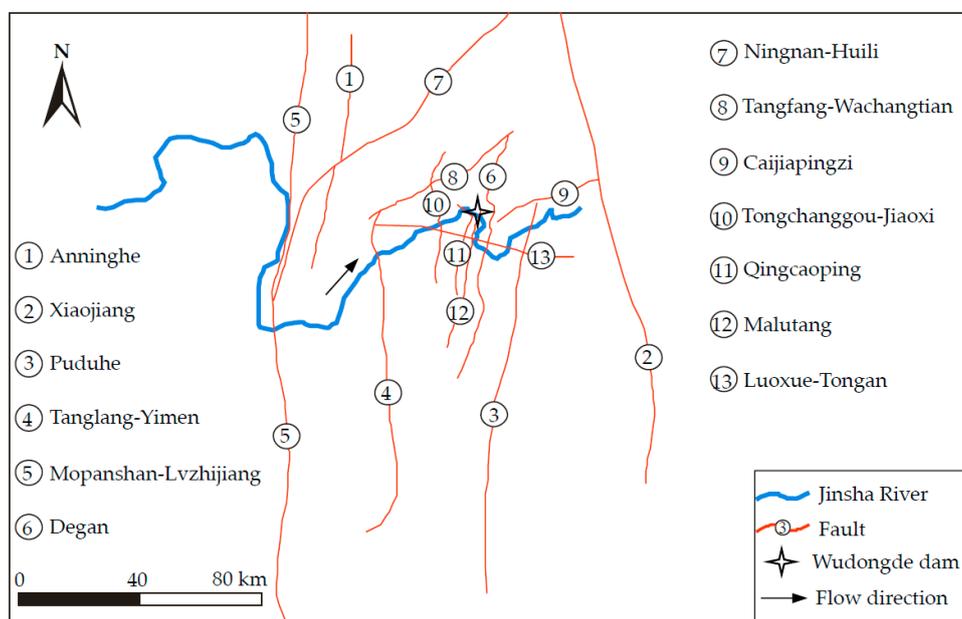


Figure 2. Regional tectonic framework map [33].

The geology is comprised of two major components; a pre-Sinian crystalline basement and a Sinian-Cretaceous sedimentary cover. The former is mainly composed of a range of metamorphic rocks (phyllite, slate and schist), which widely outcrop along the Jinsha River. The latter consists of limestone, sandstone, mudstone, and shale formations and is also widespread.

The geomorphological settings reflect the complex interplay between the geological and structural conditions of the area. Steep topographical characteristics predominate this region, with the average slope angle ranging from 30° to 45° . The valleys present a mountain canyon geomorphology. Geomorphic features include cliffs, ridges, gorges, rocky slopes, and Quaternary deposits along the river valleys. The distribution and extension of the river network and ridges is controlled by the structures to some extent. The effect of high relief and structural control is also well reflected by deep gorges and narrow valleys carved by numerous channels.

Phyllite, slate, schist, shale, and mudstone are strongly weathered and fractured and are prone to slope failures. Quaternary deposits composed of alluvial and eluvial deposits often outcrop as a cover layer on the riverbed and gentle slopes and are affected by transitional or rotational landslides triggered by the undercutting of the Jinsha River.

2.2. Landslide Identification

Landslide identification was carried out using SPOT5 remote sensing images and field surveys. The SPOT5 images, which were obtained by the SPOT5 satellite launched in 2002 with a multispectral resolution of 10 m and a panchromatic resolution of 2.5 m, have been largely utilized in geological hazard surveys. A series of field investigations, aimed at the confirmation of the mapped landslides through remote sensing images and the investigation of the link between the occurrence of landslides and environmental factors, were conducted. In general, there are three main types of landslides present in the study area; translational movements (55 locations), rock falls (26 locations), and debris flows (239 locations). Each type of landslide has different mechanism requiring separate study for the spatial

prediction of landslides. Therefore, in this study, only the 55 translational landslides (Figure 3) were used for analysis. At the time of fieldwork, 44% of the slope movements were classified as active, while the class of dormant movements corresponds to 33%, and 23% of the slope movements were considered stabilized. The highest frequency of landslide phenomena was recorded in mudstones and metamorphic rock formations (phyllite, slate, and schist). The smallest landslide covered an area of 0.03 km², while the largest one was 4.2 km². Figure 4 gives some examples of the landslides.

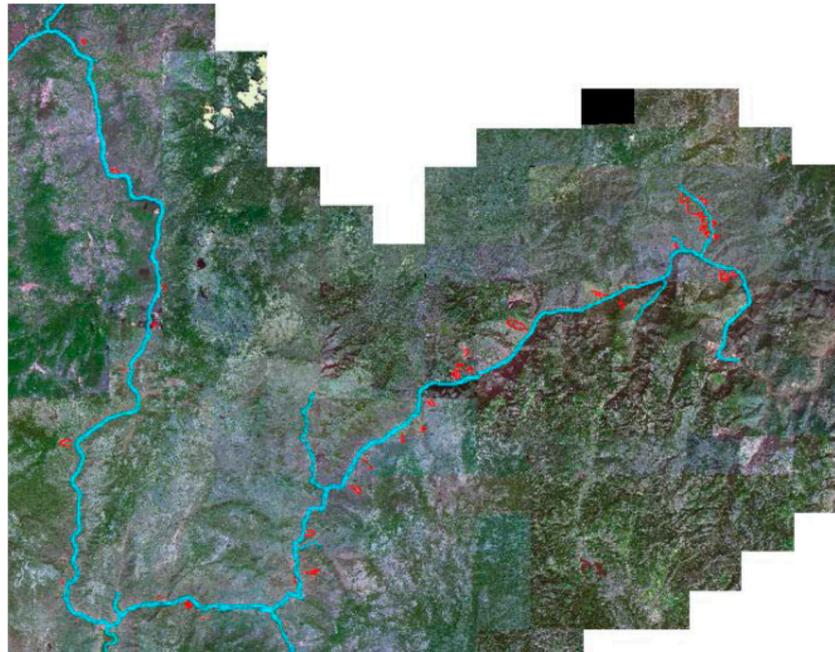


Figure 3. Landslide map of the study area.

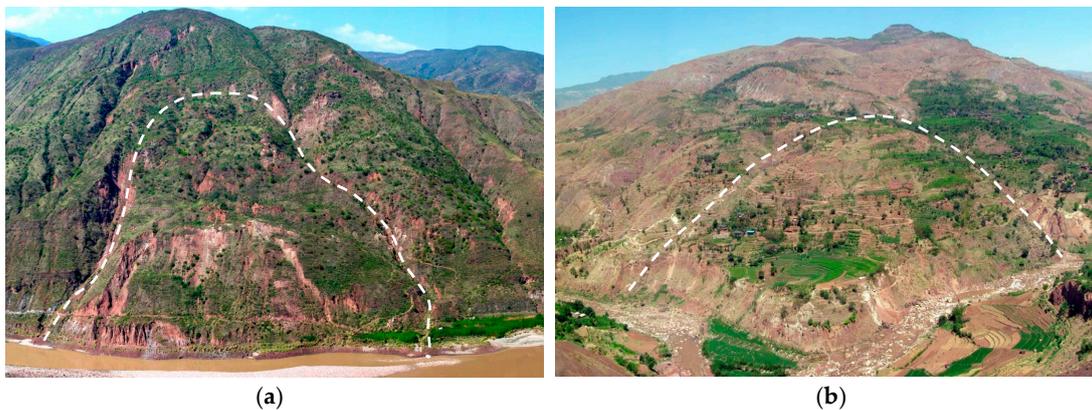


Figure 4. Examples of slope failures in the study area. (a) Xiaochatou landslide; (b) Daopo landslide.

3. Influencing Factors

Landslide hazard assessment is often performed based on the assumption that future landslides will occur in the areas where environmental conditions are similar to those of past and present failures [34]. Therefore it relies strongly on identifying landslide scars, characterizing the properties of failed sites, and confirming the link between landslide distribution and spatial variation of parameters [35]. There is no standard principle for selecting the influencing parameters because they often vary from one area to another. However it is commonly recognized that the parameter selection depends on the environmental conditions of the study area, the mechanisms of failed slopes,

and the scale of analyses [17]. It should be implemented according to the data availability and the significance of data in relation to the problem at question [23].

Based on the review of previous literature [13,16,17,25] and field surveys, seven factors that were identified as causative and triggering factors for landslide activity in the study area were selected to predict the failure potential of slopes. They are lithology, slope angle, slope aspect, slope height, slope structure, distance from faults, and land use. This research selected a single slope as a computing unit [16], and values of the influence factors were derived for each slope in the database. Table 1 presents the factors and their classes used in this study. The frequency distribution of landslide occurrence in each class is shown in Figure 5.

Table 1. Landslide influencing factors and their classifications.

Description	Class ID
1. Lithology	
Limestones and massive sandstones	1
Sandstones, shale, schist and mudstones	2
Mixed layers, and Quaternary deposits	3
2. Slope angle	
0°–10°	1
10°–20°	2
20°–30°	3
30°–40°	4
>40°	5
3. Slope height	
<100 m	1
100–200 m	2
200–300 m	3
>300 m	4
4. Slope aspect	
315°–45° (N)	1
225°–315° (W)	2
45°–135° (E)	3
135°–225° (S)	4
5. Slope structure	
Anti-dip	1
Insequent	2
Transverse	3
Dip-bedded	4
6. Distance from faults	
>1500 m	1
1000–1500 m	2
500–1000 m	3
0–500 m	4
7. Land use	
Barren land	1
Agriculture	2
Residential	3

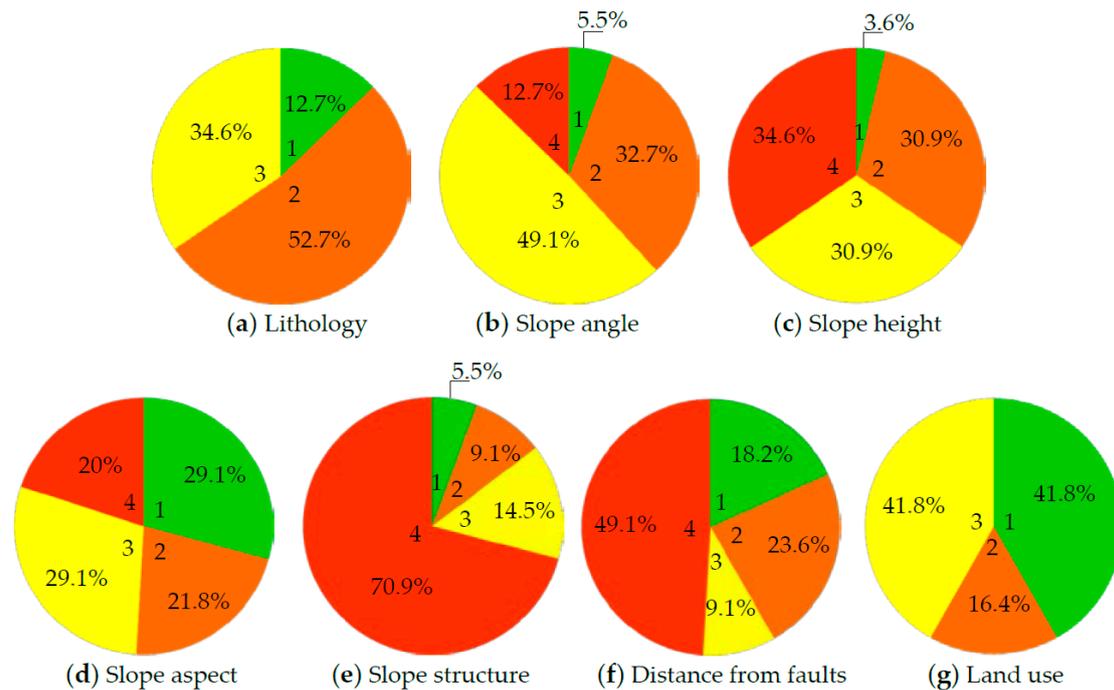


Figure 5. Frequency of landslide occurrence in each category. (a) Lithology; (b) Slope angle; (c) Slope height; (d) Slope aspect; (e) Slope structure; (f) Distance from faults; (g) Land use.

3.1. Lithology

Lithology has been considered to be one of the main parameters influencing the stability of slopes. Since different lithologic units have different slope stability performances, they are very important for landslide prediction [4]. The large surface development of formations, such as schist formations, flysch sediments, and metamorphic rock formations with anisotropic geomechanical behavior, facilitates the manifestation of abundant slope failures. The lithology was derived from a lithologic map at a 1:100,000 scale and was grouped into three categories (Table 1) according to Peng et al. [25] and Fourniadis et al. [36]. Figure 5a shows that approximately 87% of the landslides occurred on Quaternary deposits, strongly weathered sandstones, mudstones, and metamorphic rocks (phyllite, slate and schist).

3.2. Slope Angle

Slope angle is an essential factor controlling the stability of a slope. The shear stress induced by gravity increases with the increase of the slope angle; therefore landslides tend to occur more frequently on steeper slopes [24]. The slope angle was measured from the slope profile that was derived from the digital elevation model (DEM). Note that a slope angle of 10° was taken as the threshold because very few slopes less than that angle have failed [16]. The slope angle of the study area was divided into five classes (Table 1): 0° – 10° , 10° – 20° , 20° – 30° , 30° – 40° , and $>40^\circ$. Figure 5b illustrates that the largest numbers of landslides were recorded in the classes of 20° – 30° (49.1%) and 10° – 20° (32.7%).

3.3. Slope Height

Slope height denotes the difference value between the head and tail elevations of a slope. A slope with higher height often indicates a higher probability of experiencing a landslide event [37,38]. The slope height was directly derived from the DEM. As listed in Table 1, the slope height was partitioned into four classes according to Zhao et al. [38] and field surveys. The distribution of landslides in the height classes is shown in Figure 5c.

3.4. Slope Aspect

Aspect controls some microclimatic parameters, such as exposure to sunlight and winds, rainfall intensity, and soil moisture [17]. It was obtained from the DEM and was categorized into four classes according to Conforti et al. [17], as shown in Table 1. The relationship between lithology and failed slopes is presented in Figure 5d.

3.5. Slope Structure

Slope structure represents the spatial relationship between the rock strata and slope face. Slopes in the study area were measured from field investigations and were classified into four structural categories (Table 1): anti-dip, insequent, transverse, and dip-bedded. Field investigations demonstrate that dip-bedded slopes are prone to failures because of the erosion of slope toes caused by the Jinsha River. Figure 5e indicates that slope failures predominantly occurred on dip-bedded slopes, occupying 70.9% of the landslides in the area. However only 5.5% of the landslides occurred on anti-dip slopes.

3.6. Distance from Faults

Field evidence suggests a strong influence of tectonic settings on the occurrence of landslides [39]. The distance from faults was obtained from the geological map of the region (1:100,000). Approximately 49.1% (Figure 5f) of slope failures occurred in the sites with a minimum distance of less than 500 m from faults.

3.7. Land Use

Land use has been used as a predisposing factor in landslide hazard assessment. It was divided into three classes (Table 1) based on field surveys. Figure 5g shows that barren land and residential land have significant influence on the occurrence of landslides, and both of them occupy 41.8% of slope failures.

4. Relevance Vector Machine

RVM is based on a Bayesian learning framework and can output the probabilities of class membership. It has the same functional form as the support vector machine (SVM) and shares many of the characteristics of SVM whilst avoiding the SVM's main limitations [40]. The structure of the RVM is represented by the sum of product of weights and kernel functions, which is expressed as follows [41]:

$$y(\mathbf{x}; \mathbf{w}) = \sum_{i=1}^N w_i K(\mathbf{x}, \mathbf{x}_i) + w_0 \quad (1)$$

where w_i is the weight and $K(\mathbf{x}, \mathbf{x}_i)$ is a kernel function. The commonly used kernel is the Gaussian kernel $K(\mathbf{x}, \mathbf{x}_i) = \exp(-\|\mathbf{x} - \mathbf{x}_i\|^2 / \sigma^2)$, where σ is the kernel parameter controlling the sensitivity of the kernel. This function is not sensitive to outliers and can handle the case in which the relationship between class labels and attributes is nonlinear [13].

In this paper, the basic theory of RVM classification is briefly introduced. For further details of RVM, readers can refer to Tipping [29,41] and Bishop [40]. For two-class (binary) classification, RVM is used to predict the posterior probability of class membership of one of the classes, given the input \mathbf{x} . By applying the logistic sigmoid function $\sigma(y) = 1/(1 + e^{-y})$ to $y(\mathbf{x})$ and adopting the Bernoulli distribution for $p(\mathbf{t} | \mathbf{w})$, the likelihood can be written as [40]:

$$p(\mathbf{t} | \mathbf{w}) = \prod_{i=1}^N \sigma\{y(\mathbf{x}_i; \mathbf{w})\}^{t_i} [1 - \sigma\{y(\mathbf{x}_i; \mathbf{w})\}]^{1-t_i} \quad (2)$$

where the targets $t_i \in \{0, 1\}$. In this study, 0 and 1 denote the stable and failed cases, respectively. The weights \mathbf{w} cannot be analytically obtained and so are denied the closed-form expression for either

the weight posterior $p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha})$ or the marginal likelihood $p(\mathbf{t} | \boldsymbol{\alpha})$, with a hyper-parameter vector $\boldsymbol{\alpha}$. Therefore, according to Tipping [40], the following approximation procedure proposed by MacKay [42] and based on Laplace's method is utilized:

(1) For the current fixed values of $\boldsymbol{\alpha}$, the most probable weights \mathbf{w}_{MP} are found, giving the location of the mode of the posterior distribution. Since $p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}) \propto p(\mathbf{t} | \mathbf{w}) p(\mathbf{w} | \boldsymbol{\alpha})$, this is equivalent to finding the maximum, over \mathbf{w} , of

$$\log\{p(\mathbf{t} | \mathbf{w}) p(\mathbf{w} | \boldsymbol{\alpha})\} = \sum_{i=1}^N \left[t_i \log y_i + (1 - t_i) \log(1 - y_i) - \frac{1}{2} \mathbf{w}^T \mathbf{A} \mathbf{w} \right] \quad (3)$$

where $y_i = \sigma\{y(\mathbf{x}_i; \mathbf{w})\}$ and $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_N)$ for the current values of $\boldsymbol{\alpha}$. Since Equation (3) is a penalized, logistic, log-likelihood function, iterative maximization is required. The iteratively reweighted least-square algorithm is adopted to find \mathbf{w}_{MP} in the following procedure.

(2) Equation (3) is differentiated twice to give:

$$\nabla_{\mathbf{w}} \nabla_{\mathbf{w}} \log p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}) |_{\mathbf{w}_{MP}} = -(\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A}) \quad (4)$$

where $\mathbf{B} = \text{diag}(\beta_0, \beta_1, \beta_2, \dots, \beta_N)$ is diagonal matrix with elements $\beta_i = \sigma\{y(\mathbf{x}_i)\}[1 - \sigma\{y(\mathbf{x}_i)\}]$, and $\boldsymbol{\Phi} = [\varphi(\mathbf{x}_1), \varphi(\mathbf{x}_2), \dots, \varphi(\mathbf{x}_N)]^T$ is the $N \times (N + 1)$ design matrix with $\varphi(\mathbf{x}_i) = [1, K(\mathbf{x}_i, \mathbf{x}_1), K(\mathbf{x}_i, \mathbf{x}_2), \dots, K(\mathbf{x}_i, \mathbf{x}_N)]^T$. This is then negated and inverted to give the covariance Σ for a Gaussian approximation to the posterior over weights centered at \mathbf{w}_{MP} :

$$\Sigma = (\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A})^{-1} \quad (5)$$

(3) Using the statistics Σ and \mathbf{w}_{MP} of the Gaussian approximation, MacKay's approach is used to update the hyper-parameter $\boldsymbol{\alpha}$ by

$$\alpha_i^{new} = \frac{1 - \alpha_i \Sigma_{ii}}{w_{MP}^2} \quad (6)$$

where Σ_{ii} is the i th diagonal element of the covariance Σ , and $\mathbf{w}_{MP} = \Sigma \boldsymbol{\Phi}^T \mathbf{B} \mathbf{t}$.

During the optimization process, many α_i will have large (in principle infinite) values; thus the values of the corresponding weights will tend toward zero. Therefore these weights and the corresponding basis functions are removed from the model and thus play no role in making predictions for new inputs. Those examples remaining with $w_i \neq 0$ are termed relevance vectors.

5. Results and Discussions

5.1. Data Processing

RVM models require that each environmental parameter should be represented by a vector of real numbers. In this work, an m-digit number is used to represent an m-category attribute [16]. Only one of the digits is one and the others are zero. For example, a four-category attribute (e.g., slope height in Table 1) is represented by (1,0,0,0), (0,1,0,0), (0,0,1,0) or (0,0,0,1). Therefore the seven influencing factors of each slope can be converted into a vector of 27 bites.

5.2. Preparation of Training and Testing Data

For a two-class RVM, 'stable' sites are necessary. In this study, 20 'stable' sites were randomly chosen from the non-landslide sites where slope angles are less than 10° , a value which was taken as the threshold because very few slopes less than that angle have failed. The landslide data and non-landslide data were randomly divided into two groups [16]: (1) group 1, which contained 80% of the landslides (45 cases) and in which the same ratio of the non-landslides (16 cases) formed the training dataset; and (2) group 2, which contained the remaining 20% of the landslides and in which non-landslides was used as testing dataset for verifying the performance of the model. The target

values for landslides were given a value of 1, while a 0 value was assigned to those of non-landslides in the training dataset.

5.3. Parameter Determination

In an RVM, the controlling parameter (i.e., the kernel parameter σ) needs to be defined. Considering that there is a lack of information in most related papers concerning the determination of the optimal σ for RVM, a procedure of its estimation was conducted, in which several different values of σ were checked to determine which one of them gives better classification capabilities. As landslide hazard assessment is linearly non-separable problem, a cross-validation approach is adopted for the parameter search [13,30]. The accuracy is defined as the percentage of correct predictions for slope failure probabilities among all testing slopes [30]. In this research, “correct” could be defined as follows: for a failed slope, the model’s prediction is considered “correct” if the model’s output is greater than 0.5; on the other hand, for a stable slope, it is considered “correct” if the model’s output is less than 0.5. In this work, the 4-fold cross-validation technique is used by dividing the training dataset into 4 folds [30]; one fold (25%) is used as the validation set and the combination of the remaining three folds (75%) is used for training. By iterating each fold as the validation set and the aggregation of other folds as the training set, four validation accuracies can be obtained. The average of these accuracies is taken as the cross-validation accuracy (CVA) for a given σ . According to Wang et al. [30], several different values of σ are considered (Table 2). The value which yields the highest CVA is deemed to be the optimal value of σ . As shown in Table 2, the RVM provides the highest CVA value of 98.3% when $\sigma = 4$, which is in accordance with the results obtained by Wang et al. [43]. Therefore the ideal value of σ is 4.

Table 2. Validation results obtained by the 4-fold cross-validation approach.

σ	1	2	3	4
CVA	93.3%	97.9%	97.5%	98.3%

5.4. Training and Testing of the RVM Model

Figure 6 presents a flow chart for the training and testing process. The training dataset was used to train and develop the RVM model. During the training process, many weights will have values very close to zero, and their corresponding basis functions are removed from the model. The examples remaining with $w_i \neq 0$ are termed relevance vectors and are used for slope failure prediction. Table 3 lists the relevance vectors and their corresponding weights. As shown in Table 3, the RVM model employs 27 cases (45%) of the training dataset as relevance vectors, including 17 failed cases and 10 stable cases. It is worth mentioning here that the relevance vectors used in the RVM model represent prototypical examples. These examples exhibit the essential features of the information of the data, and thus are able to transform the input data into the specified targets [44].

The goal of the RVM model is to correctly classify a given slope to the target observation of either failed slope ($t = 1$) or stable slope ($t = 0$). For a given slope with particular environmental factors described by a vector \mathbf{x}_i , its failure probability is represented by $p(t = 1 | \mathbf{x}_i)$, which is computed by Equation (3). The classification rule of the RVM used in this study is the following: if $p(t = 1 | \mathbf{x}_i) > 0.5$, the slope is classified into the occurrence of a landslide; if $p(t = 1 | \mathbf{x}_i) < 0.5$, it is divided into the absence of a landslide.

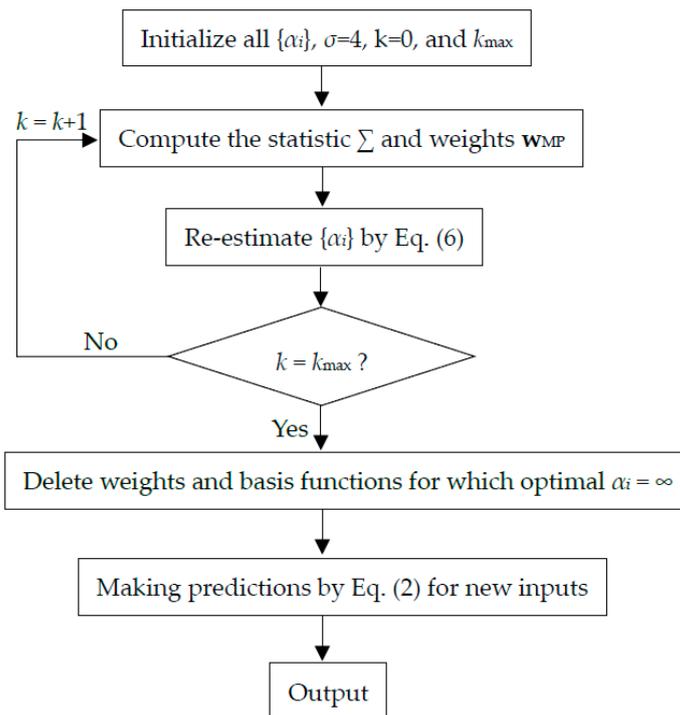


Figure 6. Flow chart of the training and testing process.

Table 3. Relevance vectors and their weights.

Slope	Status	Parameters							Weight
		Lithology	Slope Angle	Slope Height	Slope Aspect	Slope Structure	Distance from Faults	Land Use	
R1	Failed	2	2	4	3	4	1	3	0.4
R2	Failed	3	3	4	4	4	4	3	5.69
R3	Failed	1	3	4	4	3	4	1	0.28
R4	Failed	2	4	4	1	4	2	1	0.06
R5	Failed	2	3	3	3	4	4	1	18.18
R6	Failed	2	2	3	1	4	4	3	2.69
R7	Failed	3	3	4	1	4	4	1	0.13
R8	Failed	2	3	3	1	4	2	1	0.46
R9	Failed	2	3	3	3	4	2	1	0.54
R10	Failed	3	3	4	3	4	4	1	0.03
R11	Failed	2	2	3	1	4	4	3	2.69
R12	Failed	2	2	4	3	4	4	3	0.08
R13	Failed	3	2	3	3	4	4	3	0.01
R14	Failed	2	2	3	1	4	4	3	2.69
R15	Failed	2	4	2	4	4	4	2	0.67
R16	Failed	2	4	4	4	2	4	2	0.03
R17	Failed	2	3	4	3	3	4	3	0.06
R18	Stable	3	1	1	2	1	2	3	-2.46
R19	Stable	2	1	1	3	2	4	1	-0.24
R20	Stable	2	1	1	3	1	3	1	-12.81
R21	Stable	2	1	1	3	3	2	3	-1.99
R22	Stable	1	1	1	4	2	2	1	-0.42
R23	Stable	2	1	1	2	2	3	2	-4.36
R24	Stable	1	1	1	2	2	1	2	-1.4
R25	Stable	1	1	1	3	3	1	2	-6.83
R26	Stable	1	1	1	3	3	4	1	-1.35
R27	Stable	2	1	1	1	1	2	1	-0.02

The RVM model was applied to predict the failure probabilities of the 60 slopes in the training dataset. The results (Figure 7) show that all the 45 failed slopes have the failure probabilities of more than 0.8, a value which is high enough for the occurrence of a landslide. In addition, the values of failure probability for the 15 stable slopes are all less than 0.25, suggesting that these slopes are impossible for suffering from slope failures. Therefore the prediction accuracy for the 60 slopes reaches 100%.

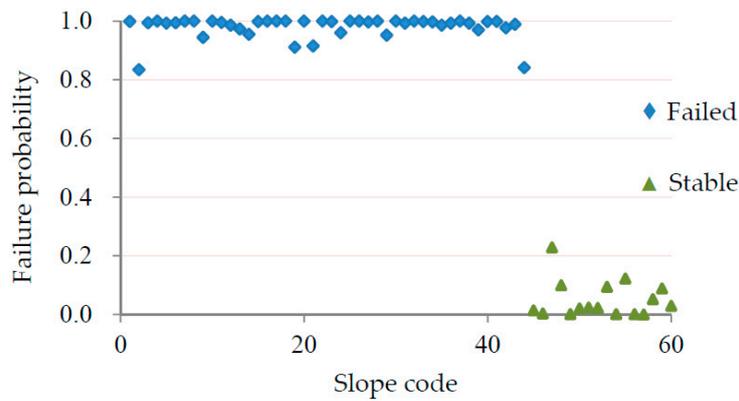


Figure 7. Failure probabilities for slopes in the training dataset.

To test the performance of the model, it was then applied to the testing dataset. The failure probability for each slope is listed in Table 4. For the 15 failed slopes, one slope has a probability of approximately 0.83 to fail and the others have failure probabilities of more than 0.95, which is very close to 1. All the stable slopes have failure probabilities of less than 0.25. Thus the accuracy of the classification of the slopes in the testing dataset is also 100%. The results demonstrate that the RVM model can provide a stable reliable prediction for landslide hazards in the study area.

Table 4. Failure probabilities for slopes in the testing dataset.

Slope	Status	Parameters							Failure Probability
		Lithology	Slope Angle	Slope Height	Slope Aspect	Slope Structure	Distance from Faults	Land Use	
T1	Failed	2	3	2	2	2	3	1	0.826
T2	Failed	2	2	2	1	4	2	3	0.997
T3	Failed	2	2	4	1	4	1	3	0.998
T4	Failed	3	2	2	4	4	3	3	0.985
T5	Failed	3	5	3	1	4	1	3	0.998
T6	Failed	2	3	4	1	4	3	2	0.992
T7	Failed	2	3	4	2	4	4	1	1.000
T8	Failed	3	2	2	2	4	3	3	0.959
T9	Failed	2	3	3	3	3	4	3	1.000
T10	Failed	3	3	4	2	4	4	2	0.999
T11	Failed	3	2	4	1	4	1	2	0.988
T12	Stable	1	1	1	3	3	3	3	0.003
T13	Stable	1	1	1	4	4	2	1	0.243
T14	Stable	3	1	1	2	1	3	2	0.001
T15	Stable	1	1	1	2	2	3	1	0.003

5.5. Evaluation of the Model

In this section, 28 landslides mapped in the upper reaches of the Jinsha River, close to the Rimian dam (Figure 1), where environmental conditions are similar to those of the study area [45], were selected as an example to evaluate the performance of the model. The distribution of the landslides is shown in Figure 8. Similar to the study area, this area is also located in the Chuandian N-S tectonic belt, with steep topology controlled by several large-scale faults. The main rock types include limestone, phyllite, schist, and shale. The mean annual precipitation and evaporation range from 350 to 650 mm, and from 750 to 2000 mm, respectively. The annual average temperature varies between 13.8 and 19.2 °C.

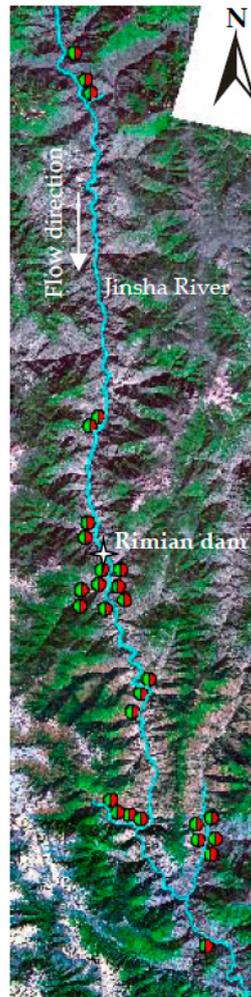


Figure 8. Distribution of landslides close to the Rimian dam site.

Table 5 lists the parameters of these slopes. The failure probabilities of these slopes were predicted using the RVM model, and the results are shown in Table 5. Among the 28 failed slopes, there are only two slopes that were wrongly partitioned into the group of stable ones, whose failure probabilities as calculated by the RVM model are less than 0.5. The prediction accuracy among the 28 slopes is approximately 92.9%, which is deemed satisfactory. Therefore the RVM model with seven triggering factors has a good prediction performance and could be extended to other sites where environmental conditions are similar to those of the study area.

Table 5. Prediction results for failed slopes close to the Rimian dam site.

Slope	Parameters							Failure Probability
	Lithology	Slope Angle	Slope Height	Slope Aspect	Slope Structure	Distance from Faults	Land Use	
U1	3	3	3	2	4	4	1	1.000
U2	2	4	4	2	4	4	1	0.999
U3	2	4	2	2	4	4	1	0.999
U4	2	4	4	2	4	4	1	0.999
U5	2	2	3	2	4	4	1	1.000
U6	1	3	2	3	2	4	1	0.994
U7	2	3	3	4	1	4	1	1.000
U8	2	3	2	3	1	4	1	0.998
U9	2	4	4	1	1	4	3	0.985
U10	1	3	3	3	1	4	1	0.999
U11	3	3	3	4	3	4	3	1.000
U12	3	3	4	4	3	4	3	0.998
U13	2	2	4	3	4	4	3	1.000
U14	2	3	1	4	2	4	1	0.992
U15	3	3	2	4	4	4	1	1.000
U16	3	2	2	3	1	4	3	0.959
U17	3	2	2	3	1	3	3	0.407
U18	2	3	2	4	1	1	2	0.788
U19	3	3	1	3	1	1	1	0.293
U20	2	3	2	4	1	1	3	0.935
U21	3	4	2	2	1	1	1	0.461
U22	3	3	4	2	4	1	1	0.996
U23	3	3	2	4	2	2	1	0.966
U24	3	4	3	3	1	4	1	0.994
U25	3	3	4	3	1	3	1	0.739
U26	1	3	2	2	3	4	3	0.975
U27	3	4	2	2	4	3	1	0.908
U28	1	2	1	1	4	3	3	0.740

6. Conclusions

This paper presents a novel approach for assessing the failure probability of slopes based on a relevance vector machine (RVM). The landslides mapped in the lower reaches of the Jinsha River were used to train and test the RVM model. Seven parameters, namely lithology, slope angle, slope height, slope aspect, slope structure, distance from faults, and land use, were selected as influencing factors of slope failures. The trained RVM model with seven factors is shown to be effective in classifying slopes into groups of stable ones and failed ones. The accuracies of the model in predicting the failure potential of slopes, using both training and testing data sets, are all very high and deemed satisfactory. To evaluate the model's performance, it was applied to landslide sites identified in the lower reaches of the Jinsha River, where environmental conditions are similar to those of the study area. An accuracy of approximately 92.9% was obtained, indicating that the model has a good generalization performance.

The study demonstrates that the RVM model is stable and reliable and could be used to predict the occurrence of translational landslides in hazard mitigation and guarding systems. Another important advantage of the model is that it can provide a probabilistic prediction for slope failures.

Acknowledgments: This work was financially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB10030100), the China Postdoctoral Science Foundation funded project (No. 2015M580135), the National Natural Science Foundation of China (NSFC) (Nos. 41372324 and 41602327) and the Chinese Special Funds for Major State Basic Research Project (No. 2014CB046901).

Author Contributions: All authors were responsible for different parts of this paper. Yanyan Li and Jianping Chen conducted field investigations and wrote the whole paper; Yanjun Shang revised the paper.

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

References

1. Aleotti, P.; Chowdhury, R. Landslide hazard assessment: Summary review and new perspectives. *Bull. Eng. Geol. Environ.* **1999**, *58*, 21–44. [[CrossRef](#)]

2. Pappalardo, G.; Mineo, S.; Rapisarda, F. Rockfall hazard assessment along a road on the Peloritani Mountains (northeastern Sicily, Italy). *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 2735–2748. [[CrossRef](#)]
3. Zhou, S.; Chen, G.; Fang, L.; Nie, Y. GIS-Based integration of subjective and objective weighting methods for regional landslides susceptibility mapping. *Sustainability* **2016**, *8*, 334.
4. Carrara, A.; Merenda, L. GIS techniques and statistical models in evaluating landslide hazard. *Earth Surf. Proc. Land.* **1991**, *16*, 427–445. [[CrossRef](#)]
5. Saroglou, H.; Marinou, V.; Marinou, P.; Tsiambaos, G. Rockfall hazard and risk assessment: An example from a high promontory at the historical site of Monemvasia, Greece. *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 1823–1836. [[CrossRef](#)]
6. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* **2013**, *51*, 350–365. [[CrossRef](#)]
7. Günther, A.; Reichenbach, P.; Malet, J.P.; Van Den Eeckhaut, M.; Hervás, J.; Dashwood, C.; Guzzetti, F. Tier-Based approaches for landslide susceptibility assessment in Europe. *Landslides* **2012**, *10*, 529–546. [[CrossRef](#)]
8. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, central Japan. *Geomorphology* **2005**, *65*, 15–31. [[CrossRef](#)]
9. Manzo, G.; Tofani, V.; Segoni, S.; Battistini, A.; Catani, F. GIS techniques for regional scale landslide susceptibility assessment: The Sicily (Italy) case study. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 1433–1452. [[CrossRef](#)]
10. Liang, W.J.; Zhuang, D.F.; Dong, J.; Pan, J.J.; Ren, H.Y. Assessment of debris flow hazards using a Bayesian Network. *Geomorphology* **2012**, *171–172*, 94–100. [[CrossRef](#)]
11. Song, Y.Q.; Gong, J.H.; Gao, S.; Wang, D.C.; Cui, T.J.; Li, Y.; Wei, B.Q. Susceptibility assessment of earthquake-induced landslides using Bayesian network: A case study in Beichuan, China. *Comput. Geosci.* **2012**, *42*, 189–199. [[CrossRef](#)]
12. Guzzetti, F.; Reichenbach, P.; Ardizzone, F.; Cardinali, M.; Galli, M. Estimating the quality of landslide susceptibility models. *Geomorphology* **2006**, *81*, 166–184. [[CrossRef](#)]
13. Yao, X.; Tham, L.G.; Dai, F.C. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [[CrossRef](#)]
14. Lee, S.; Ryu, J.; Won, J.; Park, H. Determination and application of weights for landslide susceptibility mapping using an artificial neural network. *Eng. Geol.* **2004**, *71*, 289–302. [[CrossRef](#)]
15. Ermini, L.; Catani, F.; Casagli, N. Artificial neural network to landslide susceptibility assessment. *Geomorphology* **2005**, *66*, 327–343. [[CrossRef](#)]
16. Lin, H.M.; Chang, S.H.; Wu, J.H.; Juang, C.H. Neural network-based model for assessing failure potential of highway slopes in the Alishan, Taiwan Area: Pre- and post-earthquake investigation. *Eng. Geol.* **2009**, *104*, 280–289. [[CrossRef](#)]
17. Conforti, M.; Pascale, S.; Robustelli, G.; Sdao, F. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (Northern Calabria, Italy). *Cetena* **2014**, *113*, 236–250. [[CrossRef](#)]
18. Chang, T.C.; Chien, Y.H. The application of genetic algorithm in debris flows prediction. *Environ. Geol.* **2007**, *53*, 339–347. [[CrossRef](#)]
19. Liu, S.H.; Lin, C.W.; Tseng, C.M. A statistical model for the impact of the 1999 Chi-Chi earthquake on the subsequent rainfall-induced landslides. *Eng. Geol.* **2013**, *156*, 11–19. [[CrossRef](#)]
20. Saito, H.; Nakayama, D.; Matsuyama, H. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi Mountains, Japan. *Geomorphology* **2009**, *109*, 108–121. [[CrossRef](#)]
21. Wan, S. A spatial decision support system for extracting the core factors and thresholds for landslide susceptibility map. *Eng. Geol.* **2009**, *108*, 237–251. [[CrossRef](#)]
22. Samui, P. Slope stability analysis: A support vector machine approach. *Environ. Geol.* **2008**, *56*, 255–267. [[CrossRef](#)]
23. Marjanović, M.; Kovačević, M.; Bajat, B.; Voženílek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **2011**, *123*, 225–234. [[CrossRef](#)]
24. Xu, C.; Dai, F.C.; Xu, X.W.; Lee, Y.H. GIS-Based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* **2012**, *145–146*, 70–80. [[CrossRef](#)]

25. Peng, L.; Niu, R.Q.; Huang, B.; Wu, X.L.; Zhao, Y.N.; Ye, R.Q. Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. *Geomorphology* **2013**, *101*, 572–582. [[CrossRef](#)]
26. Kanungo, D.P.; Arora, M.K.; Sarkar, S.; Gupta, R.P. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Eng. Geol.* **2006**, *85*, 347–366. [[CrossRef](#)]
27. Vahidnia, M.H.; Alesheikh, A.A.; Alimohammadi, A.; Hosseinali, F. A GIS-based neuro-fuzzy procedure for integrating knowledge and data in landslide susceptibility mapping. *Comput. Geosci.* **2010**, *36*, 1101–1114. [[CrossRef](#)]
28. Bui, D.T.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. *Comput. Geosci.* **2012**, *45*, 199–211.
29. Tipping, M.E. The relevance vector machine. In *Advances in Neural Information Processing Systems*; Solla, S.A., Leen, T.K., Muller, K.R., Eds.; MIT Press: Cambridge, MA, USA, 2000; Volume 12, pp. 652–658.
30. Wang, X.D.; Ye, M.Y.; Duanmu, C.J. Classification of data from electronic nose using relevance vector machines. *Sens. Actuators B Chem.* **2009**, *140*, 143–148. [[CrossRef](#)]
31. Zhang, W.; Li, H.Z.; Chen, J.P.; Zhang, C.; Xu, L.M.; Sang, W.F. Comprehensive hazard assessment and protection of debris flows along Jinsha River close to the Wudongde dam site in China. *Nat. Hazards* **2011**, *58*, 459–477. [[CrossRef](#)]
32. Chen, H.; Hu, J.M.; Qu, H.J.; Wu, G.L. Early Mesozoic structural deformation in the Chuandian N-S Tectonic Belt, China. *Sci. China Ser. D* **2011**, *54*, 1651–1664. [[CrossRef](#)]
33. 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](#)]
34. Rozos, D.; Pyrgiotis, L.; Skias, S.; Tsagaratos, P. An implementation of rock engineering system for ranking the instability potential of natural slopes in Greek territory. An application in Karditsa County. *Landslides* **2008**, *5*, 261–270. [[CrossRef](#)]
35. Che, V.B.; Kervyn, M.; Suh, C.E.; Fontijn, K.; Ernst, G.G.J.; del Marmol, M.-A.; Trefois, P.; Jacobs, P. Landslide susceptibility assessment in Limbe (SW Cameroon): A field calibrated seed cell and information value method. *Catena* **2012**, *92*, 83–98. [[CrossRef](#)]
36. Fourniadis, I.G.; Liu, J.G.; Mason, P.J. Landslide hazard assessment in the Three Gorges area, China, using ASTER imagery: Wushan—Badong. *Geomorphology* **2007**, *84*, 126–144. [[CrossRef](#)]
37. Dai, F.C.; Lee, C.F. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* **2002**, *42*, 213–228. [[CrossRef](#)]
38. Zhao, W.H.; Huang, R.Q.; Ju, N.P.; Zhao, J.J. Assessment model for earthquake-triggered landslides based on quantification theory I: Case study of Jushui River basin in Sichuan, China. *Nat. Hazards* **2014**, *70*, 821–838. [[CrossRef](#)]
39. Mineo, S.; Pappalardo, G.; Rapisarda, F.; Cubito, A.; Di Maria, G. Integrated geostructural, seismic and infrared thermography surveys for the study of an unstable rock slope in the Peloritani Chain (NE Sicily). *Eng. Geol.* **2015**, *195*, 225–235. [[CrossRef](#)]
40. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
41. Tipping, M.E. Sparse Bayesian learning and relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
42. Mackay, D.J.C. The evidence framework applied to classification networks. *Neural Comput.* **1992**, *4*, 720–736. [[CrossRef](#)]
43. Wang, F.; Gou, B.; Qin, Y. Modeling tunneling-induced ground surface settlement development using a wavelet smooth relevance vector machine. *Comput. Geotech.* **2013**, *54*, 125–132. [[CrossRef](#)]
44. Samui, J.; Lansivaara, T.; Kim, D. Utilization relevance vector machine for slope reliability analysis. *Appl. Soft Comput.* **2011**, *11*, 4036–4040. [[CrossRef](#)]
45. Zhou, F.J. Early Intelligence Identification Methods of Stability and Fuzzy Synthetic Prediction of Hazard for Landslides in Rimian Hydropower Station. Ph.D. Thesis, Jilin University, Changchun, China, 2013; p. 225.

