

Article

# Landslide Susceptibility Modeling Based on GIS and Novel Bagging-Based Kernel Logistic Regression

Wei Chen <sup>1,\*</sup>, Himan Shahabi <sup>2,\*</sup>, Shuai Zhang <sup>1</sup>, Khabat Khosravi <sup>3</sup>, Ataollah Shirzadi <sup>4</sup>, Kamran Chapi <sup>4</sup>, Binh Thai Pham <sup>5</sup>, Tingyu Zhang <sup>6</sup>, Lingyu Zhang <sup>1</sup>, Huichan Chai <sup>7</sup>, Jianquan Ma <sup>1</sup>, Yingtao Chen <sup>1</sup>, Xiaojing Wang <sup>1</sup>, Renwei Li <sup>1</sup> and Baharin Bin Ahmad <sup>8</sup>

<sup>1</sup> College of Geology & Environment, Xi'an University of Science and Technology, Xi'an 710054, China; 16209062016@stu.xust.edu.cn (S.Z.); 16209062017@stu.xust.edu.cn (L.Z.); majianquan@xust.edu.cn (J.M.); chenyt@xust.edu.cn (Y.C.); wangjing@xust.edu.cn (X.W.); 17209074018@stu.xust.edu.cn (R.L.)

<sup>2</sup> Departments of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj 66177-15175, Iran

<sup>3</sup> Department of Watershed Management, Faculty of Natural Resources, Sari Agricultural Sciences and Natural Resources University, Sari 48181-68984, Iran; khabat.khosravi@gmail.com

<sup>4</sup> Department of Rangeland and Watershed Management, Faculty of Natural Resources, University of Kurdistan, Sanandaj 66177-15175, Iran; A.Shirzadi@uok.ac.ir (A.S.); k.chapi@uok.ac.ir (K.C.)

<sup>5</sup> Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam; phambinhgvt@gmail.com

<sup>6</sup> School of Earth Science and Resources, Chang'an University, Xi'an 710064, China; 2016027003@chd.edu.cn

<sup>7</sup> School of Earth and Environment, Anhui University of Science & Technology, HuaiNan 232001, China; haiyin310@126.com

<sup>8</sup> Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia (UTM), Johor Bahru 81310, Malaysia; baharinahmad@utm.my

\* Correspondence: chenwei0930@xust.edu.cn (W.C.); h.shahabi@uok.ac.ir (H.S.); Tel.: +98-9186658739 (H.S.)

Received: 29 October 2018; Accepted: 6 December 2018; Published: 7 December 2018



**Abstract:** Landslides cause a considerable amount of damage around the world every year. Landslide susceptibility assessments are useful for the mitigation of the associated potential risks to local economic development, land use planning, and decision makers. The main aim of this study was to present a novel hybrid approach of bagging (B)-based kernel logistic regression (KLR), named the BKLR model, for spatial prediction of landslides in the Shangnan County, China. We first selected 15 conditioning factors for landslide susceptibility modeling. Then, the prediction capability of all conditioning factors was evaluated using the least square support vector machine method. Model validation and comparison were performed based on the area under the receiver operating characteristic curve and several statistical-based indexes, including positive predictive rate, negative predictive rate, sensitivity, specificity, kappa index, and root mean square error. Results indicated that the BKLR ensemble model outperformed and outclassed the KLR and the benchmark support vector machine model. Our findings overall confirmed that a combination of the meta model with a decision tree classifier based on a functional algorithm can decrease the over-fitting and variance problems of data, which could enhance the prediction power of the landslide model. The resultant susceptibility maps could be useful for hazard mitigation in the study area and other similar landslide-prone areas.

**Keywords:** landslide; meta classifier; prediction power; China

## 1. Introduction

Landslides are one of the most important geological hazards worldwide [1]. Natural hazards and risks have become an important issue affecting human safety; the damage caused by these events

to human life and the environment cannot be ignored [2]. Due to the potential economic, social, environmental, and health impacts of landslides, policy makers pay attention to landslide hazard zonation maps to identify sensitive areas and sustainable locations for future development [3].

Globally, around 66 million people are living within the areas at high risk of landslides [4]. For example, in the United States, the annual financial losses by landslides is estimated about USD \$1.5 billion. Most areas in China are mountainous, so are susceptible to landslides, and direct and indirect economic losses induced by landslides account for more than ¥20 billion every year, endangering the lives of local people [5].

Due to the damage caused by landslides, the first step in landside hazard assessment is preparation of landslide susceptibility maps (LSMs) on a regional scale [6]. Statistical models are most commonly used in landslide susceptibility mapping, which are based on the analysis of the relationships between influencing factors and existing landslides [7]. In these statistical approaches, bivariate and multivariate statistical techniques are used for landslide susceptibility mapping throughout the world, including frequency ratio [8–10], index of entropy [11–15], bivariate statistical analysis [16], multivariate adaptive regression spline [17], analytical hierarchy process [18,19], statistical index [20,21], weight of evidence [13,21], evidential belief function [22,23], certainty factor [24,25], and logistic regression [26–29].

In addition to the above mentioned development of LSMs, various data mining techniques have been introduced for LSM, for example, neuro-fuzzy [30–32], artificial neural network [33,34], kernel logistic regression [33,35], multivariate adaptive regression spline [19,36], decision trees [37–40], support vector machines [41,42], random forest [23,43], adaptive neuro-fuzzy inference system [44–46], and naive Bayes [47,48]. However, the best method for creating LSMs is still under discussion.

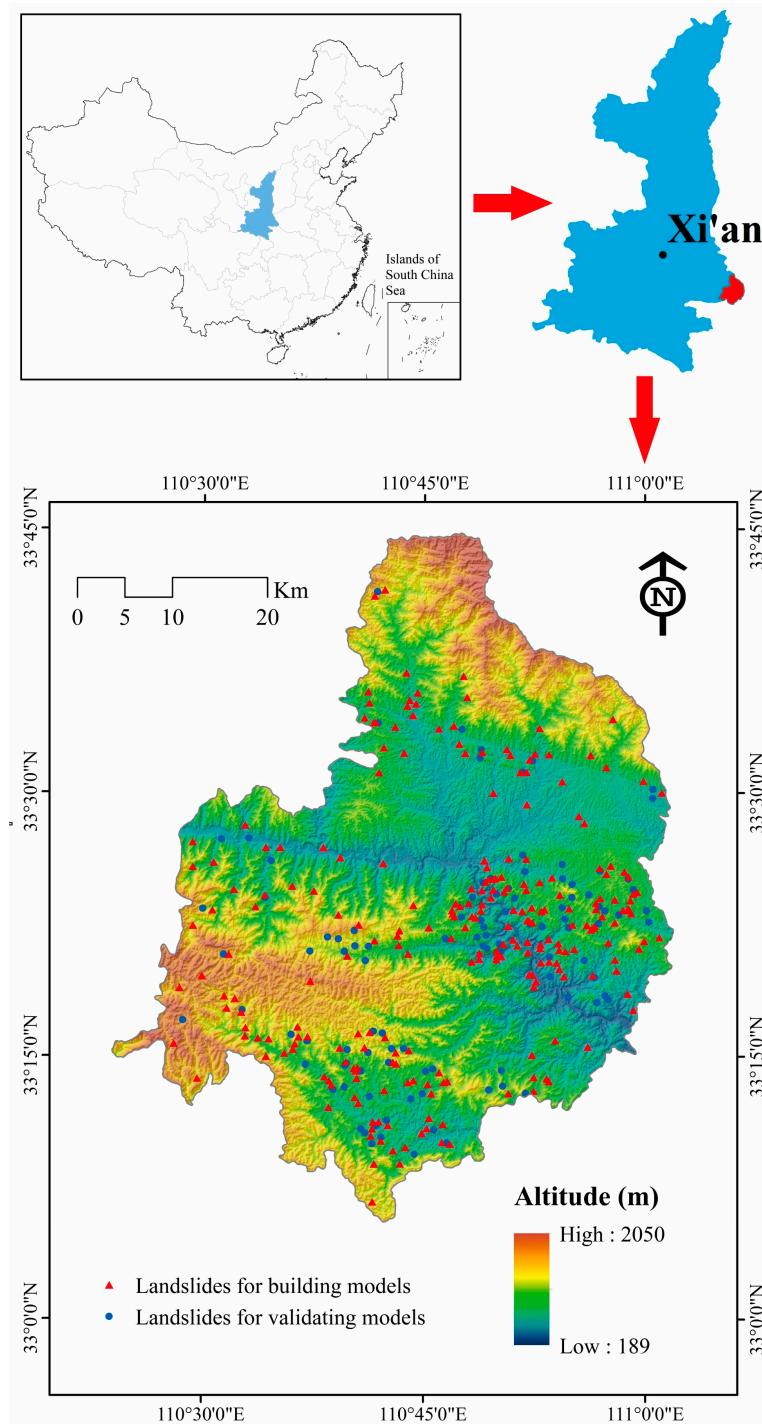
Researchers have reported that hybrid models produce better outcomes than individual machine learning techniques, so they are thought to be best for creating the best LSMs [49,50]. Therefore, the aim of this study was to introduce a new hybrid bagging-based two-class kernel logistic regression (BKLR) for landslides susceptibility mapping. The performance of this BKLR model is compared with the single KLR and the benchmark support vector machines (SVM).

## 2. Study Area and Data Used

Shangnan County, Shaanxi Province, China (Figure 1) regularly suffers a great deal of damage from landslides, so was selected as a suitable site for evaluating landslide susceptibility models. The altitude in the area ranges from 189 to 2050 m above sea level (a.s.l.), and the area and maximum slope angle are approximately 2307 km<sup>2</sup> and 65°, respectively. The average temperature for the whole county is 14.6 °C, ranging between 11.1 and 15.0 °C for various regions. The spatial distribution of air temperature is affected by altitude. The temperature is low in the northern and southern mountainous areas, and high in the central and Danjiang River areas.

Topographically, Shangnan County is a mountainous terrain in the Eastern Qinling Mountains. The altitudes in the northern and southwestern areas are higher than in the middle and southeastern areas. The study area can be divided into three geomorphological units: (I) Hilly areas with altitudes below 500 m and a relative height difference of less than 200 m; (II) low-relief areas with altitudes between 500 and 1000 m and a relative height difference between 200 and 500 m; and (III) mid-mountain zones with altitudes above 1000 m and a relative height difference greater than 500 m.

Geologically, the study area is located at the border zone of the North China and the Yangtze plates, and spans multiple stratigraphic regions from north to south. Therefore, the lithologies of the stratum are obviously different. The strata in the region are dominated by the Archean to Ordovician periods, and Devonian and Carboniferous are partially outcropped (Figure 2). Five major faults divide the study area into distinct structural zones: (1) Shangnan-Danfeng fault (NW–SE direction); (2) Zhulinguan-Taifenglou large fault (W–E direction); (3) Banyan-Yaolinghe fault (NW–SE direction); (4) Bailuchu-Weiijiatai fault (NW–SE direction); and (5) Shiliping-Sanguanmiao fault (W–E direction).

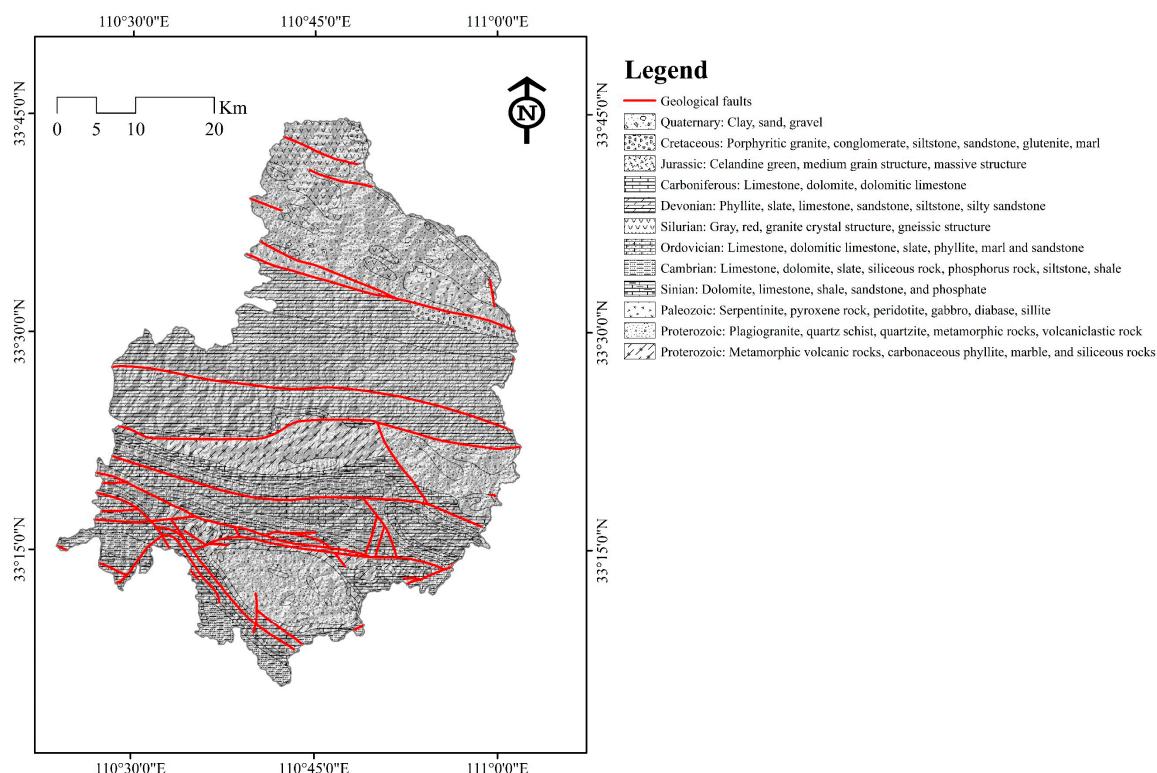


**Figure 1.** Location of the study area and landslide inventory.

Landslide inventory is the basis for LSM. Reliable and accurate landslide inventory data are crucial for LSM [51]. To enhance the reliability and accuracy of landslide inventory maps, a total of three techniques were used in the study: Historical reports, interpretation of aerial photographs, and extensive global positioning system (GPS) field surveys. According to landslide inventory maps in the region, a total of 348 landslides were identified (Figure 1). From these locations, 244 (70%) were randomly selected for building the models, and 104 (30%) were used for validating models.

Following the development of a landslide inventory, the factors necessary to create the landslide susceptibility maps must be determined [52]. The factors used in the studies evaluating landslide susceptibility can be categorized into three major groups: Topographical, environmental, and geological.

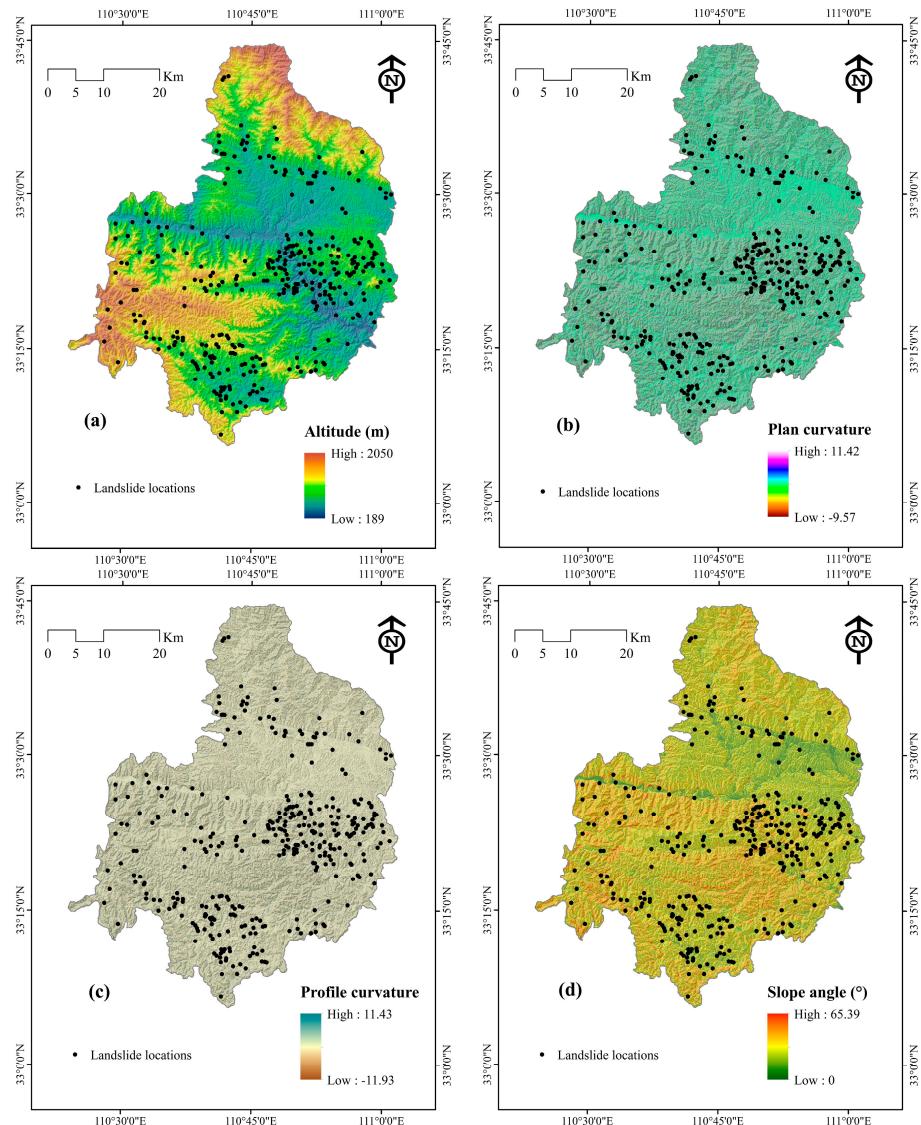
Fifteen factors (11 continuous and 4 categorical) were chosen and produced from the digital topographical maps at a 1:50,000 scale, LANDSAT-8 satellite images with a spatial resolution of 30 m, the compilation of geological maps on a 1:200,000 scale, the compilation of land use maps on a 1:100,000 scale, and soil maps on a 1:1,000,000 scale. The list of the factors used in this study are shown in Table 1. Altitude, plan curvature, profile curvature, slope angle, slope aspect, topographic wetness index (TWI), stream power index (SPI), sediment transport index (STI), distance to roads, and distance to rivers factors were determined from topographical maps. Distance to fault and lithology factors were produced from the geological maps. Normalized difference vegetation index (NDVI) was produced from the LANDSAT-8 satellite images. Land use and soil factors were extracted from land use and soil maps, respectively. Finally, a 30 m pixel size was chosen for all the factor maps (Figure 3).



**Figure 2.** Geological map of the area.

**Table 1.** Landslide conditioning factors.

Category	Factors	GIS Data Type	Scale or Resolution
Topographic factors	Altitude	ARC/INFO GRID	30 × 30 m
	Plan curvature	ARC/INFO GRID	30 × 30 m
	Profile curvature	ARC/INFO GRID	30 × 30 m
	Slope angle	ARC/INFO GRID	30 × 30 m
	Slope aspect	ARC/INFO GRID	30 × 30 m
	TWI	ARC/INFO GRID	30 × 30 m
	SPI	ARC/INFO GRID	30 × 30 m
	STI	ARC/INFO GRID	30 × 30 m
Environmental factors	Distance to rivers	ARC/INFO GRID	30 × 30 m
	Distance to roads	ARC/INFO GRID	30 × 30 m
	NDVI	ARC/INFO GRID	30 × 30 m
	Land use	ARC/INFO polygon coverage	1:100,000
	Soil	ARC/INFO polygon coverage	1:1,000,000
Geological factors	Lithology	ARC/INFO polygon coverage	1:200,000
	Distance to faults	ARC/INFO GRID	30 × 30 m

**Figure 3. Cont.**

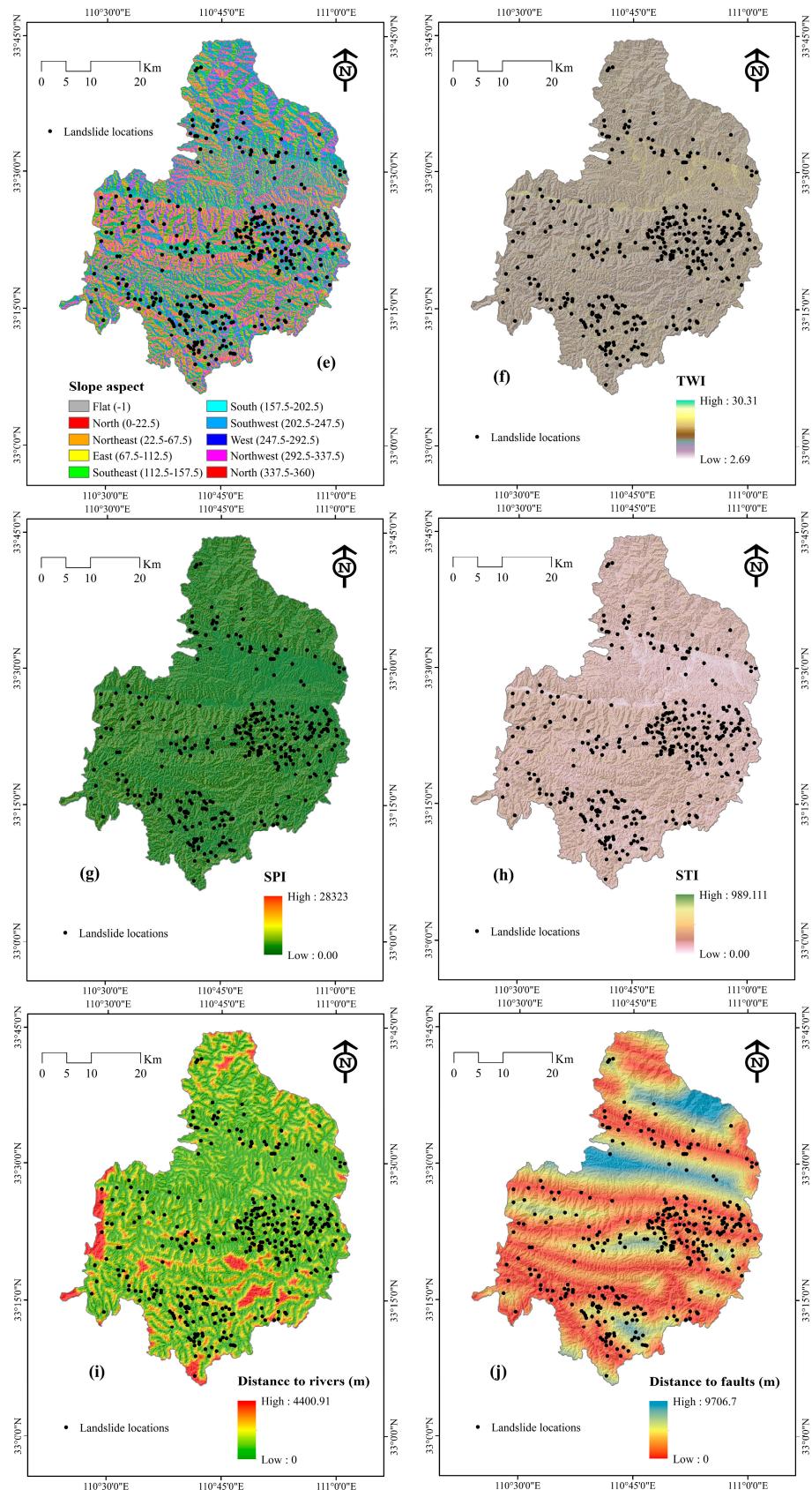
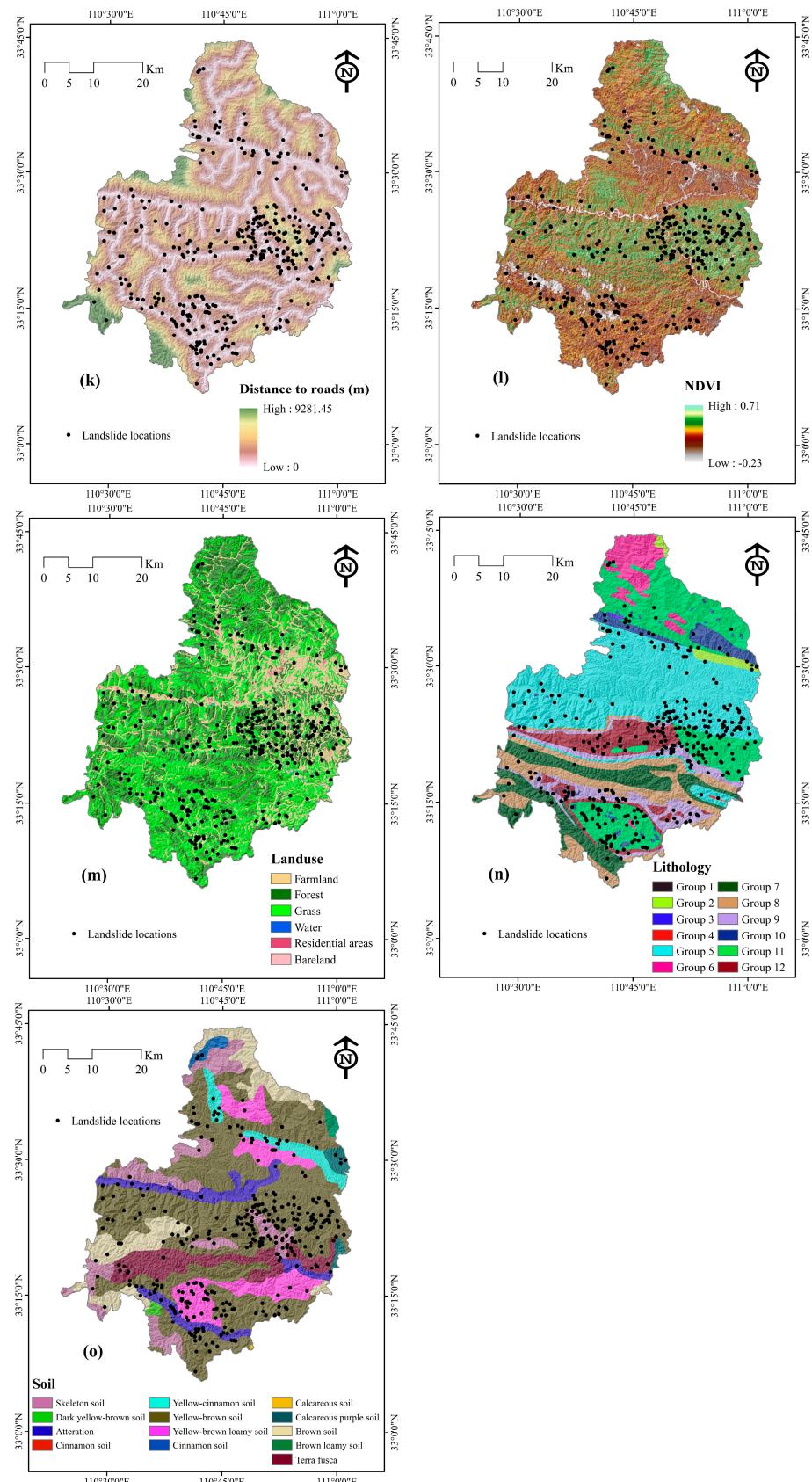


Figure 3. Cont.



**Figure 3.** Landslide conditioning factors: (a) Altitude; (b) plan curvature; (c) profile curvature; (d) slope angle; (e) slope aspect; (f) topographic wetness index (TWI); (g) stream power index (SPI); (h) sediment transport index (STI); (i) distance to rivers; (j) distance to faults; (k) distance to roads; (l) normalized difference vegetation index (NDVI); (m) land use; (n) lithology and (o) soil.

### 3. Modeling Approaches

#### 3.1. Bagging

Bagging is based on the concepts of bootstrapping and aggregating, and is one of the most popular ensemble algorithms [53]. In bagging, the training set is randomly sampled  $n$  times with replacement, producing  $n$  training sets with sizes equal to the original training set [53].

In this study, we considered a learning set  $C$  consisting of  $n$  independent observations [54,55]. Here, they are landslide conditioning factors, so we have  $C = \{(X_i, Y_i), i = 1, 2, \dots, n\}$ . Firstly, set  $C_b$  ( $b = 1, 2, \dots, n$ ) denotes the  $b$ -th bootstrap sample of the training set  $C$  obtained by drawing with replacement  $n$  elements of  $C$ . Secondly, compute the bootstrapped estimator  $g^*(\cdot)$  by the plug-in principle:  $g^*(\cdot) = h_n((X_1^*, Y_1^*), \dots, (X_n^*, Y_n^*))(\cdot)$ . Finally, repeat the above step  $m$  times, where the  $m$  is often defined as 50 or 100, yielding  $g^{*k}(\cdot)$  ( $k = 1, 2, \dots, m$ ). So, the bagged estimator is

$$g_{Bag}(\cdot) = \frac{\sum_{k=1}^m g^{*k}(\cdot)}{m}.$$

In theory, the bagged estimator is described as follows:

$$g_{Bag}(\cdot) = {}^* [g^*(\cdot)], \quad (1)$$

where the theoretical quantity corresponds to  $m \rightarrow \infty$  and the finite number  $m$  in practice governs the accuracy of the Monte Carlo approximation.

#### 3.2. Kernel Logistic Regression

KLR is a powerful and flexible discriminative method, which possesses the ability to provide the confidence of class prediction [56]. A conventional logistic regression model was constructed in high-dimensional feature space and proposed by a Mercer kernel [57]. Given the labelled training data:

$$D = \{(x_i, y_i)\}, i = 1, x_i \in X \in R^d, y_i \in [0, 1], \quad (2)$$

where  $x_i$  is the landslide conditioning factor, and  $K : X \times X \rightarrow R$  evaluates the inner product between the images of input vectors in the feature space.

The kernel function used in this study is the isotropic radial basis function (RBF):

$$K(x, x') = e^{\{-\xi \|x - x'\|^2\}} \quad (3)$$

Then, a conventional logistic regression model was constructed in the feature space as follows:

$$\log it\{h(x)\} = \omega \cdot \varphi(x) + c, \log it(p) = \log \frac{p}{1-p}, \quad (4)$$

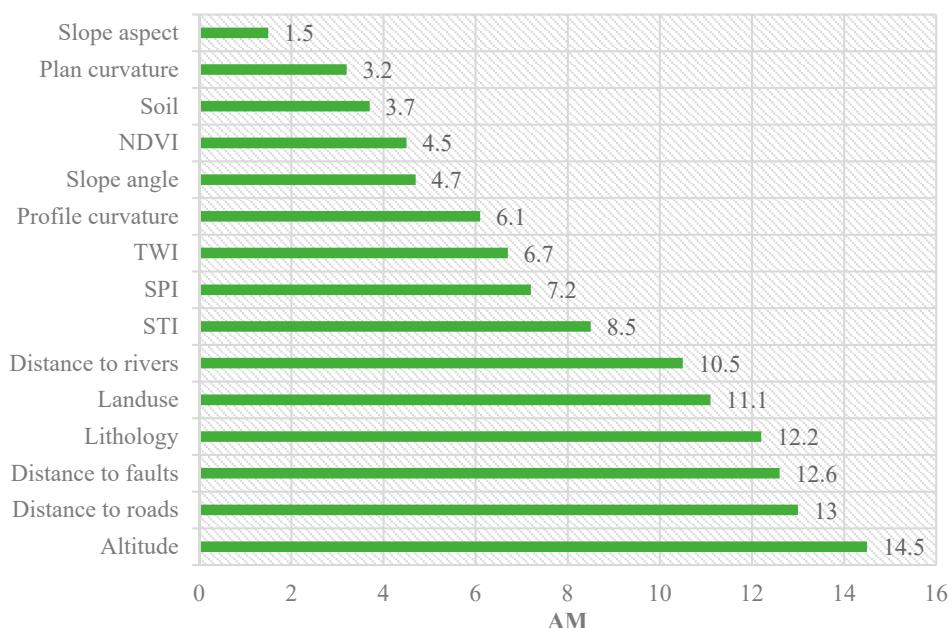
where  $(\omega, c)$  are the optimal model parameters.

## 4. Results

#### 4.1. Selection of Landslide Conditioning Factors

There is no global guideline for the selection of landslide conditioning factors, but they should be selected based on the characteristics of the case study area, data availability, and literature review [58]. One of the most important steps in spatial prediction modeling is the selection of appropriate and effective conditioning factors among all available factors. Not all conditioning factors have an equal impact on landslide occurrences; thus, the most and least important conditioning factors must be identified and the least effective factors should be removed from modeling, as they can reduce the prediction power of the models.

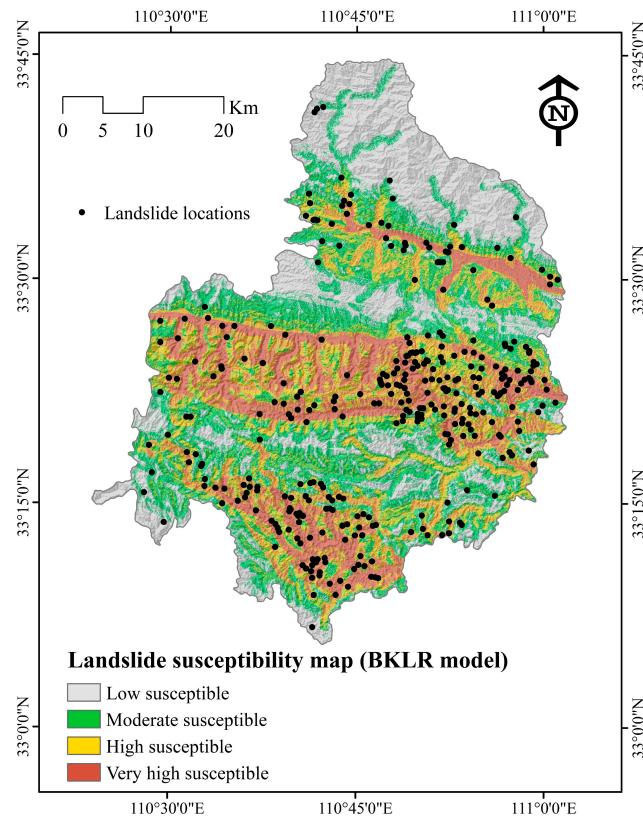
In the present study, the least support vector machine (LSSVM) [59] and average merit (AM) criteria [60,61] were applied for the selection of the proper conditioning factors as well as determining their importance. The results showed that all selected conditioning factors have an impact on landslide occurrences in the Shangnan area, China. Altitude has the greatest impact on landslide occurrences ( $AM = 14.5$ ), followed by distance to road ( $AM = 13$ ), distance to fault ( $AM = 12.6$ ), lithology ( $AM = 12.6$ ), land use ( $AM = 11.1$ ), distance to river ( $AM = 10.5$ ), STI ( $AM = 8.5$ ), SPI ( $AM = 7.2$ ), TWI ( $AM = 6.7$ ), profile curvature ( $AM = 6.1$ ), slope angle ( $AM = 4.7$ ), NDVI ( $AM = 4.5$ ), soil ( $AM = 3.7$ ), plan curvature ( $AM = 3.2$ ), and slope aspect ( $AM = 1.5$ ) (Figure 4). Thus, the most effective conditioning factors are altitude, distance to road, and distance to fault, which aligns with results of Tien Bui et al. [33] and Chen et al. [39]. As all 15 landslide conditioning factors positively contribute in the study area, all were used for further analysis.



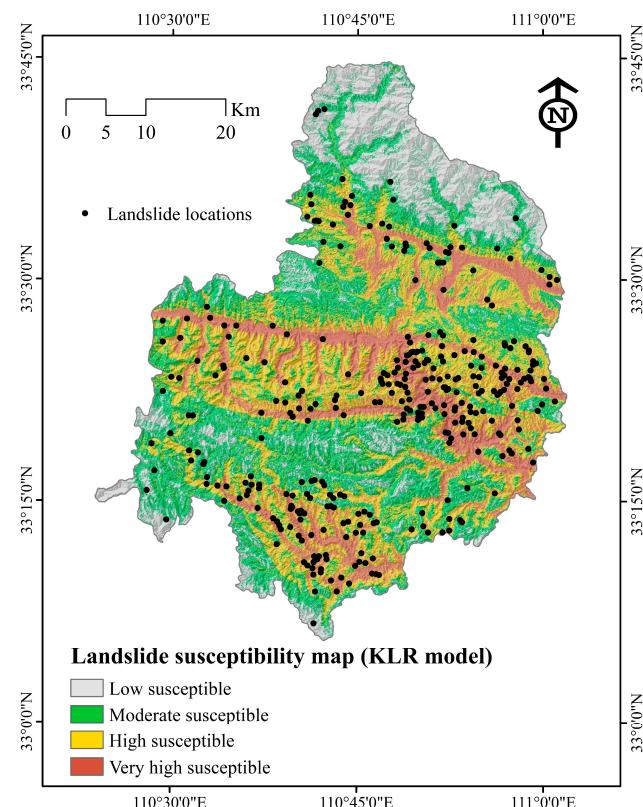
**Figure 4.** Importance of conditioning factors.

#### 4.2. Generation of Landslide Susceptibility Maps

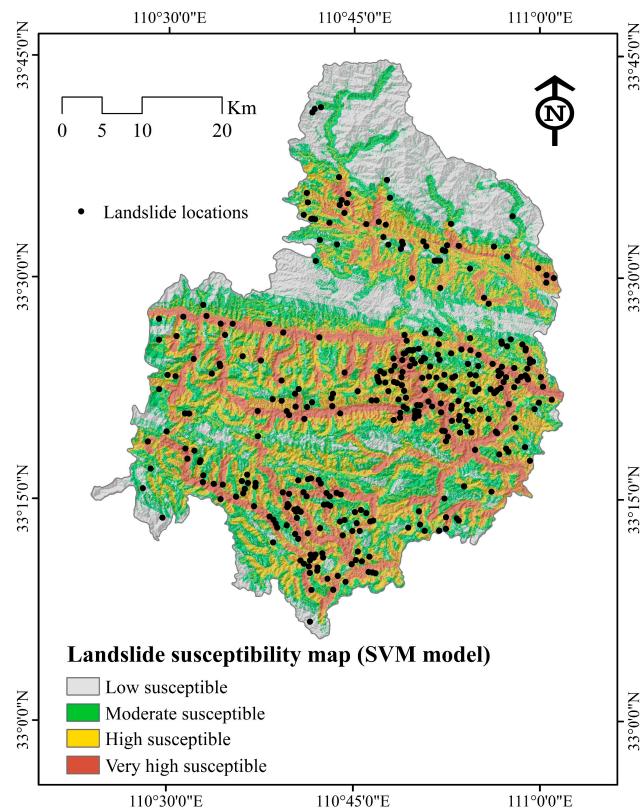
The final model results produce a landslide susceptibility map, which can be used as an effective tool for managing future landslide occurrences. These maps were constructed following these steps. Firstly, the whole study area was converted to pixels using ArcGIS (Esri, CA, USA) software. Then, the training dataset was trained in a training phase, then all pixels were predicted based on the learned trend and assigned in unique indexes. Next, all indexes were classified based on the natural breaks (Jenks) classification scheme [62,63]. Finally, these indexes for the three models were classified into four classes: Low, moderate, high, and very high susceptibility (Figures 5–7). The percentage of area covered by low, moderate, high, and very high susceptibility for the three models is shown in Figure 8. For the BLRK model, these areas are 28.94%, 26.86%, 23.58%, and 20.63%, respectively. For the KLR model, they are 17.68%, 32.03%, 32.53%, and 17.77%, respectively, and for SVM, the area percentages are 24.59%, 28.29%, 27.50%, and 19.62%, respectively.



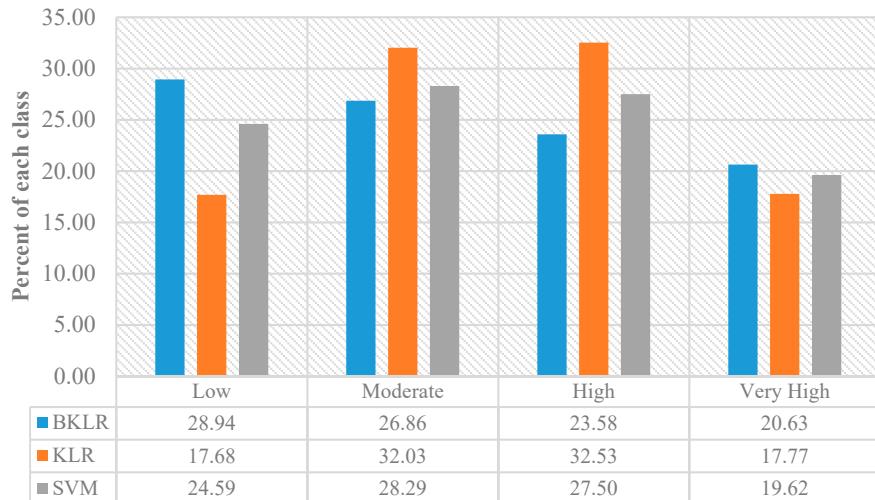
**Figure 5.** Landslide susceptibility map using the bagging-based two-class kernel logistic regression (BKLR) model.



**Figure 6.** Landslide susceptibility map using the kernel logistic regression (KLR) model.



**Figure 7.** Landslide susceptibility map using the support vector machines (SVM) model.



**Figure 8.** Distribution of landslide susceptibility zones.

#### 4.3. Validation and Comparison of Models

The last and final step in landslide susceptibility modeling is validation of the achieved results and maps, as the results of modeling have no scientific significance without validation [64,65]. Model performance and validation in training and testing phases were evaluated using statistical criteria (positive predictive rate (PPR), negative predictive rate (NPR), sensitivity, specificity, accuracy, kappa index, and RMSE), the results of which are shown in Tables 2 and 3. The results of the model performance evaluation shows that BKLR performed the best ( $PPR = 0.785$ ) in terms of positive predictive rate, followed by SVM ( $PPR = 0.718$ ) and KLR ( $PPR = 0.692$ ). In the case of the negative predictive rate, BKLR ( $NPR = 0.777$ ) was superior to SVM ( $NPR = 0.699$ ) and KLR ( $NPR = 0.681$ ). The highest sensitivity was found for BKLR (0.774), with 77.4% of the landslide positions were classified

as landslides, followed by SVM (0.686) and KLR (0.672). BKLR had the highest specificity (0.788), followed by SVM (0.730) and KLR (0.701). The BKLR model classified 78.8% of the non-landslide locations as non-landslide classes. In terms of accuracy, BKLR (0.781) was superior to SVM (0.708) and KLR (0.686). The BKLR model correctly classified 78.1% of the landslide and non-landslide pixels. In terms of the kappa index, the BKLR model performed best (0.56), demonstrating a substantial to almost perfect agreement between prediction and observation, followed by SVM (0.41) and KLR (0.37). In terms of the RMSE, BKLR had the lowest value (0.31), followed by SVM (0.46) and KLR (0.50). Generally, the BKLR model, in terms of all criteria, had the best performance in the training phase, followed by SVM and KLR. The model evaluation in testing phase is shown in Table 3. Generally, the results show that BKLR, for both the training and testing phases, performed the best, followed by the SVM and KLR models.

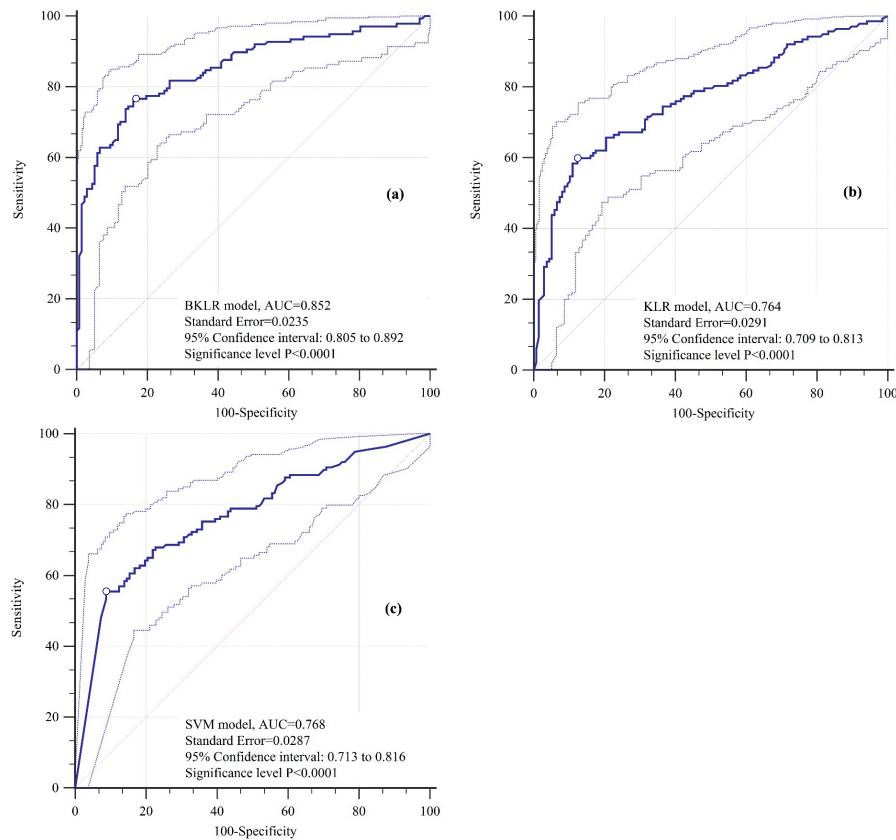
**Table 2.** Model performance comparison.

Parameter	Model		
	BKLR	KLR	SVM
True positive	189	164	167
True negative	192	171	178
False positive	52	73	66
False negative	55	80	77
Positive predictive rate (%)	0.785	0.692	0.718
Negative predictive rate (%)	0.777	0.681	0.699
Sensitivity (%)	0.774	0.672	0.686
Specificity (%)	0.788	0.701	0.730
Accuracy (%)	0.781	0.686	0.708
Kappa index	0.562	0.372	0.416
RMSE	0.391	0.500	0.463

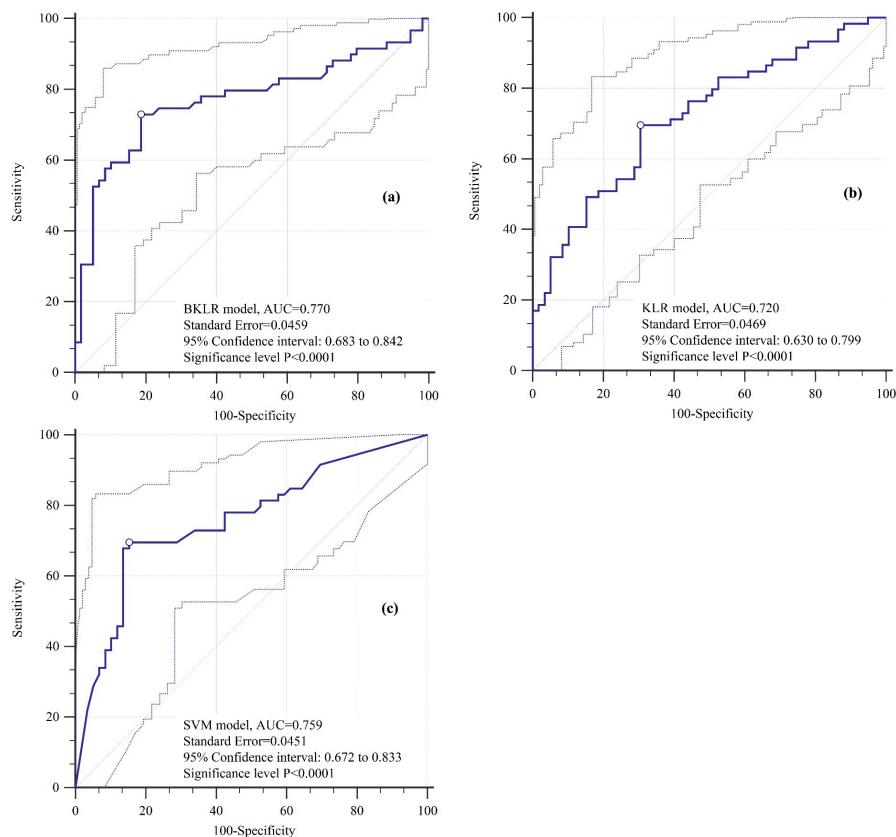
**Table 3.** Model validation.

Parameter	Model		
	BKLR	KLR	SVM
True positive	67	56	62
True negative	90	76	90
False positive	14	28	14
False negative	37	48	42
Positive predictive rate (%)	0.826	0.667	0.814
Negative predictive rate (%)	0.708	0.614	0.680
Sensitivity (%)	0.644	0.542	0.593
Specificity (%)	0.864	0.729	0.864
Accuracy (%)	0.754	0.636	0.729
Kappa index	0.509	0.371	0.458
RMSE	0.439	0.506	0.470

In the present research, the prediction capability of the three models and their results were also investigated using a popular technique: Receiver operating characteristics (ROC). The main advantage of this technique is that it determines the prediction power of the models quantitatively using area under the ROC curve (AUC) [66,67]. The AUC result using the training dataset shows that all three models have good and reasonable prediction power. The BKLR model had the highest prediction power (AUC = 0.852), followed by SVM (AUC = 0.768) and KLR (AUC = 0.764) (Figure 9). As the achieved maps were built using the training dataset, the ROC result using the training dataset cannot be considered alone for comparison. The testing dataset, which was not used for model building, must be considered for model validation and comparison. The results show that BKLR (AUC = 0.770) had the highest prediction power, followed by SVM (AUC = 0.759) and KLR (AUC = 0.720) (Figure 10). According to Yesilnacar [68], all the models have good and reasonable prediction power.



**Figure 9.** Receiver operating characteristics (ROC) curves using the training dataset: (a) BKLR model, (b) KLR model, (c) SVM model.



**Figure 10.** ROC curves using the validation dataset: (a) BKLR model, (b) KLR model, (c) SVM model.

## 5. Discussion

In this study, a novel bagging-based kernel logistic regression classification model was proposed and applied for landslide susceptibility in Shangnan County, China. A total of 15 conditioning factors were selected to create landslide susceptibility models. The results showed that all the conditioning factors had positive effects on landslide occurrence with different degrees of influence according to the AM values based on LSSVM. Altitude, distance to roads, and distance to faults received the highest AM values, in agreement with some existing reports [33,39]. Altitude comprehensive controls the local climate and vegetation development characteristics [69], which are related to the stability of slopes [70]. The AM values for lithology, land use, and distance to rivers were both larger than 10.0 as well. Generally, the strength of a slope body varies with different lithology types [71]. For areas with different land use types, the corresponding physical and mechanical characteristics of soils and rocks have notable differences [44]. Rivers can affect the hydro-geological conditions of slopes, which have strong connections with landslide occurrence [39]. On the basis of relevant studies, STI, SPI, and TWI are usually indispensable conditioning factors in landslide susceptibility modeling [23,66,72]. Curvature, which includes profile curvature and plan curvature, can reflect the geometric features of slopes, and the geometric features can influence stress distribution of slopes [30,73]. For slope angle, landslides generally occur in a certain range of slope angles, and slope angle is another critical conditioning factor [45]. The relationships between the other factors (NDVI, soil type, and slope aspect) and landslide occurrence have also been analyzed by other researchers [74–76].

To evaluate the performance of BKLR, KLR, and SVM models in landslide susceptibility mapping, ROC curves and AUC values of various models were obtained. The results confirmed that all three models were reasonable for landslide susceptibility mapping, and the BKLR model had the best accuracy and prediction capacity. The results also verified that bagging is an effective tool to increase the prediction accuracy of a single classifier by creating an ensemble learning classifier [77]. In this case, bagging improved the stability of the conventional KLR model using the RBF kernel function. The sensitivity of a single classifier to noise in a dataset could be decreased by bootstrap sampling, so the variance of the classification model decreases correspondingly [53].

There are some advantages of the BKLR: This model can produce class probabilities, whereas SVM is a deterministic classifier. BKLR is also recommended for future application to the study area with new techniques and models or new hybrid models, and the most appropriate model can be selected for future modeling. As a practical recommendation, we propose that the government and decision makers must complete an extensive field survey to confirm the results of the present study (and especially the areas that have high and very high susceptibility to landslide occurrences). According to case study characteristics, policy makers can make an informed decision for future landslide hazard mitigation.

## 6. Conclusions

The main aim of this study was to construct a new hybrid model using bagging-based kernel logistic regression (KLR), called BKLR, for spatial prediction of landslides in the Shangnan area in China. The prediction capability of 15 conditioning factors was assessed for the modeling process using the LSSVM algorithm. The performance of the new hybrid model was evaluated using several popular statistical measures, including PPR, NPR, sensitivity, specificity, the kappa index, and RMSE. A landslide susceptibility map was produced using the hybrid model and it was compared with the susceptibility maps produced by the single KLR model and SVM algorithm.

Results indicated that the BKLR ensemble model had the best goodness-of-fit and prediction power for landslide susceptibility mapping, followed by the SVM and KLR models. The BKLR ensemble model can be used as a promising technique for spatial prediction of landslides in the current study area, and can be applied to other similar landslide prone areas for better landslide susceptibility mapping. Hybrid modeling is an efficient technique for improving the predictive capability of weak individual classifiers.

**Author Contributions:** W.C., H.S., S.Z., K.K., A.S., K.C., B.T., P., T.Z., L.Z., H.C., J.M., Y.C., X.W., R.L., and B.B.A. contributed equally to the work. T.Z., L.Z., H.C., J.M., Y.C., X.W., and R.L. collected field data and conducted the landslide mapping and analysis. W.C., H.S., S.Z., K.K., A.S., and K.C. wrote the manuscript. W.C., H.S., K.K., A.S., K.C., B.T., P., and B.B.A. provided critical comments in planning this paper and edited the manuscript. All the authors discussed the results and edited the manuscript.

**Acknowledgments:** The authors wish to thank Lufei Yang (Northwest Nonferrous Survey and Engineering Company) for useful information provided. This study has received financial support from the National Natural Science Foundation of China (Grant Nos. 41807192, 41602359, 41602212), China Postdoctoral Science Foundation (Grant Nos. 2018T111084, 2017M613168), Project funded by Shaanxi Province Postdoctoral Science Foundation (Grant No. 2017BSHYDZZ07), the Doctoral Scientific Research Foundation of Xi'an University of Science and Technology (Grant No. 2013QDJ038), and the Universiti Teknologi Malaysia (UTM) based on a Research University Grant (Q.J130000.2527.17H84).

**Conflicts of Interest:** No potential conflict of interest is reported by the authors.

## References

1. Malamud, B.D.; Turcotte, D.L.; Guzzetti, F.; Reichenbach, P. Landslide inventories and their statistical properties. *Earth Surf. Process. Landf.* **2004**, *29*, 687–711. [[CrossRef](#)]
2. Glade, T. *The Temporal and Spatial Occurrence of Rainstorm-Triggered Landslide Events in New Zealand: An Investigation into the Frequency, Magnitude and Characteristics of Landslide Events and Their Relationship with Climatic and Terrain Characteristics: A Thesis Submitted [to the] Victoria University of Wellington in Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Physical Geography*; Victoria University of Wellington: Kelburn, New Zealand, 1997.
3. Mohammady, M.; Pourghasemi, H.R.; Pradhan, B. Landslide susceptibility mapping at golestan province, iran: A comparison between frequency ratio, dempster–shafer, and weights-of-evidence models. *J. Asian Earth Sci.* **2012**, *61*, 221–236. [[CrossRef](#)]
4. Sassa, K.; Canuti, P. *Landslides-Disaster Risk Reduction*; Springer Science & Business Media: Berlin, Germany, 2008.
5. Quan-min, X.; Xiang, B.; Yuan-you, X. Systematic analysis of risk evaluation of landslide hazard. *Rock Soil Mech.* **2005**, *26*, 71–74.
6. Pradhan, B.; Lee, S. Delineation of landslide hazard areas on penang island, malaysia, by using frequency ratio, logistic regression, and artificial neural network models. *Environ. Earth Sci.* **2010**, *60*, 1037–1054. [[CrossRef](#)]
7. Tien Bui, D.; Lofman, O.; Revhaug, I.; Dick, O. Landslide susceptibility analysis in the hoa binh province of vietnam using statistical index and logistic regression. *Nat. Hazards* **2011**, *59*, 1413–1444.
8. Aditian, A.; Kubota, T.; Shinohara, Y. Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of ambon, Indonesia. *Geomorphology* **2018**, *318*, 101–111. [[CrossRef](#)]
9. Ding, Q.; Chen, W.; Hong, H. Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping. *Geocarto Int.* **2017**, *32*, 619–639. [[CrossRef](#)]
10. Zhang, Z.; Yang, F.; Chen, H.; Wu, Y.; Li, T.; Li, W.; Wang, Q.; Liu, P. GIS-based landslide susceptibility analysis using frequency ratio and evidential belief function models. *Environ. Earth Sci.* **2016**, *75*, 1–12. [[CrossRef](#)]
11. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Alizadeh, M.; Chen, W.; Mohammadi, A.; Ahmad, B.; Panahi, M.; Hong, H.; et al. Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in cameron highlands, malaysia. *Remote Sens.* **2018**, *10*, 1527. [[CrossRef](#)]
12. Chen, W.; Pourghasemi, H.R.; Naghibi, S.A. A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. *Bull. Eng. Geol. Environ.* **2018**, *77*, 647–664. [[CrossRef](#)]
13. Tsangaratos, P.; Ilia, I.; Hong, H.; Chen, W.; Xu, C. Applying information theory and gis-based quantitative methods to produce landslide susceptibility maps in nancheng county, China. *Landslides* **2017**, *14*, 1091–1111. [[CrossRef](#)]
14. Youssef, A.M.; Al-Kathery, M.; Pradhan, B. Landslide susceptibility mapping at al-hascher area, jizan (saudi arabia) using GIS-based frequency ratio and index of entropy models. *Geosci. J.* **2015**, *19*, 113–134. [[CrossRef](#)]

15. Devkota, K.C.; Regmi, A.D.; Pourghasemi, H.R.; Yoshida, K.; Pradhan, B.; Ryu, I.C.; Dhital, M.R.; Althuwaynee, O.F. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in gis and their comparison at mugling–narayanghat road section in nepal himalaya. *Nat. Hazards* **2013**, *65*, 135–165. [[CrossRef](#)]
16. Roşian, G.; Csaba, H.; KingaOlga, R.; Botan, C.; Gavrilă, I.G. Assessing landslide vulnerability using bivariate statistical analysis and the frequency ratio model. Case study: Transylvanian plain (Romania). *Z. Geomorphol.* **2016**, *60*, 359–371. [[CrossRef](#)]
17. Felicísmo, Á.M.; Cuartero, A.; Remondo, J.; Quirós, E. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: A comparative study. *Landslides* **2013**, *10*, 175–189. [[CrossRef](#)]
18. Shirzadi, A.; Chapi, K.; Shahabi, H.; Solaimani, K.; Kavian, A.; Ahmad, B.B. Rock fall susceptibility assessment along a mountainous road: An evaluation of bivariate statistic, analytical hierarchy process and frequency ratio. *Environ. Earth Sci.* **2017**, *76*, 152. [[CrossRef](#)]
19. Pourghasemi, H.R.; Rossi, M. Landslide susceptibility modeling in a landslide prone area in mazandarn province, north of Iran: A comparison between GLM, GAM, MARS, and M-AHP methods. *Theor. Appl. Climatol.* **2016**, *130*, 609–633. [[CrossRef](#)]
20. Nicu, I.C. Application of analytic hierarchy process, frequency ratio, and statistical index to landslide susceptibility: An approach to endangered cultural heritage. *Environ. Earth Sci.* **2018**, *77*, 79. [[CrossRef](#)]
21. Razavizadeh, S.; Solaimani, K.; Massironi, M.; Kavian, A. Mapping landslide susceptibility with frequency ratio, statistical index, and weights of evidence models: A case study in northern Iran. *Environ. Earth Sci.* **2017**, *76*, 499. [[CrossRef](#)]
22. Althuwaynee, O.F.; Pradhan, B.; Park, H.-J.; Lee, J.H. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *Catena* **2014**, *114*, 21–36. [[CrossRef](#)]
23. Pourghasemi, H.R.; Kerle, N. Random forests and evidential belief function-based landslide susceptibility assessment in western mazandaran province, iran. *Environ. Earth Sci.* **2016**, *75*, 185. [[CrossRef](#)]
24. Prefac, Z.; Dumitru, S.; Chendeş, V.; Sîrodoev, I.; Cracu, G. Assessment of landslide susceptibility using the certainty factor model: Răcăuţa catchment (curvature subcarpathians) case study. *Carpath. J. Earth Environ. Sci.* **2016**, *11*, 617–626.
25. Chen, W.; Xie, X.; Peng, J.; Shahabi, H.; Hong, H.; Tien Bui, D.; Duan, Z.; Li, S.; Zhu, A.-X. Gis-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *CATENA* **2018**, *164*, 135–149. [[CrossRef](#)]
26. Chen, W.; Pourghasemi, H.R.; Zhao, Z. A gis-based comparative study of dempster-shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. *Geocarto Int.* **2017**, *32*, 367–385. [[CrossRef](#)]
27. Erfanian, M.; Farajollahi, H.; Souri, M.; Shirzadi, A. Comparing the efficiency of weight of evidence, logistic regression and frequency ratio methods for mapping groundwater spring potential in ghelgazi watershed, kordestan province of iran. *JWSS-Isfahan Univ. Technol.* **2016**, *20*, 59–72. [[CrossRef](#)]
28. Shahabi, H.; Hashim, M.; Ahmad, B.B. Remote sensing and gis-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central zab basin, Iran. *Environ. Earth Sci.* **2015**, *73*, 8647–8668. [[CrossRef](#)]
29. Mandal, S.; Mandal, K. Modeling and mapping landslide susceptibility zones using gis based multivariate binary logistic regression (lr) model in the rorachu river basin of eastern Sikkim Himalaya, India. *Model. Earth Syst. Environ.* **2018**, *4*, 69–88. [[CrossRef](#)]
30. 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](#)]
31. Oh, H.J.; Pradhan, B. Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. *Comput. Geosci.* **2011**, *37*, 1264–1276. [[CrossRef](#)]
32. Chen, W.; Panahi, M.; Tsangaratos, P.; Shahabi, H.; Ilia, I.; Panahi, S.; Li, S.; Jaafari, A.; Ahmad, B.B. Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. *CATENA* **2019**, *172*, 212–231. [[CrossRef](#)]

33. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378. [[CrossRef](#)]
34. Ngadisih Bhandary, N.P.; Yatabe, R.; Dahal, R.K. Logistic Regression and Artificial Neural Network Models for Mapping of Regional-Scale Landslide Susceptibility in Volcanic Mountains of West Java (Indonesia). In Proceedings of the International Symposium on Earthhazard and Disaster Mitigation: The Symposium on Earthquake and Related Geohazard Research for Disaster Risk Reduction, Bandung, Indonesia, 11–12 October 2016; p. 060001.
35. Chen, W.; Shahabi, H.; Shirzadi, A.; Hong, H.; Akgun, A.; Tian, Y.; Liu, J.; Zhu, A.-X.; Li, S. Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* **2018**, *1*–23. [[CrossRef](#)]
36. Zabihi, M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Behzadfar, M. Gis-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in iran. *Environ. Earth Sci.* **2016**, *75*, 1–19. [[CrossRef](#)]
37. Nefeslioglu, H.; Sezer, E.; Gokceoglu, C.; Bozkir, A.; Duman, T. Assessment of landslide susceptibility by decision trees in the metropolitan area of istanbul, turkey. *Math. Probl. Eng.* **2010**, *2010*. [[CrossRef](#)]
38. Hong, H.; Pradhan, B.; Xu, C.; Bui, D.T. Spatial prediction of landslide hazard at the yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena* **2015**, *133*, 266–281. [[CrossRef](#)]
39. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Tien Bui, D.; Duan, Z.; Ma, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *CATENA* **2017**, *151*, 147–160. [[CrossRef](#)]
40. Chen, W.; Xie, X.; Peng, J.; Wang, J.; Duan, Z.; Hong, H. Gis-based landslide susceptibility modelling: A comparative assessment of kernel logistic regression, naïve-bayes tree, and alternating decision tree models. *Geomat. Nat. Hazards Risk* **2017**, *8*, 950–973. [[CrossRef](#)]
41. Huang, Y.; Zhao, L. Review on landslide susceptibility mapping using support vector machines. *Catena* **2018**, *165*, 520–529. [[CrossRef](#)]
42. Lin, G.F.; Chang, M.J.; Huang, Y.C.; Ho, J.Y. Assessment of susceptibility to rainfall-induced landslides using improved self-organizing linear output map, support vector machine, and logistic regression. *Eng. Geol.* **2017**, *224*, 62–74. [[CrossRef](#)]
43. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Al-Katheeri, M.M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* **2016**, *13*, 839–856. [[CrossRef](#)]
44. Chen, W.; Pourghasemi, H.R.; Panahi, M.; Kornejady, A.; Wang, J.; Xie, X.; Cao, S. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. *Geomorphology* **2017**, *297*, 69–85. [[CrossRef](#)]
45. Nasiri Aghdam, I.; Varzandeh, M.H.M.; Pradhan, B. Landslide susceptibility mapping using an ensemble statistical index (Wi) and adaptive neuro-fuzzy inference system (ANFIS) model at alborz mountains (Iran). *Environ. Earth Sci.* **2016**, *75*, 553. [[CrossRef](#)]
46. Tien Bui, D.; 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. [[CrossRef](#)]
47. Pham, B.T.; Bui, D.T.; Pourghasemi, H.R.; Indra, P.; Dholakia, M. Landslide susceptibility assessment in the uttarakhand area (India) using GIS: A comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theor. Appl. Climatol.* **2015**, *1*–19. [[CrossRef](#)]
48. Tsangaratos, P.; Ilia, I. Comparison of a logistic regression and naïve bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. *CATENA* **2016**, *145*, 164–179. [[CrossRef](#)]
49. Tien Bui, D.; Pradhan, B.; Revhaug, I.; Nguyen, D.B.; Pham, H.V.; Bui, Q.N. A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the lang son city area (Vietnam). *Geomat. Nat. Hazards Risk* **2015**, *6*, 243–271.

50. Chen, W.; Shirzadi, A.; Shahabi, H.; Ahmad, B.B.; Zhang, S.; Hong, H.; Zhang, N. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve bayes tree classifiers for a landslide susceptibility assessment in langao county, china. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1955–1977. [[CrossRef](#)]
51. Zhou, C.; Yin, K.; Cao, Y.; Ahmed, B.; Li, Y.; Catani, F.; Pourghasemi, H.R. Landslide susceptibility modeling applying machine learning methods: A case study from longju in the three gorges reservoir area, China. *Comput. Geosci.* **2018**, *112*, 23–37. [[CrossRef](#)]
52. Süzen, M.L.; Kaya, B.S. Evaluation of environmental parameters in logistic regression models for landslide susceptibility mapping. *Int. J. Digit. Earth* **2012**, *5*, 338–355. [[CrossRef](#)]
53. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
54. Mbogning, C.; Broët, P. Bagging survival tree procedure for variable selection and prediction in the presence of nonsusceptible patients. *BMC Bioinform.* **2016**, *17*, 1–21. [[CrossRef](#)] [[PubMed](#)]
55. Bühlmann, P. Bagging, boosting and ensemble methods. In *Handbook of Computational Statistics*; Springer: Berlin, Germany, 2012; pp. 985–1022.
56. Sugiyama, M.; Simm, J. A computationally-efficient alternative to kernel logistic regression. In Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing, Tokyo, Japan, 29 August–1 September 2010; pp. 124–129.
57. Mercer, J. Functions of positive and negative type, and their connection with the theory of integral equations. *Philos. Trans. R. Soc. London Ser. A Contain. Pap. A Math. Phys. Character* **1909**, *209*, 415–446. [[CrossRef](#)]
58. Chen, W.; Panahi, M.; Pourghasemi, H.R. Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. *CATENA* **2017**, *157*, 310–324. [[CrossRef](#)]
59. Isabelle, G.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
60. Witten, I.H.; Frank, E.; Mark, A.H. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan kaufmann: Burlington, MA, USA, 2011.
61. Chen, W.; Li, H.; Hou, E.; Wang, S.; Wang, G.; Panahi, M.; Li, T.; Peng, T.; Guo, C.; Niu, C.; et al. Gis-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Sci. Total Environ.* **2018**, *634*, 853–867. [[CrossRef](#)]
62. Pourghasemi, H.R.; Rahmati, O. Prediction of the landslide susceptibility: Which algorithm, which precision? *CATENA* **2018**, *162*, 177–192. [[CrossRef](#)]
63. Pham, B.T.; Prakash, I.; Bui, D.T. Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* **2018**, *303*, 256–270. [[CrossRef](#)]
64. Chung, C.-J.F.; Fabbri, A.G. Validation of spatial prediction models for landslide hazard mapping. *Nat. Hazards* **2003**, *30*, 451–472. [[CrossRef](#)]
65. Chen, W.; Zhang, S.; Li, R.; Shahabi, H. Performance evaluation of the gis-based data mining techniques of best-first decision tree, random forest, and naïve bayes tree for landslide susceptibility modeling. *Sci. Total Environ.* **2018**, *644*, 1006–1018. [[CrossRef](#)]
66. Chen, W.; Yan, X.; Zhao, Z.; Hong, H.; Bui, D.T.; Pradhan, B. Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive bayes and rbfnetwork models for the long county area (China). *Bull. Eng. Geol. Environ.* **2018**, *1*–20. [[CrossRef](#)]
67. Chen, W.; Shahabi, H.; Shirzadi, A.; Li, T.; Guo, C.; Hong, H.; Li, W.; Pan, D.; Hui, J.; Ma, M.; et al. A novel ensemble approach of bivariate statistical-based logistic model tree classifier for landslide susceptibility assessment. *Geocarto Int.* **2018**, *33*, 1398–1420. [[CrossRef](#)]
68. Yesilnacar, E. The Application of Computational Intelligence to Landslide Susceptibility Mapping in Turkey. Ph.D. Thesis, Department of Geomatics, University of Melbourne, Victoria, Australia, 2005; p. 423.
69. Gruber, S.; Haeberli, W. Permafrost in steep bedrock slopes and its temperature-related destabilization following climate change. *J. Geophys. Res. Earth Surf.* **2007**, *112*. [[CrossRef](#)]
70. Buma, J.; Dehn, M. A method for predicting the impact of climate change on slope stability. *Environ. Geol.* **1998**, *35*, 190–196. [[CrossRef](#)]
71. Zeng, Z.; Wang, H. Recognition of lithology and its use in identification of landslide-prone areas using remote sensing data. In *Landslide Disaster Mitigation in Three Gorges Reservoir, China*; Wang, F., Li, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 487–496.

72. Chen, W.; Peng, J.; Hong, H.; Shahabi, H.; Pradhan, B.; Liu, J.; Zhu, A.X.; Pei, X.; Duan, Z. Landslide susceptibility modelling using gis-based machine learning techniques for chongren county, Jiangxi province, China. *Sci. Total Environ.* **2018**, *626*, 1121–1135. [[CrossRef](#)] [[PubMed](#)]
73. Ayalew, L.; Yamagishi, H.; Ugawa, N. Landslide susceptibility mapping using GIS-based weighted linear combination, the case in tsugawa area of agano river, Niigata prefecture, Japan. *Landslides* **2004**, *1*, 73–81. [[CrossRef](#)]
74. Sasaki, Y. *Shallow Landslide Process and Hazard Mapping Using a Soil Strength Probe*; Engineering Geology for Society and Territory—Volume 2; Lollino, G., Giordan, D., Crosta, G.B., Corominas, J., Azzam, R., Wasowski, J., Sciarra, N., Eds.; Springer International Publishing: Cham, French, 2015; pp. 957–960.
75. Magliulo, P.; Di Lisio, A.; Russo, F.; Zelano, A. Geomorphology and landslide susceptibility assessment using gis and bivariate statistics: A case study in southern italy. *Nat. Hazards* **2008**, *47*, 411–435. [[CrossRef](#)]
76. Gokceoglu, C. Discussion on “combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using aster images and gis” by choi et al. (2012), engineering geology, 124, 12–23. *Eng. Geol.* **2012**, *129–130*, 104–105. [[CrossRef](#)]
77. Pham, B.T.; Prakash, I. A novel hybrid model of bagging-based naïve bayes trees for landslide susceptibility assessment. *Bull. Eng. Geol. Environ.* **2017**. [[CrossRef](#)]



© 2018 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/>).