

Article

Improving the Accuracy of Landslide Detection in “Off-site” Area by Machine Learning Model Portability Comparison: A Case Study of Jiuzhaigou Earthquake, China

Qiao Hu ^{1,2}, Yi Zhou ^{1,*}, Shixing Wang ¹, Futao Wang ¹ and Hongjie Wang ^{1,2}

¹ Airspace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; huqiao@radi.ac.cn (Q.H.); WangSX@irsa.ac.cn (S.W.); wangft@aircas.ac.cn (F.W.); wanghj@radi.ac.cn (H.W.)

² University of Chinese Academy of Science, Beijing 100049, China

* Correspondence: zhouyi@radi.ac.cn; Tel.: +86-136-7116-9922

Received: 14 September 2019; Accepted: 28 October 2019; Published: 29 October 2019



Abstract: The rising machine learning (ML) models have become the preferred way for landslide detection based on remote sensing images, but the performance of these models in a sample-free area are rarely concerned in many studies. In this study, we used a cross-validation method (training model in one area and validation in another) to compare the model portability of trained ML models applied in an “off-site” area, as a consideration of the landslide detection ability of these models in sample-free areas. We integrate nighttime light imagery, multi-seasonal optical Landsat time-series and digital elevation data, and we employed support vector machines (SVM), artificial neural networks (ANN) and random forest (RF) models to classify the satellite imagery and identify landslides. Samples of two scenarios generated from two subareas of the Jiuzhaigou disaster-stricken region are used for the cross-application and accuracy evaluation of three ML models. The results revealed that when the trained models are applied in areas outside those in which they were developed, the landslide identification accuracy of these three models has declined. Especially for the SVM and ANN models, the accuracy is greatly reduced and there appears a seriously imbalanced user’s and producer’s accuracy. However, although the performance of the RF model is lower than that of SVM and ANN models in their local area, the RF model exhibits stable portability, and retains the original performance and achieves a satisfactory balance between overestimation and underestimation in “off-site” areas. An additional validation from a new area proved that the landslide detection performance of the RF model with stable portability is higher than that of the SVM and ANN models in “off-site” areas. The results suggest that evaluating the model portability through cross-application can be a useful way to determine the most suitable model for landslide detection in “off-site” areas with a similar geographic environment to model development areas, so as to maximize the accuracy of landslide detection based on limited samples.

Keywords: landslide detection; model portability; Landsat imagery; nighttime light; support vector machines (SVM); artificial neural networks (ANN); random forest (RF); Jiuzhaigou Earthquake

1. Introduction

Landslides represent a serious hazard in many areas of the world, and result in enormous casualties and severe economic loss every year [1,2]. The compilation of landslide inventories is important for landslide prevention [3,4], geomorphological and erosion studies [5,6] and landslide susceptibility and risk assessment [1,7–9].

Conventional techniques for producing reliable landslide inventories mainly relies on field surveys [10–12] and the visual interpretation of aerial photos [1,13], but these ways are time-costly and labor-intensive [1]. Also, it is difficult to make a complete observation of landslides in densely vegetated areas [14]. With the development of earth observation data, various satellite images are widely considered as the most accessible data providing critical information necessary for supporting landslide detection [15], and the emergence of satellite image processing techniques means that landslides can be detected more accurately, completely and rapidly than ever before [1,16–18]. Landslide detection based on remote sensing images is actually a binary classification of the image units to be processed to determine whether the units belong to the landslide category or non-landslide category; these units can be independent image pixels or image objects composed of pixels with similar features such as spectrum and texture. Therefore, landslide detection techniques on the basis of satellite images can be categorized into being semi-automatic pixel-based [19,20] and object-based [21–23]. The object-based high resolution image classification takes into account the geometry and texture of partitioned objects, and it can achieve high overall accuracy (above 85%) in landslide detection [21,23]. However, the spatial resolution of these images is usually less than 1 m (Quickbird and GeoEye images), and different sequences of applying the multiple rules may lead to different segmentation results, which are hard to define [24]. For a large-scale area, the number of images to be processed is large and time-consuming [25], and that makes the object-based method impractical to apply [26]. In contrast, repetitive observations with dense satellite time-series, such as Landsat (30 m) imagery, are favored by researchers in pixel-based landslide detection [27,28]. Although the shortcomings that the Landsat-like imagery usually lead to a moderate classification accuracy [24,29], this can be effectively overcome by the applications of multi-temporal imagery, which extracts time-series trajectories; incorporating multiple seasonal features, it is used as auxiliary information. For instance, the Normalized Difference Vegetation Index (NDVI) time-series allows separation between permanently non-vegetated and post-event landslide areas in different geographic settings [29–31], and the Normalized Difference Water Index (NDWI), derived from multi-spectral imagery, is commonly used to differentiate landslides from water bodies of mountainous areas [27,32]. In addition, Chen et al. [25] also confirmed that the multi-seasonal change characteristics of the nighttime light index can also separate landslides from anthropogenic activities interlinked with light use.

As far as specific image processing algorithms are concerned, either pixels or objects are used as underlying image units for the purpose of classifying remotely-derived imagery. The information contained within and among these units (e.g., spectral, textural, etc.) can be subjected to a variety of classification algorithms [33], especially the emerging machine learning (ML) methods [15,34], which are considered effective approaches for remote sensing applications with emphasis on image classification and object recognition [35], such as SVM [36,37], ANN [38,39], RF [21,23,40], logistic regression [39,41,42], Bayesian theory [43], Dempster-Shafer theory [14] and neuro-fuzzy classifier [27]. All of these approaches above aim to increase the quality of the landslide detection and limit classification errors which are typical of the landslide maps obtained from the classification of satellite images [1], and the above literature review shows that these models have achieved convincing results in different areas. However, in most studies of the performance of these models, the model training area (model development area) and model prediction area (model application area) are the same; that is, model training and validation samples are generated in the same area; little attention has been paid to the predictive performance of these models to detect landslides in areas outside the model development zone (model portability) [12]. Therefore, one question that needs to be examined is how these models compare with each other in terms of their ability to identify landslides and their portability to areas outside those in which they were developed.

The aim of this study is to compare the widely used ML model with respect to model performance and portability, so as to determine the optimal ML model for “off-site” landslide detection and further improve the detection accuracy of landslides. In view of this, we propose a hypothesis that the model with good portability can also achieve satisfactory landslide detection accuracy in “off-site” areas. In order to achieve these goals, the SVM, ANN and RF models, which have been widely used and demonstrated to provide reliable results of landslide detection, were chosen to represent ML models. There were two subareas in Jiuzhaigou County selected as test regions because of their high landslide hazard, following the Ms 7.0 earthquake (8 August, 2017) and similar geomorphological settings and disaster-causing environments. Cross-applications of the three ML models between the two subareas were conducted to evaluate their respective prediction performance and portability. An additional set of validation data was prepared to validate our hypothesis. If the hypothesis proposed is proven, this cross-application strategy can provide a feasible idea for model portability evaluation, and also recommend the optimal ML model for landslide detection in “off-site” areas with a similar geographic environment as model development areas under the condition of limited time, limited resources and high quality.

2. Study Area

Jiuzhaigou County is located in Sichuan Province, southwestern China, where this area is in the transition zone of the western margin of the Sichuan Basin and the Qinghai–Tibet Plateau, which is part of the Songpan–Ganzi geosynclinal area. The exposed strata in Jiuzhaigou are neritic-littoral carbonate sediments of the Middle Devonian to Middle Triassic age [44]. On 8 August 2017, an Ms = 7.0 earthquake with a focal depth of 20 km struck the Jiuzhaigou World Natural and Cultural Heritage site. The epicenter was in Bimang Valley, at 33.2°N, 103.82°E. The earthquake released a massive amount of energy in the park, and was responsible for triggering thousands of landslides [45].

The study area is located in the most serious disaster zone of Jiuzhaigou County (Figure 1a), and the area lies between latitudes 33.11°N and 33.33°N and longitudes 103.71°E and 103.90°E, covering a total area of 208 km² (Figure 1b). Based on the digital elevation model (DEM) data with a resolution of 30 m, the elevation of this area ranges from 2191 m to 4453 m above sea level, and the maximum and minimum slopes are 75.68° and 0°, respectively, with a mean of 29.68° and standard deviation of 11.26°. Forest and grassland are the main types of landform, and probably account for 56% and 27% of the study area, respectively. The earthquake-induced landslides in this area are mainly of moderate- and small-scales, and also with a few large landslides and avalanches. The most common types of these coseismic landslides are shallow debris flows (Figure 1e,g) and small-scale rock falls (Figure 1h); the shallow debris flows are mostly formed by the further expansion of the activities of the existing landslides.

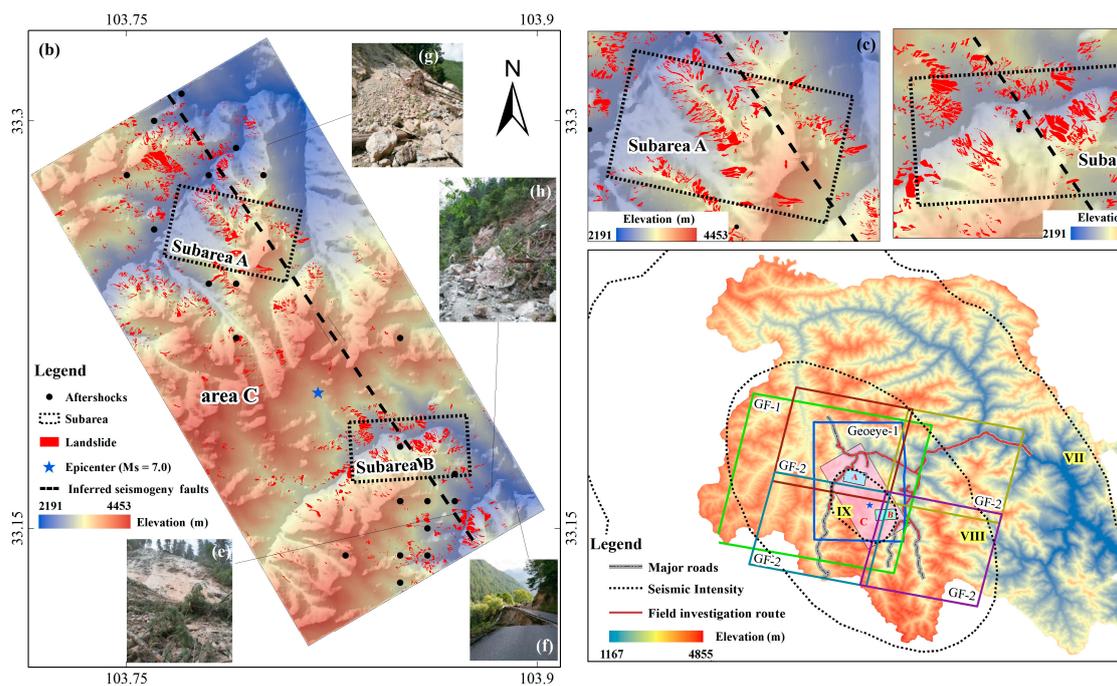


Figure 1. Study area and landslide event-inventory of the Jiuzhaigou earthquake. (a) Image coverage and field investigation route of Jiuzhaigou earthquake; (b) location of study area and landslide; (c) subarea A; (d) subarea B; (e) bedrock-weathered loose layer landslide; (f) subgrade landslide; (g) accumulated landslide; (h) small-scale collapse landslide.

3. Materials and Methods

In this study, we assume that model portability can be used as an important reference for selecting the most suitable prediction model for landslide detection in “off-site” areas, and the cross-application of these models can be used as an effective way for evaluating model portability. Therefore, this study has carried on the following experimental designs (Figure 2). First, we constructed Random Forest (RF), support vector machines (SVM) and artificial neural networks (ANN) models as classifiers of remote sensing images for landslide detection (Part A in Figure 2, Section 3.1). Second, we processed multi-source remote sensing images, and extracted several image features related to landslide identification from Landsat time-series images, the digital elevation model (DEM) and nighttime light data (Part B in Figure 2, Section 3.2.1). Third, we compiled landslide inventory to generate landslide positive samples (Part C1 in Figure 2, Section 3.2.2), and conducting sensitivity analysis to generate a corresponding number of non-landslide negative samples, thus constituting model training and validation data sets (Part C2 in Figure 2, Section 3.3). Fourth, three models were independently trained and validated based on the sample set of subarea A and subarea B, and then the optimal performance model is determined for “off-site” application (Part D1 in Figure 2, Section 4.1). Fifth, the samples in subarea A and subarea B were merged, and the merged sample set is used to train and validate three machine learning (ML) models, after which we then we determined the best model for landslide detection in “off-site” areas (Part D2 in Figure 2, Section 4.2). Sixth, the sample data sets in subareas A and B were prepared for cross-validation and portability evaluation of the RF, SVM and ANN models, respectively, and then the model with the best portability was chosen for landslide detection in the “off-site” area (Part D3 in Figure 2, Section 4.3). Finally, we used the additional validation samples generated from area C to test the predictive performance of the models determined in different sample scenarios to prove our hypothesis that the model with better portability can also perform well for landslide detection in “off-site” areas (Part D4 in Figure 2, Section 4.4).

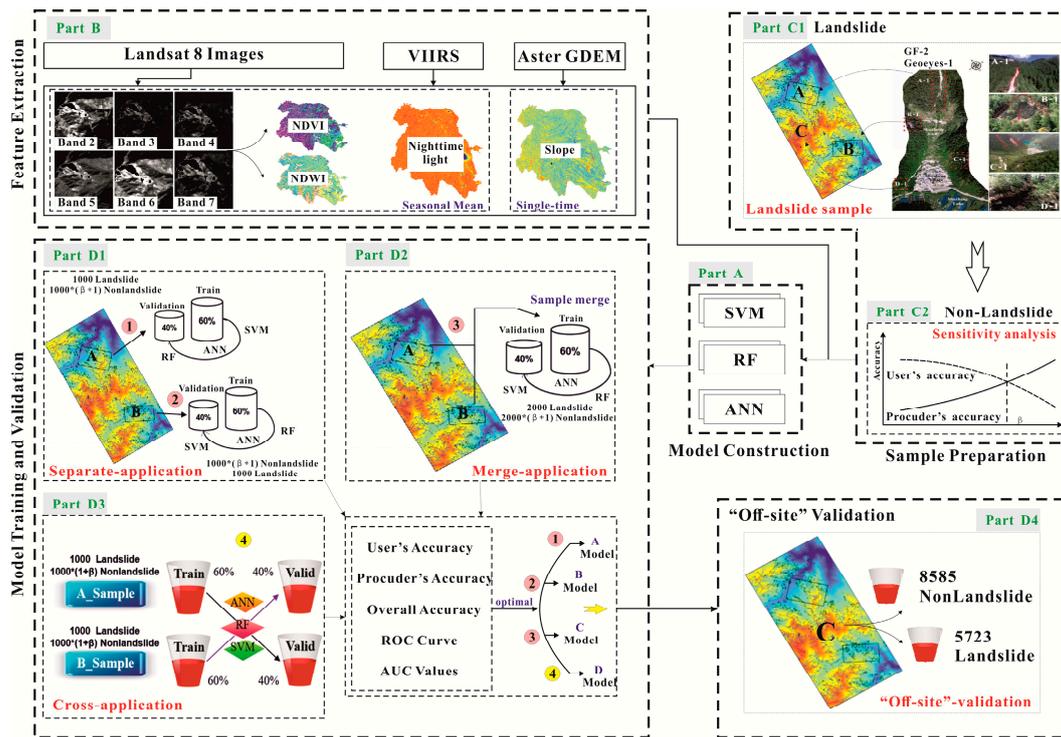


Figure 2. Overall methodology adopted for the analysis.

3.1. Applied ML Model

3.1.1. SVM Model

The SVM model is a machine learning technique based on the discrimination of classes in a high-dimensional feature space, which is generated via a nonlinear transformation of the predictors [46], where the optimal classification hyperplane and the kernel function are the two main aspects of SVM, and the kernel function enables the linear classification of input samples through their transformation into a high-dimensional space [36].

In a landslide detection, the training data consist of the input matrix $X = x_i$ (30 landslide recognition features), $i = 1, 2, 3 \dots 30$, and output vector $Y = y_j$, $y_j =$ landslide, non-landslide (final category). Then, SVM maps the input vector X to a high-dimensional feature space through nonlinear mapping $\phi(x)$, as expressed by Equation (1).

$$f(x) = w \phi(x) + \theta \tag{1}$$

where $f(x)$ is the regression function of SVM, w is the weight vector, and θ is a bias term. By introducing the Lagrange multipliers and optimality constraints, $f(x)$ can also be obtained using Equation (2).

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) \times K(x, x_i) + \theta \tag{2}$$

where a_i and a_i^* are the Lagrange multipliers, and $K(x, x_i)$ is a kernel function, such as a sigmoid, polynomial, linear, or radial basis function (RBF) [47,48].

Detailed discussions of SVM have been provided by Huang and Zhao [49]. With the application of the SVM model in this study, the RBF was selected as the kernel function (Equation (3)), and the penalty factor (C) and the gamma (γ) parameters were set as 1.0 and 0.3.

$$K(x_i, x_j) = \exp(-\delta \|x_i - x_j\|^2) \tag{3}$$

where δ is the width parameter of the RBF kernel function.

3.1.2. ANN Model

The ANN model is an algorithm model that imitates the behavioral characteristics of a biological neural network, and it is characterized by the features of the independent statistical distribution of the data, self-learning and associative memory [50]. Briefly, this model comprises an input layer, a hidden layer and an output layer [39]. In landslide detection, 30 landslide recognition features are regarded as input layers, and the output layer is the results of the binary classification of pixels (landslide or non-landslide). The ANN model can be applied as Equation (4) [51,52].

$$y_i = f(\text{nets}) = f\left(\sum_{i=0}^n w_i x_i\right) \quad (4)$$

where x represents the image features, w is the weight of these features, and f is a nonlinear activation function. The ANN model can process information by adjusting the initial weights of a large number of interconnected nodes (neurons) [53]. Next, each hidden and output layer neuron processes its inputs by multiplying each input x by a weight w . Finally, the final result y is produced by summing the product and then processing the sum using f .

For parameterizing the ANN model, the hidden layers, training cycle, learn rate and momentum of the used ANN model were set as 61, 200, 0.1 and 0.9, respectively, in this application.

3.1.3. RF Model

The RF model was introduced by Breiman [54], and it belongs to a non-parametric ML classifier that has proven to accurately differentiate spectrally complex classes [55]. An RF model is an ensemble classifier that grows multiple decision trees, and is trained using bagging, thereby letting the trees determine the probability of the class membership [25].

In the RF algorithm, a random vector i_k (landslide recognition features) is generated independently of the previous random vectors and distributed to all trees. Each tree is grown using the training sample and random vector i_k , resulting in a collection of tree-structured classifiers $h(x, i_k)$, $k = 1, 2, \dots, n$ for input vector x . Typically in the RF algorithm, the generalization error (GE) is expressed as Equation (5) [56].

$$GE = P_{x, y}(mg(x, y) < 0) \\ mg(x, y) = av_k I(h_k(x) = y) - \max_{j \neq y} av_k I(h_k(x) = j) \quad (5)$$

where x and y are landslide conditioning factors that indicate the probability over x and y space, mg is the margin function, and I is the indicator function.

For parameterizing the RF classifier, we used 100 trees and took the square root of the number of layers as a split criterion at each node.

3.2. Data Used

Feature construction and sample (landslide sample and non-landslide sample) preparation are two indispensable aspects of ML model applications. In this study, we extracted landslide recognition features and mapped landslide/non-landslide samples based on multi-source remote sensing images.

3.2.1. Image Features for Landslide Detection

In this study, three datasets—multi-seasonal daytime optical Landsat time-series, DEM from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), and nighttime light imagery from the Visible Infrared Imaging Radiometer Suite (VIIRS)—were utilized to derive features for landslide identification. All of the features derived from these datasets were resampled to 30 m pixel size using the nearest neighbor method.

The DEM with a resolution of 30 m was derived from the National Basic Science Data Sharing Service Platform, Chinese Academy of Sciences (<http://www.gscloud.cn/>). The elevation and slope degree features were obtained from a single-time product of DEM with the support of ArcGIS 10.2 software.

All available Landsat-8 imagery (cloud cover < 20%) from August 9, 2017 to February 28, 2019 for the Landsat tiles (path/row 130/37) covering whole areas were obtained from the United States Geological Survey (USGS, <https://earthexplorer.usgs.gov/>). We used the quality assessment (QA) band (the QA band makes it easier for users to identify “bad” and “good” pixels; users can screen out valuable pixels more accurately with the help of QA band) to remove middle- and high-confidence cloud pixels and then we carried out some conventional processing (radiation calibration, atmospheric correction, etc.) on these images. Moreover, we utilized the blue (Band 2), green (Band 3), red (Band 4), near infrared (Band 5), first (Band 6) and second (Band 7) shortwave infrared bands as spectral features for landslide identification. The Normalized Difference Vegetation Index (NDVI) ($[(\text{Band } 5 - \text{Band } 4)] / [(\text{Band } 5 + \text{band } 4)]$) [57] and Normalized Difference Water Index (NDWI) ($[(\text{Band } 3 - \text{Band } 6)] / [(\text{Band } 3 + \text{Band } 6)]$) [58] for each single-date image were calculated. Then, the seasonal mean of the band values [59] and the indices of NDVI and NDWI were calculated in winter (December to February), summer (June to August), and fall (autumn) (September to November) as auxiliary information for landslide detection. Landsat images from March to May (Spring) were not considered in this study, because they are heavily affected by cloud cover, so that too few effective pixels can be used.

Nighttime light data are tightly associated with population density, income level and economic activities [60,61]; Chen et al. (2019) demonstrated the effectiveness of nighttime light imagery in improving the accuracy of landslide detection. In this study, nighttime light (VIIRS) data from August 9, 2017 to February 28, 2019 were obtained from the National Oceanic and Atmospheric Administration (NOAA, <https://www.ngdc.noaa.gov/eog/>). VIIRS data are available with a spatial resolution of 450 m approximately and can be provided on a monthly basis. Before calculating the seasonal mean values of nighttime light, we removed outliers by forcing values >40 and <0.2. Consistent with Landsat images, we also calculated the mean of the light index by season.

After the above processing, we extracted 30 variables as follows: Elevation, slope degree, seasonal mean values of winter, summer and fall of spectral (bands of blue, green, red, NIR, SWIR1, SWIR2), NDVI and NDWI, seasonal mean values of winter, summer, spring and fall of nighttime light indices. Then, Pearson’s correlation coefficient, which is defined as the covariance of two conditioning factors divided by the product of their standard deviations, between these predisposing factors, was calculated. Finally, the 30 variables mentioned above with weak correlations were retained as image features for the training and landslide prediction of the ML model.

3.2.2. Landslide Inventory

Sufficient and credible landslide samples are the premise for ML model application. After constructing the image features, the next step was to label the pixels that represent the landslide. In Jiuzhaigou County, a large number of coseismic landslides were triggered in the earthquake-affected area. On the basis of images for compiling landslide inventory before and after the earthquake, which are close to the seismic origin time and have low cloud coverage and optimal resolution, we chose the series of pre-seismic images on the Google Earth (GE) Platform, 0.5 m-resolution Geoeye-1 (shot on August 14 2017) post-seismic images as the data source.

Then the processed post-seismic Geoeye-1 image can overlies on the GE Platform, so that the image can not only provide a stereo-like perspective view, but also facilitate the comparison of the pre- and post-earthquake images. Next, the areas with bright tones, flow-textures and no vegetable-coverage were identified as coseismic landslides and were delineated as polygons through visual interpretation and contrast of aforementioned two kinds of images (Figure 1a). Finally, combined with GaoFen (GF) images (GF-1, shot on August 16 2017; GF-2, shot on August 09 2017), field surveys, repeated reviews

and validations were done to reduce errors and ensure the objectivity and precision of the landslide inventory prepared [62,63].

According to the landslide inventory mapped by the China Earthquake Administration, we converted planar landslide data into grid pixels of $30\text{ m} \times 30\text{ m}$, and some vector points are used to represent the location of landslide occurrence. Then, a total of 8005 points (pixels) were defined as landslide samples. Finally, combining the extracted image features for landslide detection, we construct sample data of two scenarios. Both in subarea A and subarea B, 1000 landslide pixels were randomly selected as the positive sample data in the corresponding regions; then the 1000 landslide pixels were divided into training samples and validation samples according to the unified 0.6:0.4 partition ratio. The remaining 5723 landslides pixels in area C served as additional positive validation samples to determine the prediction ability of our selected ML model to detect landslides in “off-site” areas.

3.3. Sensitivity Analysis

In addition to landslide (positive) samples, the selection of a non-landslide sample (negative) is also an indispensable part of both the model training set and validation set. Usually, the class distribution of the landslide sample and non-landslide sample is unbalanced, and the number of non-landslides is more than the number of landslides. Therefore, this sample imbalance will inevitably introduces a bias of classification, and usually causes overestimating the non-landslide class and underestimating the landslide class [64]. On this problem of sample bias, Stumpf and Kerle [21] and Chen et al. [25] proposed a sensitivity analysis approach to evaluate the impact of unbalanced class distribution on the final landslide detection accuracy. The core of this approach is to find an optimal balance between the user’s accuracy (represents overestimation) and the producer’s accuracy (represents underestimation) by tests of different class ratios (β , the ratio of landslide and non-landslide samples) to define the optimal ratio to be used. To help the three ML models better learn the complexity of non-landslide activities, this study ensured that the sample number of landslides in subarea A and B is always 500, then the β_i value started from $\beta=1.0$ and increased by a 0.1 step size, until reaching a non-landslide, five-fold class distribution ($\beta_i=5.0$). This means that the sample size for a non-landslide is $500 \times \beta_i$ in each iteration. Although finite iteration did not solve class bias entirely, the strategy provided a significantly better balance between user’s and producer’s accuracies than could be achieved with the natural class distribution or an ad hoc balanced training sample [21].

Finally, by iteratively increasing β_i , the curves of user’s and producer’s accuracy were plotted, and the β corresponding to the balance (curve intersection) of user’s and producer’s accuracy was determined. In order to reflect the curves of six cases (three ML models applied in each of two subareas) more clearly, we using the difference curve between user’s accuracy and producer’s accuracy to show the results of sample sensitivity analysis in six scenarios (Figure 3). A difference of value 0 represents the intersection point of user’s accuracy and producer’s accuracy curves, i.e., the balance between user’s accuracy and producer’s accuracy has been achieved.

On the basis of sensitivity analysis, it was found that the accuracy balance between user’s and producer’s does not show uniform characteristics in the six scenarios. In each case, there are several β values satisfying the balance characteristics (difference = 0) from which user’s and producer’s accuracy are generated. Inconsistent β values in various scenarios lead to incomparability among multiple models. Therefore, we introduced the distance index and the standard deviation to determine the uniform β value applicable in all six scenarios.

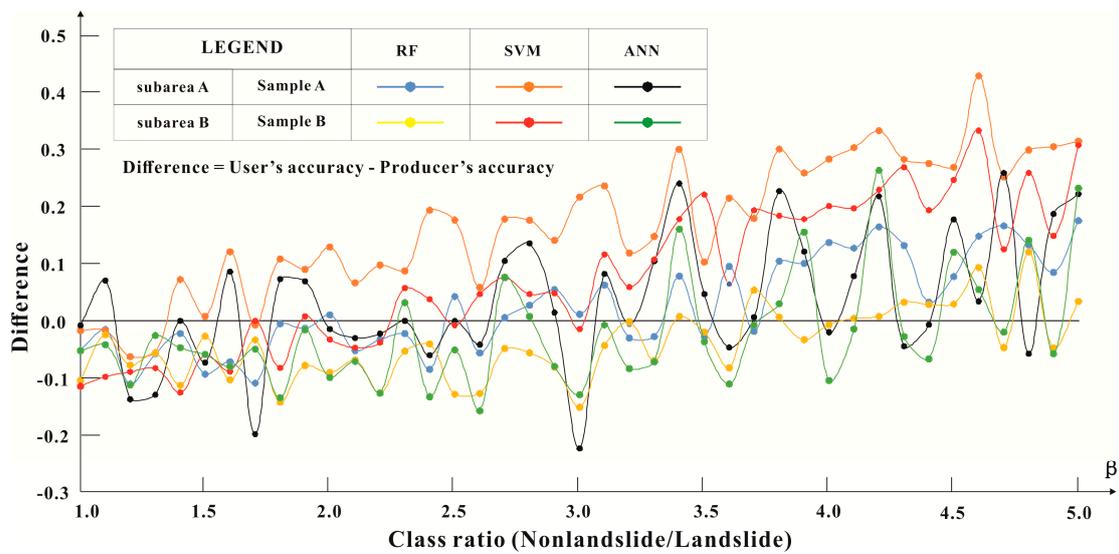


Figure 3. Estimates of the optimal class ratio (β).

The smaller the distance index, the closer the balance between user's accuracy and producer's accuracy, the smaller the standard deviation is, and the smaller the influence of β on six scenarios is, according to the statistics of distance index and standard deviation (Table A1). Ultimately, we regard $\beta = 1.5$ as the optimal class ratio of non-landslide to landslide.

4. Results

Through sensitivity analysis of positive and negative samples, the number of landslides and non-landslides contained in sample sets in subareas A and B can be determined. Then, according to the determined β value and combined with landslide samples in subarea A and subarea B, we can construct independent sample sets in subarea A and subarea B for the training and validation of three ML models. When comparing and analyzing the predictive performance of the three ML models using the samples generated from subarea A and subarea B, respectively, a model with the best performance will be recommended in area A and area B for landslide detection in "off-site" areas (Section 4.1), and when a new sample set, formed by merging the samples in subareas A and B, is used to train and validate the three models, it also recommended a model with the best performance to detect landslides in an "off-site" area (Section 4.2). In addition, in order to prove that the portable model determined by cross-application (Section 4.3) is superior to other models for landslide detection in "off-site" areas, an additional sample generated from "off-site" area C was used to test our hypothesis (Section 4.4).

In all application scenarios, the samples (landslide/non-landslide) were evaluated as one of four validation categories: True positive (TP), true negative (TN), false negative (FN), false positive (FP). TPs represent spatially and temporally correct landslide pixels, TNs are correct non-landslide pixels, FNs are reference landslide pixels missing in the detection and FPs are pixels identified as landslides, but are not present in the reference map. Then the model performance was evaluated by the indices of user's accuracy (UA, $TP/(TP + FP)$), producer's accuracy (PA, $TP/(TP + FN)$), F1 score ($2 \times PA \times UA/(PA + UA)$), AUC (area under the Receiver Operating Characteristic curves) value.

4.1. Separated-Application

In subarea A and subarea B, on the basis of established β (1.5), 1000 landslide samples (pixels) and 1500 non-landslide samples were prepared to form sample sets, respectively. Then, the sample set was divided into training samples and validation samples using a 60/40% ratio and the RF, SVM and ANN models were trained and validated in subareas A and B, respectively. The training and validation results of these RF, ANN and SVM models in subareas A and B are shown in Table 1 (Columns 1–6).

Table 1. Performance comparison of Random Forest (RF), artificial neural networks (ANN) and support vector machines (SVM) models in subareas A and B.

Models	Indices	Separated-Application				Merged-Application		Cross-Application	
		A _t A _t	A _t A _v	B _t B _t	B _t B _v	M _t M _t	M _t M _v	A _t B _v	B _t A _v
RF	UA	90.61%	87.98%	98.81%	96.44%	92.83%	90.54%	95.47%	77.90%
	PA	99.67%	97.00%	97.00%	94.75%	98.17%	95.75%	84.25%	85.50%
	AUC	0.993	0.983	0.999	0.994	0.992	0.984	0.974	0.918
	F1-score	0.949	0.923	0.979	0.956	0.954	0.931	0.895	0.815
ANN	UA	94.71%	91.38%	97.64%	96.46%	96.50%	94.61%	89.55%	68.98%
	PA	98.50%	98.00%	96.33%	95.25%	96.58%	96.50%	98.50%	94.50%
	AUC	0.994	0.993	0.994	0.995	0.994	0.993	0.988	0.843
	F1-score	0.966	0.946	0.970	0.959	0.965	0.956	0.938	0.798
SVM	UA	98.68%	90.87%	99.83%	95.87%	98.83%	92.89%	95.35%	41.45%
	PA	99.67%	94.50%	99.83%	98.75%	98.67%	96.38%	61.50%	100%
	AUC	0.999	0.983	1.000	0.998	0.999	0.990	0.968	0.849
	F1-score	0.992	0.926	0.998	0.973	0.987	0.946	0.748	0.586

A_tA_t: Training accuracy with samples in subarea A; A_tA_v: Validation accuracy with samples in subarea A; B_tB_t: Training accuracy with samples in subarea B; B_tB_v: Validation accuracy with samples in subarea B; M_tM_t: Training accuracy with merged samples; M_tM_v: Validation accuracy with merged samples; A_tB_v: Validation accuracy with samples in subarea B on the basis of training model with samples in subarea A; B_tA_v: Validation accuracy with samples in subarea A on the basis of training model with samples in subarea B.

As far as model performance in subarea A is concerned, the F1 score and AUC value of the SVM model are highest in model training, followed by the ANN model and RF model. With the validation samples, the ANN model achieved the highest F1 score and AUC value, followed by the SVM and RF model. In addition, further analysis of user's and producer's accuracy shows that the SVM model achieved the most balanced user's and producer's accuracies as compared to the ANN model and RF model. Although according to the statistical results of the F1 score, AUC value, and the balance between the user's accuracy and producer's accuracy, it is shown that there is no significant difference in these indices of the RF, SVM and ANN models, and considering the overall performance of three models in subarea A, the RF model was definitely not recommended because of its poor predictive performance and its having the worst balance between the user's and producer's accuracy. However, the ANN model with highest F1 score and AUC is designated as the best-performing model.

The same analysis in subarea B shows that the SVM model has the best performance, both in model training and model validation, and the ANN performance is lower than that of the RF model in model training; but with the validation samples, the overall accuracy, AUC, F1 score and the balance of user's accuracy and producer's accuracy of the ANN model are slightly superior to that of the RF model.

Therefore, in these two separated-application scenarios, the ANN model was recommended as the preferred model on the basis of sample sets in subarea A, and the SVM model was determined as the optimal model based on the samples in subarea B. The RF model was not recommended for application in "off-site" area C because of its poor model performance.

4.2. Merged-Application

In this scenario, two sets of samples from subarea A and subarea B were merged into a new set of samples containing 2,000 landslide samples and 3,000 non-landslide samples. All samples were re-divided into training samples and validation samples according to the ratio of 0.6:0.4. Then, this merged sample set was used to train and validate our RF, ANN and SVM models. The performance of the three models is shown in Table 1 (Columns 7–8).

With the training samples, although the number of samples has doubled, the performance of the RF, ANN and SVM models still exhibits similar characteristics as in "separated-applications". The SVM model achieved the highest F1 score and AUC values, followed by the ANN model and then the RF

model. Taking an overview of the model performance based on the validation samples, the F1 score and AUC value of the ANN model are higher than the SVM and RF models. Moreover, the user's and producer's accuracy shows that the ANN model achieved the most balanced overestimation and underestimation; better than the SVM and RF models. Therefore, we can draw a conclusion that in the "merged-application" scenario, the ANN model should be recommended as the preferred model for landslide detection in "off-site" area C.

4.3. Cross-Application

Although the models with the best performance in each sample scenario have been determined in the above applications, it cannot be neglected that the three models do not show significant performance differences for landslide prediction in the local area. Usually, the slight difference in accuracy between these models may be due to the randomness of the sample selection. Therefore, it is not enough to determine the most suitable model for landslide detection in "off-site" areas simply based on the model accuracy evaluation results obtained from these local samples. The portability of a model is an important feature, because it measures the model's ability to be applied to other areas without additional training. With the cross-application of three models to evaluate their portability, subarea A and subarea B are used as development and application zones, respectively, for the cross-application of the three ML models to evaluate the landslide detection ability in an "off-site" area without additional samples.

4.3.1. Portability of the Three Models to Subarea B

In this section, subarea B was used to test the portability of three models developed for the subarea A. Therefore, according to the setting of β equal to 1.5 in sensitivity analysis, and the uniform ratio (60/40%) of training samples and validation samples, 1500 samples (600 landslide pixels and 900 non-landslide pixels) randomly generated from subarea A were used to train three models, and 1000 samples (400 landslide pixels and 600 non-landslide pixels) from subarea B had the same geographical-environmental characteristics as they did in the model development area. These 1000 samples were used to validate the performance of the three models applied in the subarea B. The training and validation results of the RF, ANN and SVM models in subareas A and B are shown in Table 1 (Columns 1–6).

For the SVM model, when this model with the highest prediction performance in the model development area was applied to a new area to be detected, its F1 score and AUC value is significantly reduced. Furthermore, the user's (95.35%) accuracy and producer's (61.5%) accuracy obtained by the SVM model in subarea B shows that the SVM model achieved the worst balanced user's and producer's accuracy. This indicates that the generalization ability of this SVM model in the "off-site" area is so poor. Even in new areas with similar geographic environment characteristics as the model training areas, the lack of additional training samples will directly lead to the decline of the prediction ability of the SVM model.

In terms of the ANN and RF model, according to the statistics of the F1 score and AUC value of these two models, the ANN and RF model have both exhibited better prediction performance and achieved a balanced user's accuracy and producer's accuracy. From the perspective of F1 score and AUC value, the prediction abilities of the ANN model and SVM model in subarea B were compared and analyzed. Consistent with the training results of subarea A, the performance of the ANN model was slightly better than that of the SVM model.

The comparison of the performance of the three models in the model application area (subarea B) show that the prediction accuracy of the ANN model and the RF model is still high, but the prediction accuracy of the SVM model dropped dramatically.

This suggests that the good performance of the ANN and RF model was maintained in the application area, and the ANN and RF have a higher portability and a more stable performance across areas. However, this also indicated the SVM model's low portability to the application area.

4.3.2. Portability of the Three Models to Subarea A

When subarea B was used as the development area and the subarea A used as the application area, 1500 samples (600 landslide pixels and 900 non-landslide pixels), randomly generated from subarea B, were used for model training, and 1000 samples (400 landslide pixels and 600 non-landslide pixels) from subarea A were used to test the prediction ability of these model in “off-site” training.

Similar results were found for the three models according to the statistics of accuracy evaluation of F1 and AUC indicators. The SVM model still shows the worst prediction performance, and its F1 score (0.586) is much lower than that of the ANN (0.797) and RF models (0.815). The producer’s accuracy of the SVM model is as high as 100%, while the user’s accuracy is only 45%. The imbalance of the user’s accuracy and producer’s accuracy suggests that the SVM model has serious “overestimation” in the application in an “off-site” area; that many pixels to be detected are finally predicted to be landslides by SVM model. Moreover, although the ANN model achieves higher prediction accuracy than the SVM model, its user’s accuracy (68.98%) and producer’s accuracy (94.50%) are also as unbalanced as those of the SVM model; the “overestimation” of the ANN model leads to more classified pixels being predicted as landslides. However, the high performance of the RF model was retained from the development area to the application area, demonstrating its validity for the development areas and its high portability to the application areas.

4.4. “Off-site” Area Application and Validation

In view of the above analysis, with the study of the portability of three models to subarea B, there is no obvious difference in the prediction performance of the ANN and RF models, so both ANN and RF models can be recommended for landslide detection in area C because of their better accuracy and good portability; however, in the study of the portability of the three models to subarea A, the ANN model tends to overestimate and its F1 score and AUC value are lower than the that of the RF model. Therefore, the RF model was finally identified as the best portability model for landslide detection in “off-site” area C by the cross-application. In addition, combined with the results of the studies of “separated-application” and “merged-application”, it can be found that in different application scenarios, the best performance models determined by the comparative analysis are different. In “separated-applications”, the SVM model and ANN model are considered to be the best models in subarea A and subarea B, respectively, while in “merged-application”, the SVM model is determined to be the best model, and is recommended for “off-site” application. On the contrary, the RF model, which is not popular in the above two application scenarios, is recommended for landslide detection in “off-site” areas because of its good portability evaluated by “cross-application”.

In order to prove which application-determined model can exhibit high landslide detection in “off-site” areas, a set of validation samples generated from area C with 5723 landslides and 8585 non-landslides was used to evaluate the predictive performance of the three models in the “off-site” area to verify our hypothesis. The performance evaluation of RF, ANN and SVM models in area C is shown in Table 2 and Figure 4.

Table 2. Performance validation of RF, ANN and SVM Models in area C.

Indices	RF	ANN	SVM
User’s accuracy	98.48%	91.61%	87.02%
Producer’s accuracy	82.72%	72.29%	68.79%
AUC	0.935	0.921	0.885
F1 Score	0.899	0.808	0.768

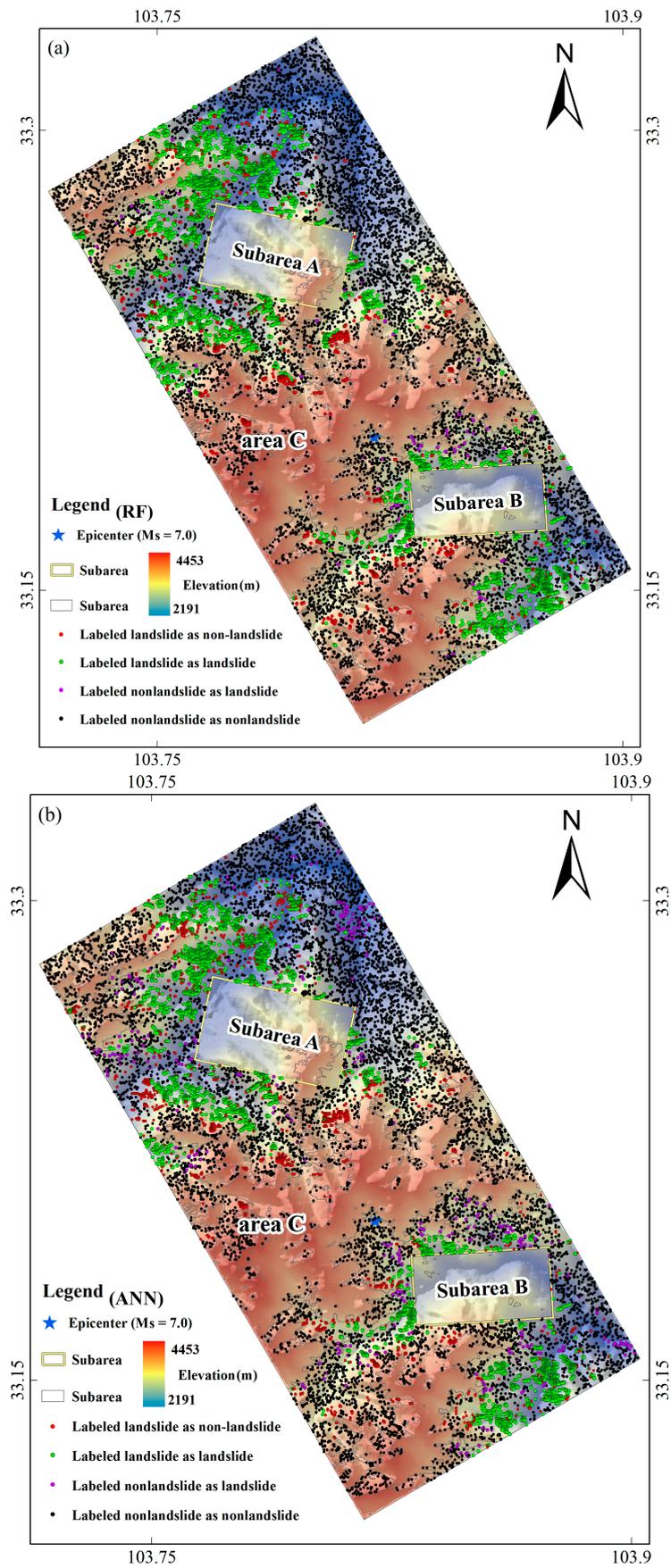


Figure 4. Cont.

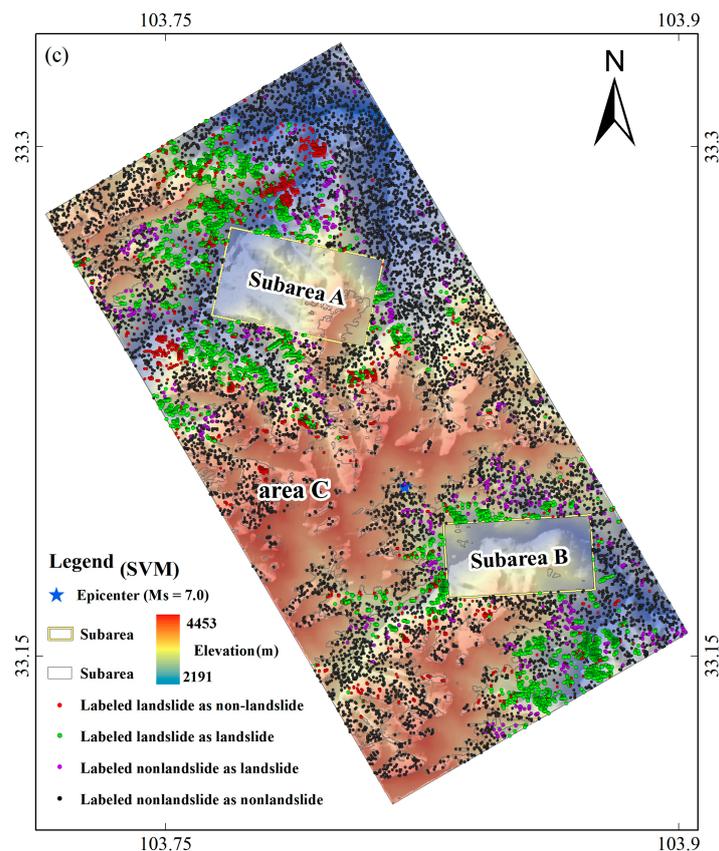


Figure 4. Performance validation of the RF, ANN and SVM Models in area C.

According to the statistics in Table 2, the overall accuracy of the three models shows that the predictive performances of the RF model are better than those of the ANN and SVM models. Comparing the AUC values in the three models, the RF model has the largest AUC value of 0.935, while the ANN model exhibits a slightly lower AUC; the AUC value of SVM model is the lowest, at 0.885. In addition, the user's and producer's accuracy of these three models are considered to reflect the model performance. Although the user's and producer's accuracy are not completely balanced in the three models, the difference between the user's accuracy and the producer's accuracy in the RF model is smaller than that in the SVM and ANN models.

Overall, the performance of the RF model in landslide detection for both development areas and application areas is stable. However, based on unified samples, the ability of the SVM model to detect landslides in "off-site" areas without additional samples is worse than that of the ANN and RF models. The SVM model shows excellent performance in the development area, but it has not achieved satisfactory results in the application area. For the ANN model, its portability falls between the ANN model and SVM model; therefore, we can conclude that the portability of RF is best, followed by the ANN and SVM models.

5. Discussions

This study mainly focused on how to use limited samples to determine optimal prediction models to detect landslides in "off-site" areas with similar geographic environmental characteristics to the model development area, and whether the high-accuracy model obtained in the development area can achieve excellent landslide prediction ability in "off-site" areas. The most significant difference between this study and previous applications (e.g., [65]) in data-scarce environments is that the limited samples in this study refer to the samples that have been collected are not distributed in the regions

to be classified/predicted, and that the regions where the samples are located and the regions to be predicted have certain similarities only in the geographic environment.

Many advanced ML models have been successful and applied widely to landslide detection with accepted accuracy [15,34]; however, none of the methods proposed so far provides a universal advantage in all situations [21], so comparing the performance of several models has become the main way in which many researchers choose the optimal model for landslide research [66]. Kalantar et al. [67] noted that landslide samples were found to have effects on ML models in terms of prediction accuracy and generalization capability, and Dhakal et al. [68] pointed out that the outcome of the hazard evaluation may depend on the samples used in the analysis. Therefore, in the traditional application, a variety of models were trained and validated by unified samples, then the model performance was judged according to the evaluation index (such as overall accuracy, AUC values, user's and producer's accuracy, etc.). However, this study found that the SVM and ANN models have higher prediction in the sample area, did not achieve optimal performance in the "off-site" area without additional training samples, and the balance between user's and producer's accuracy was also unsatisfactory, suggesting that the prediction ability of the model with superior performance in the sample area still needs further analysis in other areas.

On the other hand, although the RF model is slightly lower than the SVM and ANN models, both in "separated-application" and "merged-application", the cross-validation of the RF model in the two subareas shows that the performance of the RF model in landslide detection for both the development areas and application areas is stable, and it retains the original performance and achieves a satisfactory balance between overestimation and underestimation in areas outside those in which they were developed. The reason for the good portability of the RF model may be that the underlying idea of the RF model is to grow multiple decision trees on random subsets of the training data and related variables, and take advantage of the high variance among individual trees, letting each tree vote for class membership, then assigning the respective class according to the majority of the votes. Such ensembles demonstrate robust and accurate performance on complex datasets, and exhibit the desirable properties, such as robustness, against over-fitting the training data [21].

To our knowledge, our work differs from previous studies in that we first proved that the model with excellent portability can also achieve good landslide detection accuracy in "off-site" areas which have the similar geographical environment to the model training area. Although Zhu et al. [12] applied cross-validation to compare the portability of the logistic regression model, ANN model and expert knowledge-based model in the landslide susceptibility study, the model portability comparison in landslide detection is still really rare. What is lacking in Zhu's research is that they only compared the portability of the model, but did not further apply the results of the model portability evaluation, that is, it did not further prove that the model with best portability has the best performance in "off-site" areas without additional retraining. Pradhan et al. [39] applied the ANN model to validate and cross-validate three areas in Malaysia with similar environmental, geological and geomorphological conditions; it focused on determining the most critical landslide-inducing factors through cross-validation, then used these factors for other probabilistic and statistical models. What is neglected in Pradhan's study is that different models have their own evaluation of the importance of each landslide-inducing factor due to different algorithm mechanisms [69].

Furthermore, Mondini et al. [43] pointed out that what is particularly promising are the semi-automatic methods that allow investigators to prepare landslide event-inventories ranging from small areas to large regions, in a limited time, with limited resources, and of high quality in order to maximize the accuracy of landslide detection in a large area with only a small number of landslide samples in a subarea after an earthquake. In view of this, the study of the portability of ML models can help to determine the optimal model to realize a high accuracy of landslide identification in adjacent areas without "seeing" any landslide sample data. These developing and testing strategies proposed in this study for landslide detection can likely be replicated in one area to neighboring areas; the premise is whether the recognition features constructed in the model development area are similar

to those in the application area to be detected. This also means that the landslide types considered in model training and landslide detection are the same, and the geographic environment of the landslide induced in the development and application area is also similar. To our knowledge, this study is the first to test model portability in determining the optimal model for the study of landslides in “off-site” areas. The results not only prove that the ML model with stable portability can be recommended as the optimal one for the application of landslide detection in “off-site” areas, but also provides a feasible way for landslide detection in “off-site” areas with fewer samples.

Additionally, this study generated image features for model portability evaluation using freely-available, medium-resolution Landsat multi-seasonal imagery and nighttime light data. Using multi-temporal images to extract the trajectory characteristics of landslide-related factors (such as NDVI, NDWI) has proven to be a feasible measure to assist landslide extraction in many studies [27,30–32]. The validity of nighttime light data used as auxiliary information for landslide detection is rarely studied; only Chen et al. [25] pointed out that nighttime light imagery is helpful for increasing the detection accuracy of landslides in anthropogenic areas. In this study, according to the importance of the landslide recognition features calculated by the RF model (Table A2), the importance of NDVI (summer), NDVI (fall, i.e., autumn), and elevation to landslide detection are the three most important items, and nighttime light plays a moderately important role in landslide identification—it is more helpful for landslide detection than the spectral characteristics in summer. The findings further prove that the nighttime light imagery is helpful for increasing the detection accuracy of landslides in anthropogenic areas [25].

6. Conclusions

In this study, we explored the effectiveness of the evaluation of ML model portability to determine the optimal model to provide improved landslide detection in an “off-site” area without additional training. Cross-applications of three models (SVM, ANN and RF) between two study areas (subareas A and B) were conducted to minimize the impact of the particularity of the sample zone on model performance. The results revealed that the SVM and ANN models achieved higher accuracy in their respective development zones, but when these two models were applied in areas outside the development areas, the performance dropped dramatically. However, the high performance of the RF model was retained from the development area to the application area, demonstrating its validity for the development area and its high portability to the application area. Another set of validation samples also proved that the RF model with optimal portability is more suitable for “off-site” landslide detection.

Our findings confirmed that the ML model has a high prediction performance in the model development area, which does not mean that it also maintains a high prediction ability outside the model development area. Evaluation of the model portability through cross-application can help determine the most suitable model for landslide detection in “off-site” areas, so as to maximize the accuracy of landslide detection based on limited samples.

Author Contributions: Conceptualization, Q.H., Y.Z. and F.W.; methodology, S.W., Y.Z. and Q.H.; validation, Q.H. and F.W.; formal analysis, Q.H. and H.W.; investigation, Q.H. and F.W.; resources, Q.H., Y.Z. and F.W.; data curation, Q.H., and H.W.; writing—original draft preparation, Q.H.; writing—review and editing, Y.Z., S.W. and F.W.; visualization, Q.H.; supervision, Y.Z. and S.W. All authors read and approved the final manuscript.

Funding: This research was funded by National Key Research and Development Program of China (2017YFB0504100 and 2016YFC0803000) and Youth Innovation Promotion Association of the Chinese Academy of Sciences (2015129).

Acknowledgments: We sincerely thank Xu Chong, Institute of Geology, China Earthquake Administration, and his research team for providing the landslide inventory of Jiuzhaigou earthquake. Moreover, this paper has benefited from valuable comments and suggestions by anonymous reviewers, and our editor, whose efforts are gratefully acknowledged.

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

Appendix A

Table A1. Estimates of the optimal class ratio (β).

β	Subarea A			Subarea B			Standard Deviation	Distance Index
	RF	SVM	ANN	RF	SVM	ANN		
1.2	-0.11	-0.06	-0.14	-0.08	-0.09	-0.11	0.027	0.248
1.3	-0.06	-0.06	-0.13	-0.06	-0.08	-0.03	0.035	0.184
1.5	-0.09	0.01	-0.07	-0.03	-0.06	-0.06	0.036	0.147
1	-0.05	-0.02	-0.01	-0.10	-0.11	-0.05	0.044	0.172
2.1	-0.05	0.07	-0.03	-0.07	-0.05	-0.07	0.051	0.141
2.3	-0.02	0.09	0.00	-0.05	0.06	0.03	0.052	0.123
1.1	-0.02	-0.02	0.07	-0.02	-0.10	-0.04	0.054	0.131
1.9	-0.01	0.09	0.07	-0.08	0.01	-0.02	0.061	0.139
3.2	-0.03	0.12	-0.01	0.00	0.06	-0.08	0.071	0.160
1.4	-0.02	0.07	0.00	-0.11	-0.13	-0.05	0.074	0.191
1.7	-0.11	-0.01	-0.20	-0.03	0.00	-0.05	0.075	0.234
2.7	0.01	0.18	0.10	-0.05	0.08	0.08	0.079	0.237
2.2	-0.03	0.10	-0.02	-0.13	-0.04	-0.13	0.082	0.210
2	0.01	0.13	-0.01	-0.09	-0.03	-0.10	0.083	0.190
2.8	0.03	0.18	0.14	-0.06	0.05	0.01	0.085	0.235
2.9	0.05	0.14	0.01	-0.08	0.05	-0.08	0.085	0.195
2.6	-0.06	0.06	-0.04	-0.13	0.05	-0.16	0.088	0.226
4.5	0.08	0.27	0.18	0.03	0.25	0.12	0.095	0.430
3.7	-0.02	0.18	0.01	0.05	0.19	-0.01	0.095	0.270
3.9	0.10	0.26	0.12	-0.03	0.18	0.15	0.097	0.385
3.1	0.06	0.24	0.08	-0.04	0.12	-0.01	0.098	0.285
1.6	-0.07	0.12	0.09	-0.10	-0.09	-0.08	0.099	0.227
3.3	-0.03	0.15	0.10	-0.07	0.11	-0.07	0.099	0.234
3.5	-0.03	0.10	0.05	-0.02	0.22	-0.04	0.100	0.253
2.5	0.04	0.18	0.00	-0.13	-0.01	-0.05	0.102	0.228
5	0.17	0.31	0.22	0.03	0.31	0.23	0.103	0.572
3.4	0.08	0.30	0.24	0.01	0.18	0.16	0.106	0.459
1.8	0.00	0.11	0.07	-0.14	-0.08	-0.13	0.106	0.249
4.2	0.16	0.33	0.22	0.01	0.23	0.26	0.110	0.553
3.8	0.10	0.30	0.23	0.01	0.18	0.03	0.115	0.432
2.4	-0.08	0.19	-0.06	-0.04	0.04	-0.13	0.116	0.262
4.1	0.13	0.30	0.08	0.00	0.20	-0.01	0.120	0.391
3.6	0.09	0.21	-0.05	-0.08	0.06	-0.11	0.124	0.283
4.8	0.13	0.30	-0.06	0.12	0.26	0.14	0.125	0.459
4.4	0.03	0.27	-0.01	0.03	0.19	-0.07	0.130	0.345
4.7	0.17	0.25	0.26	-0.05	0.13	-0.02	0.131	0.419
4.9	0.08	0.30	0.19	-0.05	0.15	-0.06	0.140	0.403
4.3	0.13	0.28	-0.04	0.03	0.27	-0.03	0.144	0.415
4	0.14	0.28	-0.02	-0.01	0.20	-0.10	0.148	0.387
3	0.01	0.22	-0.22	-0.15	-0.01	-0.13	0.156	0.369
4.6	0.15	0.43	0.03	0.09	0.33	0.05	0.161	0.573

Sort in ascending order according to "Distance Index" and then in ascending order according to "Standard Deviation". The first five items of "Distance Index" denoted by bold markers, and the first five items of "Standard Deviation" are denoted by italic markers. Finally, we use $\beta = 5$ as the most appropriate class ratio.

Table A2. Weights of landslide-related factors calculated by RF model.

Landslide_Related Features	Weight Calculated by RF Model
summer_ndvi	0.29301622
autumn_ndvi	0.10382754
elevation	0.10074204
summer_band 4	0.06635309
summer_band 7	0.06287415
winter_ndvi	0.04509334
slope	0.03764887
summer_band 2	0.03555093
summer_ndwi	0.02913229
summer_band 5	0.02501019
winter_band 2	0.02171738
autumn_ndwi	0.01790977
spring_nighttime light	0.01509238
summer_nighttime light	0.01455534
winter_nighttime light	0.01375868
autumn_nighttime light	0.01205927
winter_band 4	0.01161285
summer_band 3	0.0115633
winter_ndwi	0.00959292
winter_band 3	0.00868868
summer_band 6	0.00812666
winter_band 5	0.00747098
autumn_band 2	0.00738832
autumn_band 3	0.00720624
autumn_band 5	0.00719601
autumn_band 7	0.00568052
winter_band 7	0.00559739
autumn_band 4	0.00524157
autumn_band 6	0.00519584
winter_band 6	0.00509723

References

- Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.T. Landslide inventory maps: New tools for an old problem. *Earth. Sci. Rev.* **2012**, *112*, 42–66. [[CrossRef](#)]
- Pradhan, A.M.S.; Tarolli, P.; Kang, H.S.; Lee, J.S.; Kim, Y.T. Shallow landslide susceptibility modeling incorporating rainfall statistics: A case study from the Deokjeok-ri Watershed, South Korea. *Int. J. Erosion Control. Eng.* **2016**, *9*, 18–24. [[CrossRef](#)]
- Guzzetti, F.; Cardinali, M.; Reichenbach, P.; Cipolla, F.; Sebastiani, C.; Galli, M.; Salvati, P. Landslides triggered by the 23 November 2000 rainfall event in the Imperia Province, Western Liguria, Italy. *Eng. Geol.* **2004**, *73*, 229–245. [[CrossRef](#)]
- Mondini, A.C.; Guzzetti, F.; Reichenbach, P.; Rossi, M.; Cardinali, M.; Ardizzone, F. Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. *Remote Sens. Environ.* **2011**, *115*, 1743–1757. [[CrossRef](#)]
- Guzzetti, F.; Ardizzone, F.; Cardinali, M.; Galli, M.; Rossi, M.; Valigi, D. Landslide volumes and landslide mobilization rates in Umbria, central Italy. *Earth Planet. Sci. Lett.* **2009**, *279*, 222–229. [[CrossRef](#)]
- Fiorucci, F.; Cardinali, M.; Carlà, R.; Rossi, M.; Mondini, A.C.; Santurri, L.; Ardizzone, F.; Guzzetti, F. Seasonal landslides mapping and estimation of landslide mobilization rates using aerial and satellite images. *Geomorphology* **2011**, *129*, 59–70. [[CrossRef](#)]
- Van Westen, C.J.; Castellanos, E.; Kuriakose, S.L. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Eng. Geol.* **2008**, *102*, 112–131. [[CrossRef](#)]
- 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](#)]

9. Pradhan, B.; Jebur, M.N.; Shafri, H.Z.M.; Tehrany, M.S. Data fusion technique using wavelet transform and Taguchi methods for automatic landslide detection from airborne laser scanning data and quickbird satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1610–1622. [[CrossRef](#)]
10. Keefer, D.K. Investigating landslides caused by earthquakes—A historical review. *Surv. Geophys.* **2002**, *23*, 473–510. [[CrossRef](#)]
11. Wang, Q.; Wang, Y.; Niu, R.Q.; Peng, L. Integration of information theory, k-means cluster analysis and the logistic regression model for landslide susceptibility mapping in the Three Gorges Area, China. *Remote Sens.* **2017**, *9*, 938. [[CrossRef](#)]
12. Zhu, A.X.; Miao, Y.; Wang, R.X.; Zhu, T.X.; Deng, Y.C.; Liu, J.Z.; Yang, L.; Qin, C.Z.; Hong, H.Y. A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping. *Catena* **2018**, *166*, 317–327. [[CrossRef](#)]
13. Mondini, A.C.; Chang, K.T. Combining spectral and geoenvironmental information for probabilistic event landslide mapping. *Geomorphology* **2014**, *213*, 183–189. [[CrossRef](#)]
14. Mezaal, M.R.; Pradhan, B.; Shafri, H.Z.M.; Yusoff, Z.M. Automatic landslide detection using dempster-shafer theory from lidar-derived data and orthophotos. *Geomat. Nat. Haz. Risk.* **2017**, *8*, 1935–1954. [[CrossRef](#)]
15. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.; Tiede, D.; Aryal, J. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sens.* **2019**, *11*, 196. [[CrossRef](#)]
16. Metternicht, G.; Hurni, L.; Gogu, R. Remote sensing of landslides: An analysis of the potential contribution to geo-spatial systems for hazard assessment in mountainous environments. *Remote Sens. Environ.* **2005**, *98*, 284–303. [[CrossRef](#)]
17. Keefer, D.K.; Larsen, M.C. Assessing landslide hazards. *Science* **2007**, *316*, 1136–1138. [[CrossRef](#)]
18. Kirschbaum, D.B.; Adler, R.; Hong, Y.; Kumar, S.; Peters-Lidard, C.; Lerner-Lam, A. Advances in landslide nowcasting: Evaluation of a global and regional modeling approach. *Environ. Earth Sci.* **2012**, *66*, 1683–1696. [[CrossRef](#)]
19. Khatami, R.; Mountrakis, G.; Stehman, S.V. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sens. Environ.* **2016**, *177*, 89–100. [[CrossRef](#)]
20. Keyport, R.N.; Oommen, T.; Martha, T.R.; Sajinkumar, K.; Gierke, J.S. A comparative analysis of pixel-and object-based detection of landslides from very high resolution images. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *64*, 1–11. [[CrossRef](#)]
21. Stumpf, A.; Kerle, N. Object-oriented mapping of landslides using random forests. *Remote Sens. Environ.* **2011**, *115*, 2564–2577. [[CrossRef](#)]
22. Holbling, D.; Friedl, B.; Eisank, C. An object-based approach for semi-automated landslide change detection and attribution of changes to landslide classes in northern Taiwan. *Earth Sci. Inf.* **2015**, *8*, 327–335. [[CrossRef](#)]
23. Li, X.; Cheng, X.; Chen, W.; Chen, G.; Liu, S. Identification of forested landslides using LiDAR data, object-based image analysis, and machine learning algorithms. *Remote Sens.* **2015**, *7*, 9705–9726. [[CrossRef](#)]
24. Yu, B.; Chen, F. A new technique for landslide mapping from a large-scale remote sensed image: A case study of Central Nepal. *Comput. Geosci.* **2017**, *100*, 115–124. [[CrossRef](#)]
25. Chen, T.-H.K.; Prishchepov, A.V.; Fensholt, R.; Sable, C.E. Detecting and monitoring long-term landslides in urbanized areas with nighttime light data and multi-seasonal Landsat imagery across Taiwan from 1998 to 2017. *Remote Sens. Environ.* **2019**, *225*, 317–327. [[CrossRef](#)]
26. Lang, S. Object-based image analysis for remote sensing applications: Modeling reality—dealing with complexity. In *Object-Based Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 3–27.
27. Singh, K.K.; Singh, A. Detection of 2011 Sikkim earthquake-induced landslides using neuro-fuzzy classifier and digital elevation model. *Nat. Hazards* **2016**, *83*, 1027–1044. [[CrossRef](#)]
28. Chen, T.H.; Lin, K.H.E. Distinguishing the windthrow and hydrogeological effects of typhoon impact on agricultural lands: An integrative OBIA and PPGIS approach. *Int. J. Remote Sens.* **2018**, *39*, 131–148. [[CrossRef](#)]
29. Barlow, J.; Martin, Y.; Franklin, S. Detecting translational landslide scars using segmentation of Landsat ETM+ and DEM data in the northern Cascade Mountains, British Columbia. *Can. J. Remote Sens.* **2003**, *29*, 510–517. [[CrossRef](#)]

30. Behling, R.; Roessner, S.; Kaufmann, H.; Kleinschmit, B. Automated spatiotemporal landslide mapping over large areas using rapideye time series data. *Remote Sens.* **2014**, *6*, 8026–8055. [[CrossRef](#)]
31. Golovko, D.; Roessner, S.; Behling, R.; Kleinschmit, B. Automated derivation and spatio-temporal analysis of landslide properties in southern Kyrgyzstan. *Nat. Hazards* **2017**, *85*, 1461–1488. [[CrossRef](#)]
32. Cheng, G.; Guo, L.; Zhao, T.; Han, J.; Li, H.; Fang, J. Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA. *Int. J. Remote Sens.* **2013**, *34*, 45–59. [[CrossRef](#)]
33. Duro, D.C.; Franklin, S.E.; Dubé, M.G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using Spot-5 HRG imagery. *Remote Sens. Environ.* **2012**, *118*, 259–272. [[CrossRef](#)]
34. Pal, M.; Mather, P. Support vector machines for classification in remote sensing. *Int. J. Remote Sens.* **2005**, *26*, 1007–1011. [[CrossRef](#)]
35. Roodposhti, M.S.; Aryal, J.; Bryan, B.A. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. *Environ. Model. Softw.* **2019**, *112*, 70–81. [[CrossRef](#)]
36. Bui, D.T.; 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](#)]
37. Hong, H.; Pradhan, B.; Jebur, M.N.; Bui, D.T.; Xu, C.; Akgun, A. Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines. *Environ. Earth Sci.* **2016**, *75*, 40. [[CrossRef](#)]
38. Danneels, G.; Pirard, E.; Havenith, H.-B. Automatic landslide detection from remote sensing images using supervised classification methods. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2007), Barcelona, Spain, 23–28 July 2007; IEEE: Barcelona, Spain, 2007; pp. 3014–3017. [[CrossRef](#)]
39. 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](#)]
40. Chen, W.; Li, X.; Wang, Y.; Chen, G.; Liu, S. Forested landslide detection using LiDAR data and the random forest algorithm: A case study of the Three Gorges, China. *Remote Sens. Environ.* **2014**, *152*, 291–301. [[CrossRef](#)]
41. Chang, K.T.; Chiang, S.H.; Hsu, M.L. Modeling typhoon-and earthquake-induced landslides in a mountainous watershed using logistic regression. *Geomorphology* **2007**, *89*, 335–347. [[CrossRef](#)]
42. Chen, S.C.; Chang, C.C.; Chan, H.C.; Huang, L.M.; Lin, L.L. Modeling typhoon event-induced landslides using GIS-based logistic regression: A case study of Alishan forestry railway, Taiwan. *Math. Probl. Eng.* **2013**, *2013*, 728304. [[CrossRef](#)]
43. Mondini, A.C.; Marchesini, I.; Rossi, M.; Chang, K.T.; Pasquariello, G.; Guzzetti, F. Bayesian framework for mapping and classifying shallow landslides exploiting remote sensing and topographic data. *Geomorphology* **2013**, *201*, 135–147. [[CrossRef](#)]
44. Hu, X.D.; Hu, K.H.; Tang, J.B.; You, Y.; Wu, C.H. Assessment of debris-flow potential dangers in the Jiuzhaigou Valley following the August 8, 2017, Jiuzhaigou earthquake, western China. *Eng. Geol.* **2019**, *256*, 57–66. [[CrossRef](#)]
45. Dai, L.X.; Xu, Q.; Fan, X.M.; Yang, Q.; Yang, F.; Ren, J. A preliminary study on remote sensing interpretation of landslides triggered by Jiuzhaigou earthquake in Sichuan on August 8th, 2017, and their spatial distribution patterns. *J. Eng. Geol.* **2017**, *25*, 1151–1164. (In Chinese)
46. Vapnik, V. *Statistical Learning Theory*; Wiley: New York, NY, USA, 1998.
47. 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. [[CrossRef](#)]
48. Marjanović, M.; Kovačević, M.; Bajat, B.; Voženilek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **2011**, *123*, 225–234. [[CrossRef](#)]
49. Huang, Y.; Zhao, L. Review on landslide susceptibility mapping using support vector machines. *Catena* **2018**, *165*, 520–529. [[CrossRef](#)]

50. Yilmaz, I. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: Conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ. Earth Sci.* **2010**, *61*, 821–836. [[CrossRef](#)]
51. 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](#)]
52. Polykretis, C.; Chalkias, C. Comparison and evaluation of landslide susceptibility maps obtained from weight of evidence, logistic regression, and artificial neural network models. *Nat. Hazards* **2018**, *93*, 249–274. [[CrossRef](#)]
53. Lee, S.; Ryu, J.H.; Lee, M.J.; Won, J.S. Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea. *Environ. Geol.* **2003**, *44*, 820–833. [[CrossRef](#)]
54. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
55. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
56. Masetic, Z.; Subasi, A. Congestive heart failure detection using random forest classifier. *Comput. Methods Prog. Biomed.* **2016**, *130*, 54–64. [[CrossRef](#)] [[PubMed](#)]
57. Martha, T.R.; Kerle, N.; van Westen, C.J.; Jetten, V.; Kumar, K.V. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4928–4943. [[CrossRef](#)]
58. Wu, X.; Niu, R.; Ren, F.; Peng, L. Landslide susceptibility mapping using rough sets and back-propagation neural networks in the Three Gorges, China. *Environ. Earth Sci.* **2013**, *70*, 1307–1318. [[CrossRef](#)]
59. Nichol, J.; Wong, M. Satellite remote sensing for detailed landslide inventories using change detection and image fusion. *Int. J. Remote Sens.* **2005**, *26*, 1913–1926. [[CrossRef](#)]
60. Doll, C.N.; Muller, J.P.; Morley, J.G. Mapping regional economic activity from night-time light satellite imagery. *Ecol. Econ.* **2006**, *57*, 75–92. [[CrossRef](#)]
61. Jean, N.; Burke, M.; Xie, M.; Davis, W.M.; Lobell, D.B.; Ermon, S. Combining satellite imagery and machine learning to predict poverty. *Science* **2016**, *353*, 790–794. [[CrossRef](#)]
62. Tian, Y.Y.; Xu, C.; Ma, S.Y.; Xu, X.W.; Wang, S.Y.; Zhang, H. Inventory and spatial distribution of landslides triggered by the 8th August 2017 Mw 6.5 Jiuzhaigou earthquakes, China. *J. Earth Sci.* **2019**, *30*, 206–217. [[CrossRef](#)]
63. Xu, C.; Wang, S.Y.; Xu, X.W.; Zhang, H.; Tian, Y.Y.; Ma, S.Y.; Fang, L.H.; Lu, R.Q.; Chen, L.C.; Tan, X.B. A panorama of landslides triggered by the 8 August 2017 Jiuzhaigou, Sichuan MS 7.0 earthquake. *Seismol. Geol.* **2018**, *40*, 232–259. (In Chinese)
64. Dalponte, M.; Orka, H.O.; Gobakken, T.; Gianelle, D.; Næsset, E. Tree species classification in Boreal forests with hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2632–2645. [[CrossRef](#)]
65. Lee, J.H.; Sameen, M.I.; Pradhan, B.; Park, H.J. Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. *Geomorphology* **2018**, *303*, 284–298. [[CrossRef](#)]
66. Vorpahl, P.; Elsenbeer, H.; Märker, M.; Schröder, B. How can statistical models help to determine driving factors of landslides? *Ecol. Model.* **2012**, *239*, 27–39. [[CrossRef](#)]
67. Kalantar, B.; Pradhan, B.; Naghibi, S.A.; Motevalli, A.; Mansor, S. Assessment of the effects of training data selection on the landslide susceptibility mapping: A comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomat. Nat. Haz. Risk.* **2018**, *9*, 49–69. [[CrossRef](#)]
68. Dhakal, A.S.; Amada, T.; Aniya, M. Landslide hazard mapping and the application of GIS in the Kulekhani watershed, Nepal. *Mt. Res. Dev.* **1999**, *19*, 3–16. [[CrossRef](#)]
69. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [[CrossRef](#)]

