applied sciences

MDPI

Article

Integrating InSAR Observables and Multiple Geological Factors for Landslide Susceptibility Assessment

Yan-Ting Lin [1], Yi-Keng Chen [2], Kuo-Hsin Yang [2], Chuin-Shan Chen [2], Jen-Yu Han [2,*]

[1] National Center for Research on Earthquake Engineering, National Applied Research Laboratories, Taipei 106, Taiwan; yanting@ncree.narl.org.tw
[2] Department of Civil Engineering, National Taiwan University, Taipei 106, Taiwan; ykchen0814@gmail.com (Y.-K.C.); khyang@ntu.edu.tw (K.-H.Y.); dchen@ntu.edu.tw (C.-S.C.)
* Correspondence: jyhan@ntu.edu.tw; Tel.: +886-2-33664347

Abstract: Due to extreme weather, researchers are constantly putting their focus on prevention and mitigation for the impact of disasters in order to reduce the loss of life and property. The disaster associated with slope failures is among the most challenging ones due to the multiple driving factors and complicated mechanisms between them. In this study, a modern space remote sensing technology, InSAR, was introduced as a direct observable for the slope dynamics. The InSAR-derived displacement fields and other in situ geological and topographical factors were integrated, and their correlations with the landslide susceptibility were analyzed. Moreover, multiple machine learning approaches were applied with a goal to construct an optimal model between these complicated factors and landslide susceptibility. Two case studies were performed in the mountainous areas of Taiwan Island and the model performance was evaluated by a confusion matrix. The numerical results revealed that among different machine learning approaches, the Random Forest model outperformed others, with an average accuracy higher than 80%. More importantly, the inclusion of the InSAR data resulted in an improved model accuracy in all training approaches, which is the first to be reported in all of the scientific literature. In other words, the proposed approach provides a novel integrated technique that enables a highly reliable analysis of the landslide susceptibility so that subsequent management or reinforcement can be better planned.

Keywords: landslide potential; InSAR; spatial factors; machine learning; slope unite

Citation: Lin, Y.-T.; Chen, Y.-K.; Yang, K.-H.; Chen, C.-S.; Han, J.-Y. Integrating InSAR Observables and Multiple Geological Factors for Landslide Susceptibility Assessment. *Appl. Sci.* **2021**, *11*, 7289. https://doi.org/10.3390/app11167289

Academic Editor: Francesco Fiorillo

Received: 24 June 2021
Accepted: 5 August 2021
Published: 8 August 2021

1. Introduction

In Asian subtropical monsoon regions, July to September is a season of strong typhoons. High rainfall intensity usually causes serious landslide events in mountainous areas [1]. It is necessary to predict landslide occurrence and behavior and adopt appropriate prevention policies and methods to improve disaster relief effectiveness and reduce casualties and property loss during and after disasters. Landslide prediction aims to predict the possibility of the occurrence of landslides in a specific area; available data are commonly used, including conditional factors and historical landslides. These data are collected from landslide inventories and static instruments, and their values are shown in spatial analysis [2]. However, traditional landslide prediction, such as mathematical evaluation models, lacks information about the temporal probability of landslides, i.e., time-series landslide behavior. Landslide displacement time-series data can directly reflect ground surface deformation and stability characteristics. Therefore, they have been recently used to develop landslide prediction models. Generally, these time-series data are collected from one-point survey equipment, such as surface extensometers and GPS devices [3]. However, field GPS surveying projects, which depend on only one or two temporarily installed reference stations, have many disadvantages [4]. In practice, steadily obtaining survey data using these single reference stations is often difficult because of

poor performance or failure. Therefore, the use of only the single-point method in landslide surveys would limit the cost-effectiveness.

In recent years, remote sensing technology has effectively detected large-scale landslide-sensitive areas and generated landslide inventories, which are crucial for predicting landslides before they occur or recur, especially in far or barely accessible areas [5]. In daytime satellite images without shadows and clouds, landslide positions can be identified through noticeable radiometric contrasts between land cover types [6]. Optical sensors cover the electromagnetic spectrum from 390 nm to 1 mm, including the visible and infrared bands. Such devices can measure the visual properties in the spectral characteristics of the land surface, which can then be used to detect and map landslides. Researchers can also combine time-series satellite images with digital elevation models (DEMs) to acquire 3D terrain, which can be used to visually detect and predict potential landslides.

However, affected by monsoons, typhoons, and thunderstorms, mountainous areas are usually shrouded in clouds at times; thus, the use of satellite images to monitor landslide disasters could be limited by weather conditions. Compared with optical sensors, synthetic aperture radar (SAR) sensors use a longer wavelength—microwaves; having all-weather and all-day operational capability, SAR sensors can penetrate cloud cover and reduce the limitation imposed by the atmosphere to remotely evaluate the accurate range and severity of landslide disasters in almost real-time [7]. Although some particular meteorological situations, such as thick rain cells, may disturb the backscattering coefficient, SAR remains more powerful than optical sensors for long-term landslide observation [8]. Spaceborne SAR, such as Envisat, ALOS PALSAR, RADARSAT, TerraSAR-X, and Sentinel-1, provide high spatial resolutions and can clearly observe target objects in full-time and in almost all-weather conditions.

Numerous applications of SAR data to ground displacement detection have demonstrated their usefulness for landslide characterization and mapping [9]. Differential SAR interferometry (DInSAR) is a commonly used method of ground deformation measurement, and it can efficiently generate or update landslide inventory [10], which is critical information about landslide behavior for landslide susceptibility assessment. DInSAR calculates the phase variation of two SAR images acquired in the same region at different times. Long-term InSAR observations are calculated as the deformation-induced phase shift through the backscattered microwave signal between several coherent acquisitions. The landslide behavior of time-series information, which depends on the millimetric measurement accuracy and the metric spatial resolution, is obtained under most atmospheric conditions [11].

Landslide prediction methods can be classified into three types: image analysis, mathematical evaluation models, and machine learning methods [12]. Image analysis uses geographic information systems, which can collect, store, manage, and analyze geographical data. The risk of landslides can be predicted by analyzing disaster data, such as history of landslides and land. The probability of landslides varies because it is based on the number of data layers used for analysis. Mathematical evaluation models use a single evaluation equation that is combined with the physical concepts of mechanics and hydrographic data, such as rainfall, runoff, and infiltration data, for landslide susceptibility assessment [13]. The use of such models is easy for simulation and fits a wide range of environments. However, mathematical evaluation models require detailed data of the geotechnical engineering and geological aspects of slope failure at sites [14], which makes these models costly and impractical for large-scale areas.

In recent years, machine learning and data mining techniques, such as support vector machine, artificial neural network, and decision tree (DT) models, have been applied for landslide susceptibility modeling [15]. These methods incorporate different factors that might cause landslides to evaluate the probability of landslide occurrence. Machine learning algorithms enrich the quality and accuracy of generated susceptibility maps. Researchers use and compare various machine learning models on the basis of different data

[16–19], integrate different machine learning models to improve accuracy [20–23], or develop new algorithms that are based on traditional machine learning models to strengthen landslide prediction results [24–26]. These techniques perform better than do classical methods. Most machine learning techniques achieve overall success rates of 75% to 95% [27]. Although many applications have demonstrated the feasibility of data-driven models for capturing nonlinear relationships and modeling the dynamic processes of landslides on the basis of historical model data, certain limitations remain [28]. As shown, landslide behavior involves temporal dependencies. However, common machine learning models ignore this intrinsic temporal dependency, which involves the effect of preceding actions on present actions in the model [29,30]. The solution proposed by this study is to combine spatial-temporal data, including InSAR observables, as a landslide susceptibility factor with other traditional geological and land cover factors into a model that can improve the prediction accuracy of potential landslides. To our knowledge, integrating InSAR observables and multiple geological factors for landslide susceptibility analysis is an effective and pioneered contribution for landslide potential prediction research.

2. Methods

This research method effectively estimates the landslide potential of slopes through four steps: (1) segmentation of slope units, (2) numerical indexing of related spatial factors, (3) correlation between spatial factors and slope landslides, and (4) use of machine learning methods. A displacement prediction analysis model was constructed following the above process. Finally, a confusion matrix was used to verify the results of the displacement prediction analysis. The overall research method and procedure are shown in Figure 1.

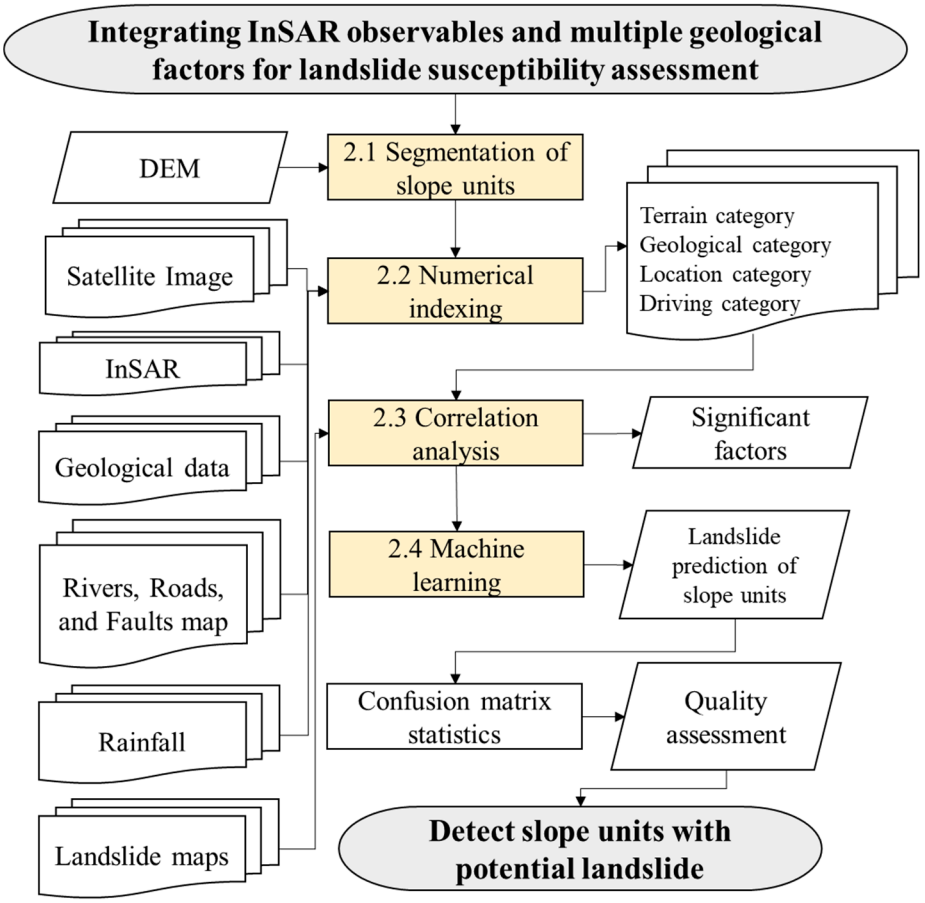


Figure 1. Flowchart of landslide susceptibility analysis based on the spatial factors with machine learning approach.

2.1. Segmentation of Slope Units

This study used the slope unit as the basis of analysis to show the topographic characteristics of each slope. These slope units serve as a framework for the subsequent geographical interpretation of environmental spatial factors. The method of slope unit segmentation refers to the catchment overlap concept proposed by Xie et al. [31], as shown in Figure 2. First, the water catchment area in a DEM is identified through the hydrology module in the software ArcGIS, and the water line is turned into a ridge line by flipping the DEM, which is divided into two slope units (left and right). When the hydrology module identifies small catchment areas, the default flow accumulation value is set to 500 as the threshold value for dividing the river area. Then, the slope units are cut out, and each area becomes less than 30 ha. With the aid of a shadow map, aspect map, slope map, river map, and satellite orthophoto overlay, the overlap between each slope unit is confirmed.

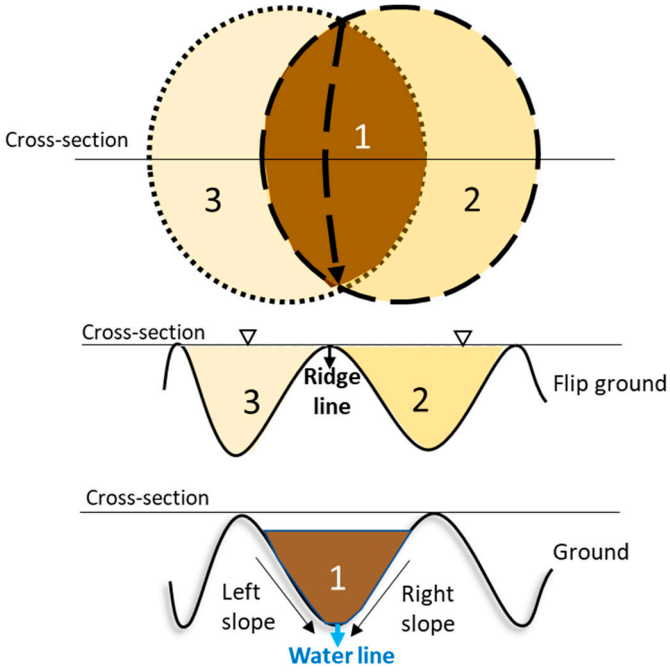


Figure 2. Schematic of dividing the slope units with the overlap method of catchment areas (modified from [31]).

2.2. Numerical Indexing of Related Spatial Factors

In this study, the spatial factors were divided into four categories: terrain, location, geological, and driving. The terrain category represents the geometric changes in surface elevation and coverage distribution, including elevation, slope, aspect, terrain roughness, profile curvature, vegetation index, and the displacement velocity gradient of InSAR. The location category shows the distance of influencing factors, including roads, rivers, and geo-faults. The geological category reflects the strength, folds, and dip slopes of rock formations. The driving category is the rainfall factor. The index calculations of these factors are described below. It should be mentioned that these spatial factors were first selected based on suggestions reported in the relevant studies in the literature [16–20]. A significance test was then performed to identify the most influential factors that have the high correlation with the landslides in the study areas. The results and discussion on the significance test of spatial factors are presented in Section 3.2.

2.2.1. Terrain Category

- Elevation, slope, and aspect

On the basis of the framework of the slope unit, the highest elevation in each unit was extracted and represented as the elevation factor, as shown in Equation (1). According to the height change caused by the horizontal movement distance, the slope factor is expressed by a tangent function on average, as indicated in Equation (2). The aspect factor refers to the direction of the maximum elevation change in the slope unit. It is calculated by the angle with the true north direction, as shown in Equation (3), where the true north direction is 0°, and the angle increases to 360° in the clockwise direction.

$$I_{elevation} = \max(Z_i) \tag{1}$$

$$I_{slope} = \tan\overline{\theta_s} = \frac{\Delta\overline{Z}}{\Delta\overline{L}} \tag{2}$$

$$I_{aspect} = \frac{180^\circ}{\pi}\tan^{-1}\left[\max\left(\theta_s\right)\right] \tag{3}$$

where Z_i is elevation, $\Delta\overline{Z}$ is the mean elevation difference, $\Delta\overline{L}$ is the mean horizontal distance, and θ_s is the main slope angle.

- Terrain roughness

Terrain roughness represents the degree of height change. When the undulating terrain faces the effect of large gravity, the smaller resistance force makes the slope have a higher possibility of landslide. The elevation standard deviation σ is used to describe the degree of elevation change in the slope unit (Equation (4)).

$$\sigma = \sqrt{\frac{\sum_i (Z_i - \overline{Z})^2}{n_s - 1}} \tag{4}$$

where $\overline{Z}$ is the average elevation in a slope unit, and n_s is the number of grids in the slope unit.

- Profile curvature

The profile curvature is expressed as the slope steepness. This study used the spatial analysis module of the software ArcMap to calculate the profile curvature of each slope unit on the basis of a 3 × 3 moving grid, which is the default grid size in ArcMap. A negative (positive) value of the curvature represents a convex (concave) slope.

- Vegetation index

Plants can effectively stabilize the rock and soil on slopes, but the exposed soil area may suffer from repeated landslide and displacement problems. Hence, the vegetation index is defined as the proportion of vegetation area in the slope unit, as shown in (Equation 5).

$$I_{veg.} = \frac{A_{veg.}}{A_s} \tag{5}$$

where $A_{veg.}$ is the area of the vegetation and A_s is the area of the slope unit.

- Annual displacement velocity gradient of InSAR

InSAR technology calculates the phase difference to estimate the displacement of the ground through more than two periods of SAR observations. The InSAR-derived ground displacement can be regarded as a direct observation of ground stability and was thus proposed as an essential index for landslide susceptibility analysis in this study. However, the original displacements from InSAR observations suffer from various influencing factors, such as vegetation changes and orbital variations of SAR satellites. In order to reduce the periodical or systematic noises due to those uncontrollable factors and to extract a

meaningful index for evaluating the ground stability, the annual velocity gradients derived from InSAR displacement fields were used in this study. First, the annual displacement information of InSAR is placed in the range from −1 to 1 by mean normalization, which is shown in Equation (6), to unify the scale and reduce the systematic error of InSAR data.

$$Z_{s_i}^{-} = \frac{Z_{S_i} - \mu}{\max\left(Z_{s_i}\right) - \min\left(Z_{s_i}\right)} \tag{6}$$

where $Z_{s_i}^{-}$ is the normalized InSAR displacement value, Z_{S_i} is the annual displacement of InSAR, and μ is the average annual displacement.

The annual displacement velocity of InSAR is obtained as the slope value in first-order linear fitting (Equation (7)). These discrete observation points are interpolated with a regular grid size of 20 m to present the field of annual displacement velocity. For highlighting the displacement positions, the field gradient is calculated with a 3 × 3 moving window, the same as for computing the profile curvatures. The index calculation is expressed as Equation (8).

$$Z_{s_i}^{-} = V\Delta t + \Delta Z \tag{7}$$

$$I_{InSAR} = \nabla_V \mathbf{f}(V) \tag{8}$$

where V is the annual displacement speed of InSAR, Δt is annual observation time, and ΔZ is the difference in annual displacement.

2.2.2. Location Category

Potential displacements are affected by the distances between slope units and location factors. In this study, three location factors were selected for analysis, namely, the river distance, road distance, and fault distance. Through each shortest distance from the centroid of the slope units to the three location factors, the formula of the location factors $I_{location}$ is expressed by Equation (9).

$$I_{location(rivers,\ roads,\ faults)} = \min\left(\sqrt{\left(X_c - X_l\right)^2 + \left(Y_c - Y_l\right)^2}\right) \tag{9}$$

where (X_c, Y_c) is the centroid coordinates of slope units, and (X_l, Y_l) is the coordinates of location factors (including rivers, roads, and faults).

2.2.3. Geological Category

- Rock Mass Strength

Rock masses with weaker strength are prone to landslides due to their difficulty in resisting the disturbance of external forces. Franklin used the degree of rock structure fracture and single compressive strength to classify the rock mass strength into seven levels [32]. In this study, the slope unit was superimposed on the environmental geological map produced by the Central Geological Survey of Taiwan, and the corresponding rock mass strength information was used as the rock mass strength index.

- Folds

When a rock is squeezed into curved folds, the fold layer becomes prone to landslides. In this study, the fold factor is defined as the number of folds in the slope unit, as shown in Equation (10).

$$I_{fold} = \sum n_f \tag{10}$$

where n_f is the number of folds in a slope unit.

- Dip Slopes

Dip slopes mean that a stratum has the same inclination as that of the slope; a slope landslide may be formed by sliding along the layer. In this study, the dip slope index is defined as the ratio of the dip slope area to the slope unit area, as shown in Equation (11).

$$I_{dip\ slope} = \frac{A_d}{A_s} \tag{11}$$

where A_d is the area of the dip slope and A_s is the area of the slope unit.

2.2.4. Driving Category (Rainfall)

The density of rainfall data collected by rainfall stations is much lower in mountainous areas than that in urban areas. Relevant studies have mostly used distance as an interpolation reference to obtain the rainfall in a whole area through grid interpolation. This study considered the distance and elevation factors of rainfall stations and added the aspect factor to construct a rainfall interpolation model, as shown in Equation (12). In this model, the elevation parameter α, distance parameter β, and aspect parameter γ are obtained through the least squares adjustment, and the parameter weight is shown in Equation (13).

$$I_{rain} = \alpha\left(\sum_{iH} W_{iH} \cdot R_i\right) + \beta\left(\sum_{iL} W_{iL} \cdot R_i\right) + \gamma\left(\sum_{i\theta} W_{i\theta} \cdot R_i\right) \tag{12}$$

$$\mathrm{W_{iH}} \propto \frac{1}{\Delta \mathrm{H}^2};\ \mathrm{W_{iL}} \propto \frac{1}{\Delta \mathrm{L}^2};\ \mathrm{W_{i\theta}} \propto \frac{1}{\Delta \theta^2} \tag{13}$$

where I_{rain} is the rainfall index, W_{iH} is the elevation weight, W_{iL} is the distance weight, $W_{i\theta}$ is the aspect weight, R_i is the rainfall observation at Station i, α is the elevation parameter, β is the distance parameter, γ is the aspect parameter, $\Delta \mathrm{H}$ is the elevation difference, $\Delta \mathrm{L}$ is the distance difference, and $\Delta\theta$ is the aspect difference.

2.3. Correlation between Spatial Factors and Slope Landslides

Significant factors were detected through the spatial factors and the displacement correlation score. The Spearman method was adopted to arrange the data in order of numerical value, thereby improving the limitation of the normal distribution assumption in the correlation analysis. The correlation coefficient γ_s is distributed between 1 and −1; a positive (negative) value indicates a positive (negative) correlation. The closer the coefficient value to 0, the more unlikely it is to affect the displacement. Its sequential linear relationship is described in Equation (14).

$$\gamma_s = 1 - \frac{6\sum \Delta_i^2}{n(n^2-1)} \tag{14}$$

where γ_s is the correlation coefficient, Δ is the difference between the spatial factor and displacement, and n is the number of samples.

Finally, a significance test was conducted through the correlation coefficient to check the significance of each factor. This test is shown in Equation (15).

$$t = 1 - \frac{\gamma_s - \rho_0}{\sqrt{\frac{1-\gamma_s^2}{n-2}}} \tag{15}$$

where ρ_0 is 0, and it is the null hypothesis (indicating no correlation). If the significance level t is greater than 0.99, the null hypothesis will be rejected; that is, the factor is correlated with the displacement.

2.4. Use of Machine Learning Methods

Machine learning is applied to establish prediction models, which are used in landslide potential and displacement prediction, by inputting the spatial factors and displacement observations. Widely used machine learning algorithms for classification prediction include naive Bayes, DT, random forest, adaptive boosting (AdaBoost), and extreme gradient boosting (XGBoost).

- Naive Bayes

As the probability model of naive Bayes assumes that the factors are independent of each other and conform to a Gaussian distribution, naive Bayes classification helps clarify a large number of complex classification problems. The early-stage spatial factors correspond to the landslide and nonlandslide slope units, and they are regarded as training samples to establish a prediction model. The later-observed spatial factors are inputted into the model to determine the landslide probability of each slope unit. The naive Bayes prediction model is based on the probability density function of the Bayesian classification method [33], as shown in Equation (16).

$$P(w_i|x) = \frac{P(x|w_i)P(w_i)}{P(x)}, \begin{cases} j \neq i \\ j = 1, 2 \end{cases} \quad (16)$$

where $P(w_i|x)$ is the probability of the classifying w_i occurring in the slope unit x, $P(x|w_i)$ is the probability of the slope unit x occurring in the classifying w_i, $P(w_i)$ is the probability of classifying w_i, and $P(x)$ is the probability of the slope unit *x*.

- DT

A DT assumes that the factors are independent of each other, and the category probability of the DT path is defined by the factor characteristics [34]. This algorithm adopts a dichotomy method, which is similar to a double-forked tree branch, to calculate the Gini coefficient value at the node. Finally, the gain value in each path is summed, and the largest accumulator will be predicted to belong to a category, as shown in Equation 17.

$$\text{gain} = \sum p_i \left(1 - \sum p_i^2\right) \quad (17)$$

where P_i is the probability. If the node has only one category, P_i will be 0. If the numbers of two categories are the same, P_i is 0.5.

- Random forest

Random forest is a collection of multiple DTs and adds the use of bagging. The observation data are taken out of the number of samples and trained as *n* types of classifiers. According to the sample difference in each DT, the random uncertainty of the data is considered. Under the same weight, the classifier uses the summed majority as the best classification tree to predict the classification [35]. Equation (18) represents the probability of the *c*-th factor in the *t*-th DT, and the average probability value *gc* of the category is obtained according to the sum of multiple DTs. Finally, the category of the slope unit *x* is determined according to the maximum g_c value (Equation (19)).

$$P\left(c|v_i(x)\right) = \frac{P\left(c|v_i(x)\right)}{\sum_l^{n=l} P\left(c_l|v_i(x)\right)} \quad (18)$$

$$g_c(x) = \frac{1}{t}\sum_{i=1}^{t} \hat{P}\left(c|v_i(x)\right) \quad (19)$$

where P is the probability, c-th is the category, v is the node, l is the number of categories, t is the number of DTs, and g_c is the average probability of the c-th category.

- AdaBoost

Boosting increases the weight of wrong data in a classification model, and the wrong information is trained to strengthen the identification. The derived new classifier will reduce the chance of early error [36]. The iterative process of the AdaBoost calculation is extremely sensitive to noise and abnormal data; therefore, these should be reduced so that the process can focus on difficult-to-classify feature factors. AdaBoost analysis initially assumes that the sample weights are equal. After the *k*-th iteration, samples are selected on the basis of the weight W_k to train the classifier *Ck*, as expressed by Equation (20).

$$\begin{cases} D=\left\{x^1, y_1, \cdots, x^n, y_n\right\} \\ W_k(i)=\dfrac{1}{n}, i=1\cdots, n \end{cases} \tag{20}$$

where D is the sample category, (x^i, y_i) is the sample information, n is the number of samples, and W_k is the weight distribution of all samples in the k-th iteration.

The classification error E_k confirms the correctness of the classification and updates the weight W_{k+1}, as shown in Equation (21). The iterative calculation of classification is completed when the error E_k is less than the preset threshold.

$$W_{k+1}(i) \leftarrow \frac{W_k(i)}{Z_k} \times \begin{cases} e^{-\frac{1}{2}\ln\frac{1-E_k}{E_k}}, & \text{if } y_k(x^i)=y_i \\ e^{\frac{1}{2}\ln\frac{1-E_k}{E_k}}, & \text{if } y_k(x^i)\neq y_i \end{cases} \tag{21}$$

where W_{k+1} is the updated weight, Z_k is the normalization coefficient, E_k is the error, and y_k is the prediction category.

- XGBoost

The XGBoost function is composed of two components: the prediction error of boosting and the complexity of DT. The feature factors are combined and branched into a DT, and a new boost function is learned from the previous calculation residuals [37]. In Equation (22), the first component calculates the error between the prediction and actual observation, and the other component indicates the complexity of the regularized DT, which covers the number of nodes and the node probability value.

$$f=\sum_{i=1}^{n} E(y_i, y_{k_i}) + \sum_{k=1}^{K} \Omega(f_k) \tag{22}$$

where E is the error between the prediction and actual observation and $\Omega(f_k)$ is the complexity of the DT.

3. Results

The experiment based on the slope unit was conducted for the following two parts of test analysis. In the first part, the correlation analysis of the spatial factor and the landslide unit was adopted to detect the significant spatial factor. In the second part, the spatial factor indicators and landslide units observed from 2007 to 2009 were applied to run the machine learning models. Then, the 2010 spatial factors were inputted into those models, and the landslide slope units were estimated. The prediction was compared with the landslide location announced by the Central Geological Survey of Taiwan's Ministry of Economic Affairs (MOEA) through a confusion matrix to verify the feasibility of this study.

3.1. Study Areas

Experimental cases in Siaolin Village and the Putunpunas River area (Kaohsiung, Taiwan) were selected to verify this study method. Both areas continued to experience a large number of landslides after the typhoon Morakot in 2009. In the Siaolin Village area, there were 128 slope units (covering 15.81 km^2), and Provincial Highway 29 is the main external traffic road. In the Putunpunas River area, there were 349 slope units (covering

61.21 km^2), and the Southern Cross-Island Highway presents a north–south vertical, as shown in Figure 3.

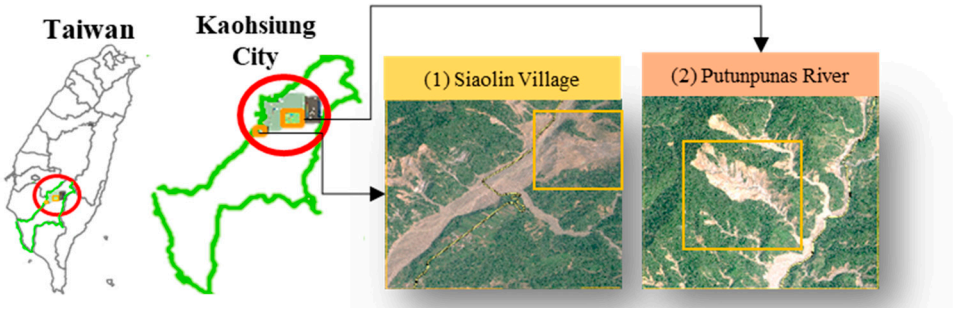
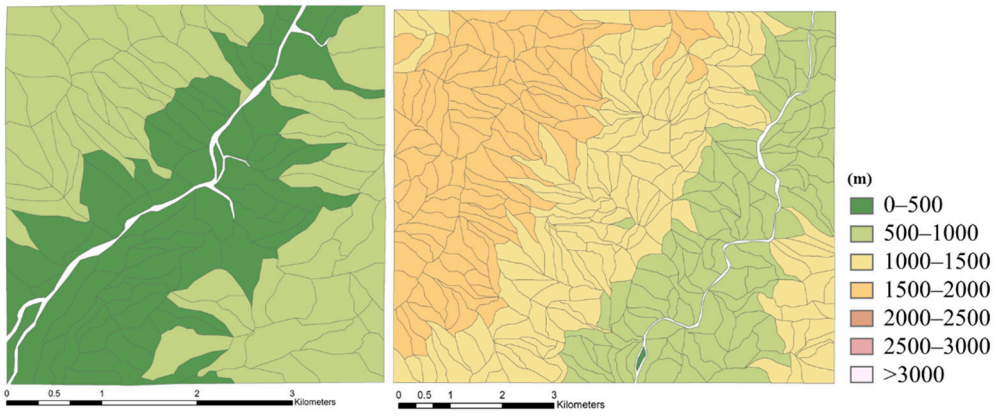
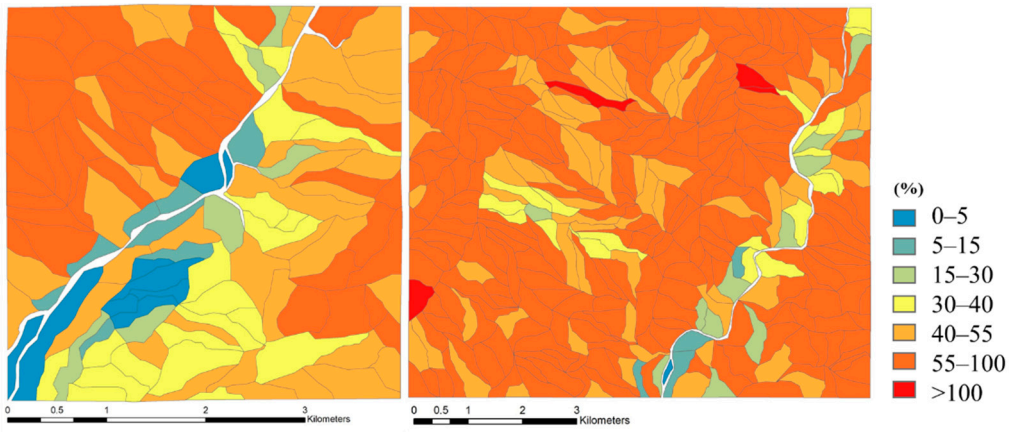


Figure 3. Geographical locations of experimental areas—(**1**) Siaolin village; (**2**) Putunpunas River.

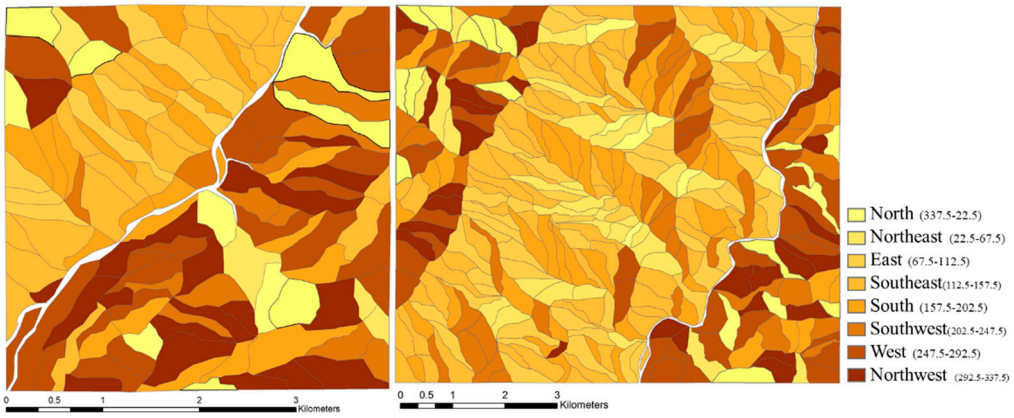
The observation time of the spatial factors ranged from hours to years. For establishing a common timescale, a year was deemed the basis of unit time, and the observed data time was a total of four years (from 2007 to 2010). The 14 spatial factors used were the elevation, slope, aspect, terrain roughness, profile curvature, vegetation index, annual displacement velocity gradient of InSAR, water distance, road distance, fault distance, rock mass strength, folds, dip slopes, and an annual rainfall, as shown in Figure 4.



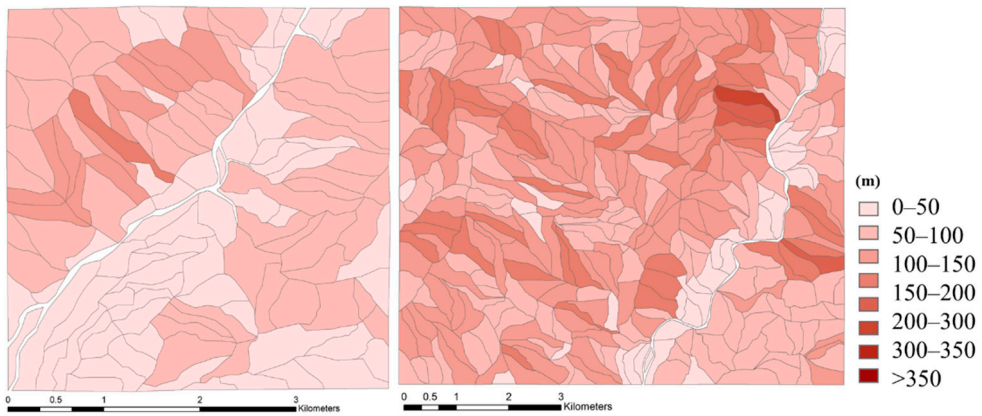
(**a**) elevation



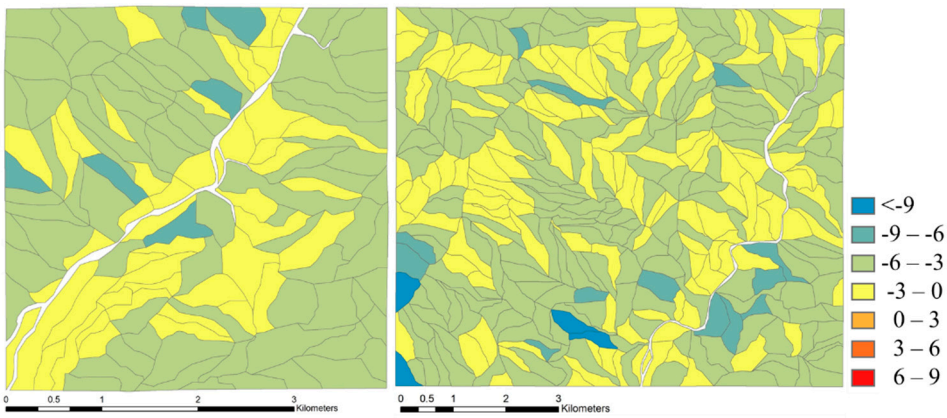
(**b**) slope



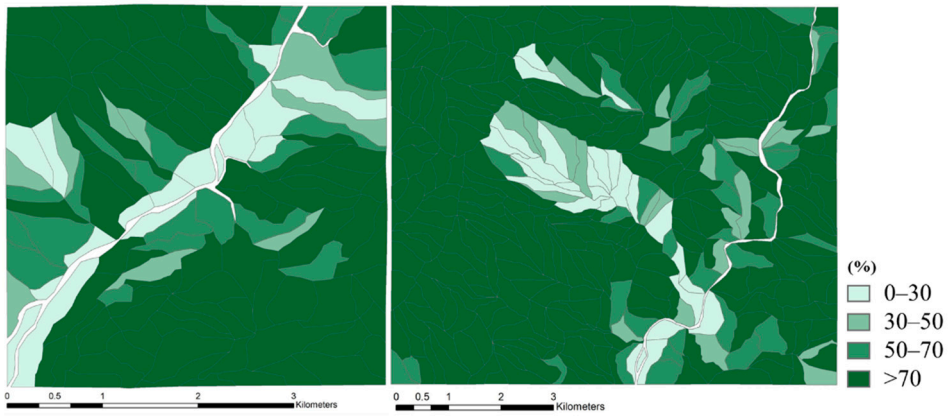
(**c**) aspect



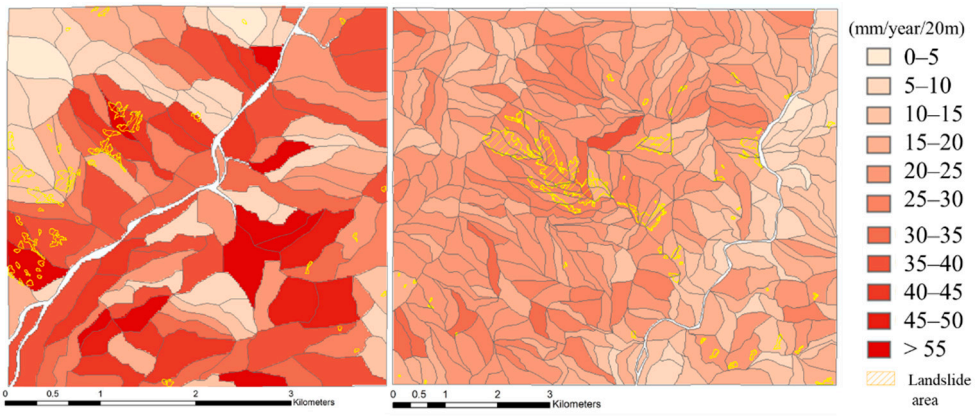
(**d**) terrain roughness



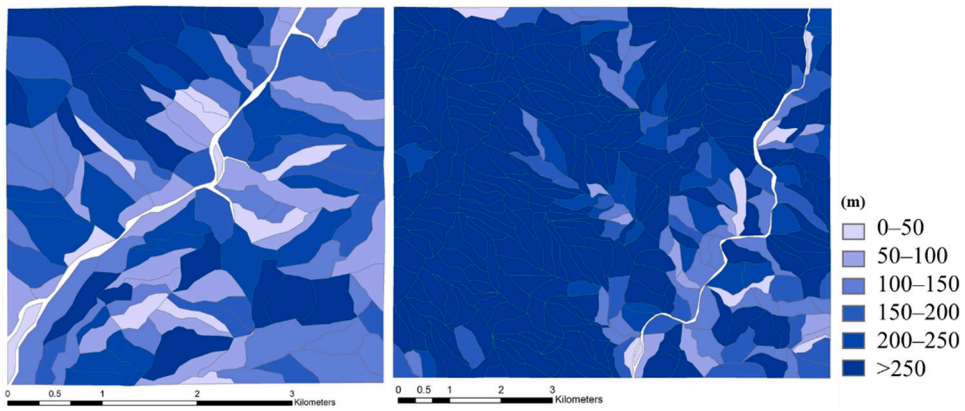
(**e**) profile curvature



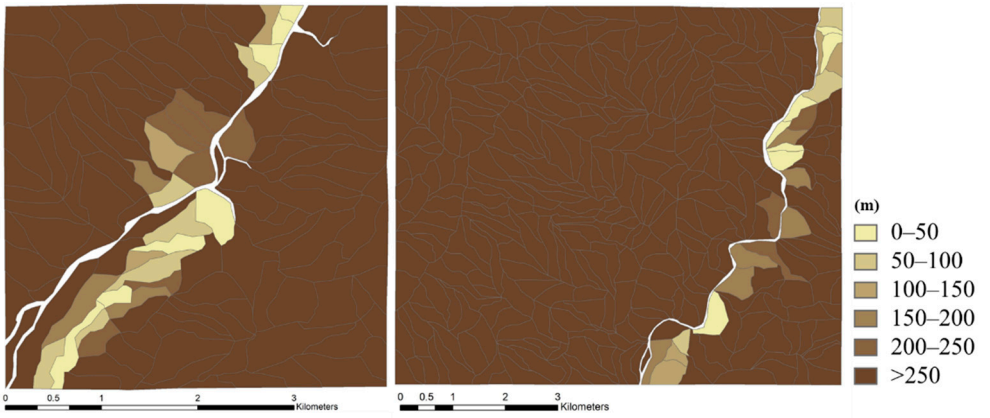
(**f**) vegetation index



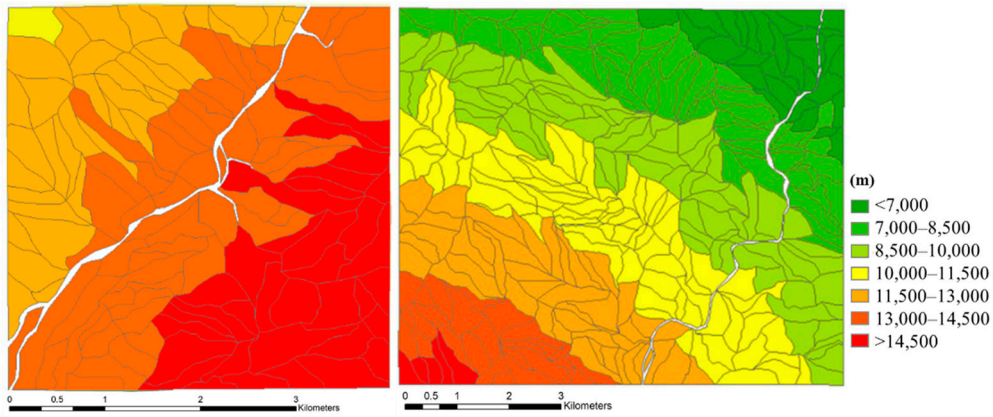
(**g**) annual displacement velocity gradient of InSAR



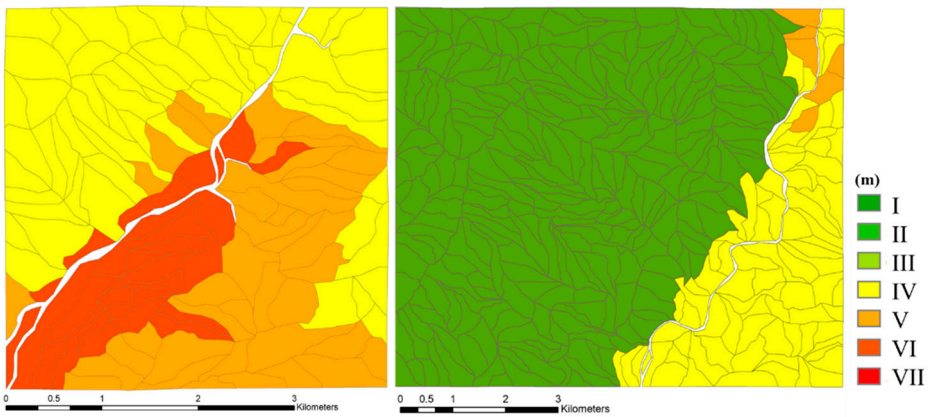
(**h**) water distance



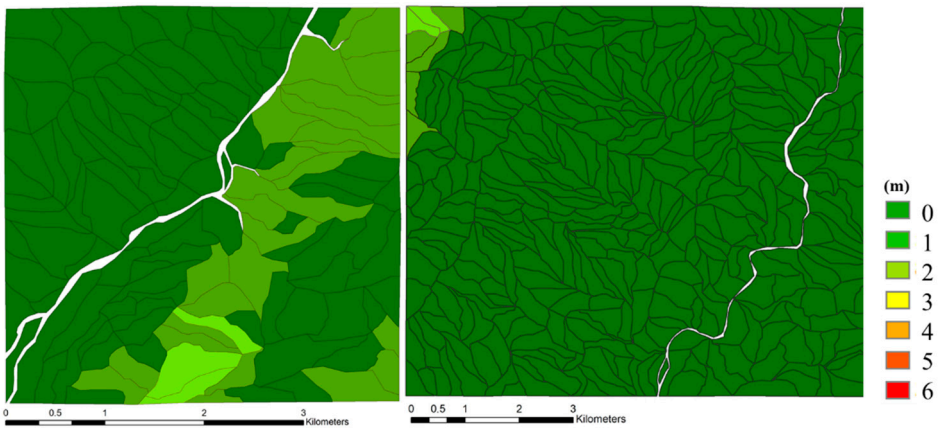
(**i**) road distance



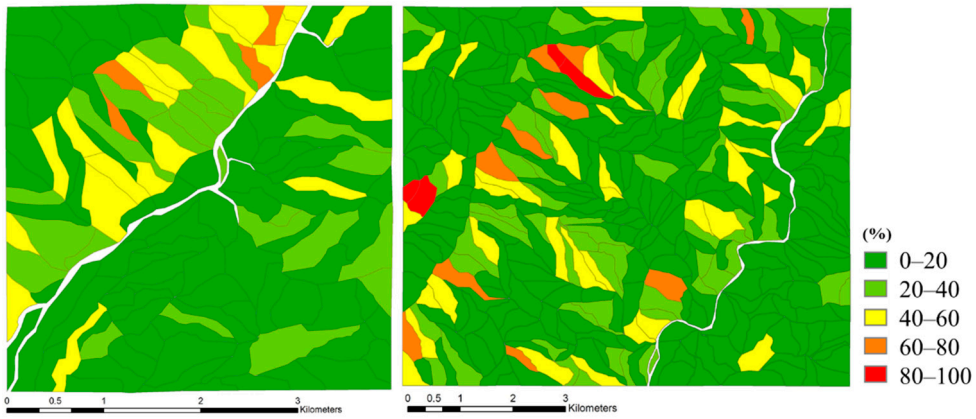
(**j**) fault distance



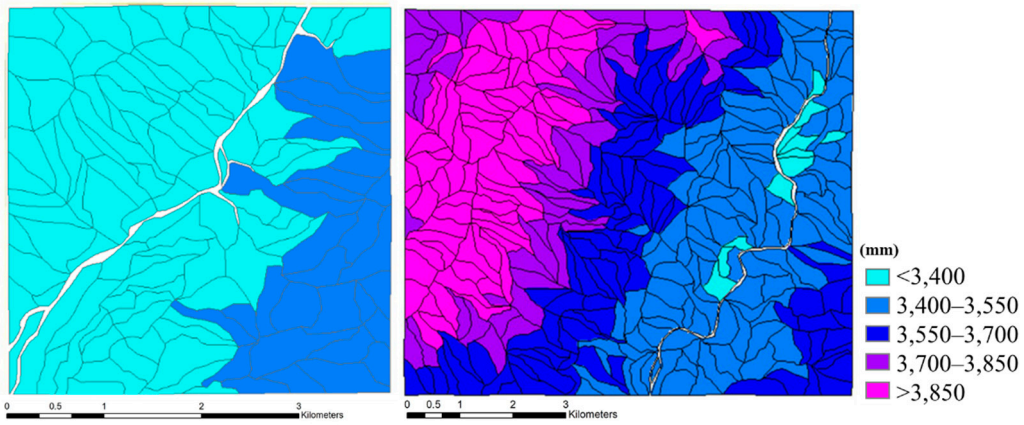
(**k**) rock mass strength



(**l**) folds



(**m**) dip slope



(**n**) annual rainfall

Figure 4. Fourteen spatial factors used in this study. These observations in 2017 are for Siaolin Village (left) and Putunpunas River (right).

3.2. Significance Test of Spatial Factors

The factor scales were unified from 1 to −1 through numerical standardization to solve the inconsistency of the factor value distribution. Then, the correlation between the spatial factors and landslides based on the slope units was examined. The correlation coefficient values were expressed as positive or negative. As seen in Figure 5, the correlation coefficients of Siaolin Village (yellow bar) were between −0.47 and 0.43, and those of Putunpunas River (dark-blue bar) were between −0.42 and 0.36. Hypothesis significance

testing was performed, and the probability of obtaining the test resulted in the *p*-value, as shown in Table 1. Then, the significant spatial factors were screened on the basis of a 99% reliability as the test threshold. There were five significant spatial factors in the Siaolin Village area (rock mass strength, aspect, terrain roughness, slope, and dip slopes) and six significant spatial factors in the Putunpunas River area (rock mass strength, aspect, vegetation index, water distance, terrain roughness, and dip slopes).

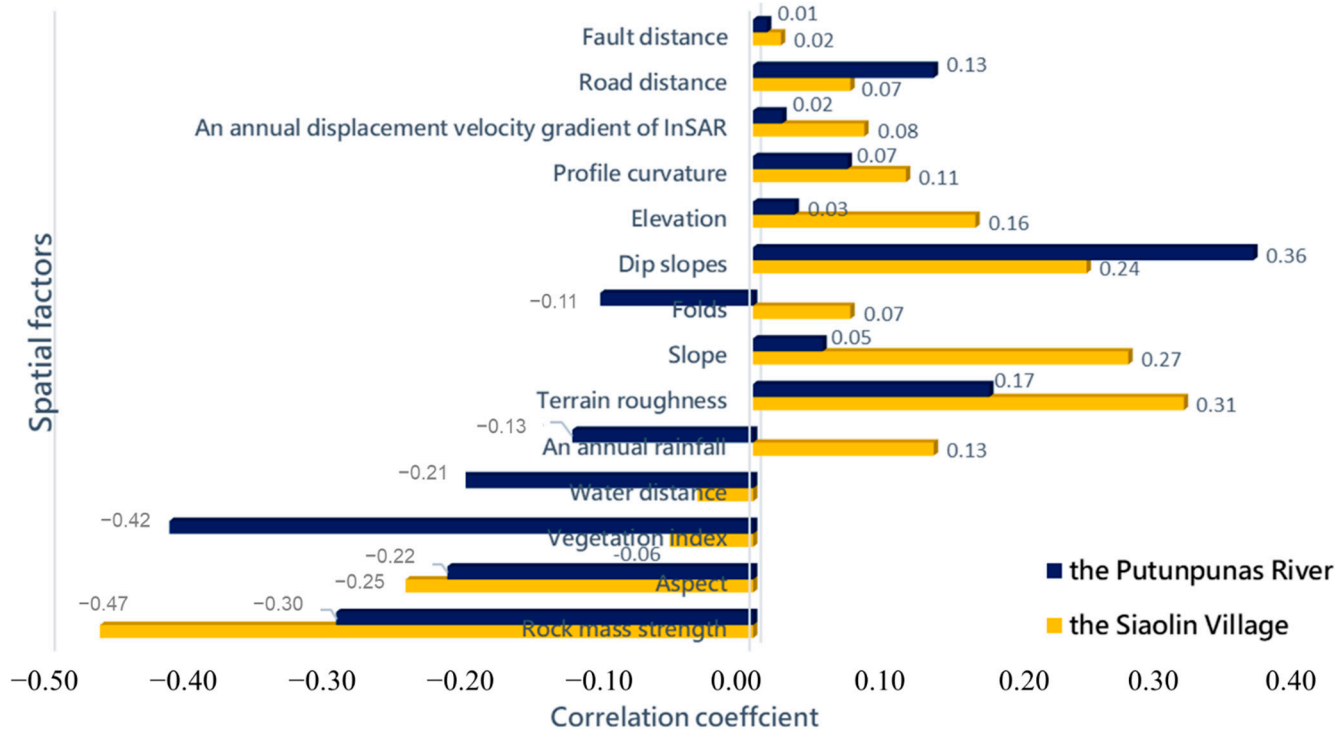


Figure 5. Histogram of correlation coefficient between landslides and the 14 spatial factors in the slope units.

Table 1. Correlation coefficients and *p*-values, quantified according to the relationship between landslides and the 14 spatial factors in the slope units.

Spatial Factor	Siaolin Village		Putunpunas River	
	Correlation Coefficient	*p*-Value	Correlation Coefficient	*p*-Value
Rock mass strength	−0.47	1.00	−0.30	1.00
Aspect	−0.25	1.00	−0.22	1.00
Vegetation index	−0.06	0.49	−0.42	1.00
Water distance	−0.04	0.37	−0.21	1.00
Annual rainfall	0.13	0.85	−0.13	0.98
Terrain roughness	0.31	1.00	0.17	1.00
Slope	0.27	1.00	0.05	0.66
Folds	0.07	0.57	−0.11	0.95
Dip slopes	0.24	0.99	0.36	1.00
Elevation	0.16	0.93	0.03	0.37
Profile curvature	0.11	0.81	0.07	0.78
Annual displacement velocity gradient of InSAR	0.08	0.62	0.02	0.29
Road distance	0.07	0.59	0.13	0.98
Fault distance	0.02	0.28	0.01	0.23

3.3. ML Prediction and Verification

According to the five machine learning methods used in this research, the relevant parameters were set as shown in Table 2. In these machine learning calculations, three years of spatial factor data (from 2007 to 2009) were used as input for learning, and the landslide prediction of the slope units was based on the 2010 spatial factors. Finally, the landslide location announced by the Central Geological Survey (MOEA, Taiwan) in 2010 was used to verify the accuracy of slope unit prediction.

Table 2. Parameters and settings required for the machine learning methods.

ML	Parameters	Values
Naive Bayes	Smoothing	10^{-9}
DT	Criterion	Gini
	The maximum of depth	20
	The minimum of samples split	10
	The minimum of samples leaf	5
Random Forest	Criterion	Gini
	The maximum of depth	20
	The minimum of samples split	2
	The minimum of samples leaf	5
	The number of estimators	100
AdaBoost	Criterion	Gini
	The maximum of depth	20
	The minimum of samples split	2
	The minimum of samples leaf	5
	The number of estimators	10
	Algorithm	SAMME
	Learning rate	0.1
XGBoost	The maximum of depth	5
	The number of estimators	1000
	Learning rate	0.1
	The minimum of child weight	1
	Gamma number	0
	Subsample number	0.8
	Colsample bytree	0.8
	Objective binary	Logistic
	nthread	4

The prediction accuracy of machine learning prediction is shown in Table 3. From the correct rate, the addition of the InSAR factor increased the accuracy of prediction by 0% to 6%. For Siaolin Village, the random forest method had the highest prediction accuracy rate (82.95%), followed by XGBoost (79.31%), AdaBoost (78.49%), naive Bayes (70.93%), and DT (68.02%). Putunpunas River showed a similar trend; the best prediction was observed from the random forest method (80.51%), followed by XGBoost (78.80%), AdaBoost (75.64%), DT (68.19%), and naive Bayes (68.19%).

Table 3. Average prediction accuracies before and after including InSAR data in different ML methods.

ML	Siaolin Village		Putunpunas River	
	with InSAR (%)	without InSAR (%)	with InSAR (%)	without InSAR (%)
Naive Bayes	70.93	70.85	68.19	68.19
DT	68.02	62.02	75.45	75.07
Random Forest	82.95	79.84	80.52	78.79
AdaBoost	78.49	77.52	75.64	75.64
XGBoost	79.31	75.97	78.80	75.80

The prediction results of the best learning method (random forest) were used to compare and evaluate the predicted classification through confusion matrixes. In Figure 6, the correctly predicted landslide slope units are colored red, and the correctly predicted noncollapsed slope units are colored cyan. In addition, the erroneously predicted landslide slope units are marked with green diagonal stripes, and the erroneously predicted nonlandslide slope units are marked with red diagonal stripes.

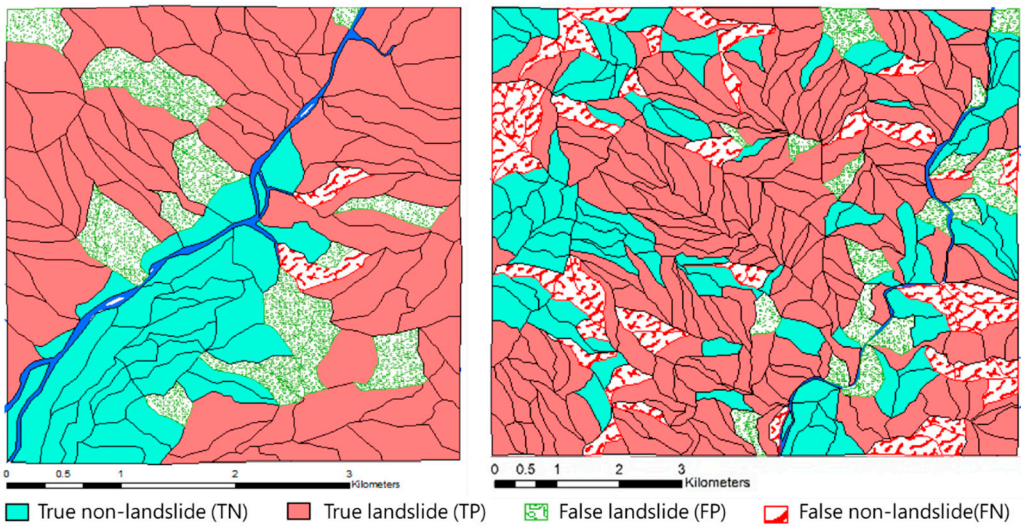


Figure 6. Visual illustration of the landslide prediction results—left: Siaolin Village; right: Putunpunas River area.

The confusion matrixes of Siaolin Village and Putunpunas River are shown in Tables 4 and 5. In Siaolin Village, the correct prediction rates of landslide and noncollapsed slope units were 78.72% and 94.30%, respectively; the average accuracy rate of the overall prediction was 82.95%. In Putunpunas River, the correct prediction rates of landslide and noncollapsed slope units were 89.67% and 66.18%, respectively; the average accuracy rate of the overall prediction was 80.52%.

Table 4. Confusion matrix for the case of the Siaolin Village analysis.

Actual \ Predictied	Lanslide	Nonlanslide	Average
Lanslide	74 (TP)	2 (FN)	-
Nonlanslide	20 (FP)	33 (TN)	-
Correct rate (%)	78.72	94.30	82.95

Table 5. Confusion matrix for the case of the Putunpunas River analysis.

Actual \ Predictied	Lanslide	Nonlanslide	Average
Lanslide	191 (TP)	46 (FN)	-
Nonlanslide	22 (FP)	90 (TN)	-
Correct rate (%)	89.67	66.18	80.52

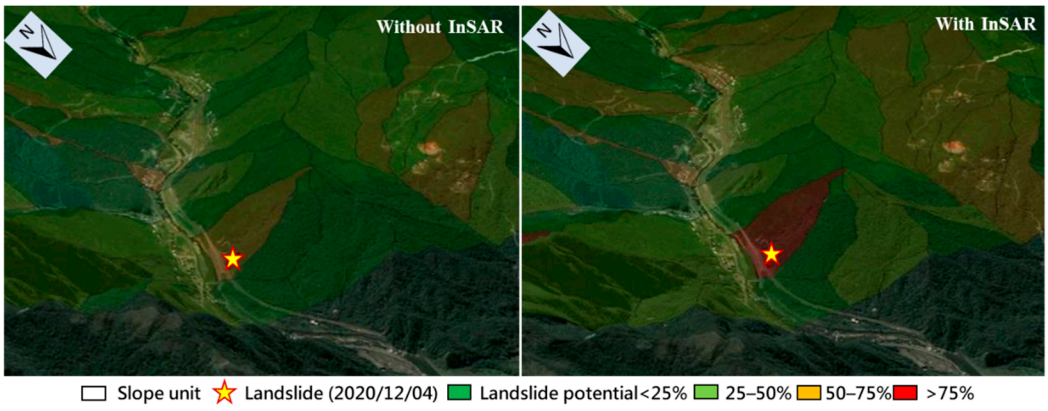
4. Discussion

Throughout the time series, the relevant spatial observation data showed changes in slopes. This study used these environmental observation data to construct the spatial factor indicators on the basis of the slope unit conditions. Significant spatial factors were then determined from the correlation analysis. According to the spatial characteristics of the slope units, the machine learning methods were applied to construct the calculation models, and the landslide potential of the slope units was evaluated.

This study was implemented with two experimental cases: Siaolin Village and Putunpunas River (Kaohsiung City, Taiwan). The experiment collected four-year spatial data (topography, locations, geology, driving categories, and landslide locations) from 2007 to 2010. Then, these data were used to construct the 14 spatial factors through indexed analysis. A common timescale (year) was established for the analysis to resolve the differences in timescales of the various spatial factors. The spatial factor datasets from 2007 to 2009 served as the input for the correlation analysis and machine learning, and the 2010 spatial factor data were used to calculate the output of potential evaluation. In the Siaolin Village area, the significant spatial factors were the rock mass strength, aspect, terrain roughness, slope, and dip slopes; the significant spatial factors in the Putunpunas River area were the rock mass strength, aspect, vegetation index, water distance, terrain roughness, and dip slopes. These significant factors in both study areas were all in the geological category, including rock mass strength, terrain roughness, and dip slopes. Obviously, the geological conditions in these areas highly influence the landslide trend.

The machine learning algorithms used in this research achieved accuracies of 60%–80% in landslide classification. Among them, the random forest method exhibited the best calculation in Siaolin Village, where it yielded a prediction accuracy rate of 82.95%; its prediction accuracy rate in Putunpunas River was 80.50%. The random forest method effectively performed independent training for high-dimensional, multi-feature factors. In addition, the random forest algorithm exhibited strong anti-interference capabilities, such as an imbalance in the number of classifications and missing parts of the feature data, so it could avoid excessive parameter setting and reduce overfitting problems. Moreover, the addition of the InSAR factor increased the accuracy of prediction up to 6%.

To further verify the proposed approach, the model established based on the training data from the two study areas was applied to another area in northern Taiwan. In December 2020, a landslide covering a slope area of around 4000 m^2 and 10,000 m^3 in earth volume occurred in this region. By feeding the local spatial factors into the model, the landslide susceptibility of each slope unit was obtained. Figure 7a,b illustrate the validation results from using 13 spatial factors (excluding InSAR data) and 14 spatial factors (including InSAR data), respectively. It shows that a medium (50%~75%) landslide potential was obtained for the landslide area if only the geological factors were considered. However, when the InSAR data were included, the model gave a high (>75%) landslide potential for that slope unit. In other words, the InSAR data provided an essential contribution for improving the prediction accuracy, as also revealed in the two study areas previously mentioned. Furthermore, it should be stressed that the model used here was established based on the training data in the two study areas in southern Taiwan, but it can still perform well in this validation case in northern Taiwan. This gives an encouraging indication that the model established based on the proposed methodology is valid not only in the study areas but could be also applicable elsewhere.



(**a**) (**b**)

Figure 7. Landslide susceptibility analysis for the 2020 landslide case in northern Taiwan: (**a**) with InSAR data; (**b**) without InSAR data.

Overall, this research reveals that InSAR observables and multiple geological factors should be integrated for landslide susceptibility analysis with machine learning technology. Future studies can refine the current timescale of annual observations into months or days to enhance the calculation accuracy. Furthermore, mechanical factors, such as fluid shearing forces and soil slippage, can be considered to improve the prediction model.

5. Conclusions

Slope instability is affected by the topography and geological conditions, and artificial construction, such as tree cutting for planting cash crops and building roads, increases the vulnerability of the landform. The prevailing extreme climate now promotes the possibility of landslide disasters in the event of short-term heavy rainfall. This study introduced the modern InSAR technology, terrain, geological, and rainfall observation data to construct spatial factors based on slope units. Through Spearman correlation analysis and verification, significant impact factors in the experimental areas were detected. More importantly, machine learning was applied for the first time to construct prediction models combining spatial factors and landslide issues. Finally, two field experiments confirmed the feasibility of the landslide susceptibility prediction analysis proposed in this study. The results prove that a better-than-80% model accuracy can be achieved by the Random Forest algorithm, and the InSAR observable is able to increase the accuracy of prediction for all training models. Relevant management will be able to follow the potential landslide slope unit to provide vegetation restoration and slope reinforcement. Eventually, this novel strategy will provide the benefits of prevention and rescue for slope landslide disasters in a forward-looking manner. Finally, it should be noted that this study only used the landslide cases in Taiwan as examples. Further studies can be conducted using the proposed methodology for the cases with various geological and climatic conditions around the world using the training data in that region.

Author Contributions: Conceptualization, Y.-T.L. and J.-Y.H.; methodology, Y.-T.L., K.-H.Y., C.-S.C. and J.-Y.H.; software, Y.-T.L. and Y.-K.C.; validation, Y.-T.L., K.-H.Y., C.-S.C. and J.-Y.H.; formal analysis, Y.-T.L. and J.-Y.H.; writing, Y.-T.L. and Y.-K.C.; visualization, Y.-T.L. and Y.-K.C.; supervision, Y.-K.C., C.-S.C. and J.-Y.H. project administration, J.-Y.H.; funding acquisition, J.-Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by CECI Engineering Consultants Inc. under contact number 06109C9007.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank the four anonymous reviewers for their constructive comments that helped to improve the original manuscript. The authors would also like to express their gratitude to the CECI Engineering Consultants Inc. in Taiwan for providing the InSAR data used in this study.

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

References

1. Lin, S.C.; Ke, M.C.; Lo, C.M. Evolution of landslide hotspots in Taiwan. *Landslides* **2017**, *14*, 1491–1501.
2. Ma, Z.; Mei, G.; Piccialli, F. Machine learning for landslides prevention: A survey. *Neural Comput. Appl.* **2020**, 1-27.
3. Benoit, L.; Briole, P.; Martin, O.; Thom, C.; Malet, J.P.; Ulrich, P. Monitoring landslide displacements with the Geocube wireless network of low-cost GPS. *Eng. Geol.* **2015**, *195*, 111–121.
4. Wang, G. GPS Landslide Monitoring: Single Base vs. Network Solutions—A case study based on the Puerto Rico and Virgin Islands Permanent GPS Network. *J. Geod. Sci.* **2011**, *1*, 191–203.
5. Ciuffi, P.; Bayer, B.; Berti, M.; Franceschini, S.; Simoni, A. Deformation Detection in Cyclic Landslides Prior to Their Reactivation Using Two-Pass Satellite Interferometry. *Appl. Sci.* **2021**, *11*, 3156.
6. 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.
7. Dabiri, Z.; Hölbling, D.; Abad, L.; Helgason, J.K.; Sæmundsson, Þ.; Tiede, D. Assessment of Landslide-Induced Geomorphological Changes in Hítardalur Valley, Iceland, Using Sentinel-1 and Sentinel-2 Data. *Appl. Sci*. **2020**, *10*, 5848.
8. Ramirez, R.; Lee, S.R.; Kwon, T.H. Long-Term Remote Monitoring of Ground Deformation Using Sentinel-1 Interferometric Synthetic Aperture Radar (InSAR): Applications and Insights into Geotechnical Engineering Practices. *Appl. Sci*. **2020**, *10*, 7447.
9. Ventisette, C.D.; Righini, G.; Moretti, S.; Casagli, N. Multitemporal landslides inventory map updating using spaceborne SAR analysis. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *30*, 238–246.
10. Shirani, K.; Pasandi, M. Detecting and monitoring of landslides using persistent scattering synthetic aperture radar interferometry. *Environ. Earth Sci.* **2019**, *78*, 42.
11. Carlà, T.; Intrieri, E.; Raspini, F.; Bardi, F.; Farina, P.; Ferretti, A.; Colombo, D.; Novali, F.; Casagli, N. Perspectives on the prediction of catastrophic slope failures from satellite InSAR. *Sci. Rep.* **2019**, *9*, 14137.
12. Utomo, D.; Chen, S.F.; Hsiung, P.A. Landslide Prediction with Model Switching. *Appl. Sci*. **2019**, *9*, 1839.
13. Chen, H.E.; Chiu, Y.Y.; Tsai, T.L.; Yang, J.C. Effect of Rainfall, Runoff and Infiltration Processes on the Stability of Footslopes. *Water* **2020**, *12*, 1229.
14. 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.
15. Al-Najjar, H.A.H.; Pradhan, B. Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. *Geosci. Front*. **2021**, *12*, 625–637.
16. Nsengiyumva, J.B.; Valentino, R. Predicting landslide susceptibility and risks using GIS-based machine learning simulations, case of upper Nyabarongo catchment. *Geomat. Nat. Hazards Risk* **2020**, *11*, 1250–1277.
17. Nhu, V.H.; Zandi, D.; Shahabi, H.; Chapi, K.; Shirzadi, A.; Al-Ansari, N.; Singh, S.K.; Dou, J.; Nguyen, H. Comparison of Support Vector Machine, Bayesian Logistic Regression, and Alternating Decision Tree Algorithms for Shallow Landslide Susceptibility Mapping along a Mountainous Road in the West of Iran. *Appl. Sci*. **2020**, *10*, 5047.
18. Shha, S.; Saha, A.; Hembram, T.K.; Pradhan, B.; Alamri, A.M. Evaluating the Performance of Individual and Novel Ensemble of Machine Learning and Statistical Models for Landslide Susceptibility Assessment at Rudraprayag District of Garhwal Himalaya. *Appl. Sci*. **2020**, *10*, 3772.
19. Lin, Y.T.; Yen, H.Y.; Chang, N.H.; Lin, H.M.; Han, J.Y.; Yang, K.H.; Chen, C.S.; Zheng, H.K.; Hsu, J.Y. Prediction of Landslides Using Machine Learning Techniques Based on Spatio-temporal Factors and InSAR Data. *J. Chin. Inst. Civ. Hydraul. Eng.* **2021**, *33*, 93–104.
20. Yu, L.; Cao, Y.; Zhou, C.; Wang, Y.; Huo, Z. Landslide Susceptibility Mapping Combining Information Gain Ratio and Support Vector Machines: A Case Study fromWushan Segment in the Three Gorges Reservoir Area, China. *Appl. Sci*. **2019**, *9*, 4756.
21. Hu, X.; Zhang, H.; Mei, H.; Xiao, D.; Li, Y.; Li, M. Landslide Susceptibility Mapping Using the Stacking Ensemble Machine Learning Method in Lushui, Southwest China. *Appl. Sci*. **2020**, *10*, 4016.
22. Chen, W.; Shahabi, H.; Zhang, S.; Khosravi, K.; Shirzadi, A.; Chapi, K.; Pham, B.T.; Zhang, T.; Zhang, L.; Chai, H.; et al. Landslide Susceptibility Modeling Based on GIS and Novel Bagging-Based Kernel Logistic Regression. *Appl. Sci*. **2018**, *8*, 2540.
23. Chen, W.; Sun, Z.; Han, J. Landslide Susceptibility Modeling Using Integrated Ensemble Weights of Evidence with Logistic Regression and Random Forest Models. *Appl. Sci*. **2019**, *9*, 171.
24. Wang, H.; Zhang, L.; Yin, K.; Luo, H.; Li, J. Landslide identification using machine learning. *Geosci. Front.* **2021**, *12*, 351–364.

25. Truong, X.L.; Mitamura, M.; Kono, Y.; Raghavan, V.; Yonezawa, G.; Truong, X.Q.; Do, T.H.; Bui, D.T.; Lee, S. Enhancing Prediction Performance of Landslide Susceptibility Model Using Hybrid Machine Learning Approach of Bagging Ensemble and Logistic Model Tree. *Appl. Sci.* **2018**, *8*, 1046.
26. Pourghasemi, H.R.; Rahmati, O. Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena* **2018**, *162*, 177–192.
27. Korup, O.; Stolle, A. Landslide prediction from machine learning. *Geol. Today* **2014**, *30*, 26–33.
28. Niu, X.; Ma, J.; Wang, Y.; Zhang, J.; Chen, H.; Tang, H. A Novel Decomposition-Ensemble Learning Model Based on Ensemble Empirical Mode Decomposition and Recurrent Neural Network for Landslide Displacement Prediction. *Appl. Sci.* **2021**, *11*, 4684.
29. Samia, J.; Temme, A.; Bregt, A.; Wallinga, J.; Guzzetti, F.; Ardizzone, F.; Rossi, M. Do landslides follow landslides? Insights in path dependency from a multi-temporal landslide inventory. *Landslides* **2017**, *14*, 547–558.
30. Wu, J.; Zeng, W.; Yan, F. Hierarchical Temporal Memory method for time-series-based anomaly detection. *Neurocomputing* **2018**, *273*, 535–546.
31. Xie, M.; Esaki, T.; Zhou, G. GIS-Based Probabilistic Mapping of Landslide Hazard Using a Three-Dimensional Deterministic Model. *Nat. Hazards* **2004**, *33*, 265–282.
32. Franklin, J.A. Safety and economy in tunneling. In *Proceedings of the 10th Canadian Rock Mechanics Symposium*; Queens University: Kingston, ON, USA, 1975; Volume 1, pp. 27–53.
33. Rish, I. An empirical study of the naive Bayes classifier. In *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*; IBM New York, USA, 2001; Volume 3, pp. 41–46.
34. Safavian, S.R.; Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **1991**, *21*, 660–674.
35. Ho, T.K. Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition IEEE, Montreal, QC, Canada, 1995; Volume 1, pp. 278–282.
36. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139.
37. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378.