Article

Development of a Novel Hybrid Intelligence Approach for Landslide Spatial Prediction

**Phong Tung Nguyen [1], Tran Thi Tuyen [2,*], Ataollah Shirzadi [3], Binh Thai Pham [4],
Himan Shahabi [5], Ebrahim Omidvar [6], Ata Amini [7], Hersh Entezami [8], Indra Prakash [9],
Tran Van Phong [10], Thao Ba Vu [11], Tran Thanh [12], Lee Saro [13,14,*] and Dieu Tien Bui [15]**

[1] Vietnam Academy for Water Resources, Hanoi 100000, Vietnam
[2] Department of Resource and Environment Management, School of Agriculture and Resources,
     Vinh University, Vinh 460000, Vietnam
[3] Department of Rangeland and Watershed Management, Faculty of Natural Resources,
     University of Kurdistan, Sanandaj 66177-15175, Iran
[4] Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
[5] Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan,
     Sanandaj 66177-15175, Iran
[6] Department of Rangeland and Watershed Management, Faculty of Natural Resources and Earth Sciences,
     University of Kashan, Kashan 87317-53153, Iran
[7] Kurdistan Agricultural and Natural Resources Research and Education Center, AREEO,
     Sanandaj 66169-49688, Iran
[8] Department of Remote Sensing and GIS, Faculty of Geography, University of Tehran,
     Tehran 14178-53933, Iran
[9] Department of Science & Technology, Bhaskarcharya Institute for Space Applications and
     Geo-Informatics (BISAG), Government of Gujarat, Gandhinagar 382007, India
[10] Institute of Geological Sciences, Vietnam Academy of Sciences and Technology, 84 Chua Lang Street,
     Dong da, Hanoi 100000, Vietnam
[11] Department of Geotechnical Engineering, Hydraulic Construction Institute, Vietnam Academy for Water
     Resources, 3/95 Chua Boc Street, Ha Noi 100000, Viet Nam
[12] NTT Hi-Tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City 700000, Vietnam
[13] Geoscience Platform Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124,
     Gwahak-ro Yuseong-gu, Daejeon 34132, Korea
[14] Department of Geophysical Exploration, Korea University of Science and Technology, 217 Gajeong-ro
     Yuseong-gu, Daejeon 34113, Korea
[15] Geographic Information System group, Department of Business and IT, University College of Southeast
     Norway, N-3800 Bø i Telemark, Norway
*   Correspondence: tuyentt@vinhuni.edu.vn (T.T.T); leesaro@kigam.re.kr (L.S)

check for updates

Abstract: We proposed an innovative hybrid intelligent approach, namely, the multiboost based naïve bayes trees (MBNBT) method for the spatial prediction of landslides in the Mu Cang Chai District of Yen Bai Province, Vietnam. The MBNBT, which is an ensemble of the multiboost (MB) and naïve bayes trees (NBT) base classifier, has rarely been applied for landslide susceptibility mapping around the world. For the modeling, we selected 248 landslide locations in the hilly terrain of the study area. Fifteen landslide conditioning factors were selected for the construction of the database based on the one-R attribute evaluation (ORAE) technique. Model validation was done using statistical metrics, namely, sensitivity, specificity, accuracy, mean absolute error (MAE), root mean square error (RMSE), and the area under the receiver operating characteristics curve (AUC). Performance of the hybrid model was evaluated and compared with popular soft computing benchmark models, namely, multiple perceptron neural network (MLPN), Support Vector Machines (SVM), and single NBT. Results indicated that the proposed MBNBT (AUC = 0.824) model outperformed the popular models, namely, the MLPN (AUC = 0.804), SVM (AUC = 0.804), and NBT (AUC = 0.800) models. Analysis

of the model results also suggested that the MB meta classifier ensemble model could enhance the prediction power of the NBT model. Therefore, the MBNBT is a suitable method for the assessment of landslide susceptibility in landslide prone areas.

1. Introduction

Landslides are a devastating natural hazard, which cause an enormous loss of properties and human life [1–3]. Many efforts have been made to correctly predict landslides, and thus, reduce landslide damages all over the world. However, landslide problems are still a great challenge to governmental agencies and hazard managers, as landslides are a phenomenon of high complexity [4] and any method for landslide prediction is dependent on the complexity of the cartographic data at the landscape level [5,6] for each particular analysis [7]. Researchers throughout the world have attempted to establish relationships between past landslide occurrences and future landslides [8–10]. Landslide prediction studies are generally carried out through mathematical models by analyzing statistical relationships between the occurrences of past landslides and landslide affecting factors [11–14].

Several models and techniques have been proposed and applied for spatial prediction of landslides all over the world. These models include quantitative and qualitative models [15]. The most popular model among quantitative models is Logistic Regression (LR), which is based on statistical theory [16,17]. This model is considered a benchmark model for comparative study [15,18–20]. Another popular quantitative model is Support Vector Machines (SVM). Among the qualitative models, Analytical Hierarchy Process (AHP), which is one of the multi-criteria decision-making (MCDM) methods, is the most popular [21]. This model is based on expert knowledge, thus the results may be biased [22].

Nowadays, some of the quantitative models that use machine learning algorithms are considered better than qualitative models [23–25], such as SVM [26,27], artificial neural networks (ANN) [28], and naïve bayes (NB) [29,30]. These models are also considered more efficient than conventional models such as frequency ratio [31,32], weight-of-evidence (WOE) [31], and AHP [33]. In addition, some of the hybrid machine learning models achieved higher performance in landslide susceptibility mapping [34,35]. Ensemble models with meta and optimization algorithms decrease the noise and over-fitting problems [18,36]. Hybrid models which are used in natural hazard studies include adaptive network-based fuzzy inference systems (ANFIS) based on the genetic algorithm (ANFIS-GA) [37,38], relevance vector machines optimized by the imperialist competitive algorithm (SVR-ICA) [39], alternating decision tree based on multiboost, bagging, rotation forest, and random subspace [36] and the radial basis function artificial neural network optimized by rotation forest (RBFRF) [15], ANFIS based on the differential evolution (ANFIS-DE) [37], ANFIS based on the biogeography-based optimization (ANFIS-BBO) and BAT algorithms (ANFIS-BAT) [40], and ANFIS based on imperialist competitive algorithm (ANFIS-ICA) and firefly algorithms (ANFIS-FA) [41].

In the present study, a novel hybrid intelligent model named MBNBT, which is combination of multiBoost (MB) and naïve bayes trees (NBT) classifier, was proposed for the spatial prediction of landslides in Mu Cang Chai District, Yen Bai Province, Vietnam. Other models based on popular machine learning techniques, namely, multiple perceptron neural network (MLPN), SVM, and single NBT were used for comparison and evaluation. Validation of the models was done using statistical metrics, namely, sensitivity, specificity, accuracy, mean absolute error (MAE), root mean square error (RMSE), and the area under the receiver operating characteristics (ROC) curve (AUC). In this study, we used ArcGIS software version 10.2 to prepare the data and maps, and Weka software version 3.9 for constructing and validating the models.

2. Study Area

The Mu Cang Chai District is located between latitudes 21°39′00″ N to 21°50′00″ N and longitudes 103°56′00″ E to 104°23′00″ E covering an approximately 1196 km² area in Yen Bai Province, Vietnam (Figure 1). The study area is occupied mainly by forest and barren, cultivable, and scrub lands. The topography of the area is hilly with elevation ranging from 280 m to 2820 m. Geologically, the area is covered by eruptive (Ngoi Thia and Tu Le complexes) and intrusive magmatic rocks (Tram Tau formation and Phu Sa Phin complex) associated with sedimentary and metamorphic rocks. Three main faults, namely, Nghia Lo, Phong Tho-Van Yen, and Nam Co-Minh are affecting the stability of rocks in the study area.
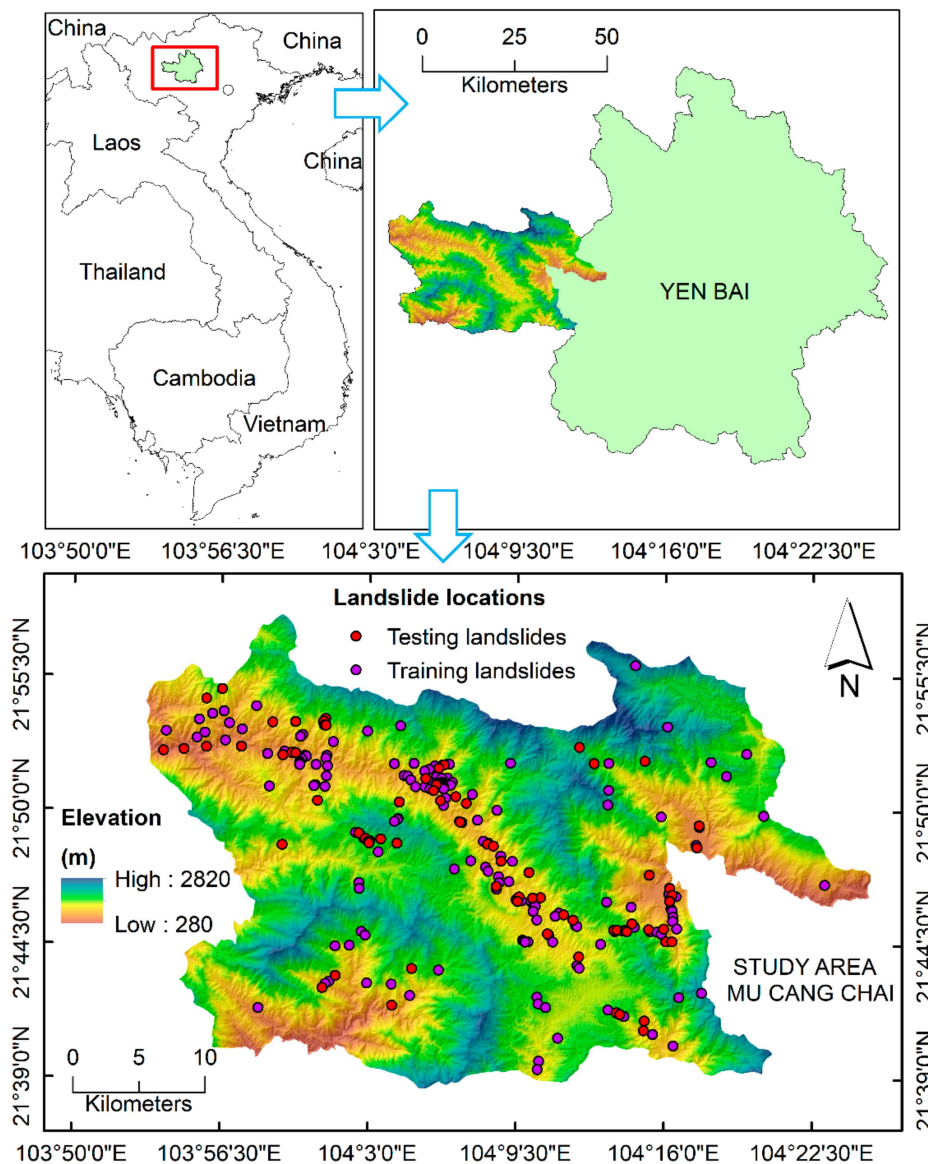
Figure 1. Location of the study area in Yen Bai Province, Vietnam and the landslide inventory map.

The area is located in a tropical monsoon region having an annual mean precipitation ranging from 3700 mm to 5490 mm mainly during the monsoon period (May to October). Mean temperature ranges from 9.7 °C (December–January) to 28°C (June–July). The annual mean temperature is 14.3 °C and the humidity is approximately 81%.

3. Data Used

A geographic information system (GIS) database including landslide inventory and affecting factors was created for the landslide spatial analysis. A landslide inventory map records the locations of landslides and other information, such as the date of occurrence and the types of ground/rock mass movements wherever available. Landslide inventory, in this study, was created from 248 historical landslide events which were identified and mapped in the study area by interpreting air photos, Landsat imageries, and Google Earth images. Field surveys were carried out under a national project in Vietnam for checking the ground truth of the occurrence of landslides (Figure 1). The largest landslide event involving 100,000 m^3 volume was recorded at the Che Cu Na commune (2011). The types of landslides, which occurred in the study area, included translational (35 events), mixed (36 events), toppling (45 events), rotational (124 events), and debris slides (8 events).

Landslide affecting factors (parameters) such as slope, aspect, profile curvature, elevation, distance to rivers, river density, curvature, distance to roads, road density, plan curvature, distance to faults, fault density, land use, lithology, and rainfall were considered for the landslide analysis as they are known as important factors to landslide occurrences in any area [42,43]. More specifically, slope affects landslide occurrences, as landslides often occur at certain critical slope angles depending on the nature of the ground mass and orientation of the sliding plane [44,45]. This aspect is related with precipitation falling on the slope, solar radiation, soil conditions, and vegetation; thus, it is considered one of the condition factors in landslides [44,45]. Profile curvature, plan curvature, and curvature represent the morphology of the surface which control the run-off and accumulation of surface water; thus, they have an effect on landslide occurrences [44]. Elevation is considered as one of the condition factors in landslides, as it is related to the weathering of soil and rocks on the slope [45]. At higher elevations, the weathering is generally much less. Distance to rivers and river density are landslide conditioning factors, as the ground mass near rivers is generally more saturated with water and a high-density drainage area drains out more surface water (run off) [44]. Distance to roads and road density affect landslide occurrences, as excavation for roads create more instability in the ground/rock mass, thus more landslides [44]. Distance to faults and fault density are considered as conditioning factors in landslides, as faults themselves cause landslides and ground/rock mass near faults are generally more fractured and vulnerable to sliding [44]. Land-use patterns greatly affect landslide occurrences. Landslides occur more in barren lands and in areas of agricultural activities [46]. Lithology is also considered a landslide conditioning factor, as the physical properties of soil and rock materials including their strength, porosity, permeability, and weathering affect sliding [45]. Rainfall is considered as one of the triggering factors in landslides, as it reduces soil cohesion and increases the pore pressure [47], and its influence on slope stability also depends on the duration and intensity of the rainfall.

In this study, the factor maps were generated from a digital elevation model (DEM) with a 20 m spatial resolution constructed from contours extracted from topographic maps at a 1:50,000 scale, a geology map at a 1:50,000 scale collected from the Vietnam Institute of Geosciences and Mineral Resources, Google Earth images, and meteorological data on the area, which were classified into different classes, as shown in Figure 2, in the raster format of 20 m resolution. Out of these, the lithological map was classified into six groups based on lithological characteristics including Group 1—acid–neutral igneous magmatic rocks and tuff; Group 2—acid–neutral intrusive magmatic rocks; Group 3—terrigenous sedimentary rocks with rich aluminosilicate components; Group 4—mafic-ultramafic magma rocks; Group 5—carbonate rocks; and Group 6—quaternary deposits [48,49]. Meteorological data were collected from global weather data for SWAT [29,50] for a 31 year period (1984–2014) to generate a rainfall map. Distance to features (i.e., roads, rivers, faults) and feature density were constructed using feature extraction from the topographic map and geological maps. In order to use the datasets for modeling, the conditioning factors were reclassified into various sub-classes [43] on the basis of the frequency analysis of landslides occurring in the area [8–10].
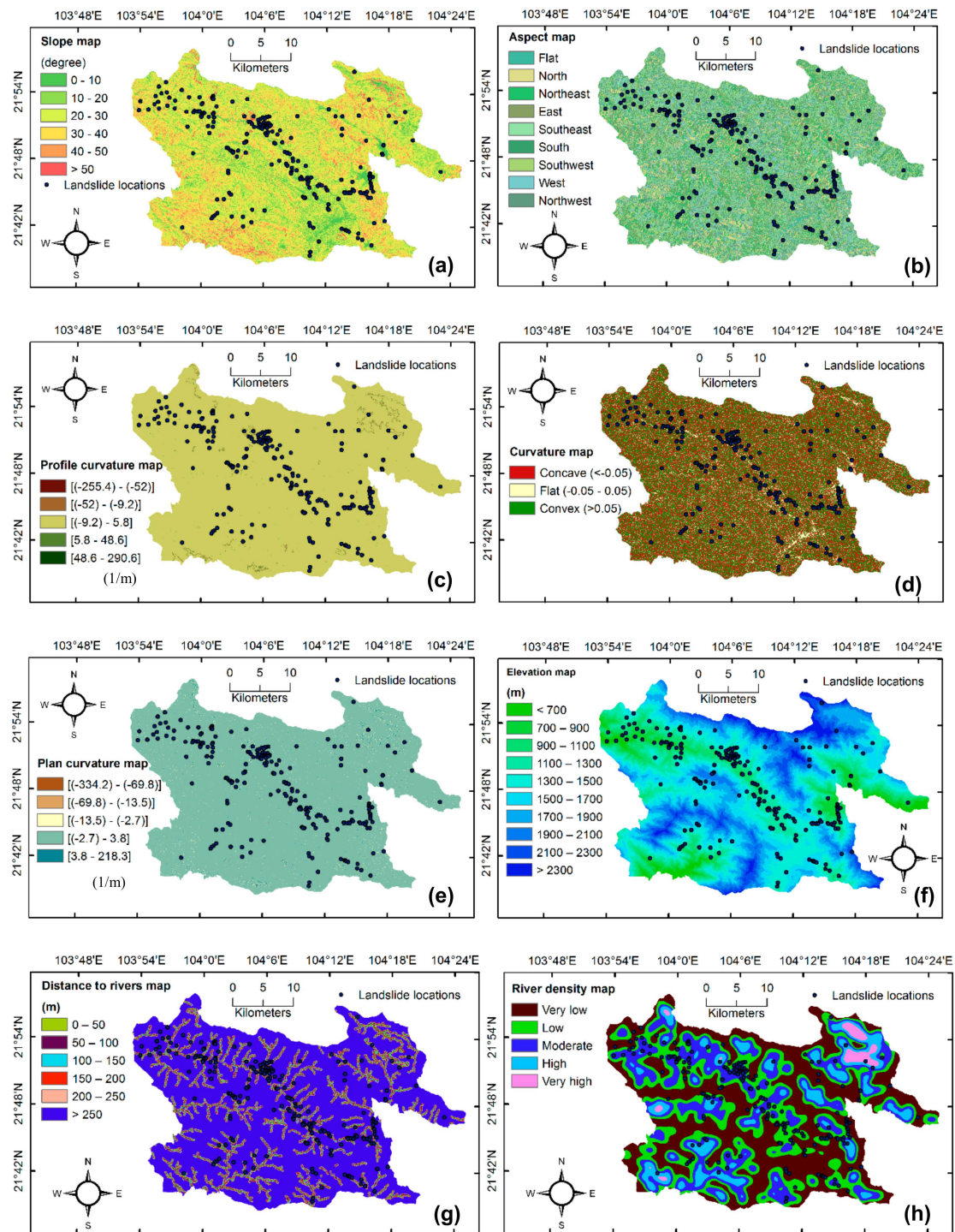
Figure 2. *Cont.*
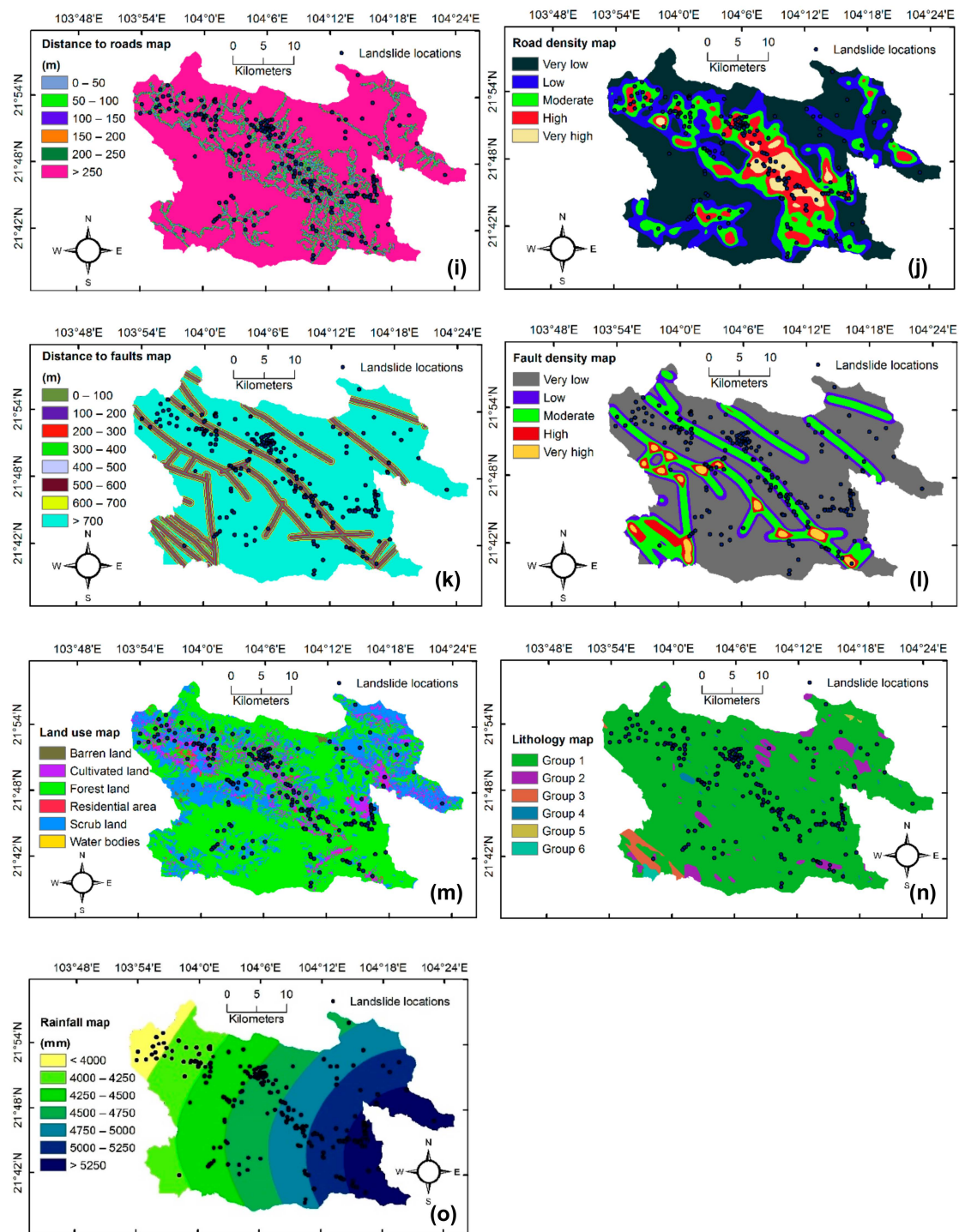
Figure 2. Maps of the landslide affecting parameters: (**a**) slope map, (**b**) aspect map, (**c**) profile curvature map, (**d**) curvature map, (**e**) plan curvature map, (**f**) elevation, (**g**) distance to rivers map, (**h**) river density map, (**i**) distance to roads map, (**j**) road density map, (**k**) distance to faults map, (**l**) fault density map, (**m**) land use map, (**n**) lithology map, and (**o**) rainfall map.

4. Methods Used

4.1. MultiBoost (MB)

The MB is considered an effective ensemble machine learning method which can help significantly in enhancing the efficiency of weaker classifiers [51]. It was proposed by Webb [52] as a combination of the adaboost ensemble and wgging technique (a variant of bagging). The main principle of the MB is to utilize the weighted aggregation of multiple classifiers generated during the selection of the bootstrap samples for classification [53]. The MB takes advantages of both wagging (reducing the variance) and adaboost (reducing bias and variance); therefore, it is more efficient than adaboost and wgging alone [52]. So far, the MB has been utilized efficiently in the fields of medical [53] and computer sciences [54]. However, the application of the MB is not popular in landslide studies.

4.2. Naïve Bayes Trees (NBT)

The NBT, which was proposed by Kohavi [55], is a hybrid intelligent approach of two machine learning methods of Naïve Bayesian (NB) and decision trees (DTs). The NBT is known as a classification tree method in which the tree structure of the NBT is constructed using the NB method at the leaves and the DTs method at the nodes [27]. The main purpose of the NBT method is to weaken the independent assumption in the NB, and deal with the fragmentation problems in the DTs method [55]. The NBT method also takes advantages of both NB and DTs. Therefore, it is known for having better classification accuracy than the single NB or DTs [55]. However, its predictive capability can be improved further in integration with ensemble techniques. Moreover, although the NBT has been utilized efficiently in various fields, namely, the computer sciences [56] and medical sciences [57], its application is still confined in the study of landslides. In the current study, the NBT method has been integrated with the MB ensemble method to construct the novel hybrid model (MBNBT) for landslide prediction.

4.3. Support Vector Machines (SVM)

The SVM, which was first developed by Vapnik [58], is a supervised strong classification method. It is based on the regression learning algorithm that works on statistical learning theory and the structured risk minimization principle [59]. At first, a hyperplane is constructed on the training dataset in order to map the original data into a high dimensional feature space [60]. Theoretically, this hyperplane separates the original input space and the four kernel functions and converts the target dataset into the two classes of landslide and non-landslide [61]. The result of the SVM modelling depends on the four kernel mathematical functions which are used for the transformation of data in the SVM, including radial base kernel function (RBF), polynomial kernel function (PF), sigmoidal kernel function (SF), and linear kernel function (LF). These functions are represented in the equations below [58,62]:

$$RBF : K(x,y) = \exp\left(-\gamma \|x - y\|^2\right) \tag{1}$$

$$PF : K(x,y) = (1 + x \cdot y)^d \tag{2}$$

$$SF : K(x,y) = \tanh(\gamma x \cdot y + r) \tag{3}$$

$$LF : K(x,y) = x \cdot y \tag{4}$$

where γ is the gamma term in the kernel function for all kernel types except linear, d is the polynomial degree term in the kernel function for the polynomial kernel, r is the bias term in the kernel function for the sigmoid kernel. γ, d, and r are user-controlled parameters, as their correct definition significantly increases the accuracy of the SVM solution [58]. The obtained rsesult from the SVM is dependent on the optimal choice of the kernel parameters. In the present study, the RBF kernel, the most used kernel function, was used to produce the landslide susceptibility map.

4.4. Multi-Layer Perceptron Networks (MLPNs)

A MLPN is one of the most important and most common ANNs [63], which is an artificial intelligence information processing system. The ANN allows for the solution of complex problems of classification, functional estimation, and optimization through the estimation of a linear or non-linear relationship between the input and output data [64–68]. It can represent and compute information from a multivariate space to another space that builds a model to generalize and predict output from input [64,69–73]. This non-linear function approximation algorithm is often used to solve classification problems [74]. In addition, the ANN is a classification of a terrain into ordinal zones of landslide susceptibility [75].

The MLPN that is often used for defining non-linear relationships, in addition to the two input and output layers, has one or more layers hidden among them [76]. Hidden layers can increase the network performance during complex functions modeling [77]. While the input layer is responsible for receiving data, the output layer determines the results of the model.

4.5. Feature Selection Based on the One-R Attribute Evaluation Technique

In landslide prediction problems, it is very important to select suitable factors, which can be used to generate the optimal input data for training and testing of the machine learning models. Feature selection is one of the techniques for this task in landslide susceptibility modeling. It helps in evaluating the importance of each factor in predicting the final results using such models on which the irrelevant or unimportant factors might be removed from the input space. Thus, it can increase the quality of input data and enhance the predictive capability of landslide models by decreasing the dimensionality of the input space, preventing redundancy, and decreasing noise and over-fitting problems [78]. There are different feature selection methods used in selecting the suitable factors for prediction modeling such as Information Gain [79,80], Forward Elimination [20], Backward Elimination [81], and One-R attribute evaluation (ORAE) [82]. Out of these methods, the ORAE, which is one of the effective filter selection methods, was selected for the first time for the landslide susceptibility modeling in this study. The main principle of the ORAE is to use statistical correlation between the output variable and a set of input factors on which it selects the most important factors for modeling. Using the ORAE, one rule (One-R) is separately constructed for each factor in the training dataset, and then the rule, which has the smallest error metric, is chosen for modelling. On the base of the smallest calculated error metrics, it will independently sort all factors according to their importance to solve prediction problems.

4.6. Validation Methods

In this study, some statistical criteria were applied; sensitivity (SEN) [61], Specificity (SPC), Accuracy (ACC), MAE, RMSE [83], and area under the ROC curve (AUC) [43,84] were used to validate the applied prediction models. In general, higher SEN, SPC, ACC, and AUC values and lower MAE and RMSE errors show better performance of the models and vice versa [43]. The SEN, SPC, ACC, and AUC are computed using four metrics including true positive (TP) (landslide correctly classified as landslide), true negative (TN) (landslide correctly classified as non-landslide), false positive (FP) (landslide incorrectly classified as landslide), and false negative (FN) (landslide incorrectly classified as non-landslide) [85,86]. According to the definition, the SEN and SPC are denoted as the fraction of landslide and non-landslide pixels that correctly and incorrectly classified [19,44]. The validation metrics can be formulated as follows:

$$SEN = \frac{TP}{TP + FN} \tag{5}$$

$$SPC = \frac{TN}{TN + FP} \tag{6}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|X_{pre.} - X_{act.}\right| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{pre.} - X_{act.})^2} \tag{9}$$

where $X_{est.}$ and $X_{obs.}$ are defined as predicted values obtained from modeling and actual values obtained from real observation, respectively, n is defined as the total number of samples used in the datasets.

Another standard statistical metric to validate the models is the ROC curve analyses [87,88]. The ROC curve is plotted based on the sensitivity and 100–specificity on the *x*- and *y*-axis, respectively. The AUC is used to judge the performance of a model in which a value of 1 indicates an accurate model, while an AUC equal to 0.5 is an inaccurate model [41]. The AUC is calulated based on the following equation:

$$AUC = \sum TP + \sum TN/P + N \tag{10}$$

where, *P* and *N* are defined as the total number of landslide and non-landslide samples, respectively.

5. Development of the MBNBT Model for Landslide Susceptibility Mapping

Landslide susceptibility assessment using the MBNBT was carried out in four main steps: (1) dataset generation, (2) model construction, (3) model validation, and (4) development of the landslide susceptibility map (LSM) (Figure 3).
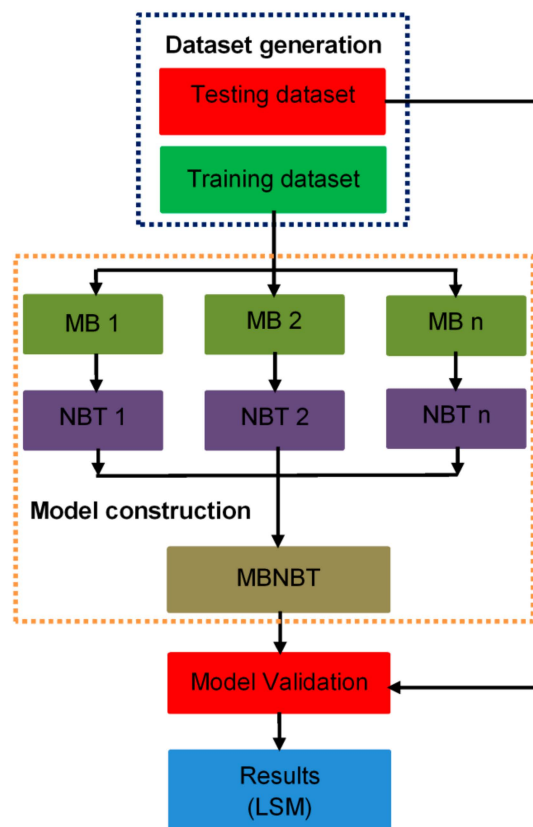

Figure 3. Methodology chart of the present study.

5.1. Generation of Datasets

In the initial step, the training and testing datasets were generated for landslide spatial prediction, where the landslides were divided into the two parts of training and testing landslides [89]. In this study, the training dataset was generated from 174 landslide locations (70% landslides) and 174 non-landslide locations, whereas the testing dataset was generated from 74 landslide locations (30% remaining landslides) and 74 other non-landslide locations. A ratio of 70/30 training and testing landslides were randomly selected in this study by using the built-in random point selection function tool in the ArcGIS software. Fifteen landslide conditioning factors were taken into account when generating the datasets. The ORAE feature selection method was applied to evaluate importance of these factors for selecting the suitable factors for model construction and validation.

5.2. Model Construction

Model construction involved two main steps: (i) optimization and (ii) classification.

(i) Optimization: In this step, the MB ensemble was utilized to optimize the original training dataset for creating optimal inputs for classification. Different sub-training datasets were first generated using different iterations. Thereafter, the optimal input data was determined with the optimal number of iterations. In this study, the optimal number of iterations was set "20".

(ii) Classification: In this step, the NBT classifier was applied to classify two variables (landslide and non-landslide) using optimal training datasets generated by the MB for the spatial prediction of landslides.

5.3. Model Validation and Comparison

In this step, the testing dataset was used to validate the performance of the models. The methods, namely, SEN, SPC, ACC, MAE, RMSE, and AUC were used to validate the models. In addition, other single benchmark models, namely, SVM, MLPN, and single NBT were selected for comparison.

5.4. Development of Landslide Susceptibility Map

In this step, the results of the trained models were used to produce the landslide susceptibility maps. These maps were generated in five susceptible classes: very high, high, moderate, low, and very low based on the susceptible indexes using the geometrical intervals method [90].

6. Results and Analysis

6.1. Importance of Landslide Conditioning Factors Using the ORAE Method

The results of the factor selection based on the ORAE method is shown in Figure 4. The most important factors for landslide occuurence were determined based on the average merit (AM) metric of this method. The results indicated that, although all 15 conditioning factors had an effect (AM > 0) on landslide incidence, road density (AM = 72.701) had the highest predictive capability because most of the observed landslides in the Mu Cang Chai District were located near roads. It was followed by elevation (AM = 63.793), distance to road (AM = 63.218), aspect (AM = 60.344), land use (AM = 58.046), river density (AM = 57.758), distance to river (AM = 56.609), lithology (AM = 53.735), curvature (AM = 52.586), fault density (AM = 49.137), slope (AM = 49.137), profile curvature (AM = 48.850), plan curvature (AM = 48.850), distance to fault (AM = 48.563), and rainfall (AM = 45.114) (Figure 4).
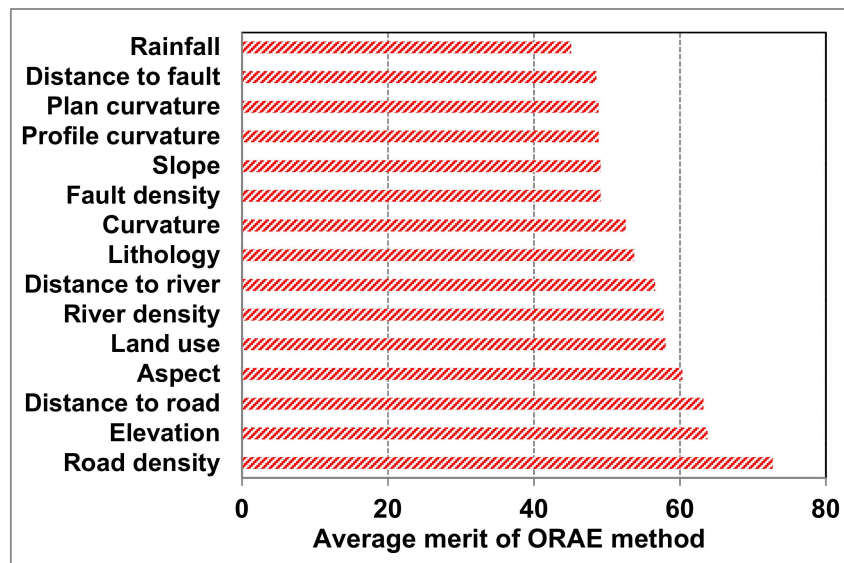
Figure 4. Factor selection using the One-R Attribute Evaluation (ORAE) method.

6.2. Model Validation and Comparison

Validation of the new proposed model and other benchmark landslide models was carried out on both the training and testing datasets. The comparative validation results of the different models are shown in Figure 5 and Tables 1 and 2, for the training and validation datasets. Out of these, the training dataset in the modelling was used for goodness-of-fit analysis. The performance or predictive capability of the models were evaluated using the validation dataset. The results of the models based on the training dataset revealed that all models predicted the spatial distribution of landslides very well. However, the results showed that the new proposed model, MBNBT, had the highest SEN (0.882), SPC (0.900), ACC (0.891), AUC (0.924), and also lower error values of MAE (0.168) and RMSE (0.224) metrics in comparison to the other models (Table 1). Similar results were obtained for the MBNBT model, based on the validation dataset: SEN (0.763), SPC (0.778), ACC (0.770), AUC (0.831), and also lower error values of MAE (0.236) metrics (Table 2). The results of the benchmark models based on the training dataset showed that although the NBT had the most goodness-of-fit (AUC = 0.831), the most predictive capability using the validation dataset was of the MLPNs (AUC = 0.810) model, followed by the NBT (AUC = 0.802) and SVM (AUC = 0.800) models (Table 2). It can be concluded that the new proposed model (MBNBT) outperformed and outclassed the other soft computing benchmark models (SVM, MLPN, and NBT) for the spatial prediction of landslides.

Table 1. Model performance using the training dataset.

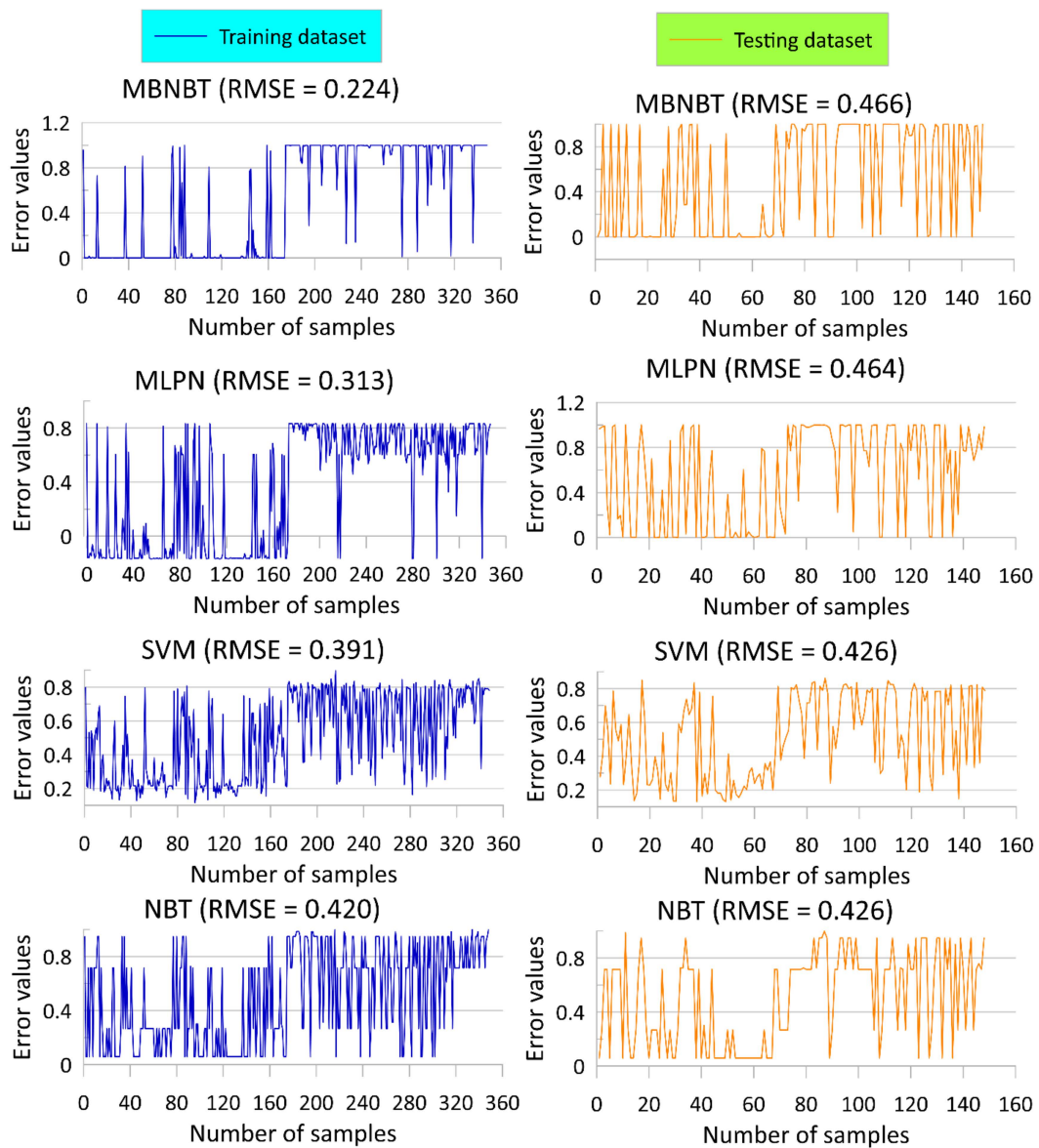| Criteria	MLPN	SVM	NBT	MBNBT
TP	148	134	132	157
TN	128	131	125	153
FP	26	40	42	17
FN	46	43	49	21
SEN (%)	0.763	0.757	0.729	0.882
SPC (%)	0.831	0.766	0.749	0.900
ACC (%)	0.793	0.761	0.739	0.891
MAE	0.307	0.302	0.340	0.168
RMSE	0.313	0.391	0.430	0.224
AUC	0.818	0.814	0.831	0.924

Figure 5. Modelling results based on the RMSE error value in the training and validation phases.

Table 2. Model performance using the validation dataset.

Criteria	MLPN	SVM	NBT	MBNBT
TP	54	59	57	58
TN	53	52	50	56
FP	20	15	17	16
FN	21	22	24	18
SEN (%)	0.720	0.728	0.704	0.763
SPC (%)	0.726	0.776	0.746	0.778
ACC (%)	0.723	0.750	0.723	0.770
MAE	0.342	0.314	0.350	0.236
RMSE	0.464	0.426	0.426	0.466
AUC	0.810	0.800	0.802	0.831

6.3. Development of Landslide Susceptibility Map

Mapping the area with landslide potential in the Mu Cang Chai District was carried out using all models and the final maps are shown in Figures 6–9, individually. The distribution of pixels were also carried out and is shown in Figures 6–9. The results showed that most of the pixels in the study area were distributed in the very low class (39.19%), followed by the very high class (23.40%), high class (16.86%), low class (10.35%), and moderate class (10.21%), respectively, whereas the landslide pixels were found mainly in the very high class (82.86%), followed by the high class (8.87%), moderate class (4.03%), very low class (2.42%), and low class (2.02%), respectively. Figure 10 shows the frequency ratio analysis of the four machine learning methods for landslide spatial prediction in the study area. Analysis of the findings indicated that the produced map was highly appropriate as most landslide pixels were found in the very high and high classes.
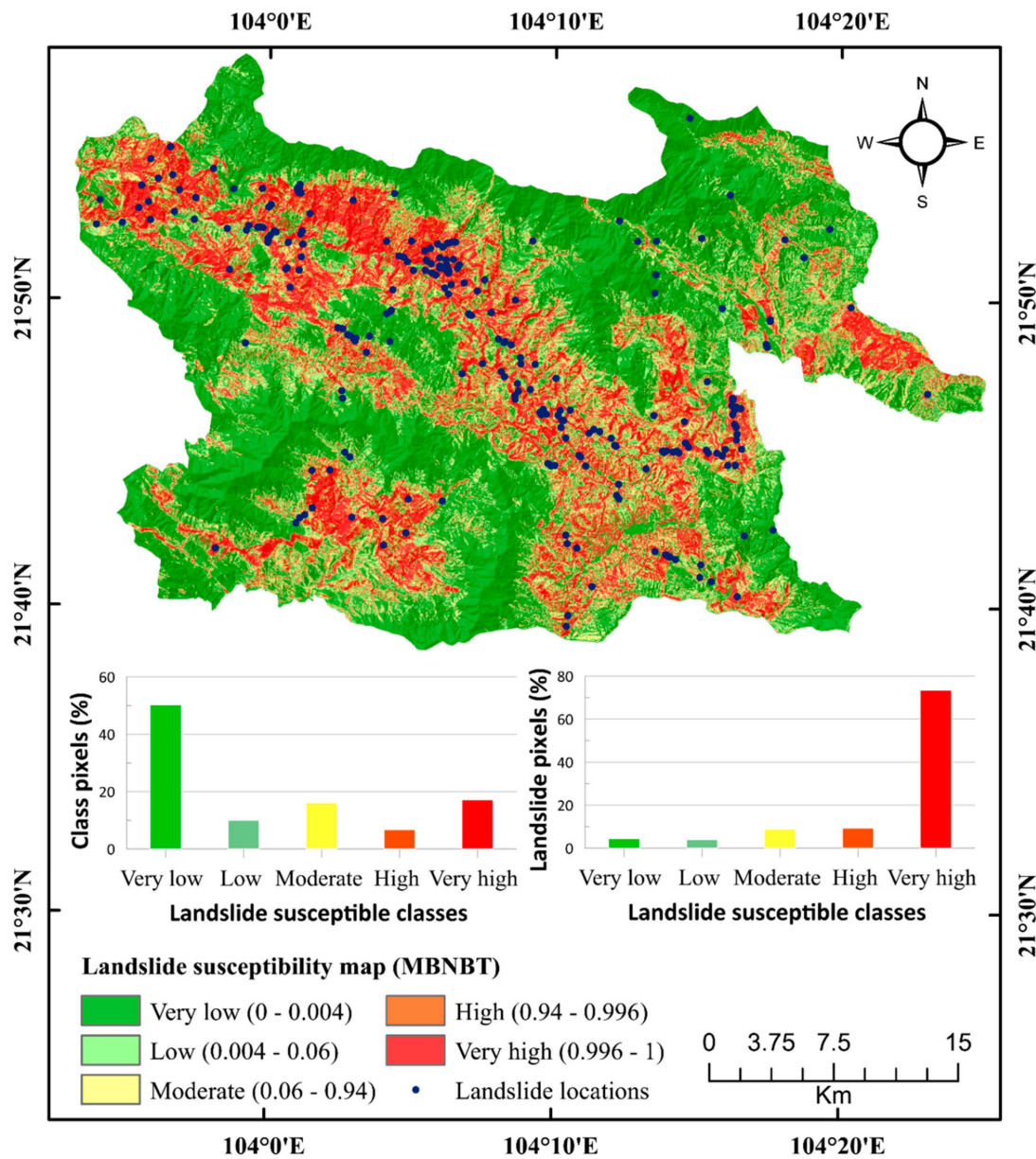

Figure 6. Landslide Susceptibility Map (LSM) of the study area based on the MBNBT model.
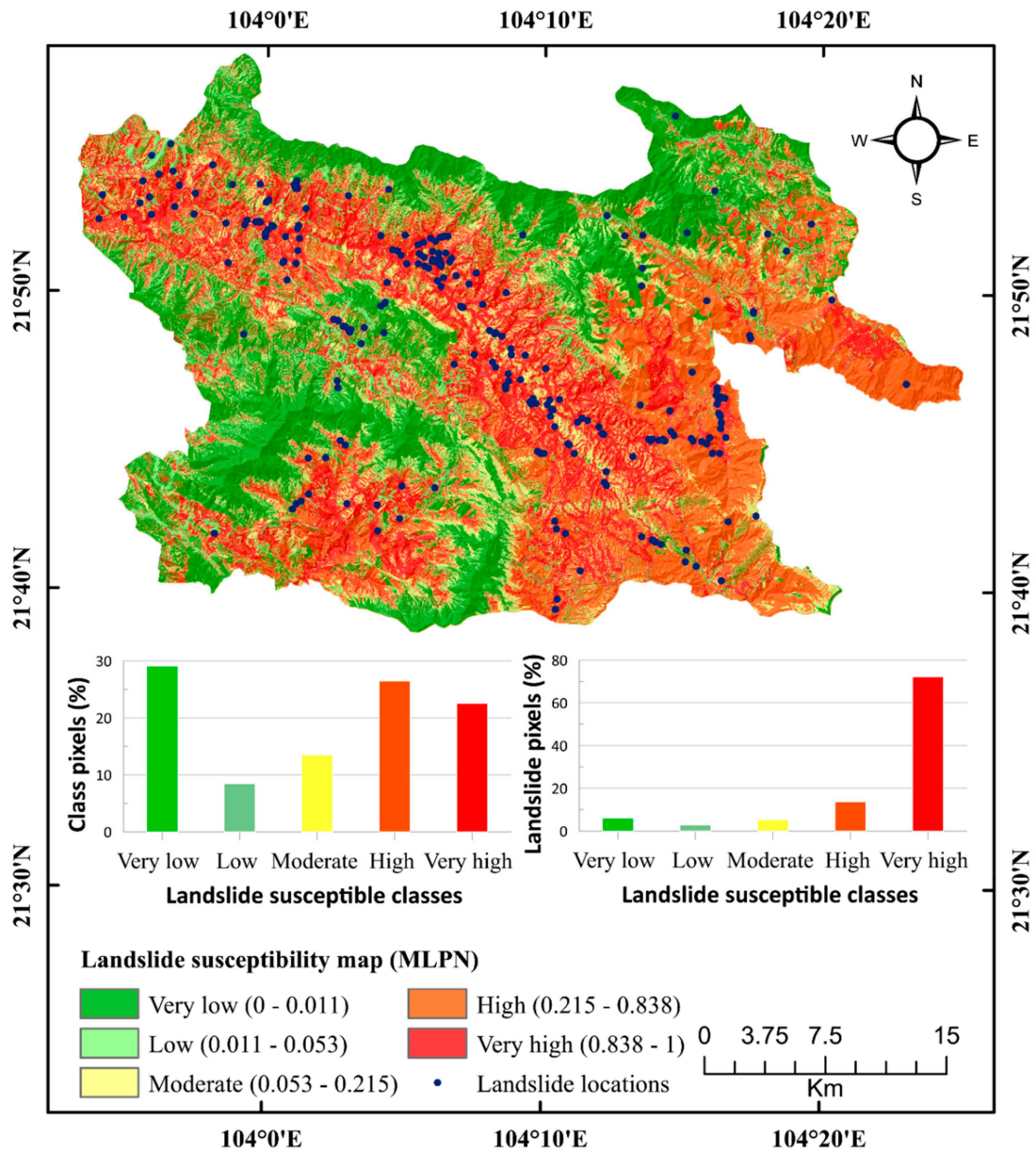
Figure 7. Landslide Susceptibility Map (LSM) of the study area based on the MLPN model.
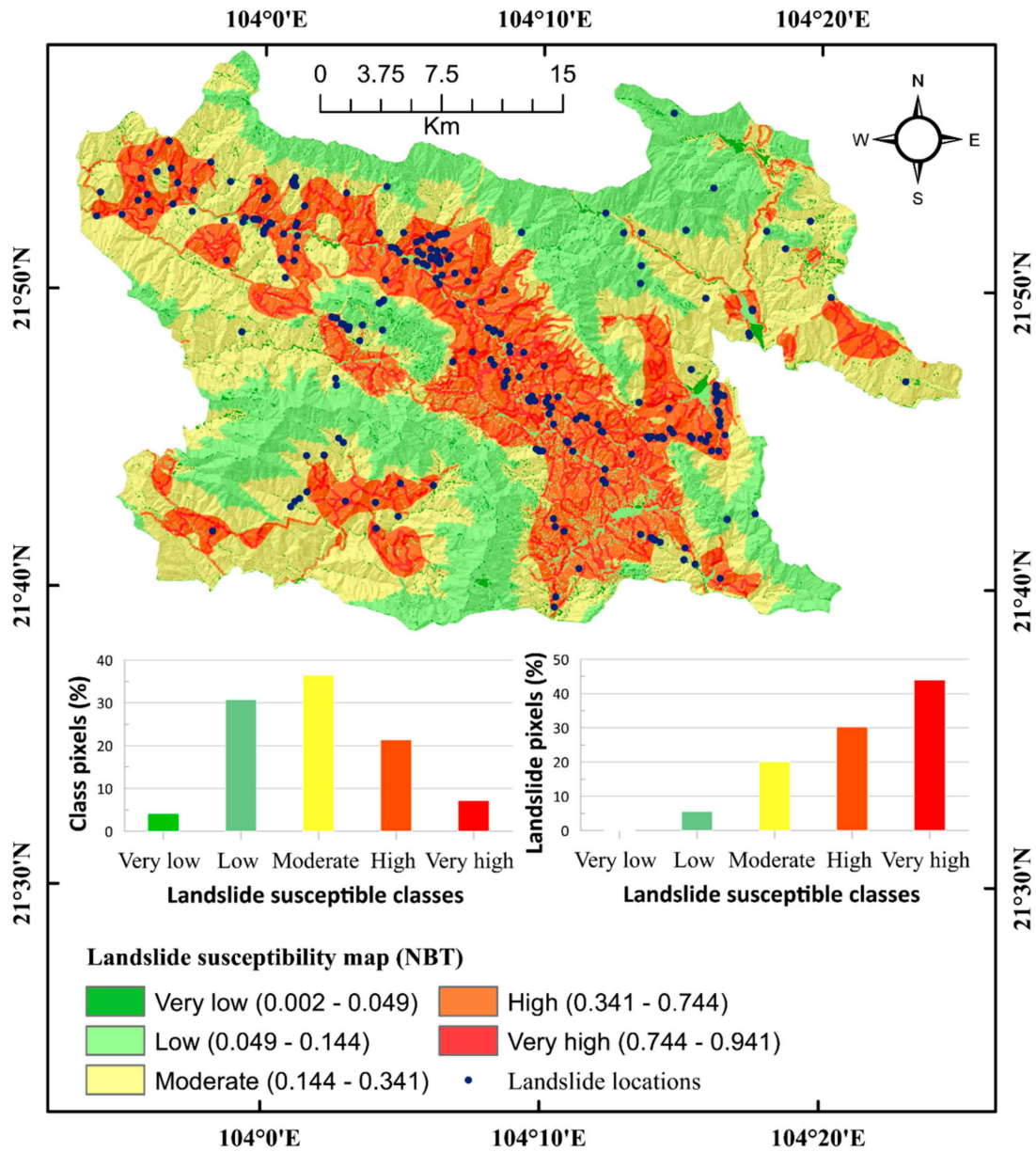
Figure 8. Landslide Susceptibility Map (LSM) of the study area based on NBT model.
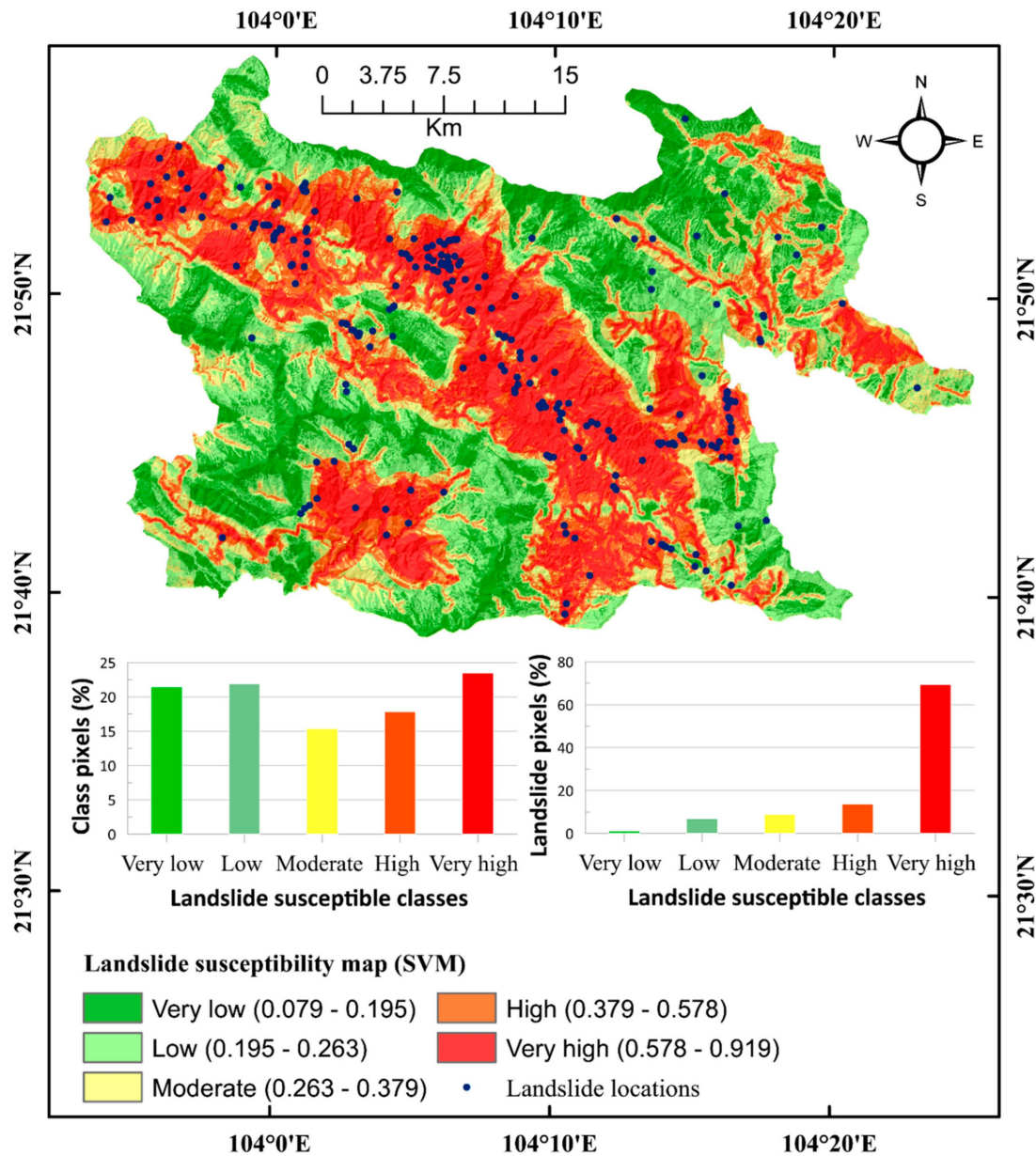
Figure 9. Landslide Susceptibility Map (LSM) of the study area based on the SVM model.

6.4. Verification of the Landslide Susceptibility Map

Validation of the efficiency of the machine learning models on producing landslide susceptibility maps was done using statistical metrics. Performance of the models' accuracy was checked and evaluated by the ROC curve and AUC methods. Training (goodness-of-fit/performance) and validation (prediction accuracy) datasets were analyzed based on the ROC curve method (Figure 11). The results of all four studied models showed high goodness-of-fit (AUC > 0.814); however, the new proposed model (MBNBT) had the best prediction accuracy (AUC = 0.825) for spatial prediction of landslides. The results also indicated that the MBNBT (AUC = 0.825) ensemble model enhanced and improved the prediction accuracy of NBT (AUC = 0.800) as a base classifier. Thus, it can be concluded that the MBNBT model outperformed and outclassed the MLPN (AUC = 0.804), SVM (0.804), and NBT (AUC = 0.800) models.
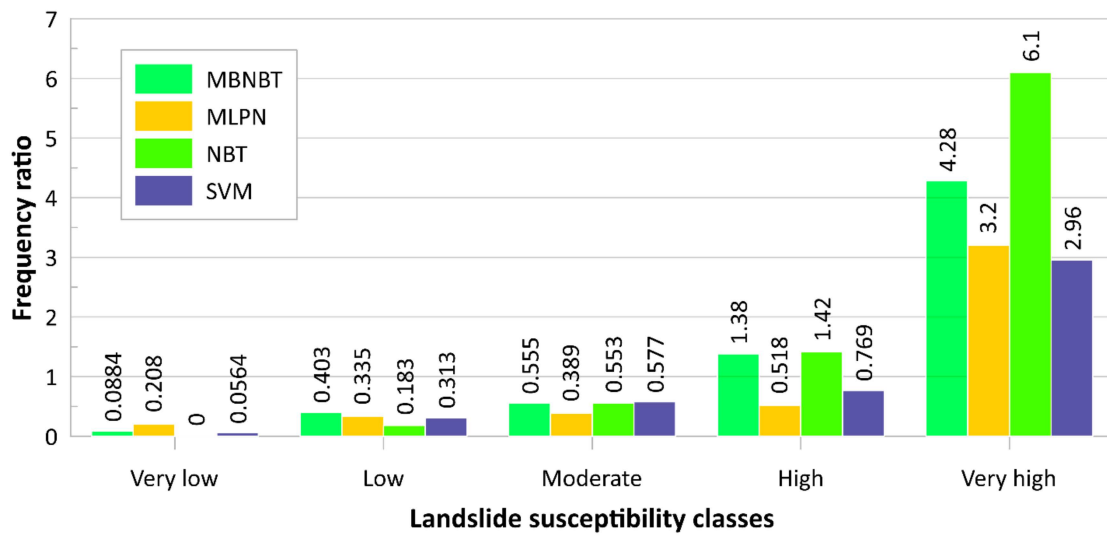
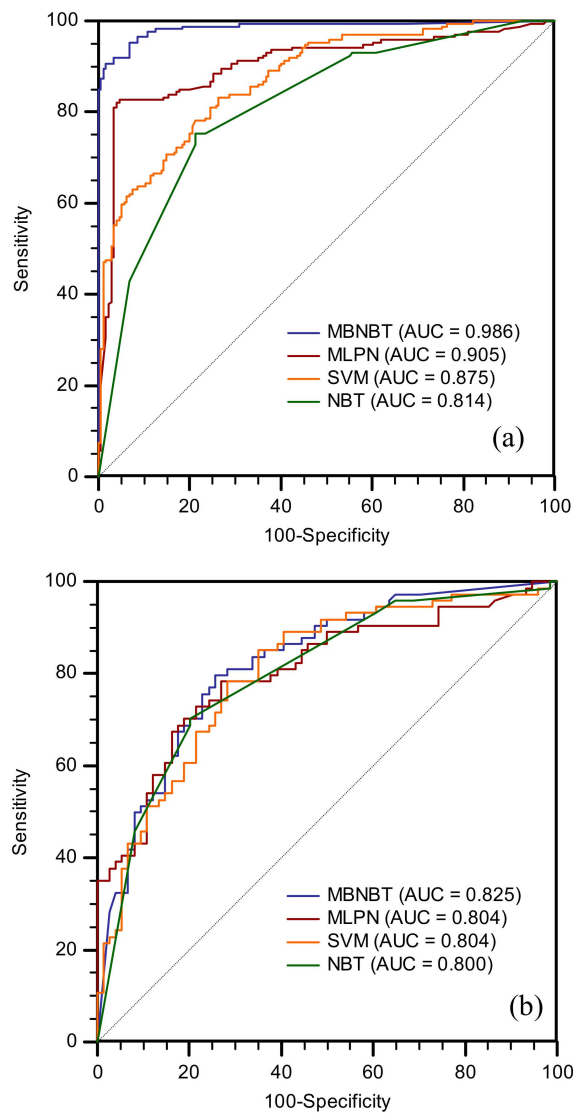**Figure 10.** Frequency ratio result of the four machine learning methods (MBNBT, MLPN, NBT, and SVM).

Figure 11. The ROC curve of the different models using the training dataset (**a**) and validating dataset (**b**).

7. Discussion

The objective of spatial landslide modeling is to generate a valid and accurate susceptibility map [44,89]. So far, many models have been built for landslide susceptibility mapping over the past few decades, out of which, machine learning algorithm-based ensemble models have received more attention in recent years [44]. In the present study, a novel hybrid intelligent model, namely, MBNBT, was introduced for the spatial prediction of landslides in the Mu Cang Chai District of Yen Bai Province (Vietnam). The performance of the MBNBT was compared with benchmark models such as MLPN, SVM, and single NBT.

Fifteen landslide conditioning factors: slope, aspect, profile curvature, curvature, plan curvature, elevation, distance to rivers, river density, distance to roads, road density, distance to faults, fault density, land use, lithology, and rainfall were selected for landslide modeling based on site condition and experience. The ORAE technique was used to select the most important conditioning factors for the landslide spatial prediction. Considering the acceptable performance of the models in the training and test stages, the ORAE can be considered a powerful technique to select important factors to enhance the power prediction capability of base/individual models while decreasing the noise and also reducing over-fitting problems. The results of the ORAE method showed that all fifteen studied factors were contributing landslide occurrences, but road density was the most important factor, as most of observed landslides in the study area were located near the roads. However, we want to make it clear that roads themselves are not responsible for landslides, but that the excavation for the roads creates instability in the surrounding ground mass, which leads to landslides [13,27,44]. Thus, the human interface in changing the existing geo-environmental conditions of the area plays an important role in landslide occurrences.

A comparative study of the predictive capability of the models using SEN, SPC, ACC, MAE, RMSE, and AUC methods indicated that the proposed novel model MBNBT was the best model for landslide spatial prediction. However, other models also gave a reasonable performance. The results of the MLPN model showed that the performance of this model was better than the SVM and NBT models, which is in agreement with the finding of Pradhan and Lee [91] and Conforti et al. [92]. Thus, the MLPN model can be used successfully for the spatial prediction of landslides. Garosi et al. [93] reported that the MLPN had good predictive performance. However, the MLPN data sampling methods employed significantly affected the performance of this model, especially when the training dataset was small [94], and this problem is considered a major deficiency of MLPNs [93]. The performance of the SVM model was reasonable in landslide susceptibility mapping [95,96]. The present study showed that the NBT had the lowest performance compared to the other models. This was because the NB-based algorithm was based on the independent assumption among predictor variables that would affect its predictive accuracy [97]; therefore, performance of NBT depends on the independence assumption [98].

Analysis of the model study results showed that the proposed MBNBT was a better model because hybrid intelligence is considered more effective than single classifiers [96]. The MBNBT takes advantages of the combination of two machine learning methods, namely, MB and NBT. More specifically, the MB used in MBNBT is known as an effective ensemble method which is able to improve the classification accuracy of single classifiers like NBT [52]. Likewise, the NBT used in MBNBT is also a good and encouraging method in landslide prediction [98] and has the advantages of both DTs and NB [55]. On the other hand, the input dataset used for MBNBT was optimized during the training process; therefore, it helped to increase the classification accuracy of the MBNBT compared with other single classifiers (MLPN, SVM, and NBT). Overall, all the studied models had reasonable efficiency for predicting the area of landslide occurrence in the study area, but the MBNBT model had the highest efficiency; thus, it can be used for better landslide susceptibility mapping.

8. Concluding Remarks

In the present study, a novel hybrid machine learning model, namely, MBNBT, was proposed for the spatial prediction of landslides in the Mu Cang Chai District of Yen Bai Province. This model is a combination of two effective machine learning techniques of the MB ensemble and the NBT base classifier. The ORAE technique was used for the selection of landslide affecting factors. Model validation was done using statistical metrics: SEN, SPC, ACC, MAE, RMSE, and AUC. Performance of the proposed model was compared with other popular models, namely, MLPN, SVM, and single NBT. Results indicated that the proposed model, MBNBT, outperformed (AUC = 0.824) the MLPN (AUC = 0.804), SVM (AUC = 0.804), and NBT (AUC = 0.800) models. Thus, the proposed novel model, MBNBT, indicates a great and promising machine learning method for landslide spatial prediction which can also be applicable for other landslide prone areas. In this study, we used a ratio of 70/30 which is a common ratio applied for generating training and testing datasets for modeling of landslide prediction. However, we propose to evaluate the performance of models with different ratios of training and testing datasets in future works for obtaining another best ratio, if any.

In the present study all the conditioning factors used for modeling had some effect on the prediction results. Thus, all these factors were considered in the analysis. However, for the removal of less important features, sensitivity analysis can be adopted depending on the requirement of the factor of safety of slope. It is also proposed to identify and classify different vegetation in the area which can prevent seepage and also act as anchor in stabilizing the ground mass.

Author Contributions: Conceptualization, B.T.P., P.T.N. and L.S.; Methodology, B.T.P., T.T.T., L.S., and I.P.; Software, B.T.P., T.V.P., L.S., A.S., and H.S.; Validation, D.T.B., H.S. and I.P.; Formal Analysis, P.T.N., T.T.T., and B.T.P.; Investigation, T.T.T., T.V.P., and T.B.V.; Resources, E.O., A.A., H.E., and P.T.N.; Data Curation, P.T.N.,T.V.P.; T.B.V., T.T., and T.T.T.; Writing—Original Draft Preparation, A.S., B.T.P.; H.S., E.O., A.A., and H.E.; Writing—Review & Editing, I.P. and B.T.P.; Visualization, T.V.P.; T.B.V., and T.T.; Supervision, B.T.P. and L.S.; Project Administration, B.T.P. and T.T.T.; Funding Acquisition, L.S.

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

References

1. Aleotti, P.; Chowdhury, R. Landslide hazard assessment: Summary review and new perspectives. *Bull. Eng. Geol. Environ.* **1999**, *58*, 21–44. [CrossRef]
2. Kanungo, D.; Sarkar, S.; Sharma, S. Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides. *Nat. Hazards* **2011**, *59*, 1491. [CrossRef]
3. Shirzadi, A.; Shahabi, H.; Chapi, K.; Bui, D.T.; Pham, B.T.; Shahedi, K.; Ahmad, B.B. A comparative study between popular statistical and machine learning methods for simulating volume of landslides. *Catena* **2017**, *157*, 213–226. [CrossRef]
4. Keefer, D.K.; Larsen, M.C. Assessing landslide hazards. *Science* **2007**, *316*, 1136–1138. [CrossRef] [PubMed]
5. Papadimitriou, F. The algorithmic complexity of landscapes. *Landsc. Res.* **2012**, *37*, 591–611. [CrossRef]
6. Papadimitriou, F. Mathematical modelling of land use and landscape complexity with ultrametric topology. *J. Land Use Sci.* **2013**, *8*, 234–254. [CrossRef]
7. Dai, F.; Lee, C.; Ngai, Y.Y. Landslide risk assessment and management: An overview. *Eng. Geol.* **2002**, *64*, 65–87. [CrossRef]
8. Sarkar, S.; Kanungo, D.P.; Patra, A.; Kumar, P. GIS based spatial data analysis for landslide susceptibility mapping. *J. Mt. Sci.* **2008**, *5*, 52–62. [CrossRef]
9. Pham, B.T.; Bui, D.T.; Prakash, I. Landslide Susceptibility Assessment Using Bagging Ensemble Based Alternating Decision Trees, Logistic Regression and J48 Decision Trees Methods: A Comparative Study. *Geotech. Geol. Eng.* **2017**, *35*, 2597–2611. [CrossRef]

10. Pham, B.T.; Khosravi, K.; Prakash, I. Application and comparison of decision tree-based machine learning methods in landside susceptibility assessment at Pauri Garhwal Area, Uttarakhand, India. *Environ. Process.* **2017**, *4*, 711–730. [CrossRef]

11. Guzzetti, F.; Reichenbach, P.; Cardinali, M.; Galli, M.; Ardizzone, F. Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* **2005**, *72*, 272–299. [CrossRef]

12. Pham, B.T.; Prakash, I. Evaluation and comparison of LogitBoost Ensemble, Fisher's Linear Discriminant Analysis, logistic regression and support vector machines methods for landslide susceptibility mapping. *Geocarto Int.* **2019**, *34*, 316–333. [CrossRef]

13. Pham, B.T.; Bui, D.T.; Prakash, I.; Nguyen, L.H.; Dholakia, M. A comparative study of sequential minimal optimization-based support vector machines, vote feature intervals, and logistic regression in landslide susceptibility assessment using GIS. *Environ. Earth Sci.* **2017**, *76*, 371. [CrossRef]

14. Pham, B.T.; Prakash, I. Machine learning methods of kernel logistic regression and classification and regression trees for landslide susceptibility assessment at part of Himalayan area, India. *Indian J. Sci. Technol.* **2018**, *11*. [CrossRef]

15. Pham, B.T.; Shirzadi, A.; Bui, D.T.; Prakash, I.; Dholakia, M. A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: A case study in the Himalayan area, India. *Int. J. Sediment Res.* **2018**, *33*, 157–170. [CrossRef]

16. Mousavi, S.Z.; Kavian, A.; Soleimani, K.; Mousavi, S.R.; Shirzadi, A. GIS-based spatial prediction of landslide susceptibility using logistic regression model. *Geomat. Nat. Hazards Risk* **2011**, *2*, 33–50. [CrossRef]

17. Shirzadi, A.; Saro, L.; Joo, O.H.; Chapi, K. A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran. *Nat. Hazards* **2012**, *64*, 1639–1656. [CrossRef]

18. Nguyen, V.V.; Pham, B.T.; Vu, B.T.; Prakash, I.; Jha, S.; Shahabi, H.; Shirzadi, A.; Ba, D.N.; Kumar, R.; Chatterjee, J.M. Hybrid Machine Learning Approaches for Landslide Susceptibility Modeling. *Forests* **2019**, *10*, 157. [CrossRef]

19. Chen, W.; Zhao, X.; Shahabi, H.; Shirzadi, A.; Khosravi, K.; Chai, H.; Zhang, S.; Zhang, L.; Ma, J.; Chen, Y. Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree. *Geocarto Int.* **2019**, 1–25. [CrossRef]

20. Pham, B.T.; Jaafari, A.; Prakash, I.; Bui, D.T. A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* **2018**, *78*, 2865–2886. [CrossRef]

21. 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]

22. Kayastha, P.; Dhital, M.R.; De Smedt, F. Application of the analytical hierarchy process (AHP) for landslide susceptibility mapping: A case study from the Tinau watershed, west Nepal. *Comput. Geosci.* **2013**, *52*, 398–408. [CrossRef]

23. Asteris, P.G.; Kolovos, K.G. Self-compacting concrete strength prediction using surrogate models. *Neural Comput. Appl.* **2017**, *31*, 409–424. [CrossRef]

24. Chen, H.; Asteris, P.G.; Jahed Armaghani, D.; Gordan, B.; Pham, B.T. Assessing Dynamic Conditions of the Retaining Wall: Developing Two Hybrid Intelligent Models. *Appl. Sci.* **2019**, *9*, 1042. [CrossRef]

25. Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhaug, I.; Prakash, I.; Bui, D.T. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci. Total Environ.* **2018**, *627*, 744–755. [CrossRef] [PubMed]

26. Yao, X.; Tham, L.; Dai, F. Landslide susceptibility mapping based on support vector machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [CrossRef]

27. Pham, B.T.; Bui, D.T.; Prakash, I.; Dholakia, M. Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS. *J. Geomat.* **2016**, *10*, 71–79.

28. Yilmaz, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat—Turkey). *Comput. Geosci.* **2009**, *35*, 1125–1138. [CrossRef]

29. Pham, B.T.; Tien Bui, D.; Pourghasemi, H.R.; Indra, P.; Dholakia, M.B. Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: A comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theor. Appl. Climatol.* **2015**, *122*, 1–19. [CrossRef]

30. Thai Pham, B.; Bui, D.T.; Prakash, I. Landslide susceptibility modelling using different advanced decision trees methods. *Civ. Eng. Environ. Syst.* **2019**, 1–19. [CrossRef]

31. Mohammady, M.; Pourghasemi, H.R.; Pradhan, B. Landslide susceptibility mapping at Golestan Province, Iran: A comparison between frequency ratio, Dempster–Shafer, and weights-of-evidence models. *J. Asian Earth Sci.* **2012**, *61*, 221–236. [CrossRef]

32. Pham, B.T.; Tien Bui, D.; Indra, P.; Dholakia, M. Landslide susceptibility assessment at a part of Uttarakhand Himalaya, India using GIS–based statistical approach of frequency ratio method. *Int. J. Eng. Res. Technol.* **2015**, *4*, 338–344.

33. Komac, M. A landslide susceptibility model using the analytical hierarchy process method and multivariate statistics in perialpine Slovenia. *Geomorphology* **2006**, *74*, 17–28. [CrossRef]

34. Tien Bui, D.; Pham, B.T.; Nguyen, Q.P.; Hoang, N.-D. Spatial prediction of rainfall-induced shallow landslides using hybrid integration approach of Least-Squares Support Vector Machines and differential evolution optimization: A case study in Central Vietnam. *Int. J. Digit. Earth* **2016**, *9*, 1077–1097. [CrossRef]

35. Shirzadi, A.; Bui, D.T.; Pham, B.T.; Solaimani, K.; Chapi, K.; Kavian, A.; Shahabi, H.; Revhaug, I. Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ. Earth Sci.* **2017**, *76*, 60. [CrossRef]

36. Shirzadi, A.; Soliamani, K.; Habibnejhad, M.; Kavian, A.; Chapi, K.; Shahabi, H.; Chen, W.; Khosravi, K.; Thai Pham, B.; Pradhan, B. Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors* **2018**, *18*, 3777. [CrossRef] [PubMed]

37. Hong, H.; Panahi, M.; Shirzadi, A.; Ma, T.; Liu, J.; Zhu, A.-X.; Chen, W.; Kougias, I.; Kazakis, N. Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Sci. Total Environ.* **2018**, *621*, 1124–1141. [CrossRef] [PubMed]

38. Tien Bui, D.; Khosravi, K.; Li, S.; Shahabi, H.; Panahi, M.; Singh, V.; Chapi, K.; Shirzadi, A.; Panahi, S.; Chen, W. New hybrids of anfis with several optimization algorithms for flood susceptibility modeling. *Water* **2018**, *10*, 1210. [CrossRef]

39. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Hoang, N.-D.; Pham, B.; Bui, Q.-T.; Tran, C.-T.; Panahi, M.; Bin Ahamd, B. A novel integrated approach of relevance vector machine optimized by imperialist competitive algorithm for spatial modeling of shallow landslides. *Remote Sens.* **2018**, *10*, 1538. [CrossRef]

40. Ahmadlou, M.; Karimi, M.; Alizadeh, S.; Shirzadi, A.; Parvinnejhad, D.; Shahabi, H.; Panahi, M. Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). *Geocarto Int.* **2018**, 1–21. [CrossRef]

41. Bui, D.T.; Panahi, M.; Shahabi, H.; Singh, V.P.; Shirzadi, A.; Chapi, K.; Khosravi, K.; Chen, W.; Panahi, S.; Li, S. Novel hybrid evolutionary algorithms for spatial prediction of floods. *Sci. Rep.* **2018**, *8*, 15364. [CrossRef] [PubMed]

42. Nohani, E.; Moharrami, M.; Sharafi, S.; Khosravi, K.; Pradhan, B.; Pham, B.T.; Lee, S.; Melesse, A.M. Landslide Susceptibility Mapping Using Different GIS-Based Bivariate Models. *Water* **2019**, *11*, 1402. [CrossRef]

43. Pham, B.T.; Tien Bui, D.; Pham, H.V.; Le, H.Q.; Prakash, I.; Dholakia, M.B. Landslide Hazard Assessment Using Random SubSpace Fuzzy Rules Based Classifier Ensemble and Probability Analysis of Rainfall Data: A Case Study at Mu Cang Chai District, Yen Bai Province (Viet Nam). *J. Indian Soc. Remote Sens.* **2016**, 1–11. [CrossRef]

44. Tien Bui, D.; Shahabi, H.; Omidvar, E.; Shirzadi, A.; Geertsema, M.; Clague, J.J.; Khosravi, K.; Pradhan, B.; Pham, B.T.; Chapi, K. Shallow landslide prediction using a novel hybrid functional machine learning algorithm. *Remote Sens.* **2019**, *11*, 931. [CrossRef]

45. Dou, J.; Yunus, A.P.; Bui, D.T.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.-W.; Khosravi, K.; Yang, Y.; Pham, B.T. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Sci. Total Environ.* **2019**, *662*, 332–346. [CrossRef] [PubMed]

46. Thai Pham, B.; Prakash, I.; Dou, J.; Singh, S.K.; Trinh, P.T.; Trung Tran, H.; Minh Le, T.; Tran, V.P.; Kim Khoi, D.; Shirzadi, A. A novel hybrid approach of landslide susceptibility modeling using rotation forest ensemble and different base classifiers. *Geocarto Int.* **2018**, 1–38. [CrossRef]

47. Dou, J.; Yunus, A.P.; Tien Bui, D.; Sahana, M.; Chen, C.-W.; Zhu, Z.; Wang, W.; Pham, B.T. Evaluating GIS-Based Multiple Statistical Models and Data Mining for Earthquake and Rainfall-Induced Landslide Susceptibility Using the LiDAR DEM. *Remote Sens.* **2019**, *11*, 638. [CrossRef]

48. Van, T.T.; Anh, D.T.; Hieu, H.H.; Giap, N.X.; Ke, T.D.; Nam, T.D.; Ngoc, D.; Ngoc, D.T.Y.; Thai, T.N.; Thang, D.V.; et al. *Investigation and Assessment of the Current Status and Potential of Landslides in Some Sections of the Ho Chi Minh Road, National Road 1A and Proposed Remedial Measures to Prevent Landslides from Threat of Safety of People, Property, and Infrastructure*; Vietnam Institute of Geosciences and Mineral Resources: Hanoi, Vietnam, 2006; p. 249.

49. Tien Bui, D. Modeling of Rainfall-Induced Landslide Hazard for the Hoa Binh Province of Vietnam. Ph.D Thesis, Norwegian University of Life Sciences, Aas, Norway, 2012.

50. NCEP. Global Weather Data for SWAT. 2018. Available online: http://globalweather.tamu.edu/home (accessed on 15 August 2018).

51. Benbouzid, D.; Busa-Fekete, R.; Casagrande, N.; Collin, F.-D.; Kégl, B. MultiBoost: A multi-purpose boosting package. *J. Mach. Learn. Res.* **2012**, *13*, 549–553.

52. Webb, G.I. Multiboosting: A technique for combining boosting and wagging. *Mach. Learn.* **2000**, *40*, 159–196. [CrossRef]

53. Kelarev, A.V.; Stranieri, A.; Yearwood, J.; Jelinek, H.F. Empirical study of decision trees and ensemble classifiers for monitoring of diabetes patients in pervasive healthcare. In Proceedings of the 2012 15th International Conference on Network-Based Information Systems (NBiS), Melbourne, Australia, 26–28 September 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 441–446.

54. Tama, B.A.; Rhee, K.H. A combination of PSO-based feature selection and tree-based classifiers ensemble for intrusion detection systems. In *Advances in Computer Science and Ubiquitous Computing*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 489–495.

55. Kohavi, R. Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid. In Proceedings of the KDD, Portland, OR, USA, 2–4 August 1996; pp. 202–207.

56. Natarajan, R.; Pednault, E. Segmented regression estimators for massive data sets. In Proceedings of the 2002 SIAM International Conference on Data Mining, Arlington, VA, USA, 11–13 April 2002; 2002; pp. 566–582.

57. Salama, M.A.; Soliman, O.S.; Maglogiannis, I.; Hassanien, A.E.; Fahmy, A.A. Rough set-based identification of heart valve diseases using heart sounds. In *Rough Sets and Intelligent Systems-Professor Zdzisław Pawlak in Memoriam*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 475–491.

58. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: Berlin/Heidelberg, Germany, 1995.

59. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

60. Choubin, B.; Moradi, E.; Golshan, M.; Adamowski, J.; Sajedi-Hosseini, F.; Mosavi, A. An Ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Sci. Total Environ.* **2019**, *651*, 2087–2096. [CrossRef] [PubMed]

61. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Alizadeh, M.; Chen, W.; Mohammadi, A.; Ahmad, B.; Panahi, M.; Hong, H. Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in Cameron Highlands, Malaysia. *Remote Sens.* **2018**, *10*, 1527. [CrossRef]

62. Lee, L.H.; Wan, C.H.; Rajkumar, R.; Isa, D. An enhanced Support Vector Machine classification framework by using Euclidean distance function for text document categorization. *Appl. Intell.* **2012**, *37*, 80–99. [CrossRef]

63. Asteris, P.; Roussis, P.; Douvika, M. Feed-forward neural network prediction of the mechanical properties of sandcrete materials. *Sensors* **2017**, *17*, 1344. [CrossRef] [PubMed]

64. Lee, S.; Ryu, J.-H.; Kim, I.-S. Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: Case study of Youngin, Korea. *Landslides* **2007**, *4*, 327–338. [CrossRef]

65. Asteris, P.G.; Nozhati, S.; Nikoo, M.; Cavaleri, L.; Nikoo, M. Krill herd algorithm-based neural network in structural seismic reliability evaluation. *Mech. Adv. Mater. Struct.* **2018**, *26*, 1146–1153. [CrossRef]

66. Asteris, P.G.; Nikoo, M. Artificial bee colony-based neural network for the prediction of the fundamental period of infilled frame structures. *Neural Comput. Appl.* **2019**, 1–11. [CrossRef]

67. Mohamad, E.T.; Hajihassani, M.; Armaghani, D.J.; Marto, A. Simulation of blasting-induced air overpressure by means of artificial neural networks. *Int. Rev. Model. Simul.* **2012**, *5*, 2501–2506.

68. Mohamad, E.T.; Faradonbeh, R.S.; Armaghani, D.J.; Monjezi, M.; Majid, M.Z.A. An optimized ANN model based on genetic algorithm for predicting ripping production. *Neural Comput. Appl.* **2017**, *28*, 393–406. [CrossRef]

69. Plevris, V.; Asteris, P.G. Modeling of masonry failure surface under biaxial compressive stress using Neural Networks. *Constr. Build. Mater.* **2014**, *55*, 447–461. [CrossRef]

70. Asteris, P.G.; Tsaris, A.K.; Cavaleri, L.; Repapis, C.C.; Papalou, A.; Di Trapani, F.; Karypidis, D.F. Prediction of the fundamental period of infilled RC frame structures using artificial neural networks. *Comput. Intell. Neurosci.* **2016**, *2016*. [CrossRef]

71. Asteris, P.G.; Plevris, V. Anisotropic masonry failure criterion using artificial neural networks. *Neural Comput. Appl.* **2017**, *28*, 2207–2229. [CrossRef]

72. Momeni, E.; Nazir, R.; Armaghani, D.J.; Maizir, H. Application of artificial neural network for predicting shaft and tip resistances of concrete piles. *Earth Sci. Res. J.* **2015**, *19*, 85–93. [CrossRef]

73. Asteris, P.; Kolovos, K.; Douvika, M.; Roinos, K. Prediction of self-compacting concrete strength using artificial neural networks. *Eur. J. Environ. Civ. Eng.* **2016**, *20*, s102–s122. [CrossRef]

74. Kawabata, D.; Bandibas, J. Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). *Geomorphology* **2009**, *113*, 97–109. [CrossRef]

75. Zare, M.; Pourghasemi, H.R.; Vafakhah, M.; Pradhan, B. Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: A comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. *Arab. J. Geosci.* **2013**, *6*, 2873–2888. [CrossRef]

76. Pijanowski, B.C.; Brown, D.G.; Shellito, B.A.; Manik, G.A. Using neural networks and GIS to forecast land use changes: A land transformation model. *Comput. Environ. Urban Syste.* **2002**, *26*, 553–575. [CrossRef]

77. Paola, J.D.; Schowengerdt, R. A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *Int. J. Remote Sens.* **1995**, *16*, 3033–3058. [CrossRef]

78. Micheletti, N.; Foresti, L.; Robert, S.; Leuenberger, M.; Pedrazzini, A.; Jaboyedoff, M.; Kanevski, M. Machine learning feature selection methods for landslide susceptibility mapping. *Math. Geosci.* **2014**, *46*, 33–57. [CrossRef]

79. Quinlan, J.R. *C4. 5: Programs for Machine Learning*; Elsevier: Amsterdam, The Netherlands, 2014.

80. Pham, B.T.; Bui, D.T.; Dholakia, M.; Prakash, I.; Pham, H.V.; Mehmood, K.; Le, H.Q. A novel ensemble classifier of rotation forest and Naïve Bayer for landslide susceptibility assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. *Geomat. Nat. Hazards Risk* **2017**, *8*, 649–671. [CrossRef]

81. Pham, B.T.; Prakash, I.; Khosravi, K.; Chapi, K.; Trinh, P.T.; Ngo, T.Q.; Hosseini, S.V.; Bui, D.T. A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. *Geocarto Int.* **2018**, 1–23. [CrossRef]

82. Holte, R.C. Very simple classification rules perform well on most commonly used datasets. *Mach. Learn.* **1993**, *11*, 63–90. [CrossRef]

83. Pham, B.T.; Pradhan, B.; Tien Bui, D.; Prakash, I.; Dholakia, M.B. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Model. Softw.* **2016**, *84*, 240–250. [CrossRef]

84. Pham, B.T.; Tien Bui, D.; Dholakia, M.B.; Prakash, I.; Pham, H.V. A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area. *Geotech. Geol. Eng.* **2016**, *34*, 1807–1824. [CrossRef]

85. Chen, W.; Shirzadi, A.; Shahabi, H.; Ahmad, B.B.; Zhang, S.; Hong, H.; Zhang, N. A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1955–1977. [CrossRef]

86. Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Bui, D.T. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **2019**, *175*, 203–218. [CrossRef]

87. Shirzadi, A.; Chapi, K.; Shahabi, H.; Solaimani, K.; Kavian, A.; Ahmad, B.B. Rock fall susceptibility assessment along a mountainous road: An evaluation of bivariate statistic, analytical hierarchy process and frequency ratio. *Environ. Earth Sci.* **2017**, *76*, 152. [CrossRef]

88. Hong, H.; Shahabi, H.; Shirzadi, A.; Chen, W.; Chapi, K.; Ahmad, B.B.; Roodposhti, M.S.; Hesar, A.Y.; Tian, Y.; Bui, D.T. Landslide susceptibility assessment at the Wuning Area, China: A comparison between multi-criteria decision making, bivariate statistical and machine learning methods. *Nat. Hazards* **2019**, *96*, 173–212. [CrossRef]

89. Shirzadi, A.; Solaimani, K.; Roshan, M.H.; Kavian, A.; Chapi, K.; Shahabi, H.; Keesstra, S.; Ahmad, B.B.; Bui, D.T. Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. *Catena* **2019**, *178*, 172–188. [CrossRef]

90. Frye, C. About the Geometrical Interval Classification Method. 2007. Available online: http://blogs.esri.com/esri/arcgis (accessed on 17 January 2019).

91. Pradhan, B.; Lee, S. Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Model. Softw.* **2010**, *25*, 747–759. [CrossRef]

92. Conforti, M.; Pascale, S.; Robustelli, G.; Sdao, F. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). *Catena* **2014**, *113*, 236–250. [CrossRef]

93. Garosi, Y.; Sheklabadi, M.; Pourghasemi, H.R.; Besalatpour, A.A.; Conoscenti, C.; Van Oost, K. Comparison of differences in resolution and sources of controlling factors for gully erosion susceptibility mapping. *Geoderma* **2018**, *330*, 65–78. [CrossRef]

94. Bui, D.T.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Landslide susceptibility assessment in the Hoa Binh province of Vietnam: A comparison of the Levenberg–Marquardt and Bayesian regularized neural networks. *Geomorphology* **2012**, *171*, 12–29.

95. Marjanovic, M.; Bajat, B.; Kovacevic, M. Landslide susceptibility assessment with machine learning algorithms. In Proceedings of the 2009 International Conference on Intelligent Networking and Collaborative Systems, Barcelona, Spain, 4–6 November 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 273–278.

96. 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.

97. Pham, B.T.; Bui, D.T.; Pourghasemi, H.R.; Indra, P.; Dholakia, M. Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: A comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theor. Appl. Climatol.* **2017**, *128*, 255–273. [CrossRef]

98. Pham, B.T.; Bui, D.T.; Prakash, I.; Dholakia, M. Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS. *Nat. Hazards* **2016**, *83*, 97–127. [CrossRef]