Article

Enhancing Prediction Performance of Landslide Susceptibility Model Using Hybrid Machine Learning Approach of Bagging Ensemble and Logistic Model Tree

Xuan Luan Truong [1,*], Muneki Mitamura [2], Yasuyuki Kono [3], Venkatesh Raghavan [2], Go Yonezawa [2], Xuan Quang Truong [4], Thi Hang Do [1,2], Dieu Tien Bui [5,*] and Saro Lee [6,7,*] 🔾

[1] Faculty of Information Technology, Hanoi University of Mining and Geology, No.14 Vien Street, Bac Tu Liem, Hanoi 10000, Vietnam; dohanghumg@gmail.com

[2] Graduate School for Creative Cities, Osaka City University, Osaka 558-8585, Japan; mitamura@sci.osaka-cu.ac.jp (M.M.); raghavan@media.osaka-cu.ac.jp (V.R.); yonezawa@media.osaka-cu.ac.jp (G.Y.)

[3] Center for Southeast Asian Studies, Kyoto University, Kyoto 606-8502, Japan; kono@cseas.kyoto-u.ac.jp

[4] Faculty of Information Technology, Hanoi University of Natural Resources and Environment, No. 14 Phu Dien, Bac Tu Liem, Hanoi 10000, Vietnam; txquang@hunre.edu.vn

[5] Geographic Information System Group, Department of Business and IT, University College of Southeast Norway, Gulbringvegen 36, N-3800 Bø i Telemark, Norway

[6] Geological Research Division, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124, Gwahak-ro, Yuseong-gu, Daejeon 34132, Korea

[7] Department of Geophysical Exploration, Korea University of Science and Technology, 217 Gajeong-ro Yuseong-gu, Daejeon 305-350, Korea

* Correspondence: truongxuanluan@humg.edu.vn (X.-L.T.); Dieu.T.Bui@usn.no (D.T.B.); leesaro@kigam.re.kr (S.L.)

check for updates

Abstract: The objective of this research is introduce a new machine learning ensemble approach that is a hybridization of Bagging ensemble (BE) and Logistic Model Trees (LMTree), named as BE-LMtree, for improving the performance of the landslide susceptibility model. The LMTree is a relatively new machine learning algorithm that was rarely explored for landslide study, whereas BE is an ensemble framework that has proven highly efficient for landslide modeling. Upper Reaches Area of Red River Basin (URRB) in Northwest region of Viet Nam was employed as a case study. For this work, a GIS database for the URRB area has been established, which contains a total of 255 landslide polygons and eight predisposing factors i.e., slope, aspect, elevation, land cover, soil type, lithology, distance to fault, and distance to river. The database was then used to construct and validate the proposed BE-LMTree model. Quality of the final BE-LMTree model was checked using confusion matrix and a set of statistical measures. The result showed that the performance of the proposed BE-LMTree model is high with the classification accuracy is 93.81% on the training dataset and the prediction capability is 83.4% on the on the validation dataset. When compared to the support vector machine model and the LMTree model, the proposed BE-LMTree model performs better; therefore, we concluded that the BE-LMTree could prove to be a new efficient tool that should be used for landslide modeling. This research could provide useful results for landslide modeling in landslide prone areas.

Keywords: landslide; bagging ensemble; Logistic Model Trees; GIS; Vietnam

1. Introduction

The problem of rainfall-induced landslides, which are triggered by high intense and long lasting precipitation, seems to be more serious in recent years in many regions around the world due to the effects of climate changes i.e., extreme rainfall events [1–8]. The rainfall-triggered landslide is especially exacerbated in countries that are located in storm centers of the world, such as Vietnam [9], Philippines [10], and China [11]. For example, the tropical typhoon of Rasmussen caused various floods and landslides with the total damages were estimated at $7 billion [12]. It anticipates that the number of landslides in the future will continue to rise due to effects of extreme rainfall events and changes of hydrological cycles [13]. Thus, landslide has become one of the hottest subject of the research community, however, accurately prediction of landslide still is a challenging real-world problem [14]. Therefore, more researches on landslide are still urgently required for deriving better detailed knowledge of slope failure and its mechanisms for designing remedial measures.

The development of a hazard map that provides detailed dimensional information of spatial distributions, temporal predictions, and destructive power of landslide is considered as an efficient tool for designing mitigation measures and management policies. However, the hazard map at the regional scale requires very detailed temporal landslide inventories that are hardly available, especially in developing countries [15]. For this context, a landslide susceptibility map (LS-map) could be alternatively employed since it helps to identify areas with high landslide probability. According to Ciampalini, et al. [16], LS-map is a valuable decision-support tool that assists local authorities in land use infrastructural planning and management

To produce susceptibility map, a variety of studying approaches has been introduced because the accuracy of the susceptibility map at regional analysis scale is controlled not only by the quality of the input maps, but also the algorithms and techniques that are employed [17]. These approaches vary from expert weighting methods to deterministic and statistical models. Evaluation of these approaches has been well presented i.e., in Chacon et al. [18] and Van Westen, et al. [19]. In recent years, new approaches that are based on advanced statistical and machine learning methods have been proposed i.e., fuzzy k-Nearest Neighbor [17]; fuzzy rule based models [20–23]; neural networks [24–30]; support vector machines [31–38]; Random Forests; metaheuristic optimized least squares support vector machines [39,40]; Cuckoo optimized relevance vector machines [41]; Chi-squared automatic interaction detection (CHAID) [42]; tree-based algorithms [43–47]; and, gene expression programming [48]. The main advantage of these methods is that they are capable of involving several to a large number of variables for reliable results, and overall, these methods are able to provide better performance models when compared to those of conventional methods [43,49,50].

In the last years, the integration of advanced machine learning algorithms and homogeneous ensemble frameworks has been explored for landslide susceptibility modeling with promising results. For example, Tien Bui, et al. [51] show that the landslide model based on a combination of functional trees with Bagging performs better than the neural network models. Pham et al. [23] concluded that the hybridization of Fuzzy Unordered Rules Induction Algorithm and Rotation forest ensemble has increased the prediction performance of the landslide model when compared to the benchmark of support vector machines model. Pham et al. [26] reported that the landslide model derived from a combinations of MultiBoost and Dagging with neural networks has significantly improved the prediction power of the landslide model using only the neural network. Thus, it could be concluded that homogeneous ensembles of machine learning are promising and should be further investigated aiming to improve the prediction capability of landslide susceptibility model.

Based on the mentioned motivation, this research aim is to expand the body knowledge of landslide modeling through introducing a new machine learning ensemble approach that combines the Logistic Model Trees (LMTree) algorithm [52] and Bagging Ensemble (BE) [53], named as BE-LMtree, for enhancing the performance of the landslide model. LMTree is a relative new and promising machine learning algorithm that was rarely explored for the landslide study, whereas Bagging ensemble is an framework that has proven efficient in landslide modeling [51,54]. Consequently, a combination of

BE and LMTree has resulted in a new powerful prediction method, and to the best of our knowledge, this is the first time that the BE-LMTree is studied for landslide susceptibility.

2. Theoretical Background of the Methods

2.1. Logistic Model Tree

Logistic Model Trees (LMTree), which is a relatively new machine learning algorithm, is developed based on the integration of tree induction algorithm and additive logistic regression [52]. The difference of LMTree when compared to the other decision tree algorithms is that the tree growing process is carried out using the LogitBoost algorithm [52,55] and the tree pruning is performed using Classification And Regression Tree (CART) [56].

Given a training dataset T = $(x_i, y_i)_{i=1}^{ds}$ with $x_i \in R^D$ is the input vector, *ds* is the number of data samples, *D* is the dimension of the training dataset, and $y_i \in (1, 0)$ is the label class. In this research context, the input vector consists of eight variables (slope, aspect, elevation, land cover, soil type, lithology, distance to fault, and distance to river), whereas the label class contains two classes, landslide (LS) and non-landslide (Non-LS). The landslide class is coded as "1" and the non-landslide is coded as "0". The objective of LMTree is to construct a tree-like structure model that is capable of classifying the training dataset into the two above classes in term of probability. The predicted numeric value to the landslide class of sample is used as susceptibility index.

Structurally, the LMTree model consists of a root node, a set of inner nodes, and a set of leaves. The aim of the training phase that includes the tree growing and the tree pruning processes is to determine the best tree structure with numbers of inner nodes and leaves. Accordingly, first, a logistic regression model Equation (1) is built at the root note using the binary LogitBoost algorithm [57] and the training dataset. In the next step, the training dataset at the root is split using the C4.5 splitting rule [58] in order to sort appropriate sub-datasets for the inner nodes, and then, logistic regression models Equation (1) for these inner nodes is built using their associated sorted datasets and the binary LogitBoost. The tree continues growing in the same procedure until it meets the stopping criterion of less than 15 samples at nodes. Finally, to prevent the LMTree model from over-fitting, the tree pruning is performed using the CART algorithm that is based on a combination of the model error and the model complexity [52].

In the LMTree building process, the binary LogitBoost algorithm [57] is used to generate logistic regression models Equation (1) for all of the inner nodes and leaves, as follows.

$$f_{LS,Non-LS}(x) = \sum_{i=1}^{D} \beta_i x_i + \beta_0 \tag{1}$$

where D is the total number of landslide input factors and β_i is the logistic coefficient.

The membership probability [52] of the landslide class at the leaves of the LMTree model is posterior probabilities derived using Equation (2) and is used as landslide susceptibility index.

$$p((LS, Non - LS)|x) = \frac{\exp f_{LS,Non-LS}(x)}{\exp f_{LS}(x) + \exp f_{Non-LS}(x)} \tag{2}$$

The complexity of the LMTree model could be estimated using the following equation [52]:

$$MC = O(dept * ds * \log n + ds * D^2 * dept + nt^2) \tag{3}$$

where MC is the model complexity; *dept* is the depth of the initial unpruned tree; *nt* is the number of nodes in the LMTree; *ds* is the number of training samples; and, *D* is the number of landslide predisposing factors.

2.2. Bagging Ensemble

Ensemble learning is a machine learning paradigm where multiple classifiers are trained and combined to enhance the prediction capability of a model. Different from popular machine learning approaches where one model is built from the training data, ensemble frameworks try to generate a set of sub-datasets from the training data, and then, each sub-dataset is used to construct a classifier, which is also called a based learner. At last, all of the based learners are combined to form the final prediction model using combination techniques i.e., averaging or majority voting [59].

Different ensemble techniques have been successfully proposed i.e., Bagging, AdaBoost, Multiboost, Stacking, and Rotation forest [60]; however, in landslide modeling, Bagging ensemble has proven robust and better than other ensembles [26,51,54], therefore, it is selected for this study.

Bagging also called Bootstrap aggregating in the full name is one of the earliest procedure for generating sub-datasets and combining based learners proposed by Breiman [53]. Using the training dataset, this technique generates bootstrap samples in which some of the samples are replicated and some samples are omitted. These bootstrap samples, which are called bootstrapped sub-datasets, are used to construct based learners using the same classification algorithm i.e., the LMTree in this work. These based learners are then combined using the majority voting strategy.

3. The Study Area and Spatial Datasets

3.1. Description of the Upper Reaches Area of Red River Basin

The study area is the Upper reaches area of the Red River Basin (URRB) (103°33′36″–104°30′50″ E, 22°05′40″–22°47′52″ N) that belongs to the Lao Cai, a north-western mountainous province in Vietnam (Figure 1). The URRB covers an area of 3273.5 km^2 with complex topography, steep slopes, and narrow valleys. The topography is highly fragmented with high mountains ranges, wide valleys, and deep streams, which result in high relief amplitudes [40]. The altitude varies from 48.1 m to 2812.6 m above sea level, with the mean and the standard deviation of 528.6 m and 484.9 m, respectively. Topographically, 61.8% of the URRB is occupied by slope angles that are higher than 15°, whereas areas with slopes less than 5° cover approximately 7.3% the total area of the URRB. The remaining 30.9% are areas located in the slope group 5–15°.
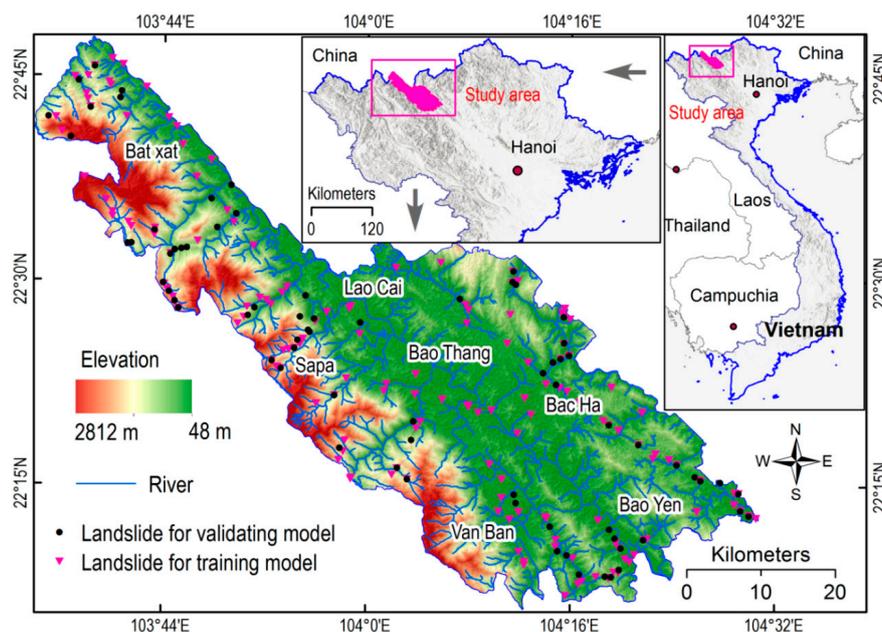

Figure 1. Location of the Upper Reaches Area of Red River Basin (Vietnam).

Hydrologically, due to the fragmentation of the terrain, the river system in the study area is dense and evenly distributed (Figure 1). These rivers are characterized by being narrow and steep, which are favorable conditions for the occurrence of flash flood and landslides. The Red River, which is the second largest river in Vietnam, is the major channel system of the URRB. This river originates from Yunnan province (China) and flows south-eastward to the study area [61].

The climate of URRB is divided into two seasons: the rainy season begins from April to October and the dry season lasts from November to March next year. The average temperature ranges range from 23 °C to 29 °C [62] and the average annual rainfall is from 1400 mm to 1900 mm [63].

The URRB is located in an active tectonic region with the relatively fast movement of the Red River fault zone that results in continuously landslide occurrences over the years [40]. It should be noted that the Red River fault zone is one of the four main tectonic features in north Vietnam that begins from Tibetan plateau (China) and extends to the Red River area of Vietnam [64,65]. Twenty seven geological formations outcrop in the basin with varied area and space distribution (Figure 2). Quaternary deposits, which consist of mainly granule, grit, breccia, pebble, boulder, and sand, cover 7.04% of the total area of the basin. Whereas, 86.68% of the basin is covered by nine geological formations, Suoi Chieng (23.62%), Ha Giang (10.96%), Nui Con Voi (10.54%), Sinh Quyen (10.43%), Ngoi Chi (8.44%), Cam Duong (8.29%), Ye Yen Sun (6.23%), Po Sen (5.96%), and Muong Hum (2.21%). The main lithologies are biotite schist, garnet-biotite gneiss, coaly shale, marble cherty shale, quartz-plagioclase-biotite schist, and two-mica schist. Detailed distribution of the lithological formations in the basin is shown in Figure 2.
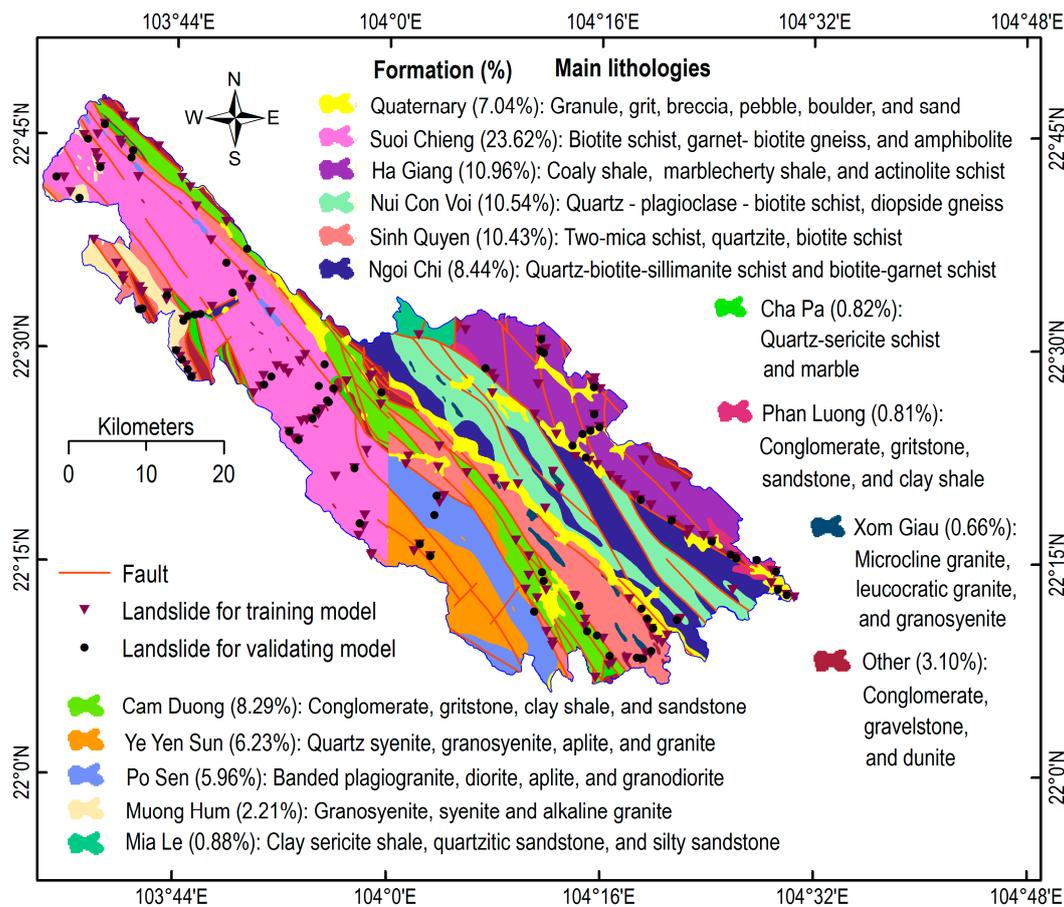

Figure 2. Geological map of the study area.

3.2. Geospatial Data

Landslide inventory map for the URRB was constructed from two main sources: (i) historic landslides from the project VAST05.02/14-15 in 2015, which was prepared by Tien Bui et al. [40]; and, (ii) landslide polygons from the State-Funded Landslide Project (SFLP) 2016 [9], a national landslide program that is carrying out in Vietnam. The SFLP project has systematically investigated and collected historic landslides for all northwest mountainous provinces in Vietnam, including the study area. Accordingly, these landslides were mainly interpreted and mapped using aerial photos and field investigations. Detailed descriptions of methods and techniques for obtaining these historic landslides in the SFLP project are present in [9].

As result, a total of 255 historic soil-mixed-boulder slides that occurred during the last two decades were registered for the landslide inventory map (Figure 1). It is noted that many rock falls were excluded out of this research because their falling mechanism are very different when compared to that of the soil-mixed-boulder slides. Analysis of the landslide inventory map showed that these slides occurred due to rainfall during tropical rainstorms [40]. Our statistical analysis of these slides showed that the largest and the smallest landslides are 116627.9 m^2 and 6.2 m^2, respectively, with the mean is 3742.5 m^2 and the standard deviation is 11467.3 m^2. Approximately 9.1% of the landslide inventories are large landslides (>10,000 m^2), whereas 9.1% of the landslide inventories are medium landslides (1000–10,000 m^2), and the remaining are landslides less than 1000 m^2. Two examples of landslide photos in the study area are shown in Figure 3.

Figure 3. Two photos of landslides in the study area: (**a**) Landslide at the Mong Sen area and (**b**) Landslide at Km 7 Lao Cai. The two photos were taken by Xuan-Luan Truong in August 2014.

Because the rainfall-trigged landslides in this study area occurred due to interactions of various geo-environmental factors, including topography, land cover, lithology, soil type, and river network [9,40,66,67], these factors were selected for this analysis. Digital elevation model (DEM) with resolution of 25 × 25 m for the URRB area was constructed using digital topographic maps 1:50,000 scale provided by the Ministry of Natural Resource and Environment of Vietnam. Using this DEM, three morphometric factors, slope, elevation, and aspect, were generated. To build the slope map (Figure 4a), seven categories were used. For the elevation map (Figure 4b), eight categories were considered. These categories were determined using Jenks natural break available in ArcGIS. For the aspect map, nice facing slopes were used (Figure 4c).

Land cover map (Figure 4d) at scale of 1:50,000 with nine classes for the URRB area was derived from the project No.02/2012/ HD-HTSP funded by Ministry of Education and Training of Vietnam. The nine classes were obtained through the classification of Landsat 8 OLI imagery in 2013 using ENVI software. Soil type map (Figure 4e) at 1:100,000 scale with 13 soil types for the URRB area was provided by Department of Agriculture and Rural Development of the Lao Cai province.

Lithological map for the URRB area was constructed based on National Geological and Mineral Resources Maps at scale of 1: 200,000, as provided by the Ministry of Natural Resource and Environment of Vietnam. Our analysis showed that more than 15 formations outcrop in the URRB area (see Figure 2). For this research, the lithological map with seven categories was constructed (Figure 4f) and these categories were separated based on clay composition, weathering characteristics, and material strength [24,68,69]. Detailed characteristics of the seven categories could be found in Tien Bui, et al. [70]. Fault is an popular factor for landslide susceptibility that was used various works i.e., in [71–73], and especially, it is an important factor for landslide modeling in areas that are affected by tectonic activities [74]. In this research, distance to fault map (Figure 4g) with seven classes [40] for the URRB area was constructed by buffering the fault lines extracted from the National Geological and Mineral Resources Maps above.
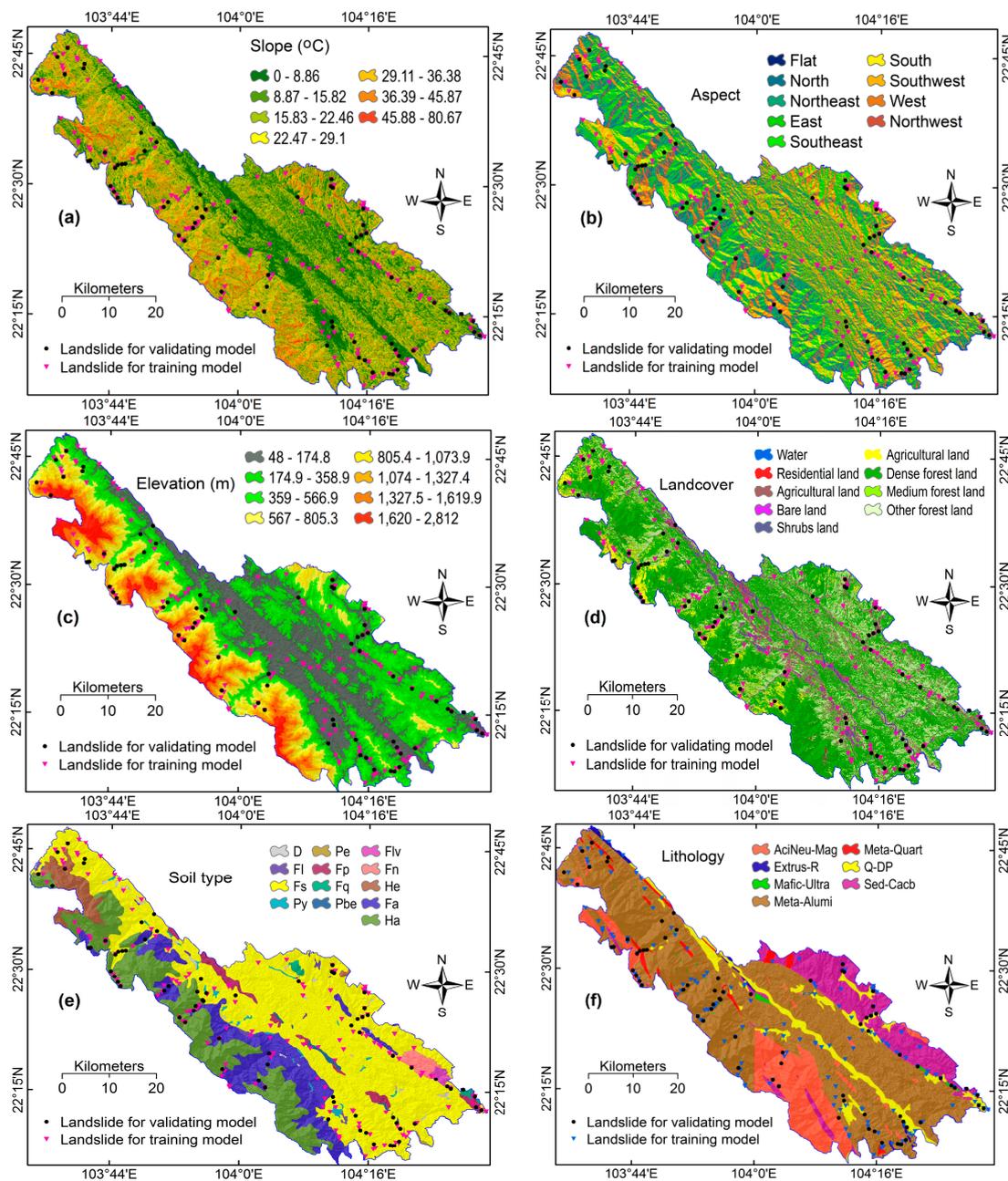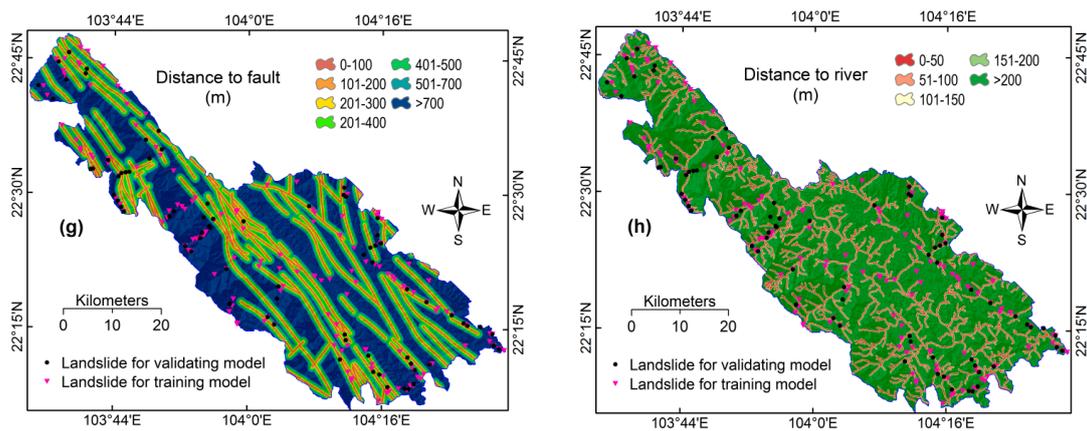

Figure 4. *Cont.*

Figure 4. Landslide predisposing factors used in this study: (**a**) Slope; (b) Aspect; (**c**) Elevation; (**d**) Land cover; (**e**) Soil type ; (**f**) Lithology; (**g**) Distance to fault; and, (**h**) Distance to river.

Soil type (e) legend: **D**: Sloping soil; **Fl**: Cultivated rice yellowish red soil; **Fs**: Yellowish red soil on claystone and metamorphic rocks; **Py**: Alluvial soil deposited by river; **Pe**: neutral-less acidic and light texture alluvial soil; **Fp**: Brown-yellowish soil on old alluvium; **Fq**: Light yellowish soil on sandstone; **Pbe**: Neutral and less acidic alluvial soil; **Flv**: Red soil on limestone; **Fn**: Brown-yellowish soil on limestone; **He**: Humus yellow red soil on claystone and metamorphic rocks; **Fa**: Yellowish red soil on acid magmatic rock; and **Ha**: Humus yellow red soil on acid igneous rock. **Lithology (f) legend**: **AciNeu-Mag**: Acid-neutral magmatic rocks; **Extrus-R**: Extrusive rocks; **Mafic-ultra**: Mafic-ultramafic rocks; **Meta-Alumi**: Metamorphic rock with aluminosilicate components; **Meta-Quart**: Metamorphic rock with rich quarts components; **Q-DP**: Quaternary deposits; and, **Sed-Cacb**: Sedimentary carbonate rocks.

4. Proposed a Hybrid Machine Learning Approach of Bagging Ensemble (BE) and Logistic Model Tree (LMTree)

In this section, the proposed hybrid machine learning approach for Landslide Susceptibility Modeling at Upper Reaches Area of Red River Basin (Viet Nam) is described and presented in the first time. Methodological concept of the proposed BE-LMT model used in this study is shown in Figure 5.
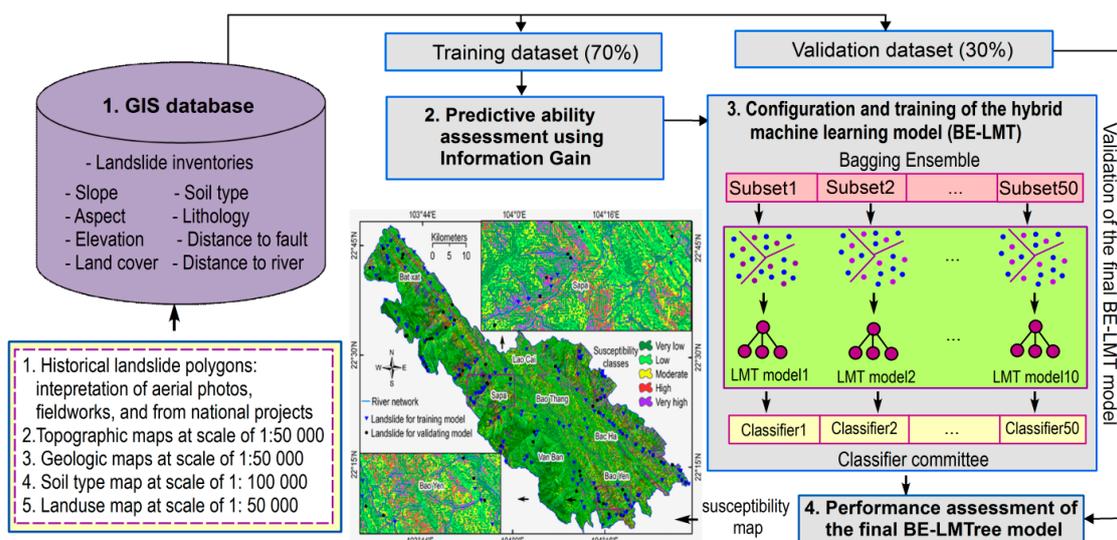

Figure 5. Methodological concept of the proposed Bagging ensemble (BE)-Logistic Model Trees (LMTree) model used in this study.

The proposed approach is a hybridization of LMTree and BE and is named as BE-LMTree. It should be noted that the data processing and coding were conducted using IDRISI Selva 17.0 (Clark University, Worcester, MA, USA, 2017) and ArcGIS 10.4 (ESRI Inc., Redlands, CA, USA 2017). The BE code is from Kuncheva [59] whereas the Logistic Model Tree algorithm is available at Weka's API [75]. The proposed BE-LMTree model was programmed by the authors in the Matlab environment.

4.1. Establishment of GIS Database, the Training Dataset and the Validation Dataset

In the first step, a GIS database for this project was designed and established using ArcCatalog software. Accordingly, the File Geodatabase format was used due to the ability to host and process very large geographic datasets with their different data types in a only one file system [76]. Accordingly, the GIS database consists of 255 landslide polygons and eight predisposing factors (slope, aspect, elevation, land cover, soil type, lithology, distance to fault, and distance to river). These landslide polygons and factors were converted to raster format with a resolution of 25 m. In this research, the categories of the eight predisposing factors were coded and normalized, as suggested in [24,77], to avoid the imbalance of categorical magnitudes [78].

In landslide modeling, cross validation [79] that has proven efficient for evaluating the model performance should be used. Accordingly, in this research, 179 landslide polygons (70%, 1006 pixels) were randomly extracted [80] and used for training the landslide models, whereas the other 76 landslides (30%, 441 pixels) were used for assessing the prediction capability of the models. Because the proposed approach in this study employs "on-off" classification, the equal amount of non-landslide pixels were also randomly sampled in the not-yet landslide areas of the basin, area with slope angles less than 5º, as suggested in [32]. Detailed discussions on sampling strategies can be found at [81]. In the next step, values of the eight predisposing factors for all of the aforementioned pixels were extracted to build the training dataset and the validation dataset. Finally, the coding process that was proposed in [17] was performed, in which the landslide pixels were assigned "1" and the non-landslide pixels were assigned "0".

Because the aforementioned partition of the landslide dataset into the training and validation datasets was randomly generated only once; therefore, a further cross validation was additionally used to ensure that the modeling result is the objective. Accordingly, 10-fold cross validation was employed in the training phase with the training dataset to build landslide models. Thus, the training dataset was randomly partitioned into 10 equally sized subsets; nine subsets were used for building the landslide model, whereas the remaining subset was used for testing the landslide model. This procedure was repeated 10 times where each subset was being used once as the testing dataset. Once the model was successfully trained using the training dataset with the 10-fold cross validation procedure, the model was again validated using the validation dataset.

4.2. Merit Evaluation of Factor

Identification of relevant features is an essential task when employing machine learning techniques for landslide susceptibility [82]. This is because landslide is a typical real-world problem that is influenced by various factors, but the contribution of these factors to the prediction model is different. If non-contribution factors are included in the model, then they may cause noises that reduce the prediction power of the final model; therefore, these factors should be excluded.

To detect non-relevant factors in this study, Pearson technique was employed to quantify the predictive power of all landslide predisposing factors. Accordingly, the meritof these features were estimated using Pearson correlation values [83] of the predisposing factors and the output using the following equation:

$$Merit_i = \frac{covr(IF_i, y)}{\sqrt{varr(IF_i) * varr(y)}} \tag{4}$$

where $Merit_i$ is the correlation value of landslide predisposing factor IFi and the label class y; $covr(.)$ is the covariance; and, $varr(.)$ is the variance.

4.3. Configuration and Training of the BE-LMTree Model

Configuration of the BE-LMTree model consists of two steps: (i) Determining the minimum number of samples (NS) that are used for growing the LMTree; and, (ii) Determining the number of bootstrap subsets (BS) used for BE. Because at least five samples are required to build a logistic regression model at a tree node [52], we varied NS from 5 to 100 with a step size of 1, and then, estimating the classification rate of the corresponding LMTree model on both the training dataset and the validation dataset. As a result, minimum of 10 samples is the best for the data at hand; therefore, NS of 10 was selected. For the case of determining the number of the bootstrap subsets, since no thumb rule is available, an empirical test was carried out by varying BS from 2 to 100, and then, compute their classification rates of the LMTree model both on the training dataset and the validation dataset. The test result revealed that the BE-LMTree with 50 tree-based classifiers provided the highest classification accuracy for the data at hand; therefore, BS of 50 is selected. Once the BE-LMTree model had been configured, the training process was carried out to derive the final BE-LMTree model.

4.4. Performance Assessment of the Final BE-LMTree Model

Because the landslide modeling in this research is considered to be a binary form of pattern recognition, therefore the performance of the final BE-LMTree model could be assessed using confusion matrix (Figure 6) [40], both on the training dataset and the validation dataset. Based on the matrix, several model measures are further derived i.e., sensitivity (SEN), specificity (SPE), positive predictive power (PP2), and negative predictive power (NP2), Kappa statistics, and classification accuracy (CLA) for the assessment, as suggested in [50]. It should be noted that a perfect landslide model will have 100% for SEN, SPE, PP2, NP2, and CLA.
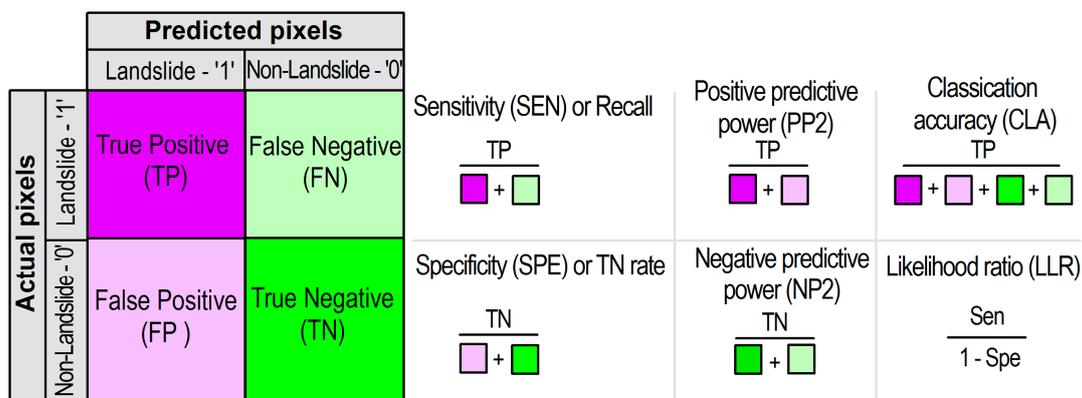

Figure 6. Confusion matrix and model measures used in this research.

For the case of CLA, although CLA provides the overall performance of the landslide model, however, a landslide model with a high CLA value may not classify the landslide pixels well. Therefore, the likelihood ratio (LLR) is additionally used [84]. LLR is a metric that assesses the trade-off of both SEN and SPE of landslide models. The higher the LLR value, the better the landslide model.

Global performance of the BE-LMTree model is summarized and assessed using the Receiver Operating Characteristic (ROC) Curve and Area Under the curve (AUC) [40,41,85]. In general, the closer the curve to the upper left corner, the better performance of the landslide model. Once the ROC curve is constructed, AUC for the model is computed and used to quantify the quality of the model. Accordingly, the performance of the model is excellent (AUC belong to 0.9–1), good (AUC belong to 0.8–0.9), fair (AUC belong to 0.7–0.8), and poor (AUC is less than 0.7) [86].

4.5. Computing Landslide Susceptibility Index

When the final BE-LMTree model is satisfied in the performance assessment check, the model is used to compute susceptibility index for all the pixels of the study area. These susceptibility indices are then converted to the ASCII raster format in ArcGIS using a Python application that was developed by the authors. Finally, the landslide susceptibility map is classified by five susceptibility classes: very high, high, moderate, low, and very low [87].

5. Results and Discussion

5.1. Predictive Ability Assessment

Result of the predictive ability evaluation of the eight predisposing factors is shown in Table 1. It is noted that the 10-fold cross validation was used to ensure the stable assessment result, as suggested in [88]. It could be seen that slope the highest predictive with the average merit (AM) is 0.225, followed by distance to river (AM of 0.171), lithology (AM of 0.148), aspect (AM of 0.129), and elevation (AM of 0.102). In contrast, soil type (AM of 0.038), distance to fault (AM of 0.055), and land cover (AM of 0.077) have low predictive ability values (Table 1).

The findings are reasonable because slope is widely recognized as the most important factor for landslide in various projects [89,90]. From the above results, it could be seen that all predisposing factors revealed predictive values to landslide model; therefore, we concluded that they are all relevant factors and are included in this analysis.

Table 1. Predictive ability of eight landslide predisposing factors using Pearson technique and 10-fold cross validation techniques.

No.	Predisposing Factors	Average Merit	Standard Deviation
1	Slope	0.225	0.008
2	Distance to river	0.171	0.008
3	Lithology	0.148	0.008
4	Aspect	0.129	0.008
5	Elevation	0.102	0.006
6	Land cover	0.077	0.008
7	Distance to fault	0.055	0.005
8	Soil type	0.038	0.005

5.2. Model Training and Evaluation

Using the eight predisposing factors, the BE-LMTree model was trained using the training dataset with the 10-fold cross validation technique. The training result is shown in Figure 7. It could be seen that the CLA of the BE-LMTree model is 93.81%, indicating a high degree of fit of the model with the dataset. Kappa statistics of 0.876 indicates the high agreement of the model and the training dataset. SEN of the BE-LMTree model is 93.02%, indicating that the proportion of the landslide pixels is correctly classified to the landslide class is 93.02%. Whereas, SPE is 94.63%, indicating that the proportion of the non-landslide pixels is correctly classified to the non-landslide class is 94.63%. PP2 is 94.72%, indicating that the probability that the BE-LMTree model correctly classifies pixels to the landslide class is 94.72%. NP2 is 92.89% indicating that the probability the BE-LMTree model correctly classifies pixels to the non-landslide class is 92.89%. Overall, these above measures have demonstrated that the BE-LMTree model performed very well with the training dataset.

To assess the contribution of landslide factors to the BE-LMT model, each factor was removed, and then, the classification accuracy (CLA) was estimated. The reduction of CLA of the BE-LMT model when one or more factors were removed indicates the contribution of these factors to the model. The result is shown in Table 2. It could be seen that when Distance to Fault and Soil type were removed from the LMT model, the CLA was reduced 2.12%. Therefore, although the average merit of Distance to

fault (0.055) and Soil type (0.038) are small (see Table 1), the two factors contributed to 2.12% increasing classification accuracy of the BE-LMT model. An even larger accuracy decrease (4.3%, see Table 2) occurred when the four most significant variables (Slope, Distance to river, Lithology, and Aspect) are used into the BE-LMT model. Overall, it is reasonable of the to keep all factors in this research.
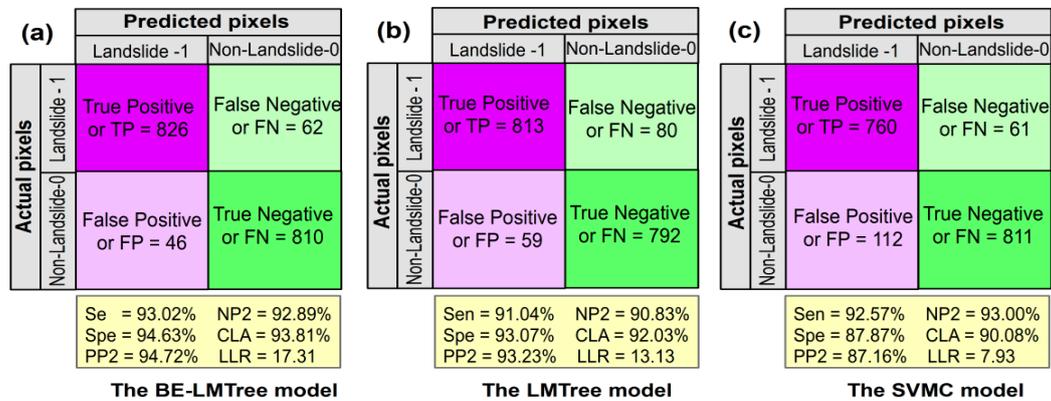

Figure 7. Confusion matrices and performance measures of the three landslide models using the training dataset: (**a**) the BE-LMTree model; (**b**) the LMTree model; and (**c**) the SVMC model.

Table 2. Contribution of the landslide predisposing factors to the BE-LMT model.

No.	Removing Factor	Classification Accuracy-CLA (%)
1	Slope	91.74
2	Aspect	92.31
3	Elevation	92.49
4	Land cover	93.60
5	Soil type	93.59
6	Lithology	91.97
7	Distance to fault	92.83
8	Distance to river	93.35
9	Distance to Fault and Soil type	91.69
10	Elevation, Land cover, Distance to fault and Soil type	89.51

The prediction performance of the BE-LMTree model is assessed using the validation dataset and the result is shown in Figure 8. It could be observed that the CLA is 87.89%, indicating a high prediction result. Kappa statistics of 0.759 indicates that the prediction performance of the model is 75.9% better than random. SEN of the BE-LMTree model is 92.25%, indicating that the proportion of the landslide pixels, which is accurately predicted, is 92.25%. SPE of the BE-LMTree model is 84.35%, indicating that the proportion of the non-landslide pixels is accurately predicted is 84.35%. PP2 of the model is 82.73%, indicating that the probability that the BE-LMTree model accurately predicts pixels to the landslide class is 82.73%. NP2 is 93.05%, indicating that the probability that the BE-LMTree model accurately predicts pixels to the non-landslide class is 93.05%.

Figure 9 shows 72 mispredicted landslide pixels (false positive) and 29 mispredicted non-landslide pixels (false negative) for the study area. We see that the 76.4% and 20.8% of the mispredicted landslide pixels were located in areas with slope angles <8.86° or slope angles from 36.39° to 5.87°, respectively. The mispredicted landslide pixels were also mainly located in elevation 174.78–358.94 m (76.4%), the lithology of sedimentary carbonate rocks (73.6%), the yellowish red soil on claystone and metamorphic rocks (87.5%), distance to fault >700 m (76.4%), and distance to river >200 m (79.2%). Distribution of the mispredicted landslide pixels in the classes in the other factors was more even. Regarding the mispredicted non-landslide pixels, they were mainly located in the distance to river >200 m (79.3%), the dense forest land (69.0%), and the yellowish red soil on claystone and metamorphic

rocks (62.1%). For the other factors, the distribution of the mispredicted non-landslide pixels in their classes was quite even.
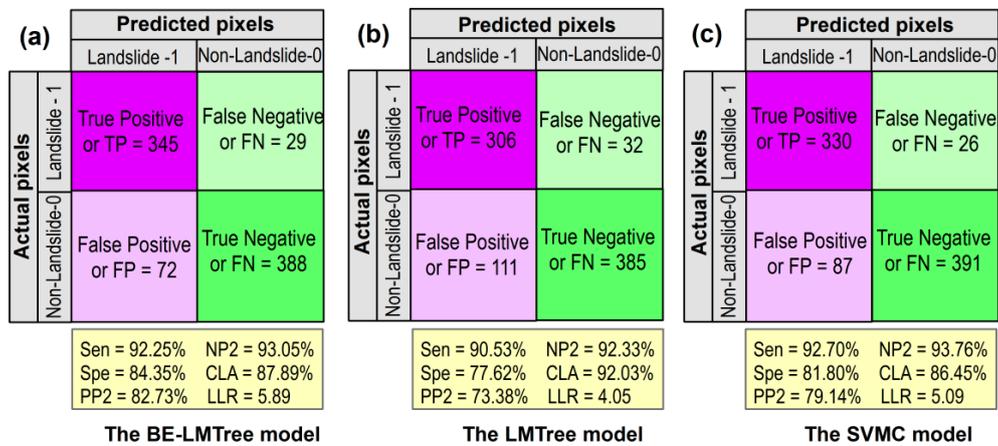

Figure 8. Confusion matrices and prediction measures of the three landslide models using the validation dataset: (**a**) the BE-LMTree model; (**b**) the LMTree model; and (**c**) the SVMC model.
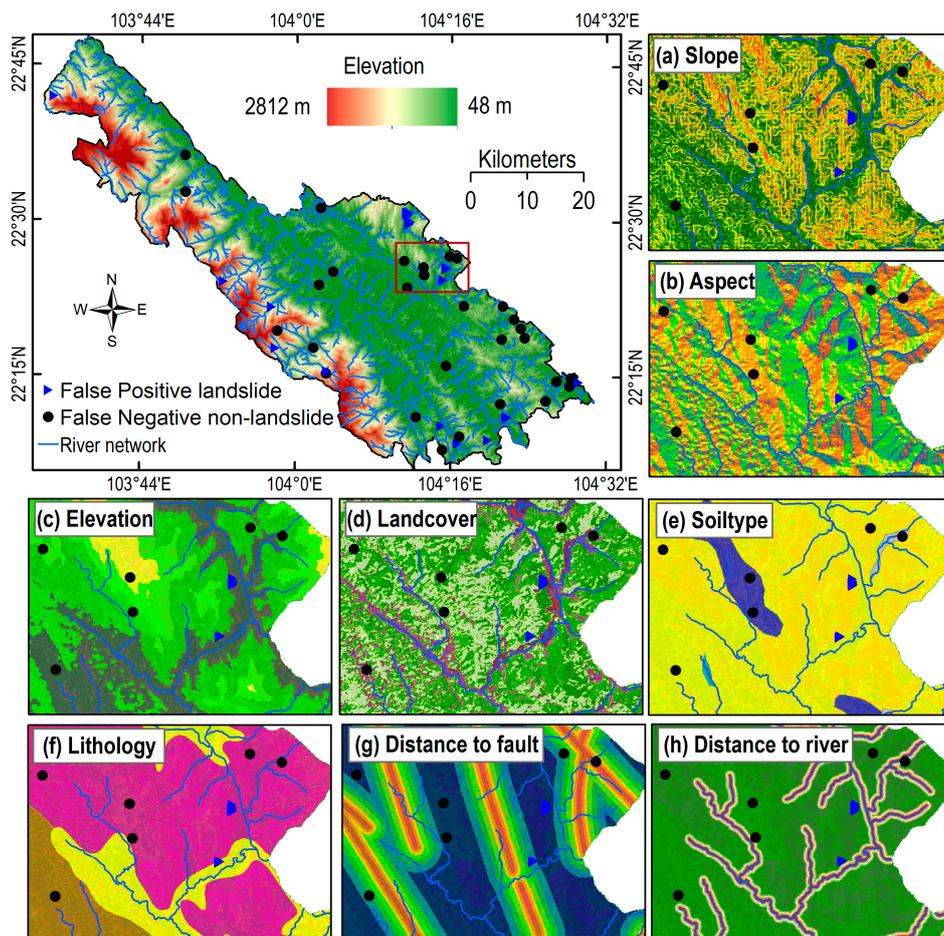

Figure 9. Mispredicted landslide pixels (false positive) and mispredicted non-landslide pixels in the validation dataset versus the eight landslide predisposing factors (legend for the eight factors was the same as in Figure 4). (**a**) Slope; (**b**) Aspect; (**c**) Elevation; (**d**) Landcover; (**e**) Soil type; (**f**) Lithology; (**g**) Distance to fault; and (**h**) Distance to river.

The global prediction capability of the BE-LMTree model is summarized and presented using the ROC curve and AUC (Figure 10). It can be seen that AUC is 0.834, indicating that the prediction capability of the proposed model is 83.4%, which is a high prediction capability.
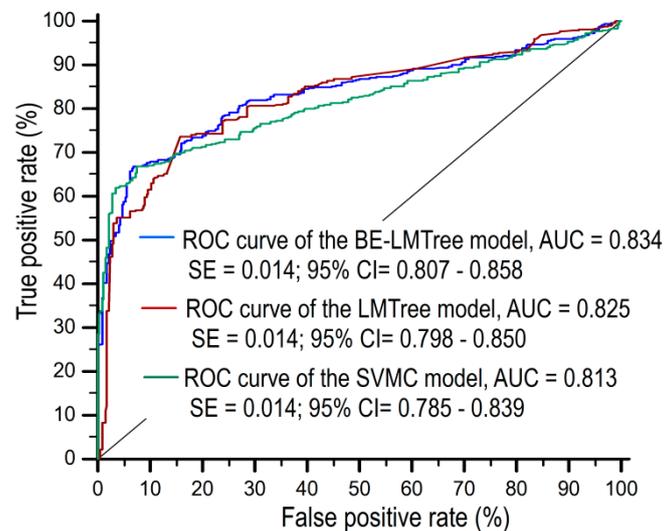

Figure 10. Receiver Operating Characteristic (ROC) curve and Area Under the curve (AUC) of the BE-LMTree model, the LMTree model, and the SVMC model using the validation dataset. SE: Standard Error; CI: Confidence Interval.

5.3. Comparison of the BE-LMTree Model with Benchmark

Because this is the first time that the BE-LMTree model is investigated for landslide modeling, the validity of the proposed model therefore was evaluated and compared with the benchmark. We select support vector machine (SVMC) as a benchmark because SVMC has proven efficient and outperforms other conventional methods [38,91]. For constructing the SVMC model, the radial basic function (RBF) kernel [41,92,93] was selected and the grid-search method [94–96] was used to derive the best the regularization (C = 9) and kernel width (γ = 0.245). In addition, the performance of the LMTree model was also included to present the merit of the proposed BE-LMTree model that is an integration of the Bagging ensemble and the LMTree.

The result is shown in Figures 7, 8, and 10. Using the training dataset, the CLA of the SVMC model (90.08%) and the LMTree model (92.03%) is slightly lower than CLA (93.81%) of the BE-LMTree model. Regarding LLR, the SVMC model (7.93) and the LMTree model (13.13) have lower values when compared to that of the BE-LMTree model (17.31). The other detailed metrics of the two models are shown in Figure 7. Overall, the BE-LMTree model performs better than the SVMC model and the LMTree model in the training dataset.

Using the validation dataset, the prediction performance of the SVMC model and the LMTree model is evaluated (Figure 8). It could be seen that the proposed BE-LMTree model (CLA = 87.98, LLR = 5.89) has a higher prediction performance when compared to those of the SVMC model (CLA = 86.45%, LLR = 5.09) and the LMTree model (CLA = 82.85%, LLR = 4.05). The global prediction capabilities of the three landslide models are assessed using the ROC curve and AUC (Figure 10). It could be been that the proposed BE-LMTree model (AUC = 0.834) is slightly higher than those of the SVMC model (AUC = 0.825) and the LMTree model (AUC = 0.813). Other detailed prediction performances of the three models are presented in Figure 8. Based on the aforementioned analysis, it could be concluded that the proposed BE-LMTree model is capable of producing the best landslide susceptibility result for this study area.

5.4. The Landslide Susceptibility Map

The final BE-LMTree model derived from the training step above was then used to compute landslide susceptibility indices for the Upper Reaches Area of Red River Basin (URRB), Vietnam. Accordingly, all of the predisposing factors in the raster maps were converted into ASCII format, and then fed to the BE-LMTree model to generate susceptibility indices. Distribution of these susceptibility indices is shown in Figure 11.
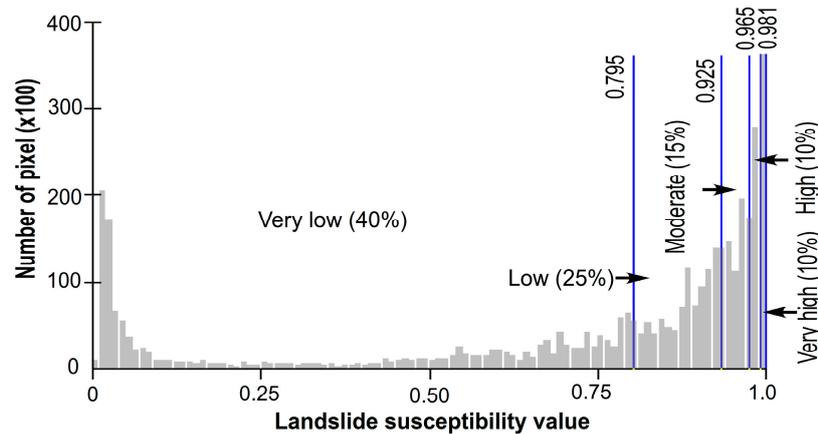

Figure 11. Distribution of these susceptibility indices versus of the five susceptibility classes.

These landslide susceptibility indices were then transformed to the raster format to manage in ArcGIS software using a python application that was programmed by the authors. Finally, the landslide susceptibility map (Figure 12) for the URRB was cartographically presented by five classes: very high (10%), high (10%), moderate (15%), low (25%), and very low (40%). To determine the thresholds for these classes, the extensively used graphic curve method has been considered to be the most suitable; a detailed explanation of it is available in [87,97,98]. The thresholds for these classes were determined based on an analysis of the susceptibility index map and the landslide inventory map, and then, the percentage of the landslide pixel versus the percentage of the susceptibility indices was calculated. At last, the four thresholds for the five classes were obtained.

Characteristics of the five landslide susceptibility classes that were derived from the BE-LMTree model the study area are shown in Table 3. Accordingly, the overall landslide frequency (OLF) proposed in [99] for the five classes was derived, and theoretically, the overall frequency should gradually grow from the very low class to the very high class [87]. It can be seen that the very high occupied only 10% of the study area, but it has the highest OLF value (4.40), followed by the high class (OLF = 1.59), the moderate class (OLF = 0.86), the low class (OLF = 0.43), and the very low class (OLF = 0.41). These confirm that the BE-LMTree model performed well with the URRB area.

Table 3. Characteristics of the landslide susceptibility classes derived from the BE-LMTree model the study area.

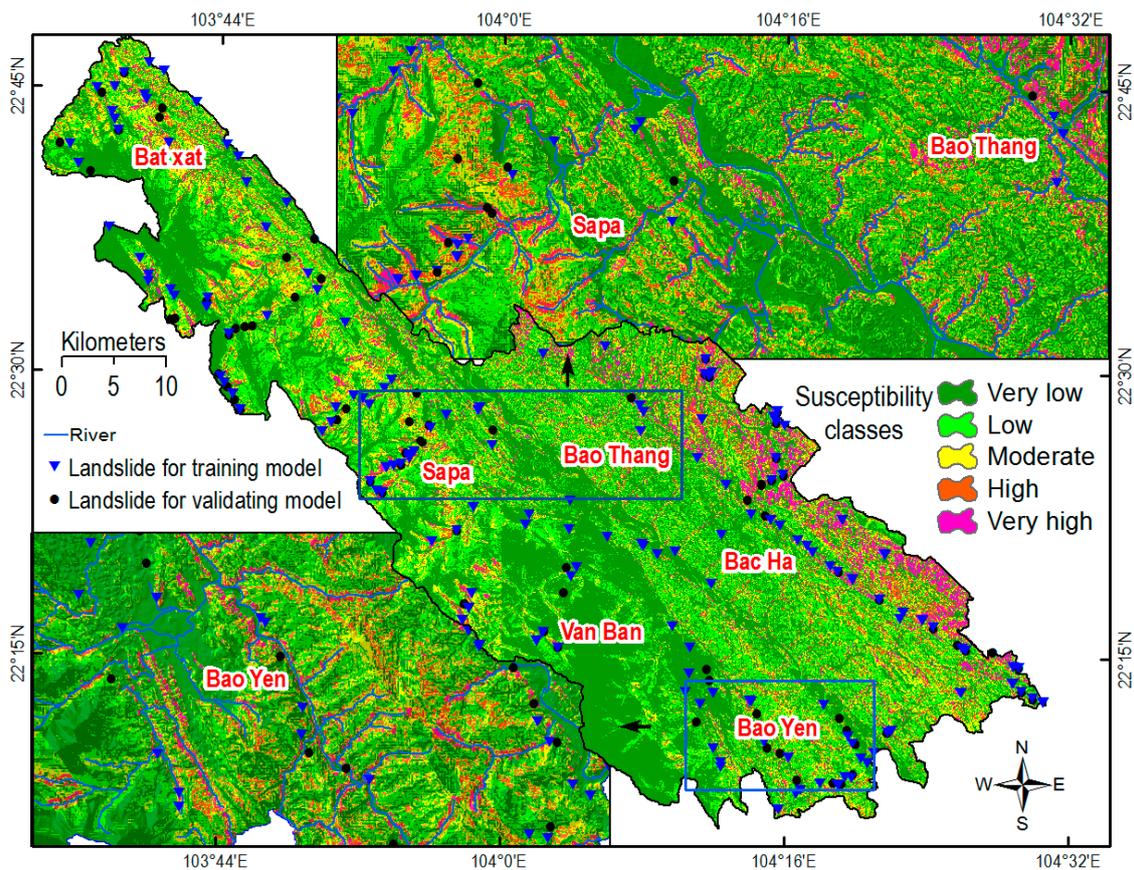| No.	Index Interval	Landslide Susceptibility (%)	Expression	Overall Landslide Frequency (OLF)	Areas (km²)
1	1.000–0.981	90–100	Very high	4.40	327.4
2	0.965–0.980	80–90	High	1.59	327.4
3	0.925–0.964	65–80	Moderate	0.86	491.0
4	0.795–0.924	40–65	Low	0.43	818.4
5	0.000–0. 794	0–50	Very low	0.41	1309.4

Figure 12. Landslide susceptibility map for the study area using the proposed BE-LMTree model.

Visual interpretation of the map (Figure 12) shows that the high probability of landslide is for areas i.e., Sapa, Bat Xat, and Bao Yen, therefore these areas should receive more attention in the development of remedial measures for the landslide prevention. Inversely, the low probability of landslide is for the Van Ban area. In fact, this area belongs to the Hoang Lien National Park, which is covered by the protected and dense tropical forest [100], therefore, having a low probability of landslide.

6. Concluding Remarks

This paper proposes a new modeling approach that is a hybrid intelligence of BE-LMTree for landslide susceptibility mapping with a case study at URRB. According to current literature, the BE-LMTree model has not been used for landslide modeling. For this purpose, the GIS database for the URRB area has been established, which contains a total of 255 historic soil-mixed-boulder slides and eight geo-environmental factors. These factors checked their merits to landslide using the Pearson correlation. The GIS database was then used to construct and verify the BE-LMTree model. Quality of the final BE-LMTree model was checked using confusion matrices and several model measures.

The results in this study point out that the new approach of the BE-LMTree could help to model landslide susceptibility with desirable prediction capability. When compared to the support vector machines (SVMC), a recognized benchmark in landslide modeling, the proposed BE-LMTree model presents a better performance. Therefore, the BE-LMTree is a new promising tool that could be used to enhance the quality of landslide susceptibility mapping.

For the case of the LMTree, this technique has been recently investigated for landslide susceptibility mapping with promising results i.e., in [50], the performance of the LMTree model in this research is lower than that of the SVMC model and the BE-LMTree model (Figures 4 and 5). Therefore, it could be concluded that the integration of the BE and the LMTree has significantly improved the

quality of the LMTree model. This is due to the stability and robustness of the BE procedures itself with the ability to reduce variances [101]. This finding agrees with [51], who concluded that the performance of the landslide model is enhanced with the use of ensemble frameworks.

The main disadvantage of the proposed approach is that the quality of the BE-LMTree model is heavily controlled by the minimum number of samples (NS) that is used for growing the LMTree and the number of bootstrap subsets (BS) used in the BE. In this research, NS and BS were determined using an empirical test. Although the NS and the BS found results in the high performance BE-LMTree model, however these do not warrant them being the optimal parameters. Therefore, the performance of the BE-LMTree model may be further enhanced if optimization algorithms are considered to integrate in the model. In addition, the BE-LMTree may create a complex forest trees i.e., 50 trees in this research. Therefore, the interpretation of the BE-LMTree model may be complicated. Despite the aforementioned limitations, the BE-LMTree can be considered as a new and valid tool for landslide susceptibility modeling.

Author Contributions: X.L.T., M.M., Y.K., V.R., G.Y., X.Q.T., and T.H.D collected data and processed input data. X.L.T., X.Q.T., D.T.B., and S.L. carried out the modeling process and wrote the paper.

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

References

1. Huggel, C.; Clague, J.J.; Korup, O. Is climate change responsible for changing landslide activity in high mountains? *Earth Surf. Process. Landf.* **2012**, *37*, 77–91. [CrossRef]
2. Uchida, T.; Sakurai, W.; Okamoto, A. Historical Patterns of Heavy Rainfall Event and Deep-Seated Rapid Landslide Occurrence in Japan: Insight for Effects of Climate Change on Landslide Occurrence. In *Advancing Culture of Living with Landslides, Proceedings of the World Landslide Forum WLF 2017, Ljubljana, Slovenia, 29 May–2 June 2017*; Springer: Cham, Switzerland, 2017; pp. 251–257.
3. Ciervo, F.; Rianna, G.; Mercogliano, P.; Papa, M. Effects of climate change on shallow landslides in a small coastal catchment in southern Italy. *Landslides* **2017**, *14*, 1043–1055. [CrossRef]
4. Sewell, R.; Parry, S.; Millis, S.; Wang, N.; Rieser, U.; DeWitt, R. Dating of debris flow fan complexes from Lantau Island, Hong Kong, China: The potential relationship between landslide activity and climate change. *Geomorphology* **2015**, *248*, 205–227. [CrossRef]
5. Gallina, V.; Torresan, S.; Critto, A.; Sperotto, A.; Glade, T.; Marcomini, A. A review of multi-risk methodologies for natural hazards: Consequences and challenges for a climate change impact assessment. *J. Environ. Manag.* **2016**, *168*, 123–132. [CrossRef] [PubMed]
6. Montz, B.E.; Tobin, G.A.; Hagelman, R.R., III. *Natural Hazards: Explanation and Integration*; Guilford Publications: New York, NY, USA, 2017.
7. Maes, J.; Kervyn, M.; de Hontheim, A.; Dewitte, O.; Jacobs, L.; Mertens, K.; Vanmaercke, M.; Vranken, L.; Poesen, J. Landslide risk reduction measures: A review of practices and challenges for the tropics. *Prog. Phys. Geogr.* **2017**, *41*, 191–221. [CrossRef]
8. Gian, Q.A.; Tran, D.-T.; Nguyen, D.C.; Nhu, V.H.; Tien Bui, D. Design and implementation of site-specific rainfall-induced landslide early warning and monitoring system: a case study at Nam Dan landslide (Vietnam). *Geomat. Nat. Hazards Risk* **2017**, *8*, 1978–1996. [CrossRef]
9. Hung, L.Q.; Van, N.T.H.; Son, P.V.; Ninh, N.H.; Tam, N.; Huyen, N.T. Landslide Inventory Mapping in the Fourteen Northern Provinces of Vietnam: Achievements and Difficulties. In *Advancing Culture of Living with Landslides: Volume 1 ISDR-ICL Sendai Partnerships 2015–2025*; Sassa, K., Mikoš, M., Yin, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 501–510.

10. Acosta, L.A.; Eugenio, E.A.; Macandog, P.B.M.; Magcale-Macandog, D.B.; Lin, E.K.-H.; Abucay, E.R.; Cura, A.L.; Primavera, M.G. Loss and damage from typhoon-induced floods and landslides in the Philippines: Community perceptions on climate impacts and adaptation options. *Int. J. Glob. Warm.* **2016**, *9*, 33–65. [CrossRef]

11. Shan, W.; Hu, Z.; Guo, Y.; Zhang, C.; Wang, C.; Jiang, H.; Liu, Y.; Xiao, J. The impact of climate change on landslides in southeastern of high-latitude permafrost regions of China. *Front. Earth Sci.* **2015**, *3*, 7. [CrossRef]

12. LeComte, D. International weather highlights 2014: Winter storms, typhoons, hurricanes, and flooding. *Weatherwise* **2015**, *68*, 20–26. [CrossRef]

13. Jiménez-Perálvarez, J.; El Hamdouni, R.; Palenzuela, J.; Irigaray, C.; Chacón, J. Landslide-hazard mapping through multi-technique activity assessment: An example from the Betic Cordillera (southern Spain). *Landslides* **2017**, *4*, 1975–1991. [CrossRef]

14. Pham, B.; Tien Bui, D.; Pourghasemi, H.; 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**, *128*, 255–273. [CrossRef]

15. Corominas, J.; van Westen, C.; Frattini, P.; Cascini, L.; Malet, J.P.; Fotopoulou, S.; Catani, F.; Van Den Eeckhaut, M.; Mavrouli, O.; Agliardi, F.; et al. . Recommendations for the quantitative analysis of landslide risk. *Bull. Eng. Geol. Environ.* **2014**, *73*, 209–263. [CrossRef]

16. Ciampalini, A.; Raspini, F.; Lagomarsino, D.; Catani, F.; Casagli, N. Landslide susceptibility map refinement using PSInSAR data. *Remote Sens. Environ.* **2016**, *184*, 302–315. [CrossRef]

17. Tien Bui, D.; Nguyen, Q.-P.; Hoang, N.-D.; Klempe, H. A Novel Fuzzy K-Nearest Neighbor Inference model with Differential Evolution for Spatial Prediction of Rainfall-Induced Shallow Landslides in a Tropical Hilly Area using GIS. *Landslides* **2017**, *14*, 1–17. [CrossRef]

18. Chacon, J.; Irigaray, C.; Fernandez, T.; El Hamdouni, R. Engineering geology maps: Landslides and geographical information systems. *Bull. Eng. Geol. Environ.* **2006**, *65*, 341–411. [CrossRef]

19. Van Westen, C.J.; Van Asch, T.W.J.; Soeters, R. Landslide hazard and risk zonation—Why is it still so difficult? *Bull. Eng. Geol. Environ.* **2006**, *65*, 167–184. [CrossRef]

20. Akgun, A.; Sezer, E.A.; Nefeslioglu, H.A.; Gokceoglu, C.; Pradhan, B. An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Comput. Geosci.* **2012**, *38*, 23–34. [CrossRef]

21. Meng, Q.; Miao, F.; Zhen, J.; Wang, X.; Wang, A.; Peng, Y.; Fan, Q. GIS-based landslide susceptibility mapping with logistic regression, analytical hierarchy process, and combined fuzzy and support vector machine methods: A case study from Wolong Giant Panda Natural Reserve, China. *Bull. Eng. Geol. Environ.* **2016**, *75*, 923–944. [CrossRef]

22. Gheshlaghi, H.A.; Feizizadeh, B. An integrated approach of analytical network process and fuzzy based spatial decision making systems applied to landslide risk mapping. *J. Afr. Earth Sci.* **2017**, *133*, 15–24. [CrossRef]

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

24. Tien Bui, D.; 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**, *171–172*, 12–29. [CrossRef]

25. 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]

26. Pham, B.T.; Tien Bui, D.; Prakash, I.; Dholakia, M.B. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* **2017**, *149 Pt 1*, 52–63. [CrossRef]

27. Gorsevski, P.V.; Brown, M.K.; Panter, K.; Onasch, C.M.; Simic, A.; Snyder, J. Landslide detection and susceptibility mapping using LiDAR and an artificial neural network approach: A case study in the Cuyahoga Valley National Park, Ohio. *Landslides* **2016**, *13*, 467–484. [CrossRef]

28. Oh, H.-J.; Lee, S. Shallow Landslide Susceptibility Modeling Using the Data Mining Models Artificial Neural Network and Boosted Tree. *Appl. Sci.* **2017**, *7*, 1000. [CrossRef]

29. 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]

30. Pascale, S.; Parisi, S.; Mancini, A.; Schiattarella, M.; Conforti, M.; Sole, A.; Murgante, B.; Sdao, F. Landslide susceptibility mapping using artificial neural network in the Urban area of Senise and San Costantino Albanese (Basilicata, Southern Italy). In *International Conference on Computational Science and Its Applications*; Springer: Berlin, Germany, 2013; pp. 473–488.

31. Yao, X.; Tham, L.G.; Dai, F.C. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [CrossRef]

32. Kavzoglu, T.; Sahin, E.; Colkesen, I. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* **2014**, *11*, 425–439. [CrossRef]

33. Kumar, D.; Thakur, M.; Dubey, C.S.; Shukla, D.P. Landslide susceptibility mapping & prediction using Support Vector Machine for Mandakini River Basin, Garhwal Himalaya, India. *Geomorphology* **2017**, *295*, 115–125.

34. Colkesen, I.; Sahin, E.K.; Kavzoglu, T. Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *J. Afr. Earth Sci.* **2016**, *118*, 53–64. [CrossRef]

35. 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]

36. Hong, H.; Pradhan, B.; Bui, D.T.; Xu, C.; Youssef, A.M.; Chen, W. Comparison of four kernel functions used in support vector machines for landslide susceptibility mapping: A case study at Suichuan area (China). *Geomat. Natl. Hazards Risk* **2016**, *8*, 544–569. [CrossRef]

37. 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**. [CrossRef]

38. Pham, B.T.; Tien Bui, D.; Prakash, I. Bagging based Support Vector Machines for spatial prediction of landslides. *Environ. Earth Sci.* **2018**, *77*, 146. [CrossRef]

39. Tien Bui, D.; Pham, T.B.; 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. Dig. Earth* **2016**, *9*, 1077–1097. [CrossRef]

40. Tien Bui, D.; Anh Tuan, T.; Hoang, N.-D.; Quoc Thanh, N.; Nguyen, B.D.; Van Liem, N.; Pradhan, B. Spatial Prediction of Rainfall-induced Landslides for the Lao Cai area (Vietnam) Using a Novel hybrid Intelligent Approach of Least Squares Support Vector Machines Inference Model and Artificial Bee Colony Optimization. *Landslides* **2017**, *14*, 447–458. [CrossRef]

41. Hoang, N.-D.; Tien Bui, D. A Novel Relevance Vector Machine Classifier with Cuckoo Search Optimization for Spatial Prediction of Landslides. *J. Comput. Civ. Eng.* **2016**, *30*, 1–10. [CrossRef]

42. Althuwaynee, O.F.; Pradhan, B.; Lee, S. A novel integrated model for assessing landslide susceptibility mapping using CHAID and AHP pair-wise comparison. *Int. J. Remote Sens.* **2016**, *37*, 1190–1209. [CrossRef]

43. Youssef, A.M.; Pourghasemi, H.R.; Pourtaghi, Z.S.; Al-Katheeri, M.M. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* **2015**, *13*, 839–856. [CrossRef]

44. Lagomarsino, D.; Tofani, V.; Segoni, S.; Catani, F.; Casagli, N. A Tool for Classification and Regression Using Random Forest Methodology: Applications to Landslide Susceptibility Mapping and Soil Thickness Modeling. *Environ. Model. Assess.* **2017**, *22*, 201–214. [CrossRef]

45. Tsangaratos, P.; Ilia, I. Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece. *Landslides* **2015**, *13*, 305–320. [CrossRef]

46. Kim, J.-C.; Lee, S.; Jung, H.-S.; Lee, S. Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea. *Geocarto Int.* **2017**, *33*, 1000–1015. [CrossRef]

47. Hong, H.; Liu, J.; Bui, D.T.; Pradhan, B.; Acharya, T.D.; Pham, B.T.; Zhu, A.X.; Chen, W.; Ahmad, B.B. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *CATENA* **2018**, *163*, 399–413. [CrossRef]

48. Hoang, N.-D.; Tien Bui, D. Spatial prediction of rainfall-induced shallow landslides using gene expression programming integrated with GIS: A case study in Vietnam. *Natl. Hazards* **2018**, *92*, 1871–1887. [CrossRef]

49. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* **2013**, *51*, 350–365. [CrossRef]

50. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378. [CrossRef]

51. Tien Bui, D.; Ho, T.-C.; Pradhan, B.; Pham, B.-T.; Nhu, V.-H.; Revhaug, I. GIS-Based Modeling of Rainfall-Induced Landslides Using Data Mining Based Functional Trees Classifier with AdaBoost, Bagging, and MultiBoost Ensemble Frameworks. *Environ. Earth Sci.* **2016**, *75*, 1101–1123. [CrossRef]

52. Landwehr, N.; Hall, M.; Frank, E. Logistic Model Trees. *Mach. Learn.* **2005**, *59*, 161–205. [CrossRef]

53. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]

54. Tien Bui, D.; Ho, T.C.; Revhaug, I.; Pradhan, B.; Nguyen, D. Landslide Susceptibility Mapping Along the National Road 32 of Vietnam Using GIS-Based J48 Decision Tree Classifier and Its Ensembles. In *Cartography from Pole to Pole*; Buchroithner, M., Prechtel, N., Burghardt, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 303–317.

55. Pham, T.D.; Bui, D.T.; Yoshino, K.; Le, N.N. Optimized rule-based logistic model tree algorithm for mapping mangrove species using ALOS PALSAR imagery and GIS in the tropical region. *Environ. Earth Sci.* **2018**, *77*, 159. [CrossRef]

56. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Chapman and Hall/CRC: New York, NY, USA, 1984.

57. Doetsch, P.; Buck, C.; Golik, P.; Hoppe, N.; Kramp, M.; Laudenberg, J.; Oberdörfer, C.; Steingrube, P.; Forster, J.; Mauser, A. Logistic Model Trees with AUC Split Criterion for the KDD Cup 2009 Small Challenge. In Proceedings of the 2009 International Conference on KDD-Cup 2009, Paris, France, 28 June–1 July 2009; pp. 77–88.

58. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: San Mateo, CA, USA, 1993.

59. Kuncheva, L.I. *Combining Pattern Classifiers: Methods and Algorithms*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2014.

60. Kotsiantis, S. Combining bagging, boosting, rotation forest and random subspace methods. *Artif. Intell. Rev.* **2011**, *35*, 223–240. [CrossRef]

61. Lu, X.X.; Oeurng, C.; Le, T.P.Q.; Thuy, D.T. Sediment budget as affected by construction of a sequence of dams in the lower Red River, Viet Nam. *Geomorphology* **2015**, *248*, 125–133. [CrossRef]

62. Do, T.; Nguyen, C.; Phung, T. *Assessment of Natural Disasters in Vietnam's Northern Mountains*; Munich University Library: Munich, Germany, 2013; p. 57.

63. Tran, T. Climate change adaptation from small and medium scale hydropower plants: A case study for Lao Cai province. *VNU J. Sci. Earth Environ. Sci.* **2011**, *27*, 32–38.

64. Jolivet, L.; Beyssac, O.; Goffe, B.; Avigad, D.; Lepvrier, C.; Maluski, H.; Thang, T.T. Oligo-Miocene midcrustal subhorizontal shear zone in Indochina. *Tectonics* **2001**, *20*, 46–57. [CrossRef]

65. Duan, B.V. The relation between fault movement potential and seismic activity of major faults in Northwestern Vietnam. *Vietnam J. Earth Sci.* **2017**, *39*, 240–255. [CrossRef]

66. Hue, T.T.; Duong, T.V.; Toan, D.V.; Nghinh, L.T.; Minh, V.C.; Pho, N.V.; Xuan, P.T.; Hoan, L.T.; Huyen, N.X.; Pha, P.D.; et al. *Investigation and Assessment of the Types of Geological Hazard in the Territory of Vietnam and Recommendation of Remedial Measures. Phase II: A Study of the Northern Mountainous Province of Vietnam*; Institute of Geological Sciences, Vietnam Academy of Science and Technology: Hanoi, Vietnam, 2004; p. 361.

67. Yem, N.T.; Thanh, N.Q.; Anh, P.L.; Chi, C.T.; Du, C.D.; Dung, N.P.; Dung, P.D.; Hai, N.P.; Hien, T.T.; Hoang, N.V.; et al. *Assessment of Landslides and Debris Flows at Some Prone Mountainous Areas Vietnam and Recommendation of Remedial Measures. Phase I: A Study of the East Side of the Hoang Lien Son Mountainous Area of Vietnam*; Institute of Geological Sciences, Vietnam Academy of Science and Technology: Hanoi, Vietnam, 2006; p. 361.

68. Van, T.T.; Tuy, P.K.; Giap, N.X.; Ke, T.D.; Thai, T.N.; Giang, N.T.; Tho, H.M.; Tuat, L.T.; San, D.N.; Hung, L.Q.; et al. *Assessment and Prediction of Geological Hazards in the 8 Coastal Provinces of Central Vietnam from Quang Binh to Phu Yen—Current Status, Causes, Prediction and Recommendation of Remedial Measures*; Vietnam Institude of Geosciences and Mineral Resoures: Hanoi, Vietnam, 2002; p. 215.

69. 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.

70. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. *Comput. Geosci.* **2012**, *45*, 199–211. [CrossRef]

71. Cevik, E.; Topal, T. GIS-based landslide susceptibility mapping for a problematic segment of the natural gas pipeline, Hendek (Turkey). *Environ. Geol.* **2003**, *44*, 949–962. [CrossRef]

72. Conforti, M.; Pascale, S.; Pepe, M.; Sdao, F.; Sole, A. Denudation processes and landforms map of the Camastra River catchment (Basilicata–South Italy). *J. Maps* **2013**, *9*, 444–455. [CrossRef]

73. Yilmaz, I. A case study from Koyulhisar (Sivas-Turkey) for landslide susceptibility mapping by artificial neural networks. *Bull. Eng. Geol. Environ.* **2009**, *68*, 297–306. [CrossRef]

74. Pachauri, A.; Pant, M. Landslide hazard mapping based on geological attributes. *Eng. Geol.* **1992**, *32*, 81–100. [CrossRef]

75. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: San Mateo, CA, USA, 2016.

76. Zeiller, M. *Modeling Our World: The ESRI Guide to Geodatabase Concepts*; ESRI Press: Redlands, CA, USA, 2010.

77. Tien Bui, D.; Hoang, N.-D. A Bayesian framework based on a Gaussian mixture model and radial-basis-function Fisher discriminant analysis (BayGmmKda V1. 1) for spatial prediction of floods. *Geosci. Model Dev.* **2017**, *10*, 3391. [CrossRef]

78. Dang, V.-H.; Dieu, T.B.; Tran, X.-L.; Hoang, N.-D. Enhancing the accuracy of rainfall-induced landslide prediction along mountain roads with a GIS-based random forest classifier. *Bull. Eng. Geol. Environ.* **2018**. [CrossRef]

79. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [CrossRef]

80. 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]

81. Erener, A.; Sivas, A.A.; Selcuk-Kestel, A.S.; Düzgün, H.S. Analysis of training sample selection strategies for regression-based quantitative landslide susceptibility mapping methods. *Comput. Geosci.* **2017**, *104*, 62–74. [CrossRef]

82. Nguyen, Q.-K.; Tien Bui, D.; Hoang, N.-D.; Trinh, P.T.; Nguyen, V.-H.; Yilmaz, I. A Novel Hybrid Approach Based on Instance Based Learning Classifier and Rotation Forest Ensemble for Spatial Prediction of Rainfall-Induced Shallow Landslides using GIS. *Sustainability* **2017**, *9*, 813. [CrossRef]

83. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.

84. Lagomarsino, D.; Segoni, S.; Rosi, A.; Rossi, G.; Battistini, A.; Catani, F.; Casagli, N. Quantitative comparison between two different methodologies to define rainfall thresholds for landslide forecasting. *Natl. Hazards Earth Syst. Sci.* **2015**, *15*, 2413–2423. [CrossRef]

85. Lucà, F.; Conforti, M.; Robustelli, G. Comparison of GIS-based gullying susceptibility mapping using bivariate and multivariate statistics: Northern Calabria, South Italy. *Geomorphology* **2011**, *134*, 297–308. [CrossRef]

86. Cantor, S.B.; Kattan, M.W. Determining the area under the ROC curve for a binary diagnostic test. *Med. Decis. Mak.* **2000**, *20*, 468–470. [CrossRef] [PubMed]

87. 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]

88. Fushiki, T. Estimation of prediction error by using K-fold cross-validation. *Stat. Comput.* **2011**, *21*, 137–146. [CrossRef]

89. Van Den Eeckhaut, M.; Vanwalleghem, T.; Poesen, J.; Govers, G.; Verstraeten, G.; Vandekerckhove, L. Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes (Belgium). *Geomorphology* **2006**, *76*, 392–410. [CrossRef]

90. Costanzo, D.; Rotigliano, E.; Irigaray, C.; Jiménez-Perálvarez, J.D.; Chacón, J. Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: Application to the river Beiro basin (Spain). *Natl. Hazards Earth Syst. Sci.* **2012**, *12*, 327–340. [CrossRef]

91. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O. Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. *Natl. Hazards* **2013**, *66*, 707–730. [CrossRef]

92. Hoang, N.-D.; Tien Bui, D. Predicting earthquake-induced soil liquefaction based on a hybridization of kernel Fisher discriminant analysis and a least squares support vector machine: A multi-dataset study. *Bull. Eng. Geol. Environ.* **2018**, *77*, 191–204. [CrossRef]

93. Hoang, N.-D.; Tien Bui, D.; Liao, K.-W. Groutability estimation of grouting processes with cement grouts using Differential Flower Pollination Optimized Support Vector Machine. *Appl. Soft Comput.* **2016**, *45*, 173–186. [CrossRef]

94. Ngoc-Thach, N.; Ngo, D.B.-T.; Xuan-Canh, P.; Hong-Thi, N.; Thi, B.H.; NhatDuc, H.; Dieu, T.B. Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study. *Ecol. Inform.* **2018**, *46*, 74–85. [CrossRef]

95. Vafaei, S.; Soosani, J.; Adeli, K.; Fadaei, H.; Naghavi, H.; Pham, T.D.; Tien Bui, D. Improving Accuracy Estimation of Forest Aboveground Biomass Based on Incorporation of ALOS-2 PALSAR-2 and Sentinel-2A Imagery and Machine Learning: A Case Study of the Hyrcanian Forest Area (Iran). *Remote Sens.* **2018**, *10*, 172. [CrossRef]

96. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Application of support vector machines in landslide susceptibility assessment for the Hoa Binh province (Vietnam) with kernel functions analysis. In *iEMSs 2012—Managing Resources of a Limited Planet, Proceedings of the 6th Biennial Meeting of the International Environmental Modelling and Software Society, Leipzig, Germany, 1 July 2012*; Brigham Young University: Provo, UT, USA, 2012; pp. 382–389.

97. Chung, C.-J.; Fabbri, A.G. Predicting landslides for risk analysis—Spatial models tested by a cross-validation technique. *Geomorphology* **2008**, *94*, 438–452. [CrossRef]

98. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I.; Dick, O.B. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *Catena* **2012**, *96*, 28–40. [CrossRef]

99. Sarkar, S.; Kanungo, D.P. An integrated approach for landslide susceptibility mapping using remote sensing and GIS. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 617–625. [CrossRef]

100. Kieu, Q.L.; Nguyen, T.T. Study on the distribution characteristics of the vegetation in high levations in Hoang Lien National park of Vietnam. *J. Vietnam. Environ.* **2015**, *6*, 84–88.

101. Mert, A.; Kılıç, N.; Akan, A. Evaluation of bagging ensemble method with time-domain feature extraction for diagnosing of arrhythmia beats. *Neural Comput. Appl.* **2014**, *24*, 317–326. [CrossRef]