Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 4310791, 15 pages
https://doi.org/10.1155/2020/4310791

Research Article

Improving Voting Feature Intervals for Spatial Prediction of Landslides

Binh Thai Pham,[1] **Tran Van Phong**,[2] **Mohammadtaghi Avand**,[3] **Nadhir Al-Ansari**,[4] **Sushant K. Singh**,[5] **Hiep Van Le**,[6] **and Indra Prakash**[7]

[1]*University of Transport Technology, Hanoi 100000, Vietnam*
[2]*Institute of Geological Sciences, Vietnam Academy of Sciences and Technology, 84 Chua Lang Street, Dong da, Hanoi 100000, Vietnam*
[3]*Department of Watershed Management Engineering, College of Natural Resources, TarbiatModares University, Tehran 14115-111, Iran*
[4]*Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, Lulea 971 87, Sweden*
[5]*Artificial Intelligence and Analytics, Health Care and Life Sciences, Virtusa Corporation, New York, NY, USA*
[6]*Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam*
[7]*DDG(R) Geological Survey of India, Gandhinagar 382010, India*

Correspondence should be addressed to Binh Thai Pham; binhpt@utt.edu.vn, Nadhir Al-Ansari; nadhir.alansari@ltu.se, and Hiep Van Le; levanhiep2@duytan.edu.vn

Received 4 June 2020; Revised 25 August 2020; Accepted 25 September 2020; Published 12 October 2020

Academic Editor: Zheng-zheng Wang

In this study, the main aim is to improve performance of the voting feature intervals (VFIs), which is one of the most effective machine learning models, using two robust ensemble techniques, namely, AdaBoost and MultiBoost for landslide susceptibility assessment and prediction. For this, two hybrid models, namely, AdaBoost-based Voting Feature Intervals (ABVFIs) and MultiBoost-based Voting Feature Intervals (MBVFIs) were developed and validated using landslide data collected from one of the landslide affected districts of Vietnam, namely, Muong Lay. Quantitative validation methods including area under the ROC curve (AUC) were used to evaluate model performance. The results indicated that both the newly developed ensemble models ABVFI (AUC = 0.859) and MBVFI (AUC = 0.839) outperformed the single VFI (AUC = 0.824) model. Thus, ensemble framework-based VFI algorithms can be used for the accurate spatial prediction of landslides, which can also be applied in other landslide prone regions of the world. Landslide susceptibility maps developed by ensemble VFI models can be used for better landslide prevention and risk management of the area.

1. Introduction

In recent years, population growth and development in unstable hilly areas have led to an increase in natural disasters such as landslides [1]. Based on 100 years of data analysis of natural hazards, after floods and earthquakes, landslides are the most frequent and important natural disaster causing casualties, financial losses, and adverse environmental impacts all over the world [2]. Landslides can cause destruction of infrastructure facilities, land use changes, erosion, and a high volume of sediment production in watersheds [3–5]. Mostly landslides occur due to gravity action on groundmass as a result of rainfall, earthquakes, soil saturation, and excavation of slopes [6, 7]. Landslide influencing factors include topography (slope, aspect, curvature, altitude, and elevation), geology (lithology, fault, and weathering crust), hydrology (rainfall and drainage), and land use [8–11]. Understanding the features and elements of landslide development and expansion helps in risk prediction and prevention of landslide damages [12].

In landslide study, it is important to identify and demarcate landslide susceptible zones [13, 14]. Landslide zoning mapping requires an assessment of the relationship between the prevailing conditions of the basin situation and the factors affecting the occurrence of the landslides [15]. In general, there are several methods of landslide susceptibility mapping and zoning [16, 17] which include mathematical/statistical and machine learning techniques [18, 19]. Mathematical modeling approach for delineating landslide hazards in watersheds was discussed in detail by Simons and Ward (http://andrewsforest.oregonstate.edu/pubs/pdf/pub2055.pdf) and Simons and Ward [20]. Corominas et al. [21] reviewed the literature and recommended methodologies for the quantitative analysis of landslide hazard, vulnerability, and risk at different spatial scales. They have also used this method for the verification and validation of the results [21].

However, it is difficult to demarcate natural boundary of transitional/gradational geological units and also continuous topographic features and factors such as elevation, slope, and topographic indices by traditional and statistical models [22, 23]. Simplification of major landslide parameters, their classes, and interactions between them can lead to incorrect results in the final map [24, 25]. These concerned led to the use of machine learning (ML) and data mining techniques in landslide studies [1, 26]. Nowadays, these methods are being used more widely for landslide susceptibility mapping due to their accuracy and speed [13, 14, 27]. Some of the prominent models used for mapping include artificial neural network (ANN) [28, 29], boosted regression trees (BRTs) [30, 31], random forest (RF) [32, 33], rotation forest (ROF) [34, 35], particle swarm optimization (PSO) [36, 37], support vector machine (SVM) [28, 38], binary logistic regression (BLR) [22], bagging [39, 40], logistic regression (LR) [33, 41], and canonical correlation forest (CCF) [42]. ML models have proven their relative superiority over bivariate and multivariate statistical models in several studies [43, 44]. In addition, to increase accuracy in dealing with complex problems and uncertainties, these models also lead to the development of new approaches to various problems [45, 46]. Although a number of ML models have been used in the landslide study, no model is perfect to be applied in all geoenvironmental conditions. Therefore, there is always scope of improvement in methodology by using different combinations of algorithms.

With this objective, a new ensemble framework-based ML models, namely, ABVFI and MBVFI, which are combination of a popular single ML model voting feature intervals (VFIs), and two effective ensemble techniques, namely, AdaBoost and MultiBoost algorithms, were proposed for the development of landslide susceptibility maps. Muong Lay district, which is one of the most landslide affected areas of Vietnam, was selected as the study area. The main contribution of this study is in the development and application of a novel hybrid approach for accurate landslide susceptibility mapping. Validation of these models was carried out using different quantitative statistical indices including area under the ROC curve and accuracy. Weka and ArcGIS software were implemented for processing the data, modeling, and mapping of landslide susceptibility.

2. Methods Used

2.1. Voting Feature Intervals. Voting feature intervals (VFIs) is one of the classification methods that is based on feature separation and works on nonincremental classification [47]. In the VFI method, the features are considered independently [48]. This method has been used successfully in various medical, computer, and natural sciences studies [49, 50]. The primary purpose of this approach is to deal with very imbalanced datasets [51].

VFI methodology involves two main steps: [1] training and [2] classification. First, in the training phase, the feature intervals are constructed around each class by calculating the lowest and highest values of each feature. In the classification stage, a feature vote is computed for each category based on each interval from each element, and then, the votes for each feature interval are united to produce one output [47]. One of the most important advantages of this algorithm is that it ignores the missing feature values at both the training and classification stages [47].

2.2. AdaBoost. AdaBoost or Adaptive Boosting is a ML algorithm devised by Yaw Freund and Robert Schapire [52]. AdaBoost is a hybrid learning technique and most well-known method of the algorithm's family. In this algorithm, models learned sequentially so that a model is trained at any one time. At the end of each time, incorrectly classified examples are identified, and their emphasis is on a new training set which can be used for the next training session for training a new model [53]. The idea is that new models should be able to compensate for errors created by previous models. In fact, AdaBoost is a meta-algorithm used to enhance performance along with other learning algorithms. Purpose of the AdaBoost algorithm is to increase learning rate of the classifiers. This algorithm combines several weak clusters to obtain a suitable boundary between two classes of data. The AdaBoost algorithm is sensitive to noise and outliers, but it is better suited to the overfitting problem in comparison to other learning algorithms [52].

If the base classifier used is better than the random classifier (50%), the algorithm's performance improves with more iteration. Even classifiers with higher error than random classifiers enhance overall performance by taking the negative coefficient [54]. In the AdaBoost algorithm, a weak classification is added at each round. At each call, weights are assigned based on the importance of the samples. With each round, the weight of misclassified samples increases, and the weight of correctly classified samples decreases, so the new classifier will focus on the more difficult-to-learn samples [55].

2.3. MultiBoost. MultiBoost is one of the ensemble learning methods developed by combining two ensemble learning algorithms, namely, AdaBoost and Wagging [56–58]. Wagging uses training samples with deferring weight, which could significantly reduce the high bias of the AdaBoost algorithm [59]. Combination of the two AdaBoost and Wagging techniques improves weak

classifications learning and transforms them into a robust classifier [56]. In case of MultiBoost technique, training of data is done in three main stages: (i) randomly, a subset is separated from the training data and used for models based on initial classification; (ii) sample weight is adjusted according to the predictive ability of the model; and (iii) the new subset is selected according to the weighted sample and is used to train the new model [60].

2.4. Validation Methods. Performance of the models was evaluated using statistical measures such as positive predictive value (PPV), area under receiver operating characteristic (ROC) curve (AUC), specificity (SPF), accuracy (ACC), negative predictive value (NPV), sensitivity (SST), root mean square error (RSME), and Kappa index (k) [61, 62]. Detail description of these indices is presented in relevant studies [4, 63–70]. Formulas of these indices are presented in Table 1.

3. Study Area

The study area of Muong Lay district is located in the northwest of Vietnam between 22°0′N and 22°5′N and 103°5′E and 103°10′E, covering 11403 km^2 is highly prone to landslides (Figure 1). The area is located, at the confluence of Da, Nam Na, and Nam Lay Rivers in a narrow and long valley [3, 4]. The elevation varies between 125 and 1778 m. The hill slopes are connected with sheered cliffs and marked by rapids. The area is tectonically active, structurally disturbed, and traversed by several faults including Chay River fault, Red River fault, and Dien Bien-Lai Chau fault zones, within the Lai Chau-Dien Bien fault zone, thus vulnerable to natural disasters such as floods and landslides. This area experiences annual average temperature ranging between 21°C and 23°C, humidity up to 84% and average number of sunshine hours ranging from 1820 to 2035 hours per year [3, 4].

4. Geospatial Database

Geospatial data of landslide inventory were obtained from the Vietnam Academy of Geosciences and Minerals official web portal (http://canhbaotruotlo.vn) and updated from Google Earth images and field surveys. In total, 271 landslide events were recorded and studied for the development of models. Landslides in the area are of rotational, translational, debris, rock falls, and mixed types. Most of the landslides occur along and adjacent to the main connecting road to the Muong Lay district, on the Highways 6 and 12 [3, 4]. For developing landsides prediction models, landslide conditioning or affecting factors such as topographical factors (aspect, slope, and curvature) were generated from digital elevation model (DEM) of 12.5 m available online (https://vertex.daac.asf.alaska.edu). Geological and topographical factors (distance to faults, distance to rivers, geology/lithology, focal flow, weathering rocks, and distance to roads) were generated and extracted from geology and topography maps (1 : 50000) collected from General Department of Geology and Minerals of Vietnam. Maps of these conditioning factors are presented in Figure 2, while the spatial analysis of past and present landslides carried out on these conditioning maps is presented in Figure 3 [3, 4]. More detailed analysis of the individual influencing factors and mechanism of landslides is presented in the published works carried out in the same area [3, 4].

5. Modeling Methodology

Major steps of the methodological framework include [1] data collection and preparation, [2] model development, [3] model validation, and [4] generation and validation of landslide susceptibility maps (Figure 4).

5.1. Data Collection and Preparation. Landslide data of 271 past landslide events were generated by identifying landslides on Google Earth images in conjunction with available landslide records. Out of these, 70% of landslide (152 locations) and nonlandslide (152 locations) data were used to generate the training dataset for building the models, whereas 30% remaining (65 landslide locations and 65 nonlandslide locations) data were used to create testing dataset for model validation. Training and testing data in the ratio of 70/30 were selected based on the experience of authors and other published work on the similar studies [72–75]. Correlation-based feature selection method [76], which is known as one of the most effective feature selection methods for landslide susceptibility modeling [77, 78], was used to select the suitable factors for landslide modeling.

5.2. Landslide Susceptibility Model Development. For the developments of models, the training dataset was used to construct the models (VFI, ABVFI, and MBVFI). In ABVFI, AdaBoost was used as an optimization technique to optimize the training dataset, which was then used as inputs for classification of landslide and nonlandslide classes using a base classifier of VFI. Similarly, in MBVFI, MultiBoost was used as an optimization technique to optimize the training dataset which was then used as inputs for classification of landslide and nonlandslide classes using a base classifier of VFI.

5.3. Landslide Susceptibility Model Validation. Validation of the models (VFI, ABVFI, and MBVFI) was carried out using the testing dataset and quantitative statistical indices, namely, AUC, ACC, SST, SPE, PPV, NPV, RMSE, and Kappa index.

5.4. Landslide Susceptibility Map Generation and Validation. Landslide susceptibility indices scores generated by the models were classified into very low, low, moderate, high, and very high susceptibility areas based on Jenks' natural break classification method [79] for map generation. Thereafter, performance of the generated maps was validated by frequency ratio analysis [80].

Table 1: Formulas of quantitative indices used for validation of the models.

No.	Quantitative indices	Formulas	
1	Positive predictive value (%)	PPV = TP/(TP + FP)	[1]
2	Negative predictive value (%)	NPV = TN/(TN + FN)	[2]
3	Sensitivity (%)	SST = TP/(TP + FN)	[3]
4	Specificity (%)	SPE = TN/(TN + FP)	[4]
5	Accuracy (%)	ACC = (TP + TN)/(TP + TN + FP + FN)	[5]
6	Kappa	$k = (R_a - R_{ept})/(1 - R_{ept})$	[6]
7	Root mean square error	$\text{RMSE} = \sqrt{(1/m)\sum_{i=1}^{m}(V_p - V_a)^2}$	[7]
8	Area under the ROC curve	AUC = ∑TP + ∑(TN/P) + N	[8]

TP, TN, FP, and FN are considered the percentage of pixels classified correctly and incorrectly as landslide and nonlandslide classes; m is the total number of instances in the datasets; V_p and V_a are predicted and actual values of outputs; R_{ept} and R_a are expected agreements and the percentage of samples predicted correctly for landslide or nonlandslide classes; N and P are the total number of landslide and nonlandslide classes, respectively [71].

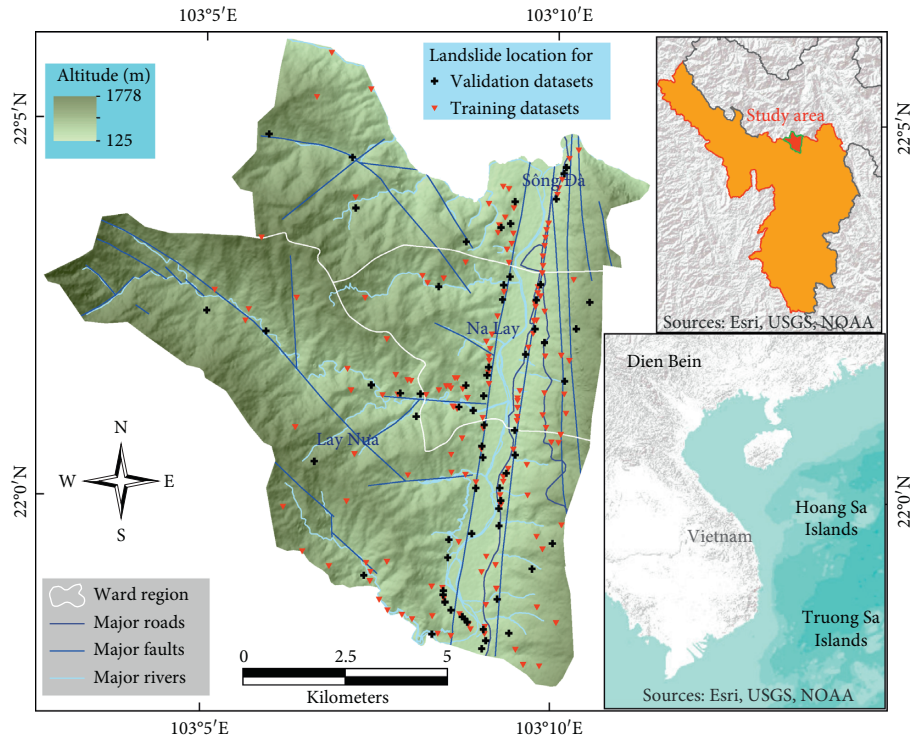


Figure 1: Spatial distribution of the landslide events in the study area.

6. Results and Analysis

6.1. Validation and Selection of Important Factors. Validation and selection of important factors was done using correlation-based feature selection [3, 77], and the results are presented in Table 2. It can be observed that distance from rivers (AM = 0.437) is the most important factor, followed by distance from roads (AM = 0.404) and distance from faults (AM = 0.336 aspect (AM = 0.226), weathering crust (AM = 0.126), geology (AM = 0.115), slope (AM = 0.076), focal flow (AM = 0.054), and curvature (AM = 0.029), respectively (Table 2).

6.2. Validation and Comparison of Landslide Susceptibility Models. Validation and comparison of landslide susceptibility models were done using PPV, NPV, SST, SPF, ACC, Kappa, and RMSE scores. The ABVFI model achieved the highest accuracy on both training (ACC = 82.12%) and testing datasets (ACC = 81.54%) compared with other models (VFI and MBVFI). This model also achieved the highest PPV (83.08%) on test data, the highest NPV on training (86.75%) and testing (80.0%) datasets. ABVFI was highly sensitive towards correctly predicting landslides in this area on both training (SST = 85.40%) and testing (SST = 80.60%) datasets. It achieved the highest SPF on the test (82.54%) dataset. ABVFI scored the highest kappa value on both training (0.624) and testing (k = 0.631) datasets. In contrast, ABVFI achieved the smallest RMSE on both training (0.367) and testing (0.390) datasets (Table 3 and Figure 5).

ABVFI model achieved the highest AUC on training (AUC = 0.897) and testing data (AUC = 0.859), followed by MBVFI on training and (AUC = 0.895) testing data (AUC = 0.839) and VFI on training (AUC = 0.845) and testing data (AUC = 0.814), respectively (Figure 6).

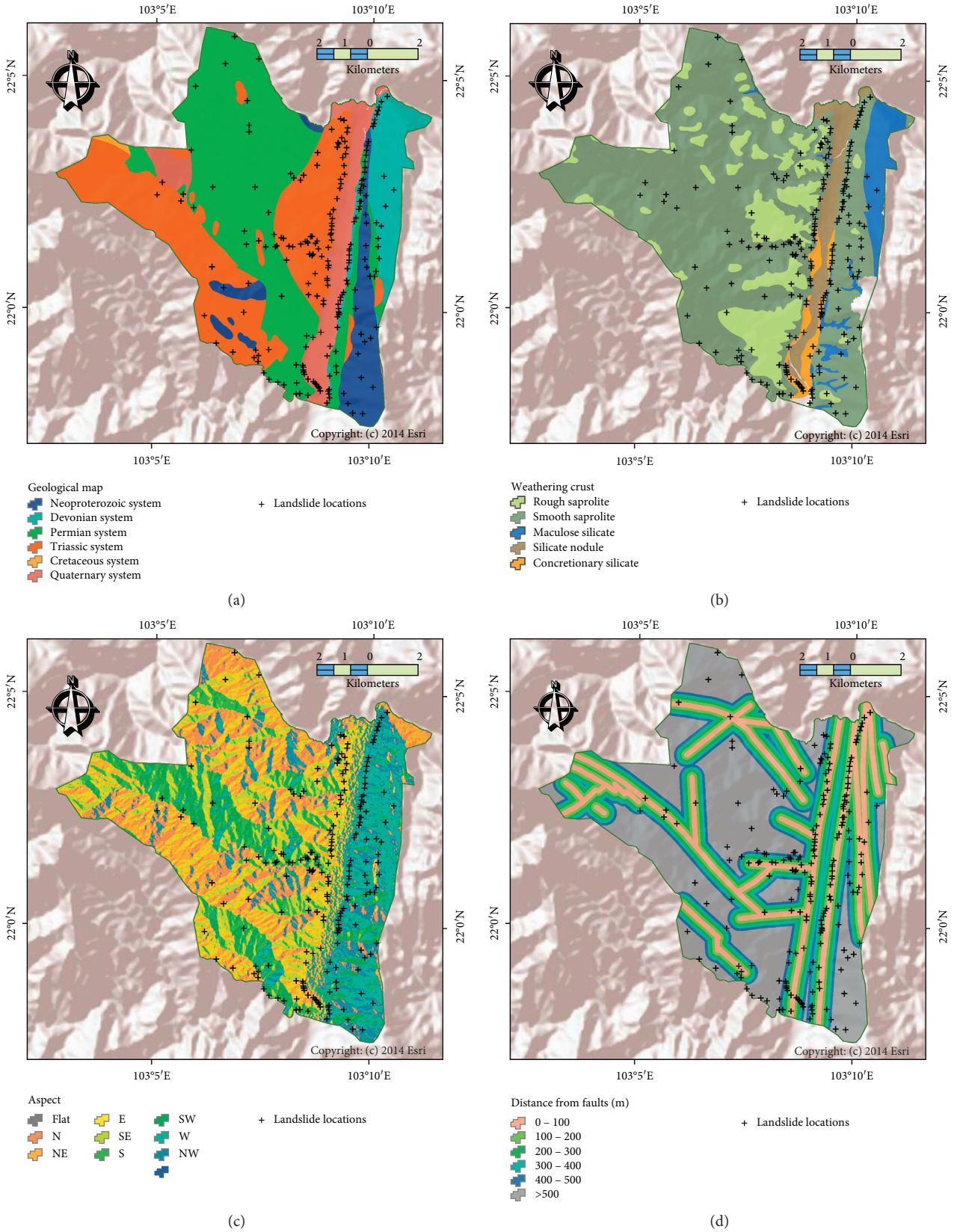


Figure 2: Continued.

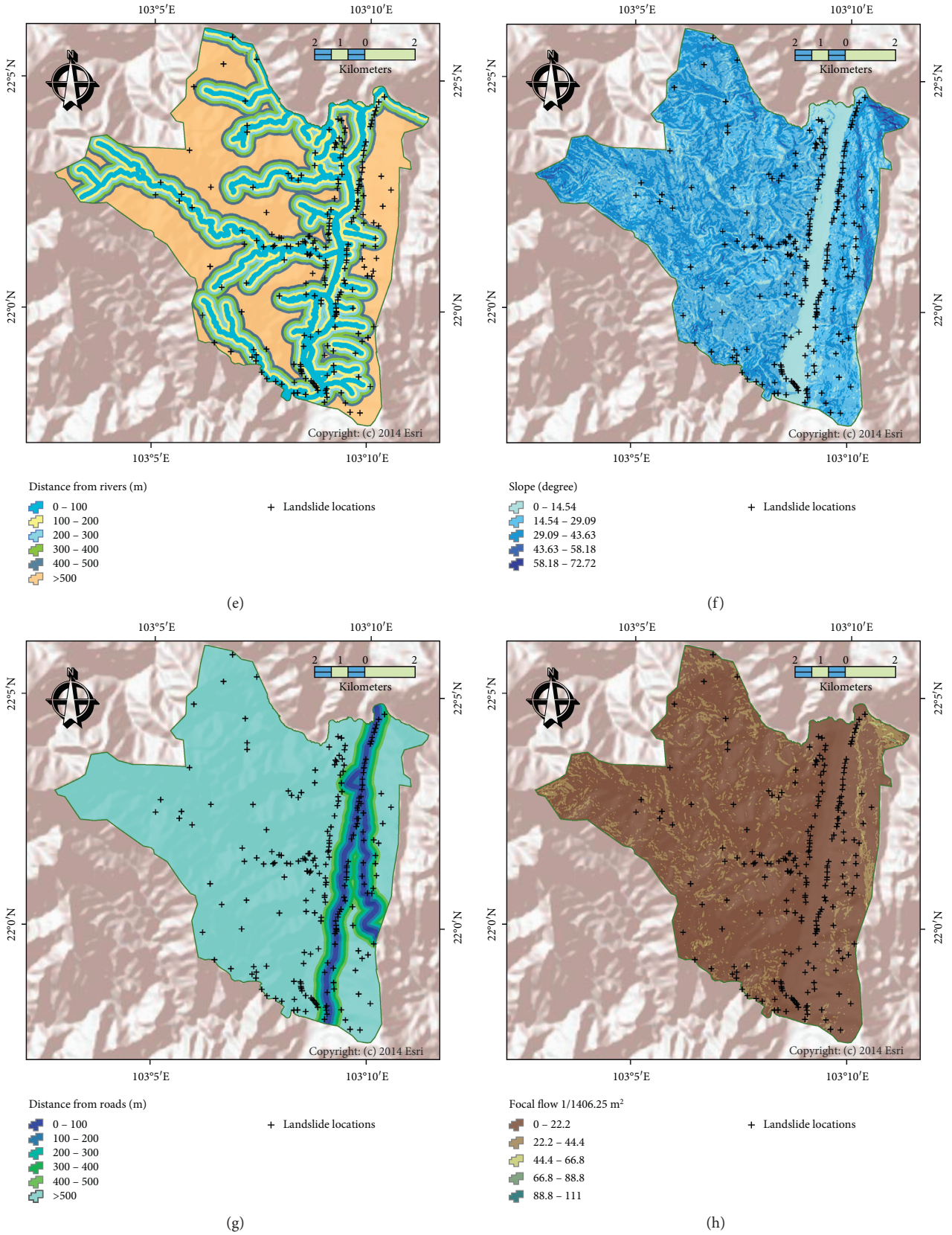
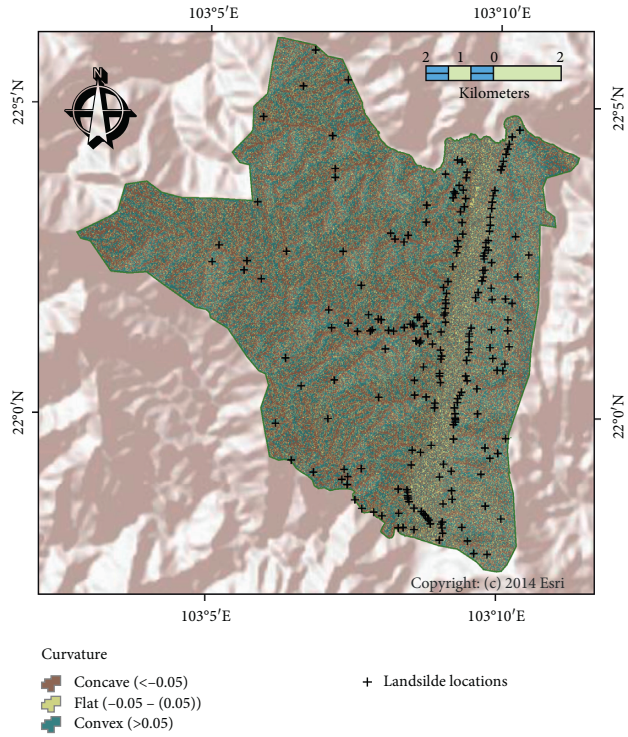


Figure 2: Continued.



(i)

Figure 2: Thematic maps: (a) geology; (b) weathering crust; (c) aspect; (d) distance from faults; (e) distance from rivers; (f) slope; (g) distance from road; (h) focal flow; (i) curvature.

In general, it is apparent that ABVFI scored the highest AUC, ACC, and kappa values and the lowest RMSE on both the train and test data; therefore, this model can be selected as the best model in terms of predictability as well as robustness. MBVFI was the second-best model followed by the VFI model.

6.3. Construction of Landslide Susceptibility Maps. Landslide susceptibility maps based on the model's study were generated into five classes: very low, low, moderate, high, and very high susceptibility areas (Figure 7). Based on the frequency analysis of each class of landslide susceptibility for each model, we found that VFI algorithm was able to predict more correctly very high and high landslide susceptible areas than the moderate and low landslide classes (Table 4). Very low landslide areas could not be predicted by VFI. MBVFI was able to predict more correctly very high landslide susceptible areas. MBVFI could equally predict high and moderate landslide susceptible areas (Table 4). Like MBVFI, ABVFI was also found to be good at predicting very high landslide-sensitive areas. Overall, ABVFI could correctly predict most of the landslide susceptible classes (Table 4).

7. Discussion

In this study, we have developed improved hybrid VFI models ensemble with AdaBoost and MultiBoost algorithms and applied them at the Muong Lay district, Dien Bien province, Vietnam, for landslide susceptibility mapping and prediction. To develop the ML models, it is important to validate and select the most suitable conditioning factors for better landslide susceptibility assessment and mapping [81]. In this study, correlation-based feature selection was applied to validate importance of the conditioning factors and accordingly select the best factors for landslide susceptibility modeling. The main principle of this method is based on the correlation analysis between the input and output variables and among input variables [3, 82]. It is a well-known feature selection method for ML applications [82]. The results indicated that distance from rivers (AM = 0.437), distance from roads (AM = 0.404), and the distance from faults (AM = 0.336) had the highest impact in the landslide susceptibility prediction in the models (Table 1), which corroborated the study of earlier workers in this region [3, 4]. Reason for greater impact of rivers on the landslide occurrences is that slope close to rivers is generally saturated with water; moreover, erosion of toe support is likely at the bottom of valleys through which river flows thus causing more landslides in river valleys. Similarly, removal of toe support while construction of roads on hilly and mountainous areas also creates instability of groundmass. Road construction also disturbs slope and surrounding rock/ground mass, which cause landslides unless protected adequately. Faults are one of the prominent slopes affecting factors, which may itself cause landslides depending on its location, orientation, and nature of infilling material. Landslides generally occur in the fault affected areas due to ongoing tectonic activities.

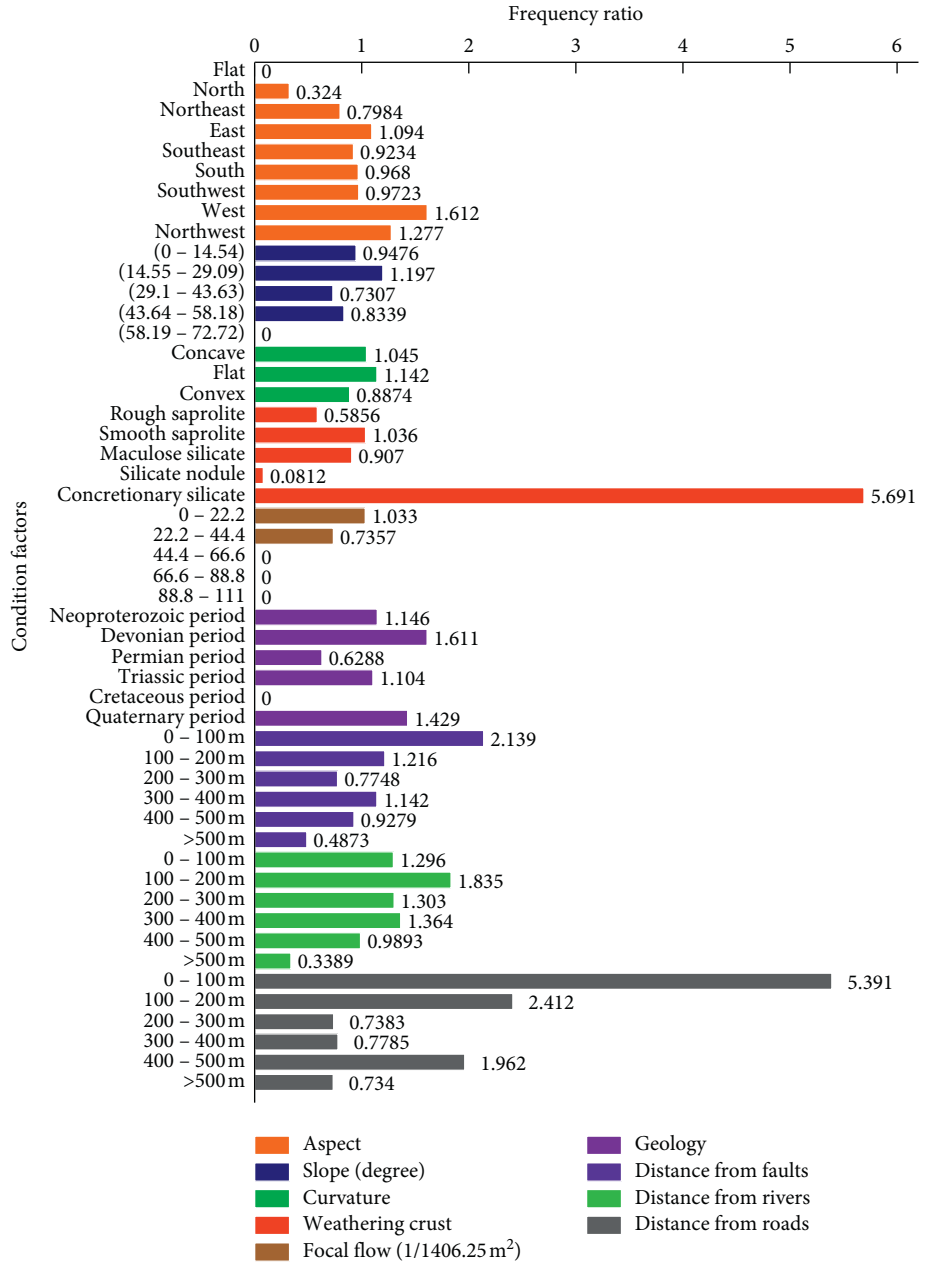


Figure 3: Frequency analysis histograms of landslide condition factors.

Validation and comparison results of the models showed that ABVFI is the most accurate and robust model on both the training and testing datasets (Table 2 and Figures 5 and 6). One of the advantages of this is that it is neither overtrained nor undertrained when compared to specifically VFI. Kappa statistics are used to evaluate the robustness of machine learning models. ABVFI and MBVFI both scored "K" greater than 0.61 on test data that makes both the models substantially robust [83, 84]. However, VFI shows a moderate kappa value of 0.446 on testing data [83, 84]. Although the RMSE value of all the three models relatively increased on testing data, it was the lowest for the ABVFI model (increase of 0.023) on training data. MBVFI scored the second lowest RMSE on testing data with an increase of 0.026 when compared to the RMSE value on training data. With the highest AUC on testing data, ABVFI scored 0.038 which is lower than it achieved on training data. On the contrary, the second-best AUC scorer MBVFI achieved 0.056, which is less AUC score on testing data than it achieved on training data. VFI achieved 0.021 which is less AUC score on the test data than it achieved on training data. In addition, it can be seen from Table 4 that the frequency ratio values of high and very high classes of the map produced by ABVFI are higher than those produced by other models (MBVFI and VFI), which proves that prediction probability of landslides of the ABVFI is higher than other models. Main reason for the better

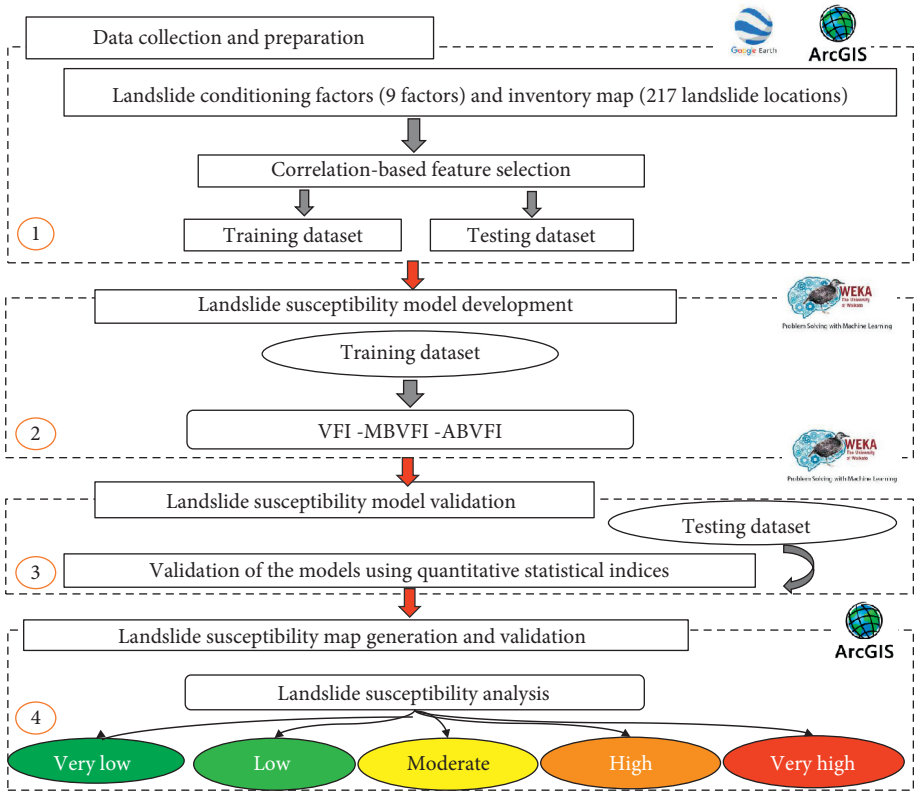


Figure 4: Methodological framework of this study.

Table 2: Importance of factors using correlation-based feature selection.

Average merit (AM)	Average rank	Landslide conditioning factors
0.437	1	Distance from rivers
0.404	2	Distance from roads
0.336	3	Distance from faults
0.226	4	Aspect
0.126	5.3	Weathering crust
0.115	5.7	Geology
0.076	7.2	Slope
0.054	8.2	Focal flow
0.029	8.6	Curvature

Table 3: Model performance using various quantitative indices.

No.	Parameters	Training model			Validation model		
		VFI	MBVFI	ABVFI	VFI	MBVFI	ABVFI
1	TP	101	120	117	47	53	54
2	TN	124	127	131	47	52	52
3	FP	50	31	34	18	12	11
4	FN	27	24	20	18	13	13
5	PPV (%)	66.89	79.47	77.48	72.31	81.54	83.08
6	NPV (%)	82.12	84.11	86.75	72.31	80.00	80.00
7	SST (%)	78.91	83.33	85.40	72.31	80.30	80.60
8	SPF (%)	71.26	80.38	79.39	72.31	81.25	82.54
9	ACC (%)	74.50	81.79	82.12	72.31	80.77	81.54
10	Kappa	0.490	0.636	0.642	0.446	0.615	0.631
11	RMSE	0.463	0.395	0.367	0.473	0.421	0.390

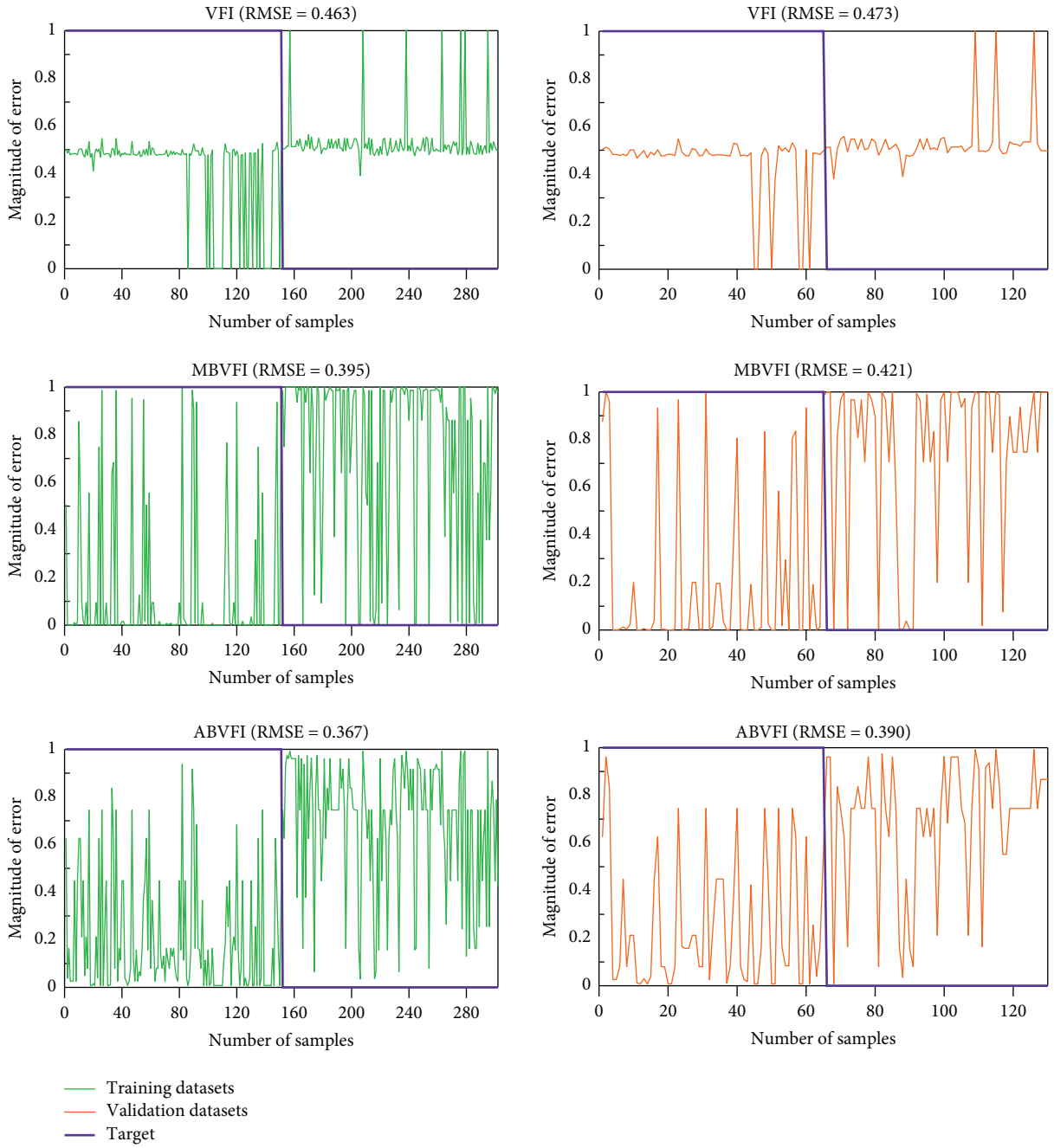


Figure 5: RMSE analysis of the models.

performance of ABVFI in comparison to other two models (MBVFI and VFI) is that it uses the AdaBoost ensemble technique, which has many advantages such as (i) it analyses large amount of data efficiently; (ii) it handles uncertainties and performs error analysis in better way; (iii) it optimizes the training dataset, selects the informative features, and provides appropriate weights to features for better data interpretation; and (iv) it is mathematically insensitive to overtraining and training error diverges to zero exponentially [85].

In general, ABVFI achieved the best performance in this study, while comparing to other models. It is noticed that this is the first time AdaBoost and MultiBoost ensemble with VFI as base classifier and were developed as hybrid models (ABVFI and MBVFI) and evaluated for the prediction of landslide susceptibility. Limitation of the study is that we have used data of available 271 landslide events for the development of models. Therefore, we suggest a larger sample size of data in future study to check and refine performance of the models.

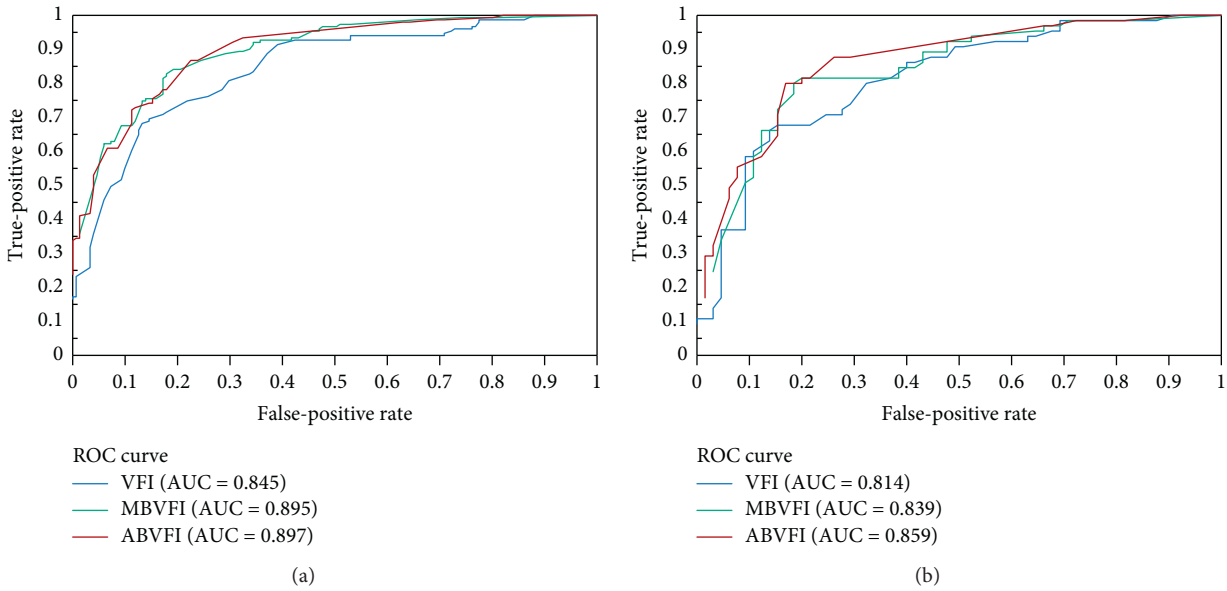


Figure 6: Validation and comparison of the models using the ROC curve: (a) training dataset; (b) validating dataset.

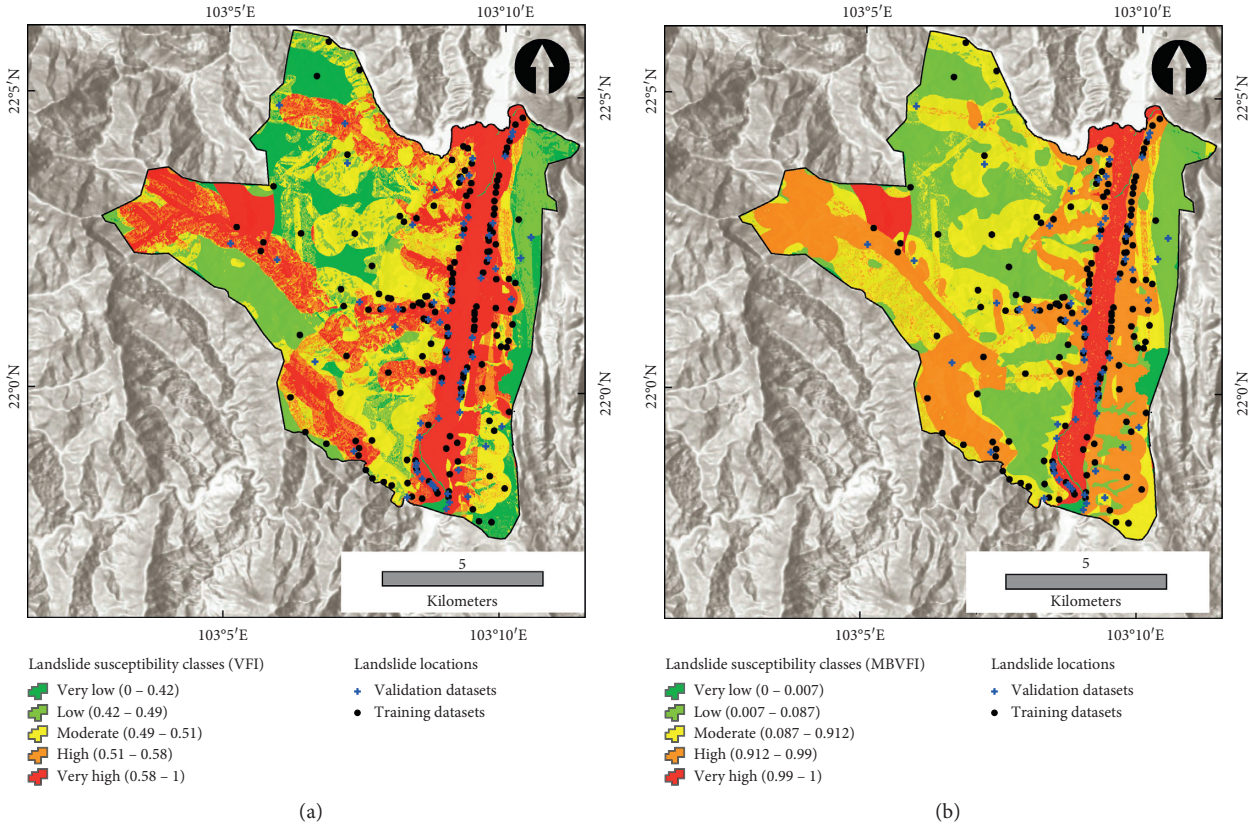
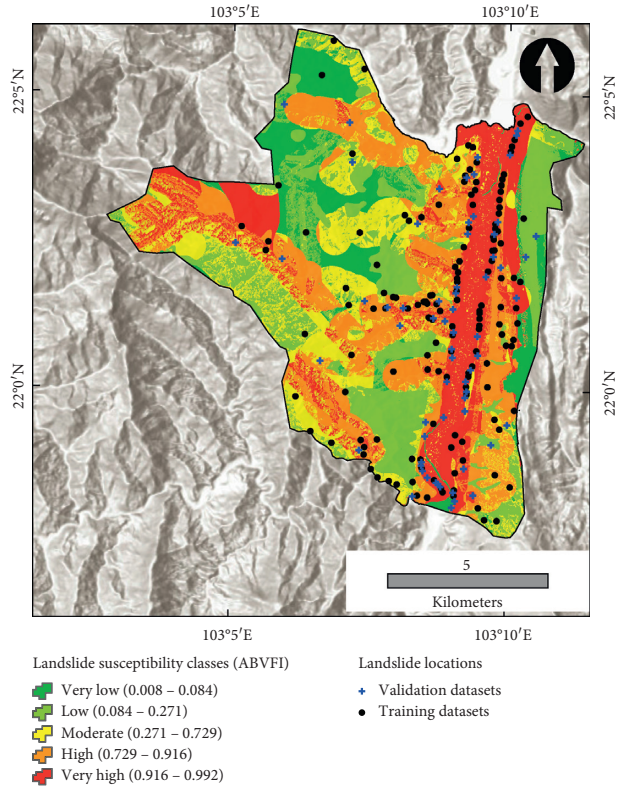


Figure 7: Continued.



(c)

Figure 7: Landslide susceptibility maps using different models: (a) VFI; (b) MBVFI; (c) ABVFI.

Table 4: Validation of landslide susceptibility maps.

Model	Class	Class pixels	Landslide pixels	% class pixels	% landslide pixels	Frequency ratio
VFI	Very low	11751	0	1.618646	0	0
	Low	196916	6	27.12428	9.230769	0.340314
	Moderate	220244	9	30.3376	13.84615	0.456402
	High	226270	35	31.16765	53.84615	1.727629
	Very high	70796	15	9.751824	23.07692	2.366421
MBVFI	Very low	99237	2	13.66944	3.076923	0.225095
	Low	108692	1	14.97182	1.538462	0.102757
	Moderate	201878	10	27.80777	15.38462	0.553249
	High	82944	5	11.42516	7.692308	0.673278
	Very high	233226	47	32.12581	72.30769	2.250766
ABVFI	Very low	79710	1	10.97969	1.538462	0.140119
	Low	167322	4	23.04784	6.153846	0.267003
	Moderate	146596	9	20.19293	13.84615	0.685693
	High	169770	14	23.38504	21.53846	0.921036
	Very high	162579	37	22.39451	56.92308	2.541832

8. Concluding Remarks

In the present study, spatial landslide susceptibility prediction models, namely, ABVFI and MBVFI with VFI as a base classifier were developed as ensemble or hybrid models, which have emerged as better decision-making tools. The hybrid novel model ABVFI (AUC = 0.897) is the best model in comparison to single VBI (AUC = 0.845) and other developed hybrid model MBVFI (0.895). Validation and statistical analysis results show that ABVFI is the most accurate and robust model on both the training and testing datasets. Accurate susceptibility maps generated by this model can be used for safe and economic construction of roads, powerhouses, and other infrastructures. Thus, the ABVFI model can be used for the proper management of landslides in the hilly areas not only in Vietnam but also other areas of the world. In future study, it is proposed to consider excessive rains and drought factors due to climate

change effects for further improvement in prediction capability of landslide susceptibility models.

Data Availability

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant no. 105.08–2019.03.

References

[1] F. Guzzetti, A. Carrara, M. Cardinali et al., "Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy," *Geomorphology*, vol. 31, no. 1–4, pp. 181–216, 1999.

[2] O. Kjekstad, "The challenges of landslide hazard mitigation in developing countries," in *Proceedings of the First North American Landslide Conference*, Vail, CO, USA, June 2007.

[3] T. V. Phong, T. T. Phan, I. Prakash et al., "Landslide susceptibility modeling using different artificial intelligence methods: a case study at Muong lay district, Vietnam," *Geocarto International*, pp. 1–24, 2019.

[4] D. V. Dao, A. Jaafari, M. Bayat et al., "A spatially explicit deep learning neural network model for the prediction of landslide susceptibility," *Catena*, vol. 188, Article ID 104451, 2020.

[5] D. T. Mafi-Gholami, B. Pradhan, O. Lofman et al., "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*, vol. 96, pp. 28–40, 2012.

[6] T. Can, H. A. Nefeslioglu, C. Gokceoglu et al., "Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses," *Geomorphology*, vol. 72, no. 1-4, pp. 250–271, 2005.

[7] H. R. Pourghasemi and N. Kerle, "Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran province, Iran," *Environmental Earth Sciences*, vol. 75, no. 3, p. 185, 2016.

[8] P. M. Atkinson and R. Massari, "Autologistic modelling of susceptibility to landsliding in the Central Apennines, Italy," *Geomorphology*, vol. 130, no. 1-2, pp. 55–64, 2011.

[9] K. C. Devkota, A. D. Regmi, H. R. Pourghasemi et al., "Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya," *Natural Hazards*, vol. 65, no. 1, pp. 135–165, 2013.

[10] D. Q. Yoshida, D. H. Nguyen, I. Prakash et al., "GIS based frequency ratio method for landslide susceptibility mapping at Da Lat city, Lam Dong province, Vietnam," *Vietnam Journal of Earth Sciences*, vol. 42, no. 1, pp. 55–66, 2020.

[11] B. T. Jaafari, I. Prakash, J. Dou et al., "A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers," *Geocarto International*, vol. 35, no. 12, pp. 1267–1292, 2020.

[12] H. R. Singh, H. R. Moradi, S. M. Fatemi Aghda, C. Gokceoglu, and B. Pradhan, "GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran)," *Arabian Journal of Geosciences*, vol. 7, no. 5, pp. 1857–1878, 2014.

[13] T. Gokceoglu, H.-B. Ly, P. T. Trinh et al., "Landslide susceptibility mapping using forest by penalizing attributes (FPA) algorithm based machine learning approach," *Vietnam Journal of Earth Sciences*, vol. 42, no. 3, 2020.

[14] B. Ghasemain, D. T. Asl, B. T. Pham et al., "Shallow landslide susceptibility mapping: a comparison between classification and regression tree and reduced error pruning tree algorithms," *Vietnam Journal of Earth Sciences*, vol. 42, no. 3, 2020.

[15] N. Micheletti, L. Foresti, S. Robert et al., "Machine learning feature selection methods for landslide susceptibility mapping," *Mathematical Geosciences*, vol. 46, no. 1, pp. 33–57, 2014.

[16] M. Leuenberger, "A landslide susceptibility model using the analytical hierarchy process method and multivariate statistics in perialpine slovenia," *Geomorphology*, vol. 74, no. 1–4, pp. 17–28, 2006.

[17] N. R. Regmi, J. R. Giardino, and J. D. Vitek, "Modeling susceptibility to landslides using the weight of evidence approach: western Colorado, USA," *Geomorphology*, vol. 115, no. 1-2, pp. 172–187, 2010.

[18] A. Ávila, F. Justino, A. Wilson, and M. Amorim, "Recent precipitation trends, flash floods and landslides in southern Brazil," *Environmental Research Letters*, vol. 11, no. 11, Article ID 114029, 2016.

[19] H. A. Bromwich, T. Y. Duman, and S. Durmaz, "Landslide susceptibility mapping for a part of tectonic kelkit valley (eastern black sea region of Turkey)," *Geomorphology*, vol. 94, no. 3-4, pp. 401–418, 2008.

[20] D. B. Simons and T. J. Ward, *Landslide Potential Delineation*, prepared for Colorado State University Experiment Station, Fort Collins, CO, USA, 1979.

[21] J. Corominas, C. van Westen, P. Frattini et al., "Recommendations for the quantitative analysis of landslide risk," *Bulletin of Engineering Geology and the Environment*, vol. 73, no. 2, pp. 209–263, 2014.

[22] H. R. Pourghasemi, H. R. Moradi, and S. M. Fatemi Aghda, "Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances," *Natural Hazards*, vol. 69, no. 1, pp. 749–779, 2013.

[23] D. G. Toll, "Artificial intelligence applications in geotechnical engineering," *Electronic Journal of Geotechnical Engineering*, vol. 1, pp. 767–773, 1996.

[24] A. Kornejady, M. Ownegh, and A. Bahremand, "Landslide susceptibility assessment using maximum entropy model with two different data sampling methods," *Catena*, vol. 152, pp. 144–162, 2017.

[25] B. Pradhan, "A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS," *Computers & Geosciences*, vol. 51, pp. 350–365, 2013.

[26] H. Hong, J. Liu, and A.-X. Zhu, "Modeling landslide susceptibility using logit boost alternating decision trees and forest by penalizing attributes with the bagging ensemble," *Science of the Total Environment*, vol. 718, Article ID 137231, 2020.

[27] B. T. Pham, T. Nguyen-Thoi, C. Qi et al., "Coupling RBF neural network with ensemble learning techniques for landslide susceptibility mapping," *Catena*, vol. 195, Article ID 104805, 2020.

[28] W. Phong, H. R. Pourghasemi, A. Kornejady, and N. Zhang, "Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques," *Geoderma*, vol. 305, pp. 314–327, 2017.

[29] X. Zhang, F. Lin, S. Zhu et al., "Mine landslide susceptibility assessment using IVM, ANN and SVM models considering the contribution of affecting factors," *PloS One*, vol. 14, no. 4, 2019.

[30] S. A. Naghibi, H. R. Pourghasemi, and K. Abbaspour, "A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS," *Theoretical and Applied Climatology*, vol. 131, no. 3-4, pp. 967–984, 2018.

[31] P. Vorpahl, H. Elsenbeer, M. Märker, and B. Schröder, "How can statistical models help to determine driving factors of landslides?" *Ecological Modelling*, vol. 239, pp. 27–39, 2012.

[32] A. Schröder, B. Pradhan, and K. Rezaei, "Assessment of landslide susceptibility using statistical-and artificial intelligence-based FR–RF integrated model and multiresolution DEMs," *Remote Sensing*, vol. 11, no. 9, p. 999, 2019.

[33] W. Lee, X. Xie, J. Wang et al., "A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility," *Catena*, vol. 151, pp. 147–160, 2017.

[34] Q. Pradhan, Z. Xu, S. Li et al., "Novel entropy and rotation forest-based credal decision tree classifier for landslide susceptibility modeling," *Entropy*, vol. 21, no. 2, 106 pages, 2019.

[35] H. Hong, J. Liu, D. T. Bui et al., "Landslide susceptibility mapping using J48 decision tree with adaboost, bagging and rotation forest ensembles in the guangchang area (China)," *Catena*, vol. 163, pp. 399–413, 2018.

[36] W. Pradhan, M. Panahi, and H. R. Pourghasemi, "Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling," *Catena*, vol. 157, pp. 310–324, 2017.

[37] H. Moayedi, M. Mehrabi, M. Mosallanezhad, A. S. A. Rashid, and B. Pradhan, "Modification of landslide susceptibility mapping using optimized PSO-ANN technique," *Engineering with Computers*, vol. 35, no. 3, pp. 967–984, 2019.

[38] C. Rashid, F. Dai, X. Xu, and Y. H. Lee, "GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang river watershed, China," *Geomorphology*, vol. 145-146, pp. 70–80, 2012.

[39] W. Lee, H. Shahabi, S. Zhang et al., "Landslide susceptibility modeling based on gis and novel bagging-based kernel logistic regression," *Applied Sciences*, vol. 8, no. 12, p. 2540, 2018.

[40] B. T. Khosravi and I. Prakash, "A novel hybrid model of bagging-based naïve bayes trees for landslide susceptibility assessment," *Bulletin of Engineering Geology and the Environment*, vol. 78, no. 3, pp. 1911–1925, 2019.

[41] S. Lee, "Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data," *International Journal of Remote Sensing*, vol. 26, no. 7, pp. 1477–1491, 2005.

[42] E. K. Sahin, I. Colkesen, and T. Kavzoglu, "A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping," *Geocarto International*, vol. 35, no. 4, pp. 341–363, 2020.

[43] W. Chen, X. Xie, J. Peng, J. Wamg, Z. Duan, and H. Hong, "GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, naïve-bayes tree, and alternating decision tree models," *Geomatics, Natural Hazards and Risk*, vol. 8, no. 2, pp. 950–973, 2017.

[44] H. Wang, H. Shahabi, A. Shirzadi et al., "Landslide susceptibility assessment at the wuning area, China: a comparison between multi-criteria decision making, bivariate statistical and machine learning methods," *Natural Hazards*, vol. 96, no. 1, pp. 173–212, 2019.

[45] J.-C. Chen, S. Lee, and H.-S. Jung, "Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang, Korea," *Geocarto International*, vol. 33, no. 9, pp. 1000–1015, 2018.

[46] S. Lee and J. Choi, "Landslide susceptibility mapping using GIS and the weight-of-evidence model," *International Journal of Geographical Information Science*, vol. 18, no. 8, pp. 789–814, 2004.

[47] G. Demiröz and H. A. Güvenir, "Classification by voting feature intervals," in *Proceedings of the European Conference on Machine Learning*, Springer, Berlin, Germany, pp. 85–92, April 1997.

[48] K. Marsolo, M. Twa, and M. A. Bullimore, "Spatial modeling and classification of corneal shape," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 2, pp. 203–212, 2007.

[49] H. A. Parthasarathy, G. Demiröz, and N. İlter, "Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals," *Artificial Intelligence in Medicine*, vol. 13, no. 3, pp. 147–165, 1998.

[50] B. T. Pham, D. T. Bui, I. Prakash et al., "A comparative study of sequential minimal optimization-based support vector machines, vote feature intervals, and logistic regression in landslide susceptibility assessment using GIS," vol. 76, no. 10, p. 371, 2017.

[51] R. Del Gaudio, G. Batista, and A. Branco, "Coping with highly imbalanced datasets: a case study with definition extraction in a multilingual setting," *Natural Language Engineering*, vol. 20, no. 3, pp. 327–359, 2014.

[52] Y. Freund and L. Mason, "The alternating decision tree learning algorithm," in *ICML '99: Proceedings of the Sixteenth International Conference on Machine Learning*, pp. 124–133, Bled, Slovenia, June 1999.

[53] H. Schwenk and Y. Bengio, "Boosting neural networks," *Neural Computation*, vol. 12, no. 8, pp. 1869–1887, 2000.

[54] Y. Freund, R. E. Schapire, and R. Schapire, "Experiments with a new boosting algorithm," in *Thirteenth International Conference on ML*, pp. 3–6, Bari, Italy, July 1996.

[55] S.-J. Wang, A. Mathew, Y. Chen et al., "Empirical analysis of support vector machine ensemble classifiers," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6466–6476, 2009.

[56] G. I. Xi, "Multiboosting: a technique for combining boosting and wagging," *Machine Learning*, vol. 40, no. 2, pp. 159–196, 2000.

[57] V.-T. Nguyen, T. H. Tran, N. A. Ha et al., "GIS based novel hybrid computational intelligence models for mapping landslide susceptibility: a case study at da lat city, Vietnam," *Sustainability*, vol. 11, no. 24, p. 7118, 2019.

[58] P. T. Ngo, D. H. Ha, H. D. Nguyen et al., "Improvement of credal decision trees using ensemble frameworks for groundwater potential modeling," *Sustainability*, vol. 12, no. 7, p. 2622, 2020.

[59] M. Van Phong, E. Benetos, C. Kotropoulos et al., "A neural network approach to audio-assisted movie dialogue detection," *Neurocomputing*, vol. 71, no. 1-3, pp. 157–166, 2007.

[60] D. Tien Bui, B. T. Pham, and Q. P. Nguyen, "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," *International Journal of Digital Earth*, vol. 9, no. 11, pp. 1077–1097, 2016.

[61] M. Hoang, S. Janizadeh, D. Tien Bui et al., "A tree-based intelligence ensemble approach for spatial prediction of potential groundwater," *International Journal of Digital Earth*, pp. 1–22, 2020.

[62] H. Pham, B. Jarihani, S. Tavakkoli Piralilou, M. Avand, and O. Ghorbanzadeh, "A semi-automated object-based gully networks detection using different machine learning models: a case study of bowen catchment, queensland, Australia," *Sensors*, vol. 19, no. 22, p. 4893, 2019.

[63] B. T. Chittleborough, T. V. Phong, T. Nguyen-Thoi et al., "Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers," *Geocarto International*, pp. 1–23, 2020.

[64] B. T. Parial, M. Avand, S. Janizadeh et al., "GIS based hybrid computational approaches for flash flood susceptibility assessment," *Water*, vol. 12, no. 3, p. 683, 2020.

[65] H.-B. Ly, P. G. Asteris, and B. Pham, "Accuracy assessment of extreme learning machine in predicting soil compression coefficient," *Vietnam Journal of Earth Sciences*, vol. 42, no. 1, 2020.

[66] B. T. Pham, S. K. Singh, and H.-B. Ly, "Using artificial neural network (ann) for prediction of soil coefficient of consolidation," *Vietnam Journal of Earth Sciences*, vol. 42, no. 1, 2020.

[67] B. T. Pham, T. Van Phong, T. Nguyen-Thoi et al., "GIS-based ensemble soft computing models for landslide susceptibility mapping," *Estudios Geológicos*, vol. 65, no. 1, pp. 49–65.

[68] B. T. Pham, A. Jaafari, M. Avand et al., "Performance evaluation of machine learning methods for forest fire modeling and prediction," *Symmetry*, vol. 12, no. 6, p. 1022, 2020.

[69] P. T. Al-Ansari, D. H. Ha, M. Avand et al., "Soft computing ensemble models based on logistic regression for groundwater potential mapping," *Applied Sciences*, vol. 10, no. 7, p. 2469, 2020.

[70] A. Jaafari, A. P. Yunus, J. Dou et al., "Machine learning methods for landslide susceptibility studies: a comparative overview of algorithm performance," *Earth-Science Reviews*, vol. 207, Article ID 103225, 2020.

[71] P. Yariyan, S. Janizadeh, T. Van Phong et al., "Improvement of best first decision trees using bagging and dagging ensembles for flood probability mapping," *Water Resources Management*, vol. 34, no. 9, pp. 3037–3053, 2020.

[72] B. T. Nguyen, I. Prakash, and A. Jaafari, "Spatial prediction of rainfall-induced landslides using aggregating one-dependence estimators classifier," *Journal of the Indian Society of Remote Sensing*, vol. 46, no. 9, pp. 1457–1470, 2018.

[73] A. Bui, K. Solaimani, M. H. Roshan et al., "Uncertainties of prediction accuracy in shallow landslide modeling: sample size and raster resolution," *Catena*, vol. 178, pp. 172–188, 2019.

[74] B. Pham, A. Shirzadi, H. Shahabi et al., "Landslide susceptibility assessment by novel hybrid machine learning algorithms," *Sustainability*, vol. 11, no. 16, p. 4386, 2019.

[75] W. Omidvar, H. R. Pourghasemi, S. A. Naghibi et al., "Prioritization of landslide conditioning factors and its spatial modeling in Shangnan ccounty, China using GIS-based data mining algorithms," *Bulletin of Engineering Geology and the Environment*, vol. 77, no. 2, pp. 611–629, 2018.

[76] M. A. Hall, *Correlation-based Feature Selection for Machine Learning*, University of Waikato, Hamilton, NewZealand, 1999.

[77] B. Thai Pham, D. Tien Bui, I. J. C. E. Prakash et al., "Landslide susceptibility modelling using different advanced decision trees methods," vol. 35, no. 1–4, pp. 139–157, 2018.

[78] B. T. Pham, B. Pradhan, D. Tien Bui, and M. B. Dholakia, "A comparative study of different machine learning methods for landslide susceptibility assessment: a case study of Uttarakhand area (India)," *Environmental Modelling & Software*, vol. 84, pp. 240–250, 2016.

[79] G. Prakash, "The data model concept in statistical mapping," *International Yearbook of Cartography*, vol. 7, pp. 186–190, 1967.

[80] A. Yalcin, S. Reis, and A. C. Aydinoglu, "A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey," *Catena*, vol. 85, no. 3, pp. 274–287, 2011.

[81] S. Park, S. Son, J. Han et al., "Effects of variable selection on the landslide susceptibility assessment using machine learning techniques," in *Proceedings of Earth Resources and Environmental Remote Sensing/GIS Applications X*, Febuary 2019, Article ID 111560M.

[82] M. Mokarram, H. R. Pourghasemi, and J. P. Tiefenbacher, "Comparison analytic network and analytical hierarchical process approaches with feature selection algorithm to predict groundwater quality," *Environmental Earth Sciences*, vol. 78, no. 21, p. 625, 2019.

[83] J. L. Fleiss and J. Cohen, "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability," *Educational and Psychological Measurement*, vol. 33, no. 3, pp. 613–619, 1973.

[84] S. K. Singh, R. W. Taylor, and M. M. Rahman, "Developing robust arsenic awareness prediction models using machine learning algorithms," *Journal of Environmental Management*, vol. 211, no. 1, pp. 25–137, 2018.

[85] M. Pradhan, M. A. Sadeghi, S. Nourizadeh et al., "Power system security assessment using adaboost algorithm," in *Proceedings of the North American Power Symposium (NAPS)*, Starkville, MS, USA, October 2009.