



Article

Hybrid BBO-DE Optimized SPAARCTree Ensemble for Landslide Susceptibility Mapping

Duc Anh Hoang¹, Hung Van Le^{1,*} , Dong Van Pham¹, Pham Viet Hoa² and Dieu Tien Bui³

¹ Faculty of Information Technology, Hanoi University of Mining and Geology, Duc Thang, Bac Tu Liem, Hanoi, Vietnam; hoanganhduc@humg.edu.vn (D.A.H.); phamvandong@humg.edu.vn (D.V.P.)

² Ho Chi Minh City Institute of Resources Geography, Vietnam Academy of Science and Technology, Mac Dinh Chi 1, Ben Nghe, 1 District, Ho Chi Minh City, Vietnam; pvhoa@hcmig.vast.vn

³ GIS Group, Department of Business and IT, University of South-Eastern Norway, Gullbringvegen 36, N-3800 Bø i Telemark, Norway; dieu.t.bui@usn.no

* Correspondence: levanhung@humg.edu.vn

Abstract: This paper presents a new hybrid ensemble modeling method called BBO-DE-STreeEns for land-slide susceptibility mapping in Than Uyen district, Vietnam. The method uses subbagging and random subsampling to generate subdatasets for constituent classifiers of the ensemble model, and a split-point and attribute reduced classifier (SPAARC) decision tree algorithm to build each classifier. To optimize hyperparameters of the ensemble model, a hybridization of biogeography-based optimization (BBO) and differential evolution (DE) algorithms is adopted. The land-slide database for the study area includes 114 landslide locations, 114 non-landslide locations, and ten influencing factors: elevation, slope, curvature, aspect, relief amplitude, soil type, geology, distance to faults, distance to roads, and distance to rivers. The database was used to build and verify the BBO-DE-STreeEns model, and standard statistical metrics, namely, positive predictive value (PPV), negative predictive value (NPV), sensitivity (Sen), specificity (Spe), accuracy (Acc), Fscore, Cohen's Kappa, and the area under the ROC curve (AUC), were calculated to evaluate prediction power. Logistic regression, multi-layer perceptron neural network, support vector machine, and SPAARC were used as benchmark models. The results show that the proposed model outperforms the benchmarks with a high prediction power (PPV = 90.3%, NPV = 83.8%, Sen = 82.4%, Spe = 91.2%, Acc = 86.8%, Fscore = 0.862, Kappa = 0.735, and AUC = 0.940). Therefore, the BBO-DE-STreeEns method is a promising tool for landslide susceptibility mapping.

Keywords: landslide; BBO; DE; SPAARC ensemble; GIS; Vietnam



Citation: Hoang, D.A.; Le, H.V.; Pham, D.V.; Hoa, P.V.; Tien Bui, D.

Hybrid BBO-DE Optimized SPAARCTree Ensemble for Landslide Susceptibility Mapping. *Remote Sens.* **2023**, *15*, 2187. <https://doi.org/10.3390/rs15082187>

Academic Editor: Bruno Adriano

Received: 5 March 2023

Revised: 14 April 2023

Accepted: 17 April 2023

Published: 20 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Landslides are a significant geohazard that continues to cause thousands of deaths and USD 100 million in damage annually worldwide [1–3]. With the increase in extreme precipitation and typhoon events [4,5], especially in developing countries' mountainous regions, the number of landslide occurrences is expected to grow [6]. Vietnam is located in Southeast Asia, which is considered as one of the world's most disaster-prone regions [7]. Therefore, the accurate spatial prediction of landslides is essential to mitigate these risks.

According to the literature, machine learning methods are preferred over statistical methods in landslide susceptibility mapping [8–14] due to the availability of geospatial data and the development of machine learning and optimization algorithms on various open-source platforms, such as Weka [15], Python [16], and Google TensorFlow [17]. This is especially true when dealing with a large number of influencing factors and limited landslide data [18]. Ensemble modeling, where multiple constituent models form the final model, has improved the reliability of landslide susceptibility mapping [19–24]. Among machine learning algorithms, decision tree-based ensemble algorithms have been found to produce superior results, making them a popular choice [10]. Additionally, these algorithms

are highly resistant to overfitting, making them reliable for this type of analysis. In fact, increasing the number of trees in a forest can improve its capacity and validation-set error rate [25,26].

Random forest [26] is a well-established decision tree-based ensemble algorithm with a reputation for high accuracy and efficient processing speed, making it a popular choice in new research. It relies on the classification and regression tree (CART) algorithm [15,27] to create its base classifiers. However, further research is needed to enhance algorithm efficiency and increase processing speed without sacrificing accuracy [28–32]. In particular, fast model building is essential for constructing optimized models that require iterative refinement to achieve the optimal solution.

The SPAARC decision tree algorithm has been proposed as a solution to reduce the computational workload of decision tree induction in a study [28]. SPAARC was tested against an implementation of the CART algorithm in Weka [15] on 14 freely available datasets from the University of California, Irvine (UCI) machine learning data repository [33]. The experimental results demonstrated that SPAARC can reduce model build times by up to 70% while maintaining classification accuracy [28]. However, the study did not investigate how SPAARC performs in an ensemble framework, and there is a need for new ensemble modeling methods in landslide susceptibility mapping that can enhance the reliability of the maps and draw reasonable conclusions.

To address this need, in this study, a new homogeneous ensemble modeling method called BBO-DE-StrEEEns is proposed. This method employs subbagging [30] and random subsampling [34] to generate subdatasets for the constituent decision trees of the ensemble. The SPAARC decision tree algorithm is then used to build the decision trees, and a hybridization of BBO [35] and DE [36] algorithms is used to optimize the hyperparameters of the ensemble model. The proposed method is verified through a case study in Than Uyen district, Lai Chau province, Vietnam, where landslides have been a recurring problem.

Four benchmark models were employed for comparison, namely, logistic regression, multi-layer perceptron neural network, support vector machine, and the SPAARC.

The remainder of the paper is organized into several sections. Section 2 provides a review of the background of the employed methods. Section 3 presents a general description of the study area and landslide database. Section 4 describes the proposed BBO-DE-StrEEEns method used to derive landslide susceptibility. The experimental results of the study are reported in Section 5, followed by a discussion. Finally, concluding remarks are presented in the last section.

2. Background of Methods Used

In this study, we investigated the effectiveness of an ensemble model of SPAARC decision trees [28] for evaluating landslide susceptibility in Than Uyen district, Lai Chau province, Vietnam. To the best of our knowledge, this is the first attempt to propose an ensemble model of SPAARC decision trees for assessing landslide susceptibility. The decision to use an ensemble model was based on the fact that ensemble learning utilizing decision trees has been shown to reduce bias and variance, as well as mitigate overfitting, without requiring pruning for each individual tree [25].

2.1. SPAARC Decision Tree Algorithm

The SPAARC decision tree algorithm is an extension of the CART algorithm, originally implemented in Weka [37]. The SPAARC algorithm integrates two key components, namely, node attribute sampling (NAS) and split-point sampling (SPS), to accelerate the decision tree induction process while maintaining classification accuracy [28]. NAS dynamically selects a subset of non-class attributes to test at each node, thereby avoiding computation of information gain for every attribute and reducing processing time. The attribute with the maximum information gain is always chosen for splitting at each node. On the other hand, SPS finds a suitable split-point for a numerical attribute at each node. In contrast to CART [31], SPS divides the attribute range into k equal-width intervals and tests only $k - 1$

possible split-points. If the number of distinct attribute values, represented by l , is less than the specified value of k , the number of possible split-points is limited to $l - 1$. Conversely, if the attribute has too many distinct values (where k is significantly less than l), SPS can assist in reducing the number of possible split-point tests. The results of experiments on 14 classification datasets from the UCI machine learning repository [32] revealed that SPAARC significantly reduced the decision tree building time and outperformed CART in terms of classification accuracy in 7 of the 14 datasets. Using $k = 10$, SPAARC reduced the total building time by more than 48% (58.499 s vs. 112.55 s) without compromising classification accuracy. The datasets used in the experiments were Mfeat-fourier, Mfeat-zernike, Page-blocks, Pen-digits, Segment, Waveform, Optical digits, Spambase, EEG eye state, Crowdsource map, Wine quality, Shuttle, Sensorless drive, and Skin segment [28].

Similar to the random forest algorithm [26], our ensemble model also includes a hyperparameter that needs to be optimized, which is referred to as the *TotalTrees* and represents the number of SPAARC decision trees within the forest.

2.2. Subbagging and Random Subspacing

It is widely recognized that when the constituent decision trees are relatively uncorrelated, the resulting ensemble model is likely to be more accurate for classification [25]. To promote such diversity within the ensemble model in this study, two techniques were employed: subbagging and random subspacing.

Subbagging (or subsample aggregating) [30] is a technique used as an alternative to bootstrap aggregating (or bagging) in the random forest algorithm [26]. Bagging is a technique used to reduce variance and overfitting in the constituent trees by generating distinct subdatasets to build decision trees through random sampling with replacement. On average, 63.2% of the examples in a subdataset will be distinct, while the rest will be duplicated. In subbagging, a proportion of examples are selected at random without replacement. For example, half-subbagging will randomly select half of the examples in the full dataset each time. The hyperparameter to be optimized for subbagging is the proportion of examples to be sampled, which we refer to as the *SizePercentage*.

Random subspacing [26] is a technique used to increase the diversity of decision trees and reduce variance. Unlike bagging, which randomly samples data with replacement, random subspacing randomly selects subsets of features to be used at each tree node. This approach helps reduce processing time by only considering a proportion of the attribute space for splitting. Additionally, it can help prevent overfitting and improve model accuracy by introducing more randomness and diversity in the trees.

In this study, we made a slight modification to the random subspacing method [26] by randomly selecting an attribute subspace for each SPAARC decision tree, rather than for each tree node. This modification was made to increase the diversity of the constituent trees in the ensemble model. As a result, we need to optimize another hyperparameter called *subSpaceSize*, which determines the proportion of the attribute space to be selected for each tree.

2.3. Hybrid BBO-DE Optimization

As mentioned earlier, our ensemble model has three hyperparameters that need to be optimized: *TotalTrees* (the number of constituent trees), *SizePercentage* (the proportion of data in Subbagging), and *subSpaceSize* (the proportion of attribute space in random subspacing). To accomplish this optimization task, we employ a hybrid BBO-DE optimization method in this study.

Differential evolution (DE) is a population-based stochastic metaheuristic, originally proposed by Storn and Price [36], that has gained popularity in solving optimization problems due to its simplicity, speed, and robustness [38]. However, DE is known to suffer from certain drawbacks such as stagnation and premature convergence [39,40]. Although DE is proficient at exploring the search space, it can be slow in locating the global optimum, indicating a better exploration ability than exploitation [41].

Biogeography-based optimization (BBO) is a bio-inspired optimization algorithm that was introduced by Simon [35] based on the biogeography theory, which is a study of how nature distributes species and optimizes environments for life. BBO has been shown to be effective for solving a wide range of optimization problems, including those typically tackled by genetic algorithms (GAs) and particle swarm optimization (PSO) [42]. To solve an optimization problem using BBO, a population of candidate solutions is created, where each solution is represented as a vector of independent variables. These variables can be thought of as suitability index variables (SIVs) in biogeography. Good solutions correspond to habitats with a high habitat suitability index (HSI), while poor solutions correspond to habitats with a low HSI. The HSI is similar to the fitness value in other bio-inspired optimization algorithms, such as GAs and PSO. BBO employs the migration operator to exchange SIVs among solutions. This sharing of SIVs allows the acceptance of new SIVs by poor solutions, which in turn may lead to an improvement in their quality. In addition to migration, BBO employs the mutation operator to model cataclysmic events that can dramatically change the HSI of a natural habitat. If a solution is selected for mutation, a randomly chosen SIV is replaced by a new, randomly generated SIV. Mutation increases population diversity and provides low and high HSI solutions with the opportunity to improve. Ma and Simon conducted a study comparing BBO with five other state-of-the-art optimization algorithms, including ant colony optimization (ACO), DE, evolution strategies (ESs), GAs, and PSO, across 25 Monte Carlo simulations [42]. The results revealed that BBO outperformed the other algorithms for 9 out of 13 benchmark functions, while PSO was the most effective for three functions. For the remaining function, both BBO and PSO performed equally well. This suggests that BBO has good exploitation ability.

Hybrid algorithms (HAs) combine the strengths of multiple algorithms into a single, more powerful optimization tool. By combining the best aspects of each constituent algorithm, HAs can often outperform individual algorithms alone [41]. In [42], several HAs were proposed that integrate BBO with DE at both the iteration and algorithm levels. By combining the exploration ability of DE with the exploitation ability of BBO, these HAs are highly effective.

The iteration-level hybridization strategy is simple. In each iteration of the proposed hybrid algorithm, DE and BBO are executed sequentially. First, DE, with its good exploration ability, is utilized to explore the search space and locate the region of the global minimum. Then, BBO, with its strong exploitation ability, is used to further exploit the identified region and find better solutions [43,44]. This iteration-level hybridization strategy allows the strengths of both DE and BBO to be combined effectively and helps overcome the weaknesses of each algorithm.

The algorithm-level hybridization involves multiple subpopulations running independently and periodically exchanging information with one another. In the hybrid BBO-DE optimization, DE subpopulations are combined using ideas from biogeography, allowing for effective information exchange among subpopulations [42]. The synergy of exploration and exploitation abilities of DE and BBO, respectively, leads to improved performance of the HAs compared to the constituent algorithms alone. Experiments have shown that the algorithm-level hybridization performs better than the iteration-level one, with the former outperforming the latter for 11 out of 13 benchmark functions [42]. This is likely due to the more effective interaction of the subpopulations in the algorithm-level hybridization. Hence, for the optimization of our ensemble model, we have utilized the hybrid BBO-DE optimization approach with algorithm-level hybridization.

3. Study Area and Landslide Data

3.1. Description of the Study Area

Than Uyen district is situated in the southeastern part of Lai Chau province, which is located in the northwest region of Vietnam. It is positioned between longitudes 103°35'E and 103°53'E, as well as latitudes 21°40'N and 22°08'N, covering an area of 792.53 km² (Figure 1).

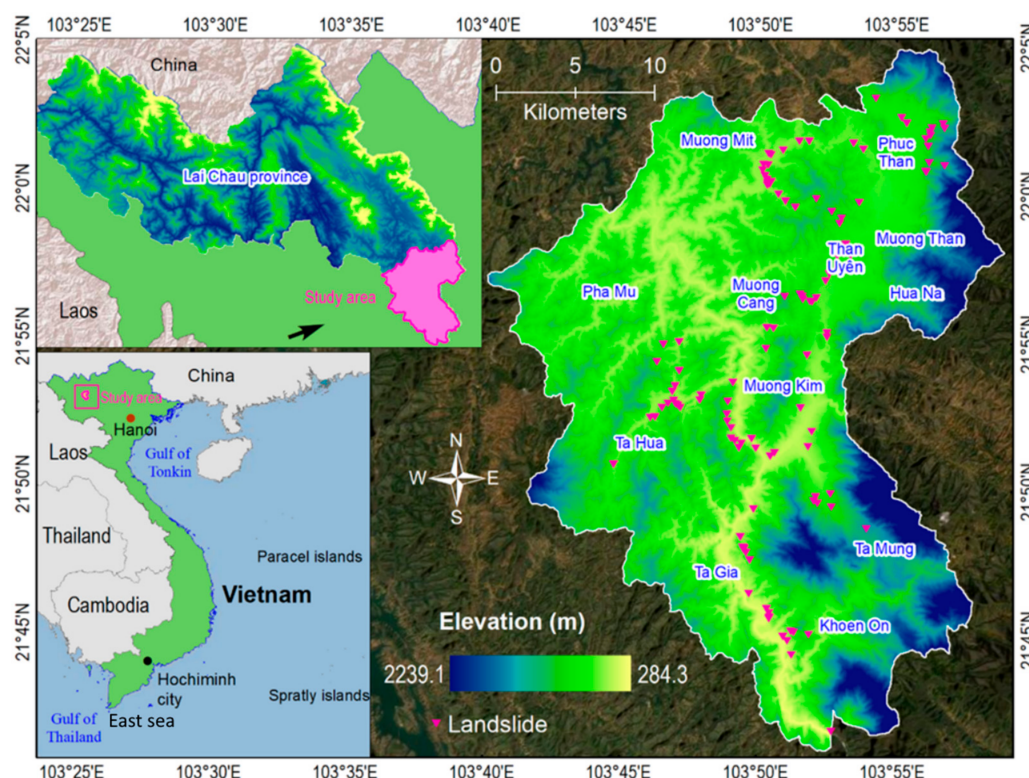


Figure 1. Location of Than Uyen district.

Than Uyen lies within the Nam Mu river basin, which is a level-1 tributary of the Da River. The district is characterized by medium-high mountains, and its terrain is notably complex and strongly dissected. To the east lies the Hoang Lien Son mountain range, while to the west are mountains that run in a northwest–southeast direction, alternating with deep valleys. The district has a relatively high river and stream density of 1.5–1.7 km/km². Than Uyen is divided into three distinct areas: the eastern area, which comprises the mountainside of the Fansipan range with rugged terrain and steep slopes; the western area, which contains the low mountains of the Pu San Cap range with elevations ranging from 600 to 1800 m above sea level; and the middle area, which is a valley composed of intermingled low hills and mountains, as well as plains with altitudes ranging from 500 to 650 m [45,46].

Than Uyen is located in a monsoonal region that experiences distinct rainy and dry seasons. The rainy season lasts from April to October, with the heaviest rainfall occurring in June and July. The dry season, on the other hand, extends from November to March of the following year. The district receives an average rainfall of 1800 to 2200 mm per year, with an average temperature ranging from 22 to 23 °C. The average humidity is around 80% [45,46].

Than Uyen has a relatively dense road network that includes several crucial routes. These include National Road 32, which connects the district to Lai Chau city and Yen Bai province, National Road 279, which connects Than Uyen to Dien Bien and Son La provinces, and Provincial Road 106, which runs from Muong Kim to Khoen On. Additionally, there are inter-commune and inter-village routes that connect residential areas. However, most of the roads in Than Uyen are winding and include many steep passes and roadside slopes, making them susceptible to landslides during the rainy season [45,46].

The study area exhibits significant and intricate tectonic activity, evident by prominent deep faults running in the northwest–southeast direction, and younger transverse faults in the north–south direction. These geological features have led to intense weathering and substantial rock disintegration, resulting in extensive zones of weak stability that extend

over hundreds of meters. These unstable areas create a high potential for landslides to occur [45,46].

Based on the petrological composition, structures, textures, physical–mechanical properties, and thickness of weathered layers, the engineering geology of the study area has been divided into the following sub-engineering geological complexes: the quaternary sedimentary complex consists of loose and poorly stable rock and soil; the terrigenous sedimentary complex formation of Yen Chau consists of partially solid and partially loose rock and soil, with stability ranging from poor to moderate; the terrigenous sedimentary complex formation of Suoi Bang consists of partially solid and partially loose rock and soil, with stability ranging from poor to moderate; the carbonate–terrigenous sedimentary complex of Pac Ma limestone consists of solid and partially solid rock with stability ranging from moderate to high; the carbonate–terrigenous sedimentary complex formation of Muong Trai consists of partially solid and partially loose rock and soil with moderate stability; the igneous rocks of the Pu Sam Cap and Phu Sa Phin complexes and Tu Le and Ngoi Thia formations consist of solid and partially solid rocks with stability ranging from moderate to high [45,46].

As of 31 December 2017, the population of Than Uyen district was 66,589, with a total of 13,838 households, out of which 3340 were classified as poor. Unfortunately, some residents of the district have built their homes along roads, directly beneath roadside slopes that pose a significant risk of landslides [47].

In recent years, Than Uyen has been one of the mountainous regions severely affected by natural disasters, particularly landslides. The causes of these landslides are multifaceted and stem from various natural, environmental, and social factors. The construction of new roads and urban areas in the region has been identified as a significant contributor to the increase in landslides caused by human activities [45,46].

3.2. Landslide Data

3.2.1. Historical Landslides

The landslide inventory map of Than Uyen district (Figure 1) used in this study was derived from the “Investigation, Assessment, and Warning Zonation for Landslides in the Mountainous Regions of Vietnam” project. This national state-funded project was conducted by the Vietnam Institute of Geosciences and Mineral Resources, Ministry of Natural Resources and Environment, and has been ongoing since 2012 [45,46].

In this project, from air-photo interpretation, 3D relief analysis based on 1:10,000 scale topographical maps, the analysis of other types (satellite and radar) of remote sensing images, and field surveys, a total of 114 landslides were identified in Than Uyen district [45,46]. The landslides were triggered by rainfall and occurred within the past decade. They were found to be concentrated along both positive and negative roadside slopes, with particularly high occurrence rates along National Road 279 from Sap Nguoi village to Khau Co pass, as well as Provincial Road 106 from Muong Kim to Khoen On, and inter-commune roads from Muong Kim to Ta Mung and from Than Uyen to Pha Mu. Importantly, we did not observe any landslide events in the district that were triggered by earthquakes during the study period.

Based on the field research and analysis, we found that the landslides in the study area were primarily triggered by heavy thunderstorms, particularly when the total daily rainfall exceeded 100 mm. During these events, the soil and rock mixture on the slopes became saturated with water, causing a reduction in shear strength and ultimately resulting in instability and failure. Figure 2 shows two photos of a landslide in the study area.



Figure 2. Two photos of the landslide on the slope wall of National Road 279, near the right bank of Nam Kim stream, Na Pa village, Muong Kim commune, Than Uyen district. Source: Vietnam Institute of Geosciences and Mineral Resources.

3.2.2. Influencing Factors

Based on the above analysis of the landslide inventory and the examination of the geo-environmental features of the study area, we identified ten factors that are believed to influence the occurrences of landslides in this area. These factors consist of elevation, slope, curvature, aspect, relief amplitude, soil type, geology, distance to faults, distance to roads, and distance to rivers. These factors have been carefully selected based on their potential impact on the stability of the terrain and the likelihood of landslide occurrences. They have been extensively employed in previous studies on landslide susceptibility analysis. Their efficacy and significance in predicting landslide occurrences have been well-documented in the literature [48–50].

Elevation is an important factor in the occurrence of landslides, as it can significantly influence slope angle, gravitational force, and the distribution of climate and vegetation cover [51]. To incorporate this factor into our analysis, we used an elevation map (Figure 3a) of the study area, which was derived from a digital elevation model (DEM) generated from the national topographic map of Vietnam at a scale of 1:50,000.

The slope gradient is a crucial factor in landslide studies, with steeper slopes being more likely to experience landslides [52]. Additionally, the curvature of the slope—which refers to changes in slope angle or direction—is an important consideration for landslide modeling, as it can significantly impact the stability of the slope material. Specifically, areas with a high absolute value of curvature, such as convex or concave slopes, are more prone to landslides than areas with a lower absolute value of curvature [53]. To account for these factors, we generated slope (Figure 3b) and curvature (Figure 3c) maps of the study area, both of which were derived from the DEM mentioned earlier.

Aspect, which refers to the compass direction of a slope and is measured in degrees from north in a clockwise direction, can influence a range of environmental factors including rainfall, sunlight, drying winds, and solar radiation. These factors, in turn, can impact soil moisture and the likelihood of landslides [54]. To account for the influence of aspect on landslide occurrence, we generated an aspect map (Figure 3d) with nine conventional classes that were extracted from the DEM used in this study.

Relief amplitude, which is defined as the maximum difference in height per unit area, is an important factor that can influence the gravitational potential energy of a rock mass and, consequently, the occurrence of landslides [55]. To account for the influence of relief amplitude on landslide occurrence in the study area, we generated a relief amplitude map (Figure 3e) using the Focal Statistic module in ArcGIS Pro software with a unit area size of 20×20 pixels.

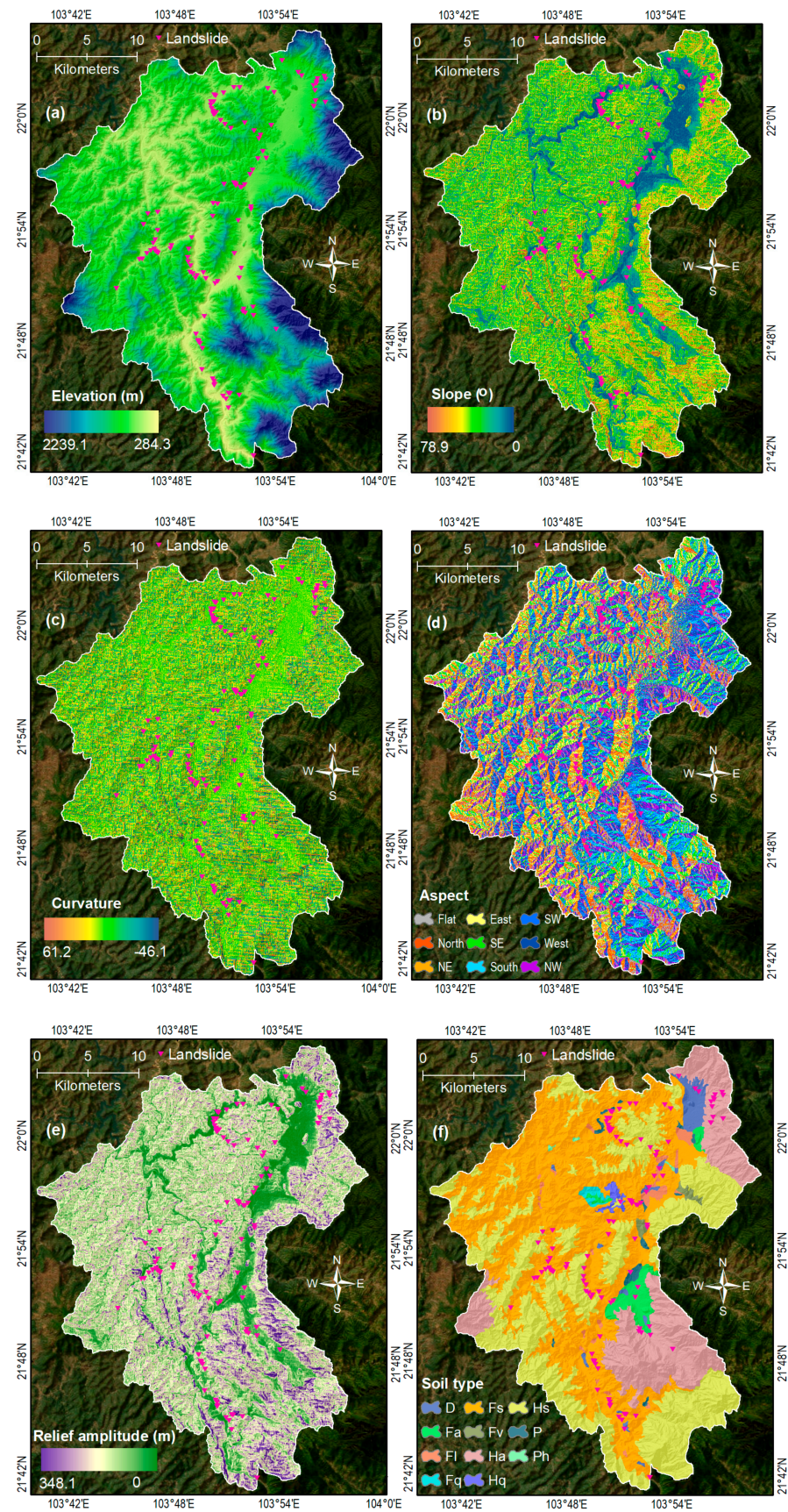


Figure 3. Cont.

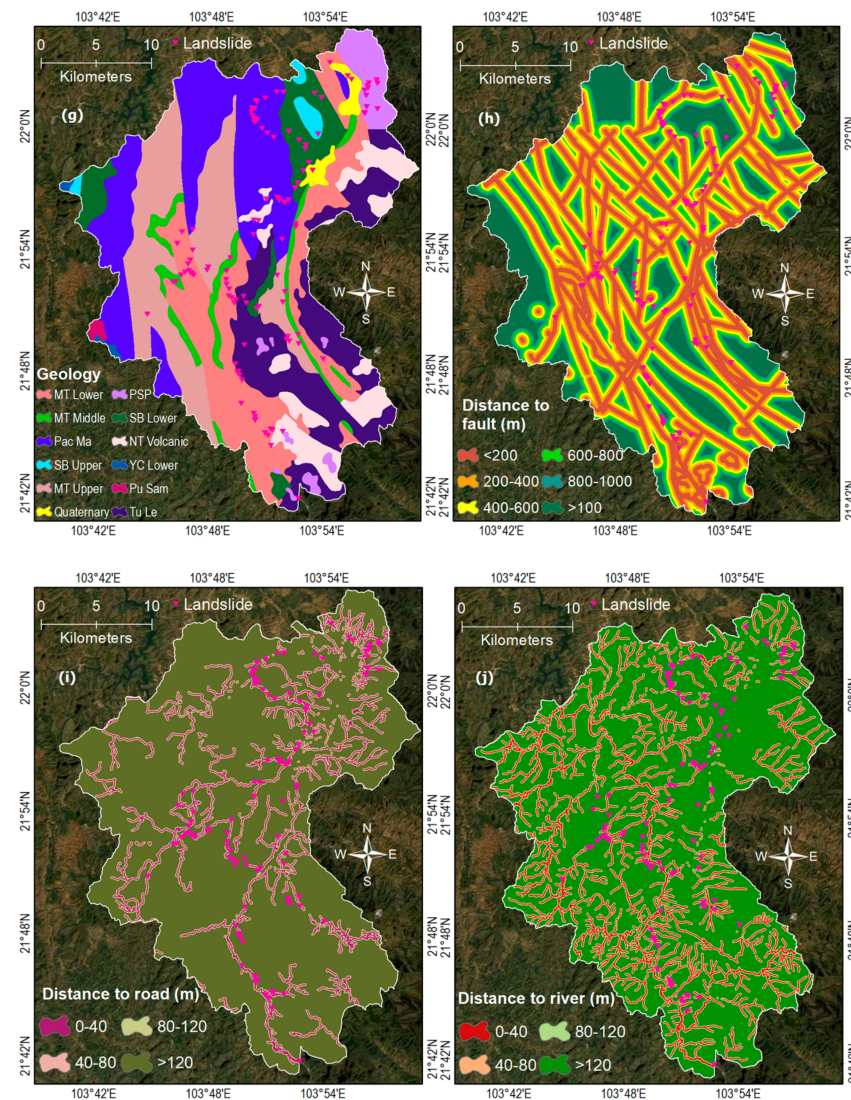


Figure 3. Landslide influencing factors: (a) elevation; (b) slope; (c) curvature; (d) aspect; (e) relief amplitude; (f) soil type; (g) geology; (h) distance to fault; (i) distance to road; and (j) distance to river.

Soil type is an important factor in landslide susceptibility mapping because it has a significant influence on soil properties such as permeability, shear strength, and density, which affect the soil's ability to hold and drain water [56]. The soil type map used in this study was derived from the national pedology map with a scale of 1:100,000, and it consisted of 11 soil types, including multi-origin diluvial soil (D), yellow–red soil on granite (Fa), yellow–red soil changed by cultivation (F1), pale yellow soil on sandstone (Fq), yellow–red soil on clay and metamorphic rocks (Fs), red–brown soil on limestone (Fv), yellow–red humus soil on granite (Ha), pale yellow humus on sandstone (Hq), yellow–red humus soil on clay rocks (Hs), alluvial soil (P), and stream alluvial soil (Ph). Figure 3f presents the distribution of soil types in the study area.

The geology setting should be considered in landslide studies because it plays a critical role in determining the type and structure of the underlying rock formations, which can significantly affect the stability of slopes [57]. The geology map (Figure 3g) of the study area was extracted from the national geological map and shows 12 geological units present in the study area, including the Muong Trai formation (MT) consisting of MT Lower, MT Middle, and MT Upper in the Middle Triassic and Late Ladinian, Pac Ma limestone in the Late Triassic and Late Carnian, Suoi Bang (SB) formation composed of SB Upper and SB Lower in the Late Triassic and Norian-Rhaetian, Quaternary sediment, Phu Sa Phin (PSP)

complex in the Late Mesozoic and Early Cenozoic, Ngoi Thia (NT volcanic) formation in the Cretaceous, Yen Chau (YC Lower) formation in the Late Cretaceous, Pu Sam Cap (PS) complex in the Paleogen, and Tu Le (TL) formation in the Cretaceous.

Faults are fractures or breaks in the Earth's crust along which movement occurs. Tectonic forces, such as the movement of tectonic plates, or other geological processes such as folding or shearing, can cause faults, and they can significantly impact slope stability [58]. To assess the potential influence of faults on landslide occurrence in the study area, we generated a distance to faults map (Figure 3h) using the geological and mineral resources map of Vietnam at a scale of 1:200,000. The buffer tool in ArcGIS Pro software was used to categorize the map into six distance buffers: 0–200 m, 200–400 m, 400–600 m, 600–800 m, 800–1000 m, and >1000 m.

Distance to roads is an important anthropogenic factor that can significantly impact landslides, as road cuts and embankments can cause slope instability [59]. In the study area, field investigations have revealed that many landslides have been triggered by road construction activities [45,46]. To generate a distance to road map for the study area, we first extracted the road network from the national topographic maps of Vietnam at a scale of 1:50,000. Next, using the Buffer tool in ArcGIS Pro software, we created a distance to road map (Figure 3i) with four buffer categories: 0–40 m, 40–80 m, 80–120 m, and >120 m.

Rivers can play a significant role in soil and rock saturation on slopes, which can increase the susceptibility of these slopes to instability and landslides. When rivers flow over or near slopes, they can deposit water into the soil and rock material, increasing its weight and reducing its shear strength. This, in turn, decreases the slope's stability and increases the risk of landslides [60,61]. Therefore, in this study, the distance to rivers was also included as a factor influencing landslides. The river network was extracted from the topographic map with a scale of 1:10,000 to create the distance to river map (Figure 3j). The buffer tool in ArcGIS Pro software was used to create four categories based on the distance from the river: 0–40 m, 40–80 m, 80–120 m, and >120 m.

4. The Proposed Hybrid BBO-DE Optimized SPAARC Tree Ensemble for Landslide Susceptibility Mapping

The flowchart for the proposed BBO-DE-STreeEns method for landslide susceptibility mapping is illustrated in Figure 4. To derive the landslide inventory and influencing factors, multisource geospatial data were processed using ArcGIS Pro 2.8 and stored in a landslide database in file geodatabase format. The influencing factors were converted to a 20×20 m grid cell and normalized within the range of [0.001–0.999]. The BBO-DE-STreeEns model was implemented by the authors in Matlab R2022a. The STreeEns model can be accessed through the Python Weka Wrapper API [15], while the Matlab code for the BBO-DE model can be found in [23].

4.1. Building the Landslide Database

In this study, 114 landslide locations were randomly divided into two subsets: a training set comprising 80 cells, which accounts for 70% of the total, and a validation set comprising 34 cells. Each of these landslide cells was labeled with a value of 1. To avoid potential bias caused by uneven proportions of landslide and non-landslide data, an equal number of grid cells were randomly sampled from areas without landslides. These non-landslide cells were labeled with a value of 0 and added to the training and validation sets. Therefore, both the training and validation sets contained 160 and 68 samples, respectively, with an equal number of landslide and non-landslide pixels in each. The training set was utilized to train the landslide models, while the validation set was used for model validation.

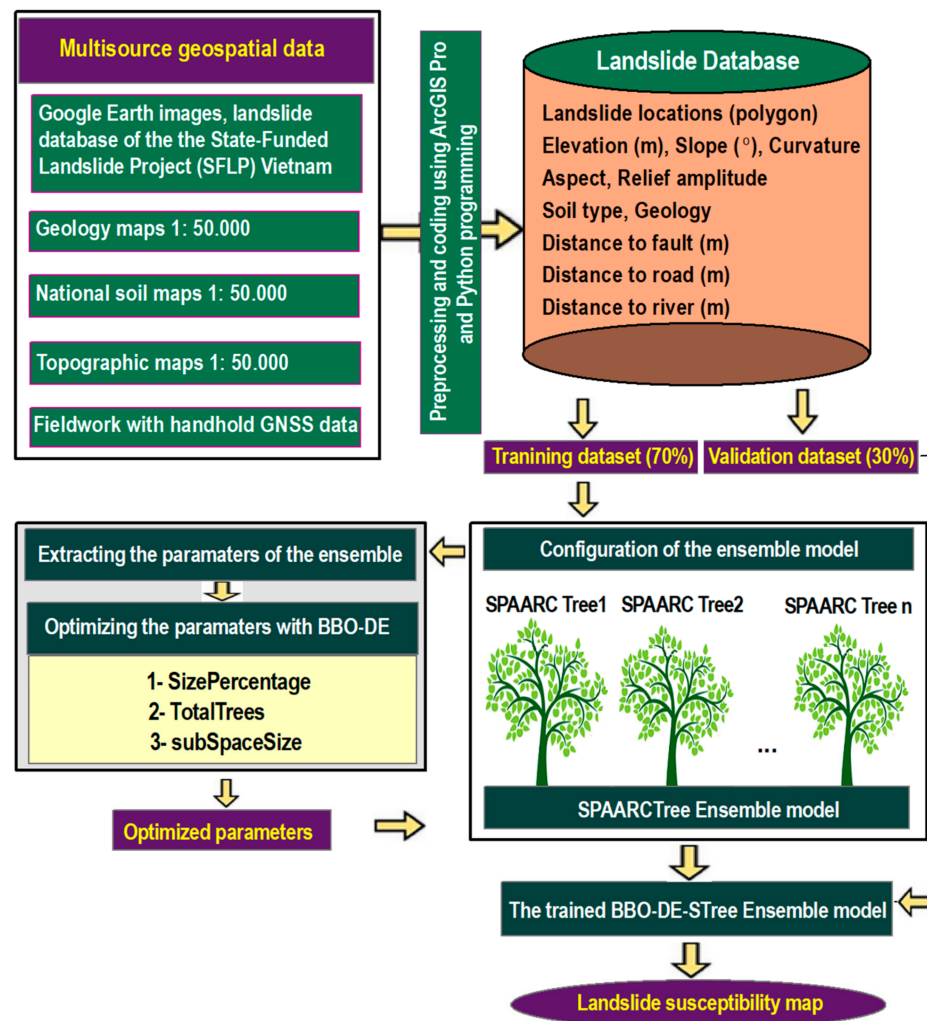


Figure 4. The flowchart of the proposed BBO-DE-STreeEns for landslide susceptible mapping.

To build the landslide database, the values of the ten influencing factors of all the pixels were extracted. The database comprises a dependent variable (label) and ten independent variables. Among the independent variables, six are categorical (aspect, soil type, geology, distance to road, distance to river, and distance to fault) and four are continuous (elevation, slope, curvature, and relief amplitude).

4.2. Cost Function and Hyperparameter Optimization

In order to achieve the best performance, the BBO-DE-STreeEns model effectiveness is dependent on the careful selection of three hyperparameters: TotalTrees, SizePercentage, and subSpaceSize. TotalTrees represents the number of SPAARC decision trees in the model, while SizePercentage indicates the proportion of data in subbagging, and subSpaceSize determines the proportion of the attribute space in random subsampling. In this study, to optimize the hyperparameters, the hybrid (algorithm-level) BBO-DE optimization was utilized, and the mean absolute error (MAE) was employed as the cost function (Equation (1)).

$$MAE = \frac{1}{n} \sum_{i=1}^n |LS_i - \hat{L}\hat{S}_i| \quad (1)$$

where LS_i represents the predicted landslide susceptibility (LS) value for the i -th sample, and $\hat{L}\hat{S}_i$ is the corresponding actual LS value, and n is the total number of samples in the training dataset.

In this study, a three-dimensional search space was defined for the hybrid BBO-DE algorithm. The first dimension is *TotalTrees*, with values ranging from 1 to 2000. The second dimension is *SizePercentage*, with values ranging from 0.1 to 1.0. The third dimension is *subSpaceSize*, with values ranging from 0.3 to 0.9.

4.3. Performance Assessment

It is customary practice to assess model performance using both training and validation datasets [62–64]. The performance of the model on the training dataset indicates how well the model is able to fit the data, while the performance on the validation dataset reflects its predictive capability.

In this study, we utilized multiple performance metrics including the receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) to evaluate the classification capacity of the landslide models. These metrics were employed to compare the performances of the different models under consideration. The ROC curve and AUC provide an overall measure of the model's predictive accuracy, while other performance metrics offer insights into the model's predictive capability.

The problem addressed in this study is binary classification of landslide samples. True positive (TP) represents the number of correctly predicted landslide samples, while false negative (FN) refers to the number of landslide samples that were predicted incorrectly. True negative (TN) is the number of non-landslide samples that are correctly predicted, and false positive (FP) is the number of non-landslide samples that are incorrectly predicted [65].

To evaluate the performance of the models, several performance metrics are computed, including positive predictive value (PPV), negative predictive value (NPV), sensitivity (Sen) (also known as true positive rate (TPR)), specificity (Spe), false positive rate (FPR), accuracy (Acc), F1 Score (Fscore), and Cohen's Kappa coefficient (Kappa). These metrics are calculated as follows [65]:

$$PPV = \frac{TP}{TP + FP}; NPV = \frac{TN}{TN + FN}; Sen = \frac{TP}{TP + FN}; Spe = \frac{TN}{TN + FP}; \quad (2)$$

$$FPR = \frac{FP}{FP + TN}; Acc = \frac{TP + TN}{TP + TN + FP + FN}; Fscore = \frac{2 \times TP}{2 \times TP + FP + FN}; \quad (3)$$

$$Kappa = \frac{2 \times (TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)} \quad (4)$$

The ROC curve is a graphical representation of the performance of a classifier at different classification thresholds. It plots the FPR on the *x*-axis and the TPR on the *y*-axis [65]. The AUC measures the overall performance of the classifier across all possible classification thresholds. A higher AUC value indicates better performance of the model, with values ranging from 0.5 to 1.0 [65]. In the context of landslide modeling, AUC is considered a standard technique for evaluating the overall performance of models. According to Peterson et al. [66], an AUC value between 0.5 and 0.6 indicates very poor performance, while values between 0.6 and 0.7, 0.7 and 0.8, 0.8 and 0.9, and 0.9 and 1.0 correspond to poor, moderate, good, and very good performance, respectively. Therefore, AUC provides a valuable measure of the model's predictive accuracy and can help to identify the threshold at which the model performs optimally.

4.4. Benchmark Models and Comparison

To assess the performance of the proposed BBO-DE-StrreeEns model and establish its merit, we compared it against four benchmark models: logistic regression (LRegr), Multi-layer perceptron neural network (MLPNeuNet), support vector machine (SVM), and SPAARC. For the MLPNeuNet model, we selected the network structure with one input layer, one hidden layer with eight neurons, and one output layer, as it demonstrated the best performance through a trial-and-error test previously described in [67]. The SVM model utilized the radial basis function kernel, and we determined the optimal values

for the C and Gamma parameters to be 0.9 and 0.185, respectively, through a grid search method outlined in Fayed and Atiya (2019). As for the SPAARC model, we utilized the default parameters, which included a minimum number of samples at the terminal node of 2 and a training size percentage of 1.0. By comparing the performance of the BBO-DE-StreeEns model with these benchmark models, we were able to validate its effectiveness and establish its superiority over the models.

To determine whether the performance of the proposed BBO-DE-StreeEns model and the benchmark models differed significantly, we used the Wilcoxon signed-rank test, which offers a pairwise comparison of landslide models [68]. The null hypothesis states that there is no difference between the two landslide models. The significance level for rejecting the null hypothesis is $\alpha = 0.05$. To assess the significance of the difference between a pair of landslide models, we calculate the p -value and z -value. If the p -value is less than the significance level and the z -value is greater than 1.96 or less than -1.96 , then we reject the null hypothesis, and the difference is considered statistically significant [69].

5. Results and Analysis

5.1. Model Results and Assessment

The five landslide susceptibility models, BBO-DE-StreeEns, LRegr, MLPNeuNet, SVM, and SPAARC, were successfully trained using the training dataset with ten-fold cross-validation.

Table 1 shows the results of training the five landslide susceptibility models using the training dataset and ten-fold cross-validation to mitigate the risk of overfitting. The hyperparameter values for the BBO-DE-StreeEns model were optimized to TotalTrees = 30, SizePercentage = 0.9, and subSpaceSize = 0.5. All models performed well with the training data, but the BBO-DE-StreeEns (AUC = 0.987, Kappa = 0.875, Fscore = 0.939, and Acc = 93.8) and SPAARC (AUC = 0.950, Kappa = 0.875, Fscore = 0.940, and Acc = 93.8) models achieved the best performance. The LRegr, MLPNeuNet, and SVM models had similar performances (Table 1).

Table 1. Performance metrics of the proposed BBO-DE-StreeEns model and the benchmarks on the training dataset.

Model	Performance Metrics											
	TP	TN	FN	FP	PPV (%)	NPV (%)	Sen (%)	Spe (%)	Acc (%)	Fscore	Kappa	AUC
BBO-DE-StreeEns	77	73	3	7	91.7	96.1	96.3	91.3	93.8	0.939	0.875	0.987
LRegr	60	63	20	17	77.9	75.9	75.0	78.8	76.9	0.764	0.538	0.855
MLPNeuNet	61	63	19	17	78.2	76.8	76.3	78.8	77.5	0.772	0.550	0.859
SPAARC	78	72	2	8	90.7	97.3	97.5	90.0	93.8	0.940	0.875	0.950
SVM	57	67	23	13	81.4	74.4	71.3	83.8	77.5	0.760	0.550	0.855

Assessing the prediction capability of landslide models is crucial to determine their effectiveness. To this end, the validation dataset was used, and the results are presented in Table 2. The performance of the five models shows a degree of variation, with the BBO-DE-StreeEns model (AUC = 0.940, Kappa = 0.735, Fscore = 0.862, and Acc = 86.8) and the SPAARC model (AUC = 0.915, Kappa = 0.676, Fscore = 0.836, and Acc = 83.5) demonstrating the highest prediction capabilities. These models achieved excellent statistical metrics, indicating their superior performance. The LRegr model (AUC = 0.853, Kappa = 0.539, Fscore = 0.750, and Acc = 76.5) performed well and followed behind. The SVM model (AUC = 0.767, Kappa = 0.529, Fscore = 0.750, and Acc = 76.5) and the MLPNeuNet model (AUC = 0.748, Kappa = 0.294, Fscore = 0.684, and Acc = 64.7) also exhibited satisfactory prediction capability (Table 2).

Table 2. Prediction performance of the proposed BBO-DE-StreeEns model and the benchmarks on the validation set.

Model	Prediction Metrics											
	TP	TN	FN	FP	PPV (%)	NPV (%)	Sen (%)	Spe (%)	Acc (%)	Fscore	Kappa	AUC
BBO-DE-STreeEns	28	31	6	3	90.3	83.8	82.4	91.2	86.8	0.862	0.735	0.940
LRegr	24	28	10	6	80.0	73.7	70.6	82.4	76.5	0.750	0.529	0.853
MLPNeuNet	26	18	8	16	61.9	69.2	76.5	52.9	64.7	0.684	0.294	0.748
SPAARC	28	29	6	5	84.8	82.9	82.4	85.3	83.8	0.836	0.676	0.915
SVM	24	28	10	6	80.0	73.7	70.6	82.4	76.5	0.750	0.529	0.767

To determine whether the BBO-DE-StreeEns model has a statistically significant difference in prediction performance compared to other models, a Wilcoxon signed rank test was conducted and the results are presented in Table 3. The test was performed on 10 pairs of models. All pairs, except for LRegr vs. SVM (p -value = 0.716 and z -value = 0.364), showed a significant difference in prediction performance. The z -values for the pairs were found to be greater than the standard value of 1.96, indicating a significant difference in the prediction performance of the models. Additionally, the p -values for the pairs were less than 0.05, further confirming the statistical significance of the results. These findings confirm that the BBO-DE-StreeEns model has the best prediction power in this study.

Table 3. Statistical tests of the proposed BBO-DE-STreeEns model and the other models.

No.	Pairwise Comparison	z -Value	p -Value	Significance
1	BBO-DE-STreeEns vs. LRegr	2.528	0.011	Yes
2	BBO-DE-STreeEns vs. MLPNeuNet	3.962	<0.001	Yes
3	BBO-DE-STreeEns vs. SPAARC	5.719	<0.001	Yes
4	BBO-DE-STreeEns vs. SVM	4.740	<0.001	Yes
5	LRegr vs. MLPNeuNet	5.635	<0.001	Yes
6	LRegr vs. SPAARC	2.719	0.005	Yes
7	LRegr vs. SVM	0.364	0.716	No
8	MLPNeuNet vs. SPAARC	4.175	<0.001	Yes
9	MLPNeuNet vs. SVM	3.919	<0.001	Yes
10	SPAARC vs. SVM	7.253	<0.001	Yes

5.2. The Role of the Landslide Influencing Factors

To determine the contribution of ten landslide influencing factors to the BBO-DE-StreeEns model, we employed the wrapper algorithm [70], using five-fold cross-validations to avoid potential bias [71]. The results are presented in Table 4 and Figure 5. Our analysis revealed that slope has the highest role (score value = 0.299), followed by distance to road (score value = 0.224) and elevation (score value = 0.142). The remaining factors had a lower contribution to the BBO-DE-StreeEns model, with score values ranging from 0.026 (distance to river) to 0.084 (distance to fault) (Table 4).

Table 4. The role of the ten influencing factors.

No.	Ranking	Score Value
1	Slope	0.299
2	Distance to Road	0.224
3	Elevation	0.142
4	Distance to Fault	0.084
5	Relief Amplitude	0.063
6	Soil Type	0.049
7	Geology	0.047
8	Curvature	0.036
9	Aspect	0.029
10	Distance to River	0.026

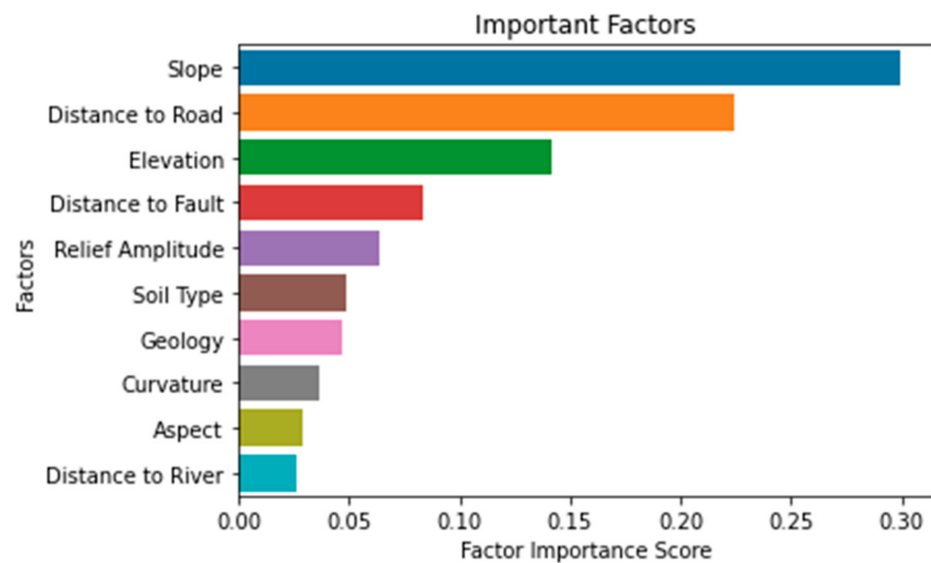


Figure 5. The role of the influencing factors.

5.3. Landslide Susceptibility Map

Given the superior performance of the BBO-DE-StreeEns model on the landslide data, it was employed to calculate the landslide susceptibility index for each pixel of Than Uyen district. The index values ranged from a minimum of 0.062 to a maximum of 0.910. The resulting output was then exported to the landslide geodatabase as described in Section 4.1.

To create a landslide susceptibility map, it is common practice to divide the map into four categories [72,73] based on the susceptibility levels: very high, high, moderate, and low. In this study, the boundaries between these categories were established by analyzing a graph presented in Figure 6. The graph was constructed by plotting the percentage of landslides against the percentage of the susceptibility map, which follows the method described in [74]. Based on the analysis of the graph, the resulting categories were assigned as follows: 10% for the low category, 20% for each of the moderate and high categories, and 50% for the very high category, covering the entire study area. The threshold values were then determined in descending order at 0.737, 0.674, and 0.502 (Figure 6). The final landslide susceptibility map for Than Uyen district, using the BBO-DE-StreeEns model, was generated based on these threshold values, and is presented in Figure 7.

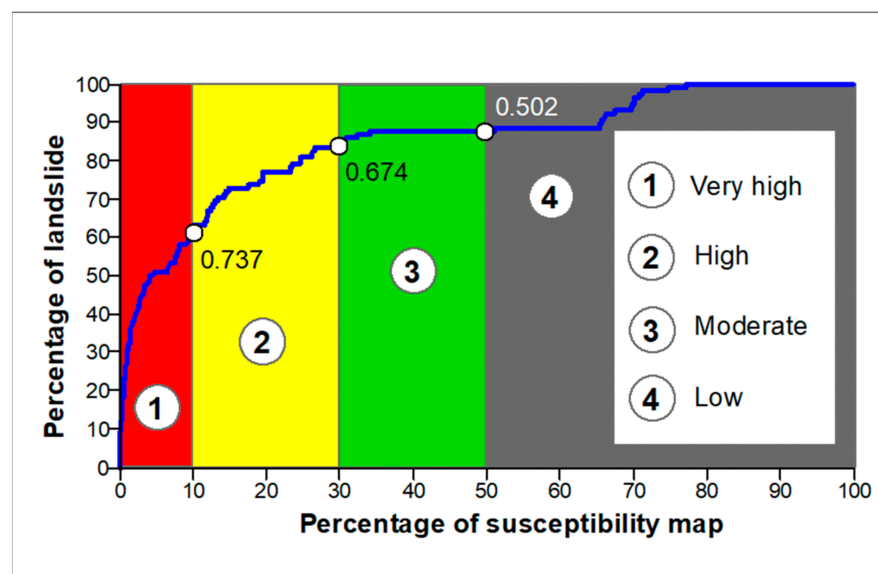


Figure 6. Percentage of landslides vs. percentage of susceptibility map for Than Uyen district.

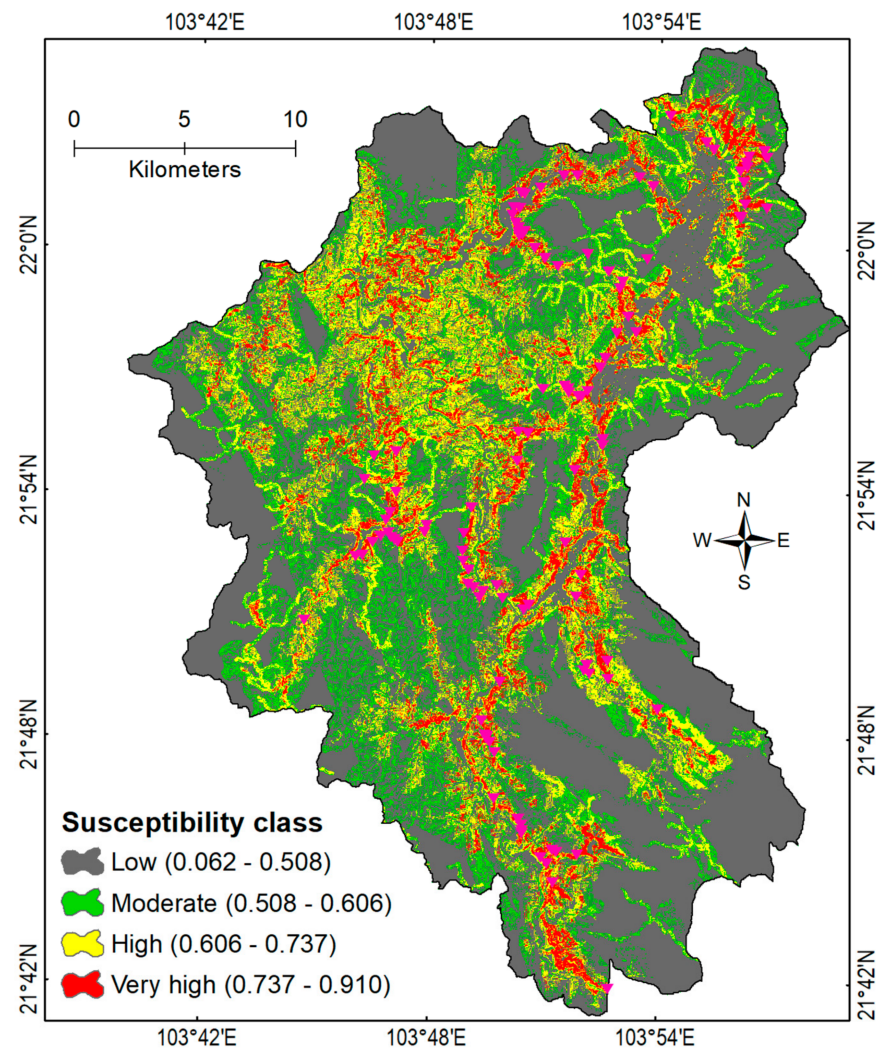


Figure 7. The landslide susceptibility map for Than Uyen district using the BBO-DE-STreeEns.

Table 5 displays the characteristics of the landslide susceptibility map divided into four categories. The results indicate that the very high, high, moderate, and low susceptibility categories cover 78.9 km², 157.8 km², 157.8 km², and 394.5 km², respectively. It is worth noting that the very high and high susceptibility categories together account for 83.33% of the total landslide locations, highlighting their significance in terms of landslide risk in the area.

Table 5. Characteristics of the four susceptibility categories of the BBO-DE-StrreeEns model.

No	Susceptibility Index	Landslide Location (%)	Verbal Description	Susceptibility Map (%)	Areas (km ²)
1	0.062–0.508	12.28	Low	50.00	394.5
2	0.508–0.606	4.39	Moderate	20.00	157.8
3	0.606–0.737	20.17	High	20.00	157.8
4	0.737–0.910	63.16	Very High	10.00	78.9

6. Discussion

Landslides are a highly destructive natural hazard that claims thousands of lives and causes economic losses estimated at USD 20 billion annually [2]. Improper land use planning and climate change are exacerbating landslide occurrences in mountainous regions across the globe [75]. Thus, accurate landslide susceptibility prediction models are

critical to mitigate its impacts. In this study, we developed and validated a novel ensemble machine learning model, the BBO-DE-StreeEns, for landslide susceptibility mapping in Than Uyen district, located in the northwest mountainous area of Vietnam, where landslides and floods are recurring problems.

The BBO-DE-StreeEns model leverages the SPAARC Tree algorithm to build trees through subbagging and random subsampling, while the hybrid BBO-DE algorithm optimizes the model's hyperparameters. Decision tree algorithms, especially when integrated into ensemble models, have been shown to be highly efficient in various spatial domains, including landslide prediction [76]. The superior predictive power of the BBO-DE-StreeEns model in this study validates this assertion.

Although the performance of the BBO-DE-StreeEns model was heavily influenced by the hyperparameters TotalTrees, SizePercentage, and subSpaceSize, there are no established guidelines on the best values for them. Therefore, the successful search for and optimization of these hyperparameters by the hybrid BBO-DE algorithm indicate its effectiveness.

The BBO-DE-StreeEns model outperformed benchmarks such as LRegr, MLPNeuNet, SVM, and SPAARC, confirming the effectiveness of combining BBO-DE, SPAARC Tree, subbagging, and random subsampling for landslide susceptibility mapping.

Of the ten landslide factors considered, slope and distance to roads emerged as the most critical factors. This finding is reasonable given that Than Uyen is a mountainous district where slope areas comprise over 90% of the total area. Landslides in this district mainly occur on slopes between 16 and 34 degrees, with many occurring near road systems. The road sections cutting through slopes are significant contributors to landslide failures.

The landslide samples in this study are predominantly distributed along roads, which may introduce incompleteness bias in the landslide inventory data. This type of bias is a common issue in landslide inventory data that can significantly impact the accuracy and reliability of landslide susceptibility models, particularly for large study areas [77–79]. To address the potential impact of incompleteness bias in landslide susceptibility modeling, various strategies have been proposed in the literature. For example, incorporating multi-source data such as field surveys, remote sensing, and historical records can help generate a more comprehensive landslide inventory, as conducted in this study. Another approach involves incorporating random effect variables in statistical models to account for the potential incompleteness bias in the inventory data [77–79]. Additionally, sensitivity analysis can be used to evaluate the robustness of modeling results to different levels of completeness in the inventory data [77]. Furthermore, a combination of non-spatial and spatial cross-validation techniques can help assess the model performance in different parts of the study area and detect the potential impact of incompleteness bias in the landslide inventory data. Comparing the results of non-spatial and spatial cross-validation can identify areas where the model is particularly sensitive to the completeness of the inventory data [77–79].

Several studies have incorporated rainfall-related factors into landslide modeling to account for the impact of rainfall on landslide occurrences [80–82]. These models have used projections of changes in the rainfall regime under various climate change scenarios to investigate how landslide susceptibility may evolve over time. However, it is crucial to acknowledge that uncertainties in rainfall predictions can potentially affect the reliability of landslide susceptibility predictions [81]. Regrettably, in the present study, we were unable to incorporate rainfall as an influencing factor in the landslide models due to the unavailability of accurate and detailed rainfall data for the study area.

7. Conclusions

In this study, we proposed a new ensemble model, the BBO-DE-StreeEns, for landslide susceptibility mapping. The model was developed and evaluated using a landslide database from Than Uyen district, which consisted of 114 landslide locations, 114 non-landslide ones, and ten influencing factors. In addition, we compared the proposed model

with four benchmark models, namely, LRegr, MLPNeuNet, SVM, and SPAARC, to assess its performance. Based on our findings, we draw the following conclusions:

- The combination of BBO-DE, SPAARC, subbagging, and random subsampling formed a powerful new ensemble model for accurate landslide susceptibility mapping.
- The BBO-DE-StrreeEns model demonstrated superior performance compared to benchmark models such as LRegr, MLPNeuNet, SVM, and SPAARC. This highlights its potential as a highly accurate solution for landslide susceptibility mapping.
- Ten landslide influencing factors, namely, elevation, slope, curvature, aspect, relief amplitude, soil type, geology, distance to faults, distance to roads, and distance to rivers, were selected based on the analysis of the landslide inventory and the geo-environmental characteristics of the study area. As all these factors had a score value of importance greater than zero and the landslide model performed well, these factors are all significant in predicting landslide occurrence in the study area.
- Among the ten factors considered, slope and distance to roads were identified as the most significant factors contributing to landslide occurrences in Than Uyen district.
- The landslide susceptibility map generated by our study provides valuable information for authorities and policymakers in Than Uyen district for land-use planning and territory management decision-making.

Author Contributions: Conceptualization, D.A.H., H.V.L. and D.T.B.; methodology, D.A.H., H.V.L. and D.T.B.; software, D.A.H., H.V.L., D.T.B. and D.V.P.; validation, H.V.L. and D.T.B.; resources, P.V.H.; data curation, D.A.H., H.V.L. and D.T.B.; writing—original draft preparation, D.A.H., H.V.L. and D.T.B.; writing—review and editing, H.V.L. and D.T.B.; visualization, D.T.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the Vietnam Ministry of Education and Training under grant number CT.2019.01.02.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

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

References

1. Petrucci, O. Landslide Fatality Occurrence: A Systematic Review of Research Published between January 2010 and March 2022. *Sustainability* **2022**, *14*, 9346. [\[CrossRef\]](#)
2. Sim, K.B.; Lee, M.L.; Wong, S.Y. A review of landslide acceptable risk and tolerable risk. *Geoenvironmental Disasters* **2022**, *9*, 3. [\[CrossRef\]](#)
3. Thirugnanam, H.; Uhlemann, S.; Reghunadh, R.; Ramesh, M.V.; Rangan, V.P. Review of landslide monitoring techniques with IoT integration opportunities. *IEEE J. Se-Lected Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5317–5338. [\[CrossRef\]](#)
4. Araújo, J.R.; Ramos, A.M.; Soares, P.M.M.; Melo, R.; Oliveira, S.C.; Trigo, R.M. Impact of extreme rainfall events on landslide activity in Portugal under climate change scenarios. *Landslides* **2022**, *19*, 2279–2293. [\[CrossRef\]](#)
5. Marc, O.; Gosset, M.; Saito, H.; Uchida, T.; Malet, J.-P. Spatial Patterns of Storm-Induced Landslides and Their Relation to Rainfall Anomaly Maps. *Geophys. Res. Lett.* **2019**, *46*, 11167–11177. [\[CrossRef\]](#)
6. Bozzolan, E.; Holcombe, E.A.; Pianosi, F.; Marchesini, I.; Alvioli, M.; Wagener, T. A mechanistic approach to include climate change and unplanned urban sprawl in landslide susceptibility maps. *Sci. Total Environ.* **2023**, *858*, 159412. [\[CrossRef\]](#)
7. He, Q.; Jiang, Z.; Wang, M.; Liu, K. Landslide and Wildfire Susceptibility Assessment in Southeast Asia Using Ensemble Machine Learning Methods. *Remote Sens.* **2021**, *13*, 1572. [\[CrossRef\]](#)
8. Liu, L.-L.; Zhang, J.; Li, J.-Z.; Huang, F.; Wang, L.-C. A bibliometric analysis of the landslide susceptibility research (1999–2021). *Geocarto Int.* **2022**, *37*, 14309–14334. [\[CrossRef\]](#)
9. Liu, S.; Wang, L.; Zhang, W.; He, Y.; Pijush, S. A comprehensive review of machine learning-based methods in landslide susceptibility mapping. *Geol. J.* **2023**. [\[CrossRef\]](#)
10. Merghadi, A.; Yunus, A.P.; Dou, J.; Whiteley, J.; ThaiPham, B.; Bui, D.T.; Avtar, R.; Abderrahmane, B. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth Sci. Rev.* **2020**, *207*, 103225. [\[CrossRef\]](#)
11. Martínez-Álvarez, F.; Bui, D.T. Advanced Machine Learning and Big Data Analytics in Remote Sensing for Natural Hazards Management. *Remote Sens.* **2020**, *12*, 301. [\[CrossRef\]](#)

12. Yaghoubzadeh-Bavandpour, A.; Bozorg-Haddad, O.; Zolghadr-Asli, B.; Martínez-Álvarez, F. Deep Learning Application in Water and Environmental Sciences. In *Computational Intelligence for Water and Environmental Sciences*; Springer: Singapore, 2022; pp. 273–290.
13. Pourghasemi, H.R.; Sadhasivam, N.; Amiri, M.; Eskandari, S.; Santosh, M. Landslide susceptibility assessment and mapping using state-of-the art machine learning techniques. *Nat. Hazards* **2021**, *108*, 1291–1316. [[CrossRef](#)]
14. Hong, H. Assessing landslide susceptibility based on hybrid Best-first decision tree with ensemble learning model. *Ecol. Indic.* **2023**, *147*, 109968. [[CrossRef](#)]
15. Beckham, C.; Hall, M.; Frank, E. WekaPyScript: Classification, Regression, and Filter Schemes for WEKA Implemented in Python. *J. Open Res. Softw.* **2016**, *4*, e33. [[CrossRef](#)]
16. Gundersen, O.E.; Shamsaliei, S.; Isdahl, R.J. Do machine learning platforms provide out-of-the-box reproducibility? *Futur. Gener. Comput. Syst.* **2021**, *126*, 34–47. [[CrossRef](#)]
17. Zenodo. TensorFlow. Available online: <https://zenodo.org/record/7604226#.ZD6-YXZBw2w> (accessed on 17 November 2022).
18. Kavzoglu, T.; Colkesen, I.; Sahin, E.K. Machine learning techniques in landslide susceptibility mapping: A survey and a case study. In *Landslides: Theory, Practice and Modelling*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 283–301.
19. Pourghasemi, H.R.; Pouyan, S.; Bordbar, M.; Golkar, F.; Clague, J.J. Flood, landslides, forest fire, and earthquake susceptibility maps using machine learning techniques and their combination. *Nat. Hazards* **2023**, *116*, 3797–3816. [[CrossRef](#)]
20. Yousef, A.M.; Mahdi, A.M.; Pourghasemi, H.R. Landslides and flood multi-hazard assessment using machine learning techniques. *Bull. Eng. Geol. Environ.* **2022**, *81*, 1–23. [[CrossRef](#)]
21. Fang, Z.; Wang, Y.; Peng, L.; Hong, H. A comparative study of heterogeneous ensemble-learning techniques for landslide susceptibility mapping. *Int. J. Geogr. Inf. Sci.* **2020**, *35*, 321–347. [[CrossRef](#)]
22. Di Napoli, M.; Carotenuto, F.; Cevasco, A.; Confuorto, P.; Di Martire, D.; Firpo, M.; Pepe, G.; Raso, E.; Calcaterra, D. Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. *Landslides* **2020**, *17*, 1897–1914. [[CrossRef](#)]
23. Chen, X.; Tianfield, H.; Du, W.; Liu, G. Biogeography-based optimization with covariance matrix based migration. *Appl. Soft Comput.* **2016**, *45*, 71–85. [[CrossRef](#)]
24. Fang, Z.; Wang, Y.; Duan, G.; Peng, L. Landslide Susceptibility Mapping Using Rotation Forest Ensemble Technique with Different Decision Trees in the Three Gorges Reservoir Area, China. *Remote Sens.* **2021**, *13*, 238. [[CrossRef](#)]
25. Russell, S.J. *Artificial Intelligence: A Modern Approach*, 4th ed.; Pearson: Hoboken, NJ, USA, 2021.
26. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
27. Loh, W.Y. Classification and regression trees. *WIREs Data Min. Knowl. Discov.* **2011**, *1*, 14–23. [[CrossRef](#)]
28. Yates, D.; Islam, M.Z.; Gao, J. SPAARC: A Fast Decision Tree Algorithm. In *Australasian Conference on Data Mining*; Springer: Brisbane, QLD, Australia, 2018.
29. Chen, J.; Li, K.; Tang, Z.; Bilal, K.; Yu, S.; Weng, C.; Li, K. A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment. *IEEE Trans. Parallel Distrib. Syst.* **2016**, *28*, 919–933. [[CrossRef](#)]
30. Yates, D.; Islam, M.Z. FastForest: Increasing random forest processing speed while maintaining accuracy. *Inf. Sci.* **2021**, *557*, 130–152. [[CrossRef](#)]
31. Latinne, P.; Debeir, O.; Decaestecker, C. Limiting the number of trees in random forests. In *International Workshop on Multiple Classifier Systems*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 178–187.
32. Oshiro, T.M.; Perez, P.S.; Baranauskas, J.A. How many trees in a random forest? In *International Workshop on Machine Learning and Data Mining in Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 154–168.
33. Asuncion, A.; Newman, D. *UCI Machine Learning Repository*; University of California, School of Information and Computer Science: Irvine, CA, USA; Available online: <http://archive.ics.uci.edu/ml/index.php> (accessed on 8 December 2021).
34. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844.
35. Simon, D. Biogeography-based optimization. *IEEE Trans. Evol. Comput.* **2008**, *12*, 702–713. [[CrossRef](#)]
36. Storn, R.; Price, K. *Differential Evolution—A Simple and Efficient Adaptive Scheme for Global Optimization Over Continuous Spaces*; International Science Computer Institute: Berkeley, CA, USA, 1995.
37. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The weka data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [[CrossRef](#)]
38. Dragoi, E.-N.; Dafinescu, V. Parameter control and hybridization techniques in differential evolution: A survey. *Artif. Intell. Rev.* **2015**, *45*, 447–470. [[CrossRef](#)]
39. Das, S. Differential evolution a survey of the state-of-the-art. *IEEE Trans. Evol. Comput.* **2011**, *15*, 4–31. [[CrossRef](#)]
40. Noman, N.; Iba, H. Accelerating Differential Evolution Using an Adaptive Local Search. *IEEE Trans. Evol. Comput.* **2008**, *12*, 107–125. [[CrossRef](#)]
41. Gong, W.; Cai, Z.; Ling, C.X. DE/BBO: A hybrid differential evolution with biogeography-based optimization for global numerical optimization. *Soft Comput.* **2011**, *15*, 645–665. [[CrossRef](#)]
42. Ma, H.; Simon, D. *Evolutionary Computation with Biogeography-Based Optimization*; John Wiley & Sons: Hoboken, NJ, USA, 2017. [[CrossRef](#)]
43. Ma, H.; Simon, D.; Fei, M.; Shu, X.; Chen, Z. Hybrid biogeography-based evolutionary algorithms. *Eng. Appl. Artif. Intell.* **2014**, *30*, 213–224. [[CrossRef](#)]

44. Ma, H.; Simon, D.; Fei, M. On the Convergence of Biogeography-Based Optimization for Binary Problems. *Math. Probl. Eng.* **2014**, *2014*, 147457. [CrossRef]
45. Vietnam Institute of Geosciences and Mineral Resources. Landslide Warning Website. Available online: <http://www.canhbaotruotlo.vn/hientrangcactinh.html> (accessed on 8 December 2021).
46. Hung, L.Q.; Van, N.T.H.; Van Son, P.; 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*; ISDR-ICL Sendai Partnerships 2015–2025; Springer: Berlin/Heidelberg, Germany, 2017; Volume 1, pp. 501–510. [CrossRef]
47. General Statistic Office. *Statistical Yearbook of Vietnam*; General Statistic Office: Hanoi, Vietnam, 2018.
48. Dai, F.; Lee, C.; Ngai, Y. Landslide risk assessment and management: An overview. *Eng. Geol.* **2002**, *64*, 65–87. [CrossRef]
49. Bui, D.T.; 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]
50. Dai, F.C.; Lee, C.F. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* **2002**, *42*, 213–228. [CrossRef]
51. Freund, C.A.; Clark, K.E.; Curran, J.F.; Asner, G.P.; Silman, M.R. Landslide age, elevation and residual vegetation determine tropical montane forest canopy recovery and biomass accumulation after landslide disturbances in the Peruvian Andes. *J. Ecol.* **2021**, *109*, 3555–3571. [CrossRef]
52. Lee, C.F.; Li, J.; Xu, Z.W.; Dai, F.C. Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. *Environ. Geol.* **2001**, *40*, 381–391. [CrossRef]
53. Pike, R.J. The geometric signature: Quantifying landslide-terrain types from digital elevation models. *J. Int. Assoc. Math. Geol.* **1988**, *20*, 491–511. [CrossRef]
54. Magliulo, P.; Di Lisio, A.; Russo, F.; Zelano, A. Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: A case study in southern Italy. *Nat. Hazards* **2008**, *47*, 411–435. [CrossRef]
55. Vergari, F.; Della Seta, M.; Del Monte, M.; Fredi, P.; Lupia Palmieri, E. Landslide susceptibility assessment in the Upper Orcia Valley (Southern Tuscany, Italy) through conditional analysis: A contribution to the unbiased selection of causal factors. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 1475–1497. [CrossRef]
56. Luino, F.; De Graff, J.; Biddoccu, M.; Faccini, F.; Freppaz, M.; Roccati, A.; Ungaro, F.; D’amico, M.; Turconi, L. The Role of Soil Type in Triggering Shallow Landslides in the Alps (Lombardy, Northern Italy). *Land* **2022**, *11*, 1125. [CrossRef]
57. Kontoes, C.; Loupasakis, C.; Papoutsis, I.; Alatza, S.; Poyiadji, E.; Ganas, A.; Psychogyiou, C.; Kaskara, M.; Antoniadi, S.; Spanou, N. Landslide Susceptibility Mapping of Central and Western Greece, Combining NGI and WoE Methods, with Remote Sensing and Ground Truth Data. *Land* **2021**, *10*, 402. [CrossRef]
58. Varnes, D.J. *Landslide Hazard Zonation: A Review of Principles and Practice*; UNESCO: Paris, France, 1984.
59. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakudayahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31. [CrossRef]
60. Gökçeoglu, C.; Aksoy, H. Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques. *Eng. Geol.* **1996**, *44*, 147–161. [CrossRef]
61. Tien Bui, D.; Lofman, O.; Revhaug, I.; Dick, O. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Nat. Hazards* **2011**, *59*, 1413–1444. [CrossRef]
62. Le, H.V.; Bui, Q.T.; Tien Bui, D.; Tran, H.H.; Hoang, N.D. A Hybrid Intelligence System Based on Relevance Vector Machines and Imperialist Competitive Optimization for Modelling Forest Fire Danger Using GIS. *J. Environ. Inform.* **2020**, *36*, 43–57.
63. Le, H.V.; Hoang, D.A.; Tran, C.T.; Nguyen, P.Q.; Hoang, N.D.; Amiri, M.; Ngo, T.P.T.; Nhu, H.V.; Van Hoang, T.; Tien Bui, D. A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas. *Ecol. Inform.* **2021**, *63*, 101300. [CrossRef]
64. Tien Bui, D.; Le, H.V.; Hoang, N.-D. GIS-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method. *Ecol. Inform.* **2018**, *48*, 104–116. [CrossRef]
65. Powers, D.M.W. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.
66. Peterson, A.T.; Papeş, M.; Soberón, J. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecol. Model.* **2008**, *213*, 63–72. [CrossRef]
67. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2015**, *13*, 361–378. [CrossRef]
68. Dang, V.-H.; Hoang, N.-D.; Nguyen, L.-M.; Bui, D.T.; Samui, P. A Novel GIS-Based Random Forest Machine Algorithm for the Spatial Prediction of Shallow Landslide Susceptibility. *Forests* **2020**, *11*, 118. [CrossRef]
69. Fix, E.; Hodges, J.L. Significance Probabilities of the Wilcoxon Test. *Ann. Math. Stat.* **1955**, *26*, 301–312. [CrossRef]
70. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324. [CrossRef]
71. Nhu, V.-H.; Hoang, N.-D.; Amiri, M.; Bui, T.T.; Ngo, P.T.T.; Hoa, P.V.; Samui, P.; Thanh, L.N.; Quang, T.P.; Bui, D.T. An approach based on socio-politically optimized neural computing network for predicting shallow landslide susceptibility at tropical areas. *Environ. Earth Sci.* **2021**, *80*, 1–18. [CrossRef]

72. Sarkar, S.; Kanungo, D. An integrated approach for landslide susceptibility mapping using remote sensing and GIS. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 617–625. [[CrossRef](#)]
73. Wu, R.; Zhang, Y.; Guo, C.; Yang, Z.; Tang, J.; Su, F. Landslide susceptibility assessment in mountainous area: A case study of Sichuan–Tibet railway, China. *Environ. Earth Sci.* **2020**, *79*, 1–16. [[CrossRef](#)]
74. Chung, C.-J.F.; Fabbri, A.G. Validation of Spatial Prediction Models for Landslide Hazard Mapping. *Nat. Hazards* **2003**, *30*, 451–472. [[CrossRef](#)]
75. Ozturk, U.; Bozzolan, E.; Holcombe, E.A.; Shukla, R.; Pianosi, F.; Wagener, T. How climate change and unplanned urban sprawl bring more landslides. *Nature* **2022**, *608*, 262–265. [[CrossRef](#)]
76. Zhou, X.; Wen, H.; Zhang, Y.; Xu, J.; Zhang, W. Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. *Geosci. Front.* **2021**, *12*, 101211. [[CrossRef](#)]
77. Steger, S.; Brenning, A.; Bell, R.; Glade, T. The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. *Landslides* **2017**, *14*, 1767–1781. [[CrossRef](#)]
78. Lima, P.; Steger, S.; Glade, T. Counteracting flawed landslide data in statistically based landslide susceptibility modelling for very large areas: A national-scale assessment for Austria. *Landslides* **2021**, *18*, 3531–3546. [[CrossRef](#)]
79. Lin, Q.; Lima, P.; Steger, S.; Glade, T.; Jiang, T.; Zhang, J.; Liu, T.; Wang, Y. National-scale data-driven rainfall induced landslide susceptibility mapping for China by accounting for in-complete landslide data. *Geosci. Front.* **2021**, *12*, 101248. [[CrossRef](#)]
80. Lin, Q.; Steger, S.; Pittore, M.; Zhang, J.; Wang, L.; Jiang, T.; Wang, Y. Evaluation of potential changes in landslide susceptibility and landslide occurrence frequency in China under climate change. *Sci. Total. Environ.* **2022**, *850*, 158049. [[CrossRef](#)] [[PubMed](#)]
81. Shou, K.-J.; Lin, J.-F. Evaluation of the extreme rainfall predictions and their impact on landslide susceptibility in a sub-catchment scale. *Eng. Geol.* **2019**, *265*, 105434. [[CrossRef](#)]
82. Shou, K.-J.; Yang, C.-M. Predictive analysis of landslide susceptibility under climate change conditions—A study on the Chingshui River Watershed of Taiwan. *Eng. Geol.* **2015**, *192*, 46–62. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.