

Article

# A New Modeling Approach for Spatial Prediction of Flash Flood with Biogeography Optimized CHAID Tree Ensemble and Remote Sensing Data

Viet-Nghia Nguyen <sup>1</sup>, Peyman Yariyan <sup>2</sup>, Mahdis Amiri <sup>3</sup>, An Dang Tran <sup>4</sup>, Tien Dat Pham <sup>5</sup>, Minh Phuong Do <sup>6</sup>, Phuong Thao Thi Ngo <sup>7</sup>, Viet-Ha Nhu <sup>8,9,\*</sup>, Nguyen Quoc Long <sup>1</sup> and Dieu Tien Bui <sup>10</sup>

<sup>1</sup> Faculty of Geomatics and Land Administration, Hanoi University of Mining and Geology, No. 18 Pho Vien, Duc Thang, Bac Tu Liem, Hanoi 10000, Vietnam; nguyenvietnghia@humg.edu.vn (V.-N.N.); nguyenuoclong@humg.edu.vn (N.Q.L.)

<sup>2</sup> Department of Surveying Engineering, Saghez Branch, Islamic Azad University, Saghez 66819-73477, Iran; P.Yariyan@iausaghez.ac.ir

<sup>3</sup> Department of Watershed & Arid Zone Management, Gorgan University of Agricultural Sciences & Natural Resources, Gorgan 4918943464, Iran; mahdisamiri94@gmail.com

<sup>4</sup> Faculty of Water Resources Engineering, Thuyloi University, 175 Tay Son, Dong Da, Ha Noi 100000, Vietnam; Antd@tlu.edu.vn

<sup>5</sup> Center for Agricultural Research and Ecological Studies (CARES), Vietnam National University of Agriculture (VNUA), Trau Quy, Gia Lam, Hanoi 100000, Vietnam; dat6784@gmail.com

<sup>6</sup> Center for Informatics and Statistics (CIS), Ministry of Agriculture and Rural Development, Ba Dinh District, Ha Noi 100000, Vietnam; dphuong@mard.gov.vn

<sup>7</sup> Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam; ngotphuongthao5@duytan.edu.vn

<sup>8</sup> Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City 70000, Vietnam

<sup>9</sup> Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City 70000, Vietnam

<sup>10</sup> GIS Group, Department of Business and IT, University of Southeast Norway, Gullbringvegen 36, N-3800 Bø i Telemark, Norway; Dieu.T.Bui@usn.no

\* Correspondence: nhuvietha@tdtu.edu.vn

Received: 28 March 2020; Accepted: 15 April 2020; Published: 26 April 2020



**Abstract:** Flash floods induced by torrential rainfalls are considered one of the most dangerous natural hazards, due to their sudden occurrence and high magnitudes, which may cause huge damage to people and properties. This study proposed a novel modeling approach for spatial prediction of flash floods based on the tree intelligence-based CHAID (Chi-square Automatic Interaction Detector) random subspace, optimized by biogeography-based optimization (the CHAID-RS-BBO model), using remote sensing and geospatial data. In this proposed approach, a forest of tree intelligence was constructed through the random subspace ensemble, and, then, the swarm intelligence was employed to train and optimize the model. The Luc Yen district, located in the northwest mountainous area of Vietnam, was selected as a case study. For this circumstance, a flood inventory map with 1866 polygons for the district was prepared based on Sentinel-1 synthetic aperture radar (SAR) imagery and field surveys with handheld GPS. Then, a geospatial database with ten influencing variables (land use/land cover, soil type, lithology, river density, rainfall, topographic wetness index, elevation, slope, curvature, and aspect) was prepared. Using the inventory map and the ten explanatory variables, the CHAID-RS-BBO model was trained and verified. Various statistical metrics were used to assess the prediction capability of the proposed model. The results show that the proposed CHAID-RS-BBO model yielded the highest predictive performance, with an overall accuracy of 90% in predicting flash floods, and outperformed benchmarks (i.e., the CHAID, the J48-DT, the logistic regression, and the multilayer perception neural network (MLP-NN) models).

We conclude that the proposed method can accurately estimate the spatial prediction of flash floods in tropical storm areas.

**Keywords:** flash flood modeling; sentinel-1; random subspace; decision tree; machine learning; Vietnam

---

## 1. Introduction

Flooding is a phenomenon in which the water level in one place is above the permitted level, which is determined by the current frequency index. Researchers and planners point out that flooding is considered a significant disaster where the flow of water can flow from any sources and can be sudden or deliberate [1]. Flash floods are the most dangerous natural occurrences among various types of floods because of their rapid occurrences in a short period of time, and they pose more risks than other floods [2]. Climate change and rapid population growth are among the main drivers of flooding [3]. Additionally, according to the Intergovernmental Panel on Climate Change (IPCC) assessment, heavy rains are forecasted to have more impact on future floods [4]. Deaths and economic damage, destruction of agricultural crops, damage to environmental ecosystems, and the spread of contagious diseases along the water route are direct effects of the floods, which can cause irreparable damage [5–8]. Considering the historical events of the floods in the period 1998–2018, about 3136 flood catastrophes worldwide have occurred, and their consequences have affected more than approximately two billion people and caused about 556 billion US\$ in economic losses [9]. Indeed, the devastating consequences of flash floods on human lives have been spotted around the world [10,11]. There are a wide range of reasons, such as changes in the urbanization process, which cause vegetation cover changes and rapidly increasing population growth, which is accompanied by land use changes, resulting in an increase of the runoff coefficient [12]. Therefore, human settlements and vital infrastructure are vulnerable to flooding, and it is likely impossible to prevent this natural disaster completely. Thus, an effective spatial prediction of such events may reduce injuries and losses [13]. However, spatial prediction of flash flooding remains challenging due to the complex environmental factors involved [14,15]. Therefore, accurate modeling and mapping of flood risks play an important role in risk management planning and preventive measures [16]. Due to the destructive effects of flash floods on the environment and their social consequences, many studies so far have attempted flood risk modeling and zoning [17–19], because identifying areas vulnerable to flooding will be one of the most effective measures to reduce flood damage and flood management [20]. However, risk modeling and flood sensitivity mapping across large areas still remain challenging, because flash floods occur largely in each region under different climate conditions, which are unpredictable [21].

The literature review shows that in the development of new technologies, precise predictive models are often required for preparing the flood risk maps, which help with decision making to minimize and to monitor these events. A vast number of studies conducted on flood risk assessment use hydrological and hydrodynamic models. For instance, Giustarini, et al. [22] attempted to map the flood risks by using the temporal correlation model combined with hydraulic variables and time in the Severn River floodplain in the UK, while Li, et al. [23] used the Urban Flood Simulation Model (UFSM) and the Urban Flood Damage Assessment Model (UFDAM) in Shanghai, China for flood simulation. Recently, Komi, et al. [24] employed the distributed and calibrated hydrological method in the River Basin in West Africa with an application of rainfall intensity analysis and frequency intensity distribution relationships in flood risk modeling. The SCS-CN (Soil Conservation Service Curve Number) method has also applied the hydrograph theory in Volvos metropolitan area, Greece [16]. However, due to the lack of hydrological data, the limitations of the forecast, and the lack of a hydrometric station to record runoff and discharge, these methods cannot be used as a basic and optimal method for risk assessment at all locations.

In recent years, multi-criteria decision-making models have also been used for mapping flood risk using six influencing factors, including rainfall, slope, elevation, river density, land use, and soil types in Sukhothai Province, Thailand [25]. Wang, et al. [26]) attempted to develop a new hybrid technique using an integration of multi-criteria decision analysis, network analytical process and Weighted Linear Composition (WLC) in Shanghai City, China. Although multi-criteria decision-making methods can be a potential approach for improving the prediction performance in environmental hazard assessment, these techniques still have critical limitations, due to differences in the weight value of each factor in different regions. Importantly, several influencing factors such as land-cover/land-use (LULC) are often obtainable from earth observation data that consist of optical and synthetic aperture radar (SAR) data. Optical remote sensing datasets, which can be acquired at a certain time throughout the year, largely affected by the cloud coverage that commonly occurs in the tropical regions [27]. On the other hand, SAR remotely sensed data could be acquired under all weather conditions and become an essential source for mapping LULC [28]. Among various SAR sensors, Sentinel-1 C band SAR, provided by the European Space Agency (ESA) with dual polarization (VH,VV) can be acquired free-of-charge with a very high temporal resolution of 6 days, which makes it possible to provide systematic continuity data for mapping LULC [29,30].

Recently, various statistical machine learning techniques have been developed, including Frequency Ratio Index (FR) for flood risk mapping in the Markam Basin of Papua New Guinea [31], and flood sensitivity modeling in part of the Middle Ganga Plain in the Ganga Land Basin [32]. A number of studies have investigated the ability, and the effectiveness, of machine learning approaches combined with various optimization techniques for forecasting flash flood risk such as a combined artificial neural network (FA-LM-ANN) model in the Bac Ha Region located in Northwest Vietnam [33] and flood prediction using a self-organized neural network (SOM) technique at Kemaman River in Malaya Peninsula [34]. Various attempts have been made to predict flood risk in the current literature. Shafapour Tehrani, Kumar and Shabani [5] employed a Support Vector Machine (SVM) model for predicting flood risk in the Brisbane River basin, Australia, whereas Jahangir, et al. [35] integrated a multilayer perceptron neural network (MLPNN) model with GIS for spatial flood analysis in Tehran Province, Iran. One of the biggest challenges of predicting the risk of flooding is the lack of data in different regions. As a result, specific models cannot be used directly in different environments. In this context, novel machine learning techniques are able to help researchers in tackling the systematic issues and improve the predictive accuracy of flooding.

Thus, this study aims to fill these gaps in the literature by developing a novel modeling framework for spatial prediction of flash floods using the random subspace (RS) ensemble and the tree intelligence-based random subspace optimization combined with biogeography optimized (the CHAID-RS-BBO model). The RS ensemble is a powerful framework that has proven efficient in various spatial domains, i.e., landslide [36], flood [37], and image classification [38], whereas the CHAID decision tree is capable of providing good classification accuracy [39–41]. For the case of the BBO, this algorithm provides a efficient solution in searching and optimizing model parameters [42–44]. The proposed method can overcome the shortcomings of recent studies on flash floods risk mapping and will provide insights for further development of techniques in monitoring flash flood in the stormy tropical regions.

## 2. Study Area and Data

### 2.1. Study Area

Luc Yen is a mountainous district of the Yen Bai province in the northwest region of Vietnam (Figure 1). It covers approximately 810.10 km<sup>2</sup>, occupying 1.2% of the total area of the Yen Bai province. It is located between latitudes of 21°55'30"N and 22°03'30"N, and between longitudes of 104°30'06"E and 104°53'30"E. In terms of morphometry, the study area has a complex terrain consisting of mountain ranges, hills, mounts, cliffs, small valleys and plains along the Chay river, connecting directly to

Thac Ba reservoir. The topography is divided into high mountainous and low-flat elevation areas. The mountainous areas have very steep slopes and sharp peaks with elevation ranging from 100 m to 1399 m, while lower elevation areas are small valleys and plains distributed along the Chay river with elevations varying from 43 m to 100 m. In addition, the study area has complex and dense small streams and springs originating from two mountain ranges (Nui Voi and Large Rock mountain) before discharging into the Chay river in a northwest to southeast direction. As a complex terrain and drain network, the study area is highly vulnerable to flash floods, taking place when rapid runoff from steep slopes discharges quickly into small streams within a short time before reaching the Chay river [45].

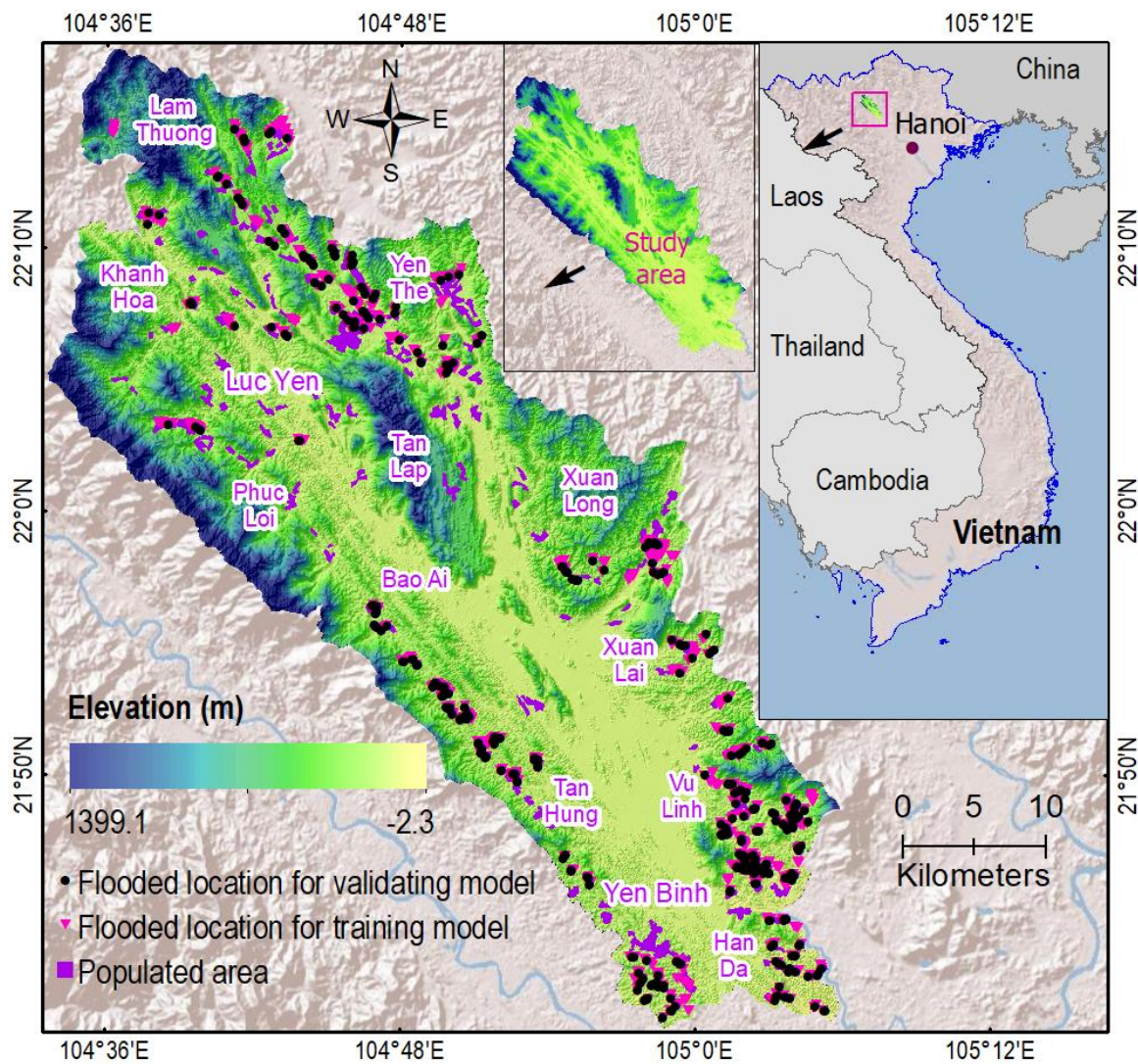


Figure 1. Location of the Luc Yen district and flash-flooded locations.

In the study area, the geology consists of six formations and outcrop complexes in the study area with an uneven distribution. Three formations account for >85% of the total study area: Song Chay (38.8%), Song Hong complex (32.6%), and Nui Chua (15.6%). The climatic condition is typically characterized as subtropical monsoonal, with two rainy seasons (April to September) and a dry season (October to March). The average yearly total rainfall ranges between 1739.3 mm and 2437.8 mm [45] and is mainly distributed in the rainy season, which accounts for 67.74–83.34% of the total annual rainfall. It is worth noting that high rainfall intensity events often occur in a short period coupled with steep slopes, and recent deforestation might cause frequent occurrences of flash floods and landslide in the study area.

## 2.2. Data

This research employed the on-off modeling approach [46] for the flash flood study, in which flash floods in the future will happen under the same conditions causing them in the past; therefore, historical flash floods must be collected. In this work, an inventory map with a total of 1866 flash flood polygons for the district was derived from the flash flood inventory map of the state-funded Project No-03/HD-KHCN-NTM of Vietnam [47]. These flash floods, which occurred during the last five years (2015–2019), were derived through the change detection techniques using multi-temporal Sentinel-1 synthetic aperture radar (SAR) imagery [33], then field surveys with handheld GPS were carried out to check and confirm the result. The largest polygon size of these flash floods is 64,064.3 m<sup>2</sup>, whereas the smallest polygon size is 912.3 m<sup>2</sup>, and the average polygon size is 6037 m<sup>2</sup>.

Because flash flood occurrences are influenced by various factors with their complex interactions, therefore, researchers have different views on this issue. However, it is common that factors are selected relating to topography, climate, soil, and human activities [48,49]. Since there are no specific rules and criteria for selecting effective flood factors in different regions, we selected ten influencing factors as the input explanatory variables in flash flood modeling in this study, based on the suggestions of various prior studies in the literature and the opinions of experts (See Table 1). These factors included land use, soil type, rock type, river density, precipitation, elevation, topographic wetness index (TWI), slope, slope direction, curvature, and aspect) (Table 1).

**Table 1.** Geospatial data sources used for the flash flood susceptibility mapping in this research.

Factor	Source	Flash Flood Relating	Reference
LULC	Sentinel-2, and ALOS-PALSAR (Advanced Land Observing Satellite-Phased Array type L-band Synthetic Aperture Radar) images [50]	Each type of land use/land cover(LULC) has a different role in the flash flood event	[51,52]
Soil type	The soil texture map 1:50,000 scale of Vietnam	Soil type has a significant influence on water infiltration	[52,53]
Lithology	Geologic map 1:50,000 scale of Vietnam	Affects water infiltration	[54]
River density	National topographic map 1:50,000 scale	The presence of rivers in any area causes floods	[55]
Rainfall	Rainfall stations	One of the most important factors is flooding	[56]
Elevation	ALOS-PALSAR DEM (Digital Elevation Model) 30 m [57]	High altitude areas connect the water flow to the rivers	[56]
TWI	ALOS-PALSAR DEM 30 m [57]	Impact on water flow accumulation rate	[5,54]
Slope	ALOS-PALSAR DEM 30 m [57]	It affects the speed and flow of water	[58]
Aspect	ALOS-PALSAR DEM 30 m [57]	It affects the direction of runoff and sunlight	[55]
Curvature	ALOS-PALSAR DEM 30 m [57]	Influence on surface infiltration.	[55,58]

### 2.2.1. Land-Use/Land-Cover (LULC)

Flash flooding begins with precipitation but depends on other factors, such as breadth, topography, and types of LULC during rainfall in the catchment [59]. Land-use type, especially vegetation compaction, has a significant impact on preventing or reducing flooding, and no matter how dense the vegetation, it will prevent severe flooding [51]. Additionally, different LULC types have different infiltration capacities and runoff coefficients, which influences significantly the time of concentration in a watershed [52,53]. Therefore, the characteristics of LULC are one of the main factors in flashflood prediction. The LULC map was interpolated using free-of-charge Sentinel-1 C band SAR data downloaded from the Copernicus open access hub of the European Space Agency (ESA) using the Sentinel Application Platform (SNAP) toolbox, with the random forest (RF) classification algorithm available on the SNAP toolbox. A total of eight types of land cover were obtained and visualized using the ArcGIS software in the study area, including bare land, crop areas, forest areas, grassland, orchard area, paddy rice, urban and built-up, and water bodies (Figure 2a). Although mountainous areas in the northern, northwest, and southern parts of the study area have different types of forest vegetation,

which may contribute to reducing flash floods, the transmitted areas from mountains to small valleys and plains consist of bare-land and grassland areas which have a high potential for flash floods taking place during or after high-rainfall-intensity events.

### 2.2.2. Soil Type

In terms of hydrology, soil types have a strong influence on the infiltration and erosion processes occurring in a watershed. This is because each soil type has different properties, which may reduce or increase runoff flow and/or erosion magnitude, and therefore have a direct relation to flash floods. For example, if the soil type is more capable of absorbing water, it can reduce runoff flow and time of water flow concentration into streams or rivers [60]. The soil layer of the study area was prepared by digitizing the soil texture map 1:50,000 scale. There are eleven soil types in the study area, in which YCMR soil occupies more than 80% of the total area, followed by WS and RM soils (Figure 2b).

### 2.2.3. Lithology

Flash flood flow often consists of different flow components, including surface flow, base flow, and groundwater flow. While soil types have a strong influence on surface flow, the type of rocks has a significant effect on base flow and ground flow system. Each type of rock has a specific permeability and density; these have different effects on infiltration and storage capacity and can influence the generation of water flow system in a watershed. For example, the resistant or impermeable rocks have less water absorption capacity, which may increase the base flow and runoff flow. Therefore, the type of rock in the region has a significant impact on flash flood risk modeling. The lithology map (Figure 2c) was obtained from the Luc Yen District Geological and Mineral Resources Map, with a scale of 1:50,000 [33]. The lithology was characterized by different types of rocks, including sedimentary, igneous, and metamorphic. The metamorphic rocks are dominant in the study area, accounting for 48%, followed by igneous and sedimentary (alluvium and recent deposits) [54]. Characteristics of lithologies in the study area were presented in previous studies [61–66] and are summarized in Table 2.

**Table 2.** Characteristics of the lithological formations in this study area.

No	Formation Structures	Main Lithology
1	Song Chay	Granit biotit, granit muscovit, granit hai mica, plagiogranit biotit.
2	Song Hong complex	Gneiss silimanite granite, granite biotite gneiss, calcite marble lenses, quartz schist silimanite granite, quartzite, gneis biotit granat, gneis biotit silimanit granat, gneis silimanit biotit, quarzit, biotite quartz slate, biotite silimanite quartz schist.
3	Quaternary	Granule, grit, breccia, boulder, pebbles, stone, cobble, sand, clay, and silt.
4	Nui Chua	Gabro, gabrodibas, horblend, Ordovician–Silurian quartzites, siliceous-sericitic schists, quartz porphyry, tuffstones, Devonian schists, limestones, sandstones, Triassic molasse sand-shaly and coarse-clastic deposits.
5	Pia Bioc Complex	Granit microclin, granit aplit, granit pegmatit.
6	Others	Rhyolite, dacite, felsite, and andesite rocks, plagioclase–granite, granophyre, granosyenite, granodiorite, diorite, and quartz–diorite.

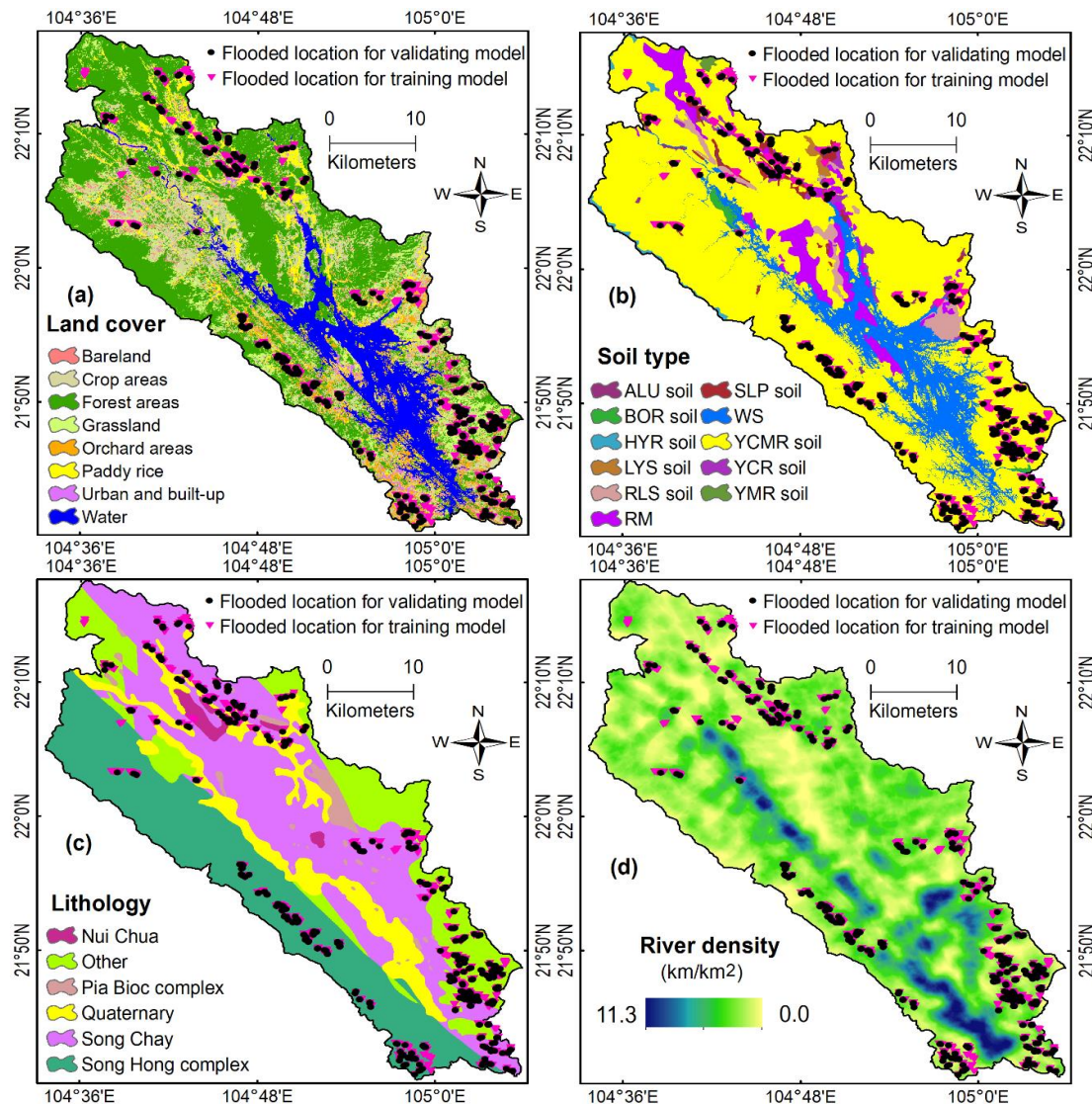
#### 1. River Density

Rivers are one of the most important factors used in flood sensitivity mapping, due to their significant impact on flood occurrence [67]. The higher the density of the water network in an area, the greater the impact on flood flow expansion [55]. In this research, river density (Figure 2d) was extracted from the above Digital Elevation Model (DEM) and river network system.

#### 2. Rainfall

One of the essential characteristics of a flash flood event is that it occurs quickly after high rainfall intensity within a short period of time (i.e., several hours) in steep mountainous areas with sparse vegetation coverage [56]. Therefore, rainfall is considered an essential factor in flood prediction, and

rainfall rate was chosen for flood risk assessment in this study. The higher the rainfall in a range, the greater the likelihood of a flood. In this research, the highest 16-day rainfall during the last 3 years at 30 stations in and around the study area was used to generate the rainfall pattern map using the Inverse Distance Weight technique [68]. The rainfall map (Figure 3a), with 142 mm in the northern areas and 620 mm in the central and southeastern areas, was interpolated through the station of the regional gauges rain in ArcGIS software.



**Figure 2.** Flash-flooded influencing factors: (a) Land cover, (b) Soil type, (c) Lithology, and (d) River density.

### 3. Elevation

Elevation and its effects play an essential role in flooding, and the lower the altitude, the greater the probability of flooding in that area [56,58]. Surface water flow often moves from high elevations towards low elevations, and therefore the low and flat area has a naturally high probability of flood occurrence [58]. The elevation map of the study area (Figure 3b) with the elevation ranging from  $-2.3$  m to 1399 m was prepared using a Digital Elevation Model (DEM) with a cell size of  $30\text{ m} \times 30\text{ m}$ . The DEM was built based on the national topographic maps available on a scale of 1:50000 obtained from Vietnam Institute of Geosciences and Mineral Resources [33].

#### 4. TWI

One of the parameters related to water flow is the topographic position index (TWI), which has been prepared through the altitude map of the study area with the following relationship [69].

$$TWI = \ln \frac{(A_s)}{(\tan \beta)} \quad (1)$$

where  $A_s$  denotes an upslope area, and  $\beta$  is the slope angle at the pixel.

Topographic moisture index is used to measure topographic control in hydrological studies [70]. *TWI* is a type of topographic property that shows the spatial distribution of moisture and cumulative water flow in response to the guiding force of water to lower areas [71]. In this area, *TWI* (Figure 3c) ranges from 142.8 to 662.1, in which the high values (>300) show the greatest density of torrential areas (30.25% of the class surface).

#### 5. Slope

Slope, as one of the environmental parameters, has a direct impact on surface water flow processes through influence on flow direction, velocity, and especially the time of water flow concentration at outfall [72]. High slopes often create faster movement and high velocity of runoff flow, as well as speeding up water flow in streams and rivers relative to lower slopes. Hence, runoff flow forming from steep slopes will cause an increase in water accumulation in low slope areas [58]. The slope layer shows a wide variation, ranging from 0 to 83.3 degrees in the study area (Figure 3d). In this area, a high slope angle in the mountainous areas has a strong effect on flash flood generation, while low slope in small valleys and plains affects the flash-flood propagation and duration (Figure 3d).

#### 6. Aspect

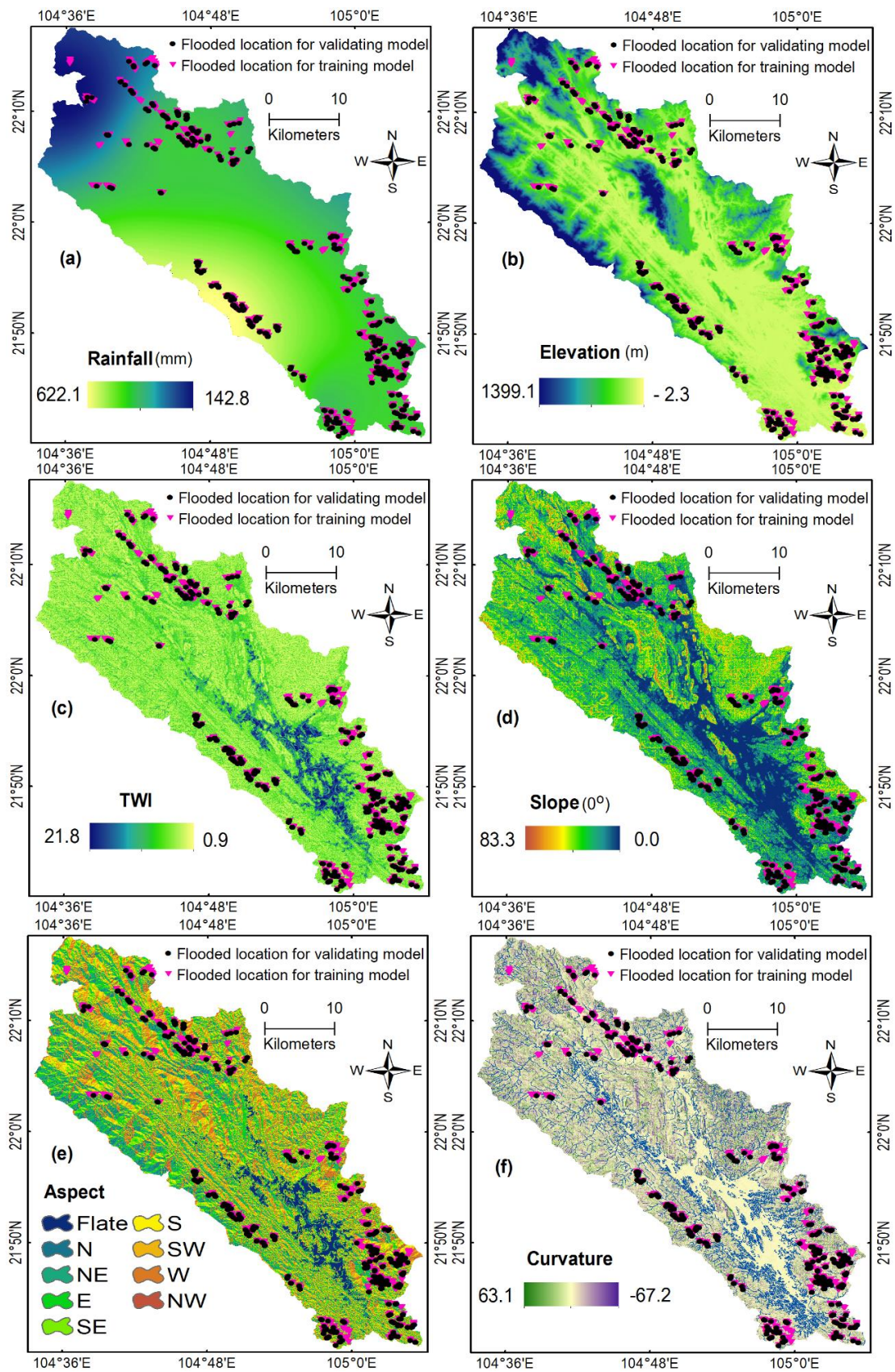
The slope aspect is one of the parameters influencing the hydrological conditions of the earth, which can affect local climate, physiographic approaches, soil moisture content and vegetation growth.

The aspect map consists of nine classes [55]: flat (−1), north (0–22.5 and 337.5–360), northeast (22.5–67.5), east (67.5–112.5), southeast (112.5–157.5), south (157.5–202.5), southwest (202.5–247.5), west (247.5–292.5) and northwest (292.5–337.5) (Figure 3e). The locations of flash floods occurring in the study is presented on the aspect map, indicating the influence of this factor on the probability of flash-flood occurrence.

#### 7. Curvature

Curvature presents the characteristic of morphometry and is obtained by intersecting a horizontal plane with the surface based on the Digital Elevation Model (30 m × 30 m). Curvature index has three states: concave (positive), convex (negative), and flat (zero), which can affect runoff processes [73]. The curvature map was prepared using altitude information on the study area. In this study area, approximately 70% of the research territory is covered by curvature values (Figure 3f). It was noted that most of the historical flash floods occurred in this area, being torrential.





**Figure 3.** Flash-flood-influencing factors: (a) Rainfall, (b) Elevation, (c) TWI, (d) Slope, (e) Aspect, and (f) curvature.

### 3. The Employed Methods

#### 3.1. Chi-Square Automatic Interaction Detection (CHAID)

The CHAID model is a classification tree technique used in many linear regressions [74]. The CHAID tree process is the division of large branches into smaller branches arranged in descending order from top to bottom, and the grouping continues based on specific factors [75]. The classification method of the CHAID algorithm was proposed by Kass [76]. This technique, as a new approach in the literature, has titles such as automatic interaction detection, classification and regression tree, artificial neural network, and genetic algorithm that can predict the required analysis [41]. The CHAID algorithm uses chi-square statistics as the separation criterion and performs the Dodge separation [77]. Thus, the classification continues as long as there is an acceptable value of chi-square between the dependent variable and the conditioning factors: that is, if the nodes with the highest chi-square value are in the first-order segmentation tree, and the nodes with the lowest chi-square value have the lowest degree. For this reason, the CHAID method chooses a statistical approach (Pearson’s square equation) that is desirable in terms of data type and the nature of the target [78].

$$X^2 = \sum_{j=1}^J \sum_{i=1}^I \frac{(n_{ij} - m_{ij})^2}{m_{ij}} \tag{2}$$

$$n_{ij} = \sum_{n \in D} f_{nI}(x_n = i \cap y_n = j), m_{ij} = \frac{n_{i.} \cdot n_{.j}}{n.} \tag{3}$$

where,  $n_{ij}$  is the frequency of observed cells,  $m_{ij}$ , is the cell frequency for  $(x_n = i, y_n = j)$ , and the  $p$  value is given by  $p = \Pr(x_{\chi^2} > x^2)$  [79].

#### 3.2. Random Subspace Ensemble (RSE)

The Random Subspace Ensemble algorithm was first developed by Hu [80]. RSE is a blended learning method in which a number of classifiers are combined and trained [81]. This algorithm, like the bagging algorithm, is randomly selected to create a training subset. The results from this technique are trees formed in earlier stages, which depend on learning differences and subcategories. The RSE algorithm is more robust than the Bagging and Adaboost algorithms.

#### 3.3. Biogeography-Based Optimization (BBO)

BBO is an evolutionary population-based search technique developed by Dan Simon [82], and was first performed on the multilayer perceptron neural network by [83]. The basic concepts of this algorithm were based on biogeography topics, including species migration, species emergence, and extinction [84]. The BBO Algorithm starts by creating habitat, then the migration and mutation steps are performed [85].

According to the BBO algorithm, the purpose of migration is to upgrade or correct the quality of existing methods [86]. Then the migration rate ( $\lambda_s$ ) is then defined to modify the suitability index variable. Therefore, due to some conditions that threaten the geographical location of the site, the habitat deviates from its optimal habitat suitability index, which is called the mutation process and is expressed as follows [87].

$$P_s^h = \begin{cases} -(\lambda_s + \mu_s)P_s + \mu_{s+1}P_{s+1}; & S = 0, \\ -(\lambda_s + \mu_s)P_s + \lambda_{s-1}P_{s-1} + \mu_{s+1}P_{s+1}; & 1 \leq S \leq S_{max} - 1, \\ -(\lambda_s + \mu_s)P_s + \lambda_{s-1}P_{s-1}; & S = S_{max} \end{cases} \tag{4}$$

where  $P_s$ ,  $\lambda_s$  and  $\mu_s$  are the possibility, the habitat migration, and the mutation, respectively;  $S_{max}$  presents the maximum Kind count.

#### 4. Proposed CHAID-RS-BBO Model for Flash Flood Susceptibility Modeling

The overall flowchart with the CHAID-RS-BBO model in this research is shown in Figure 4.

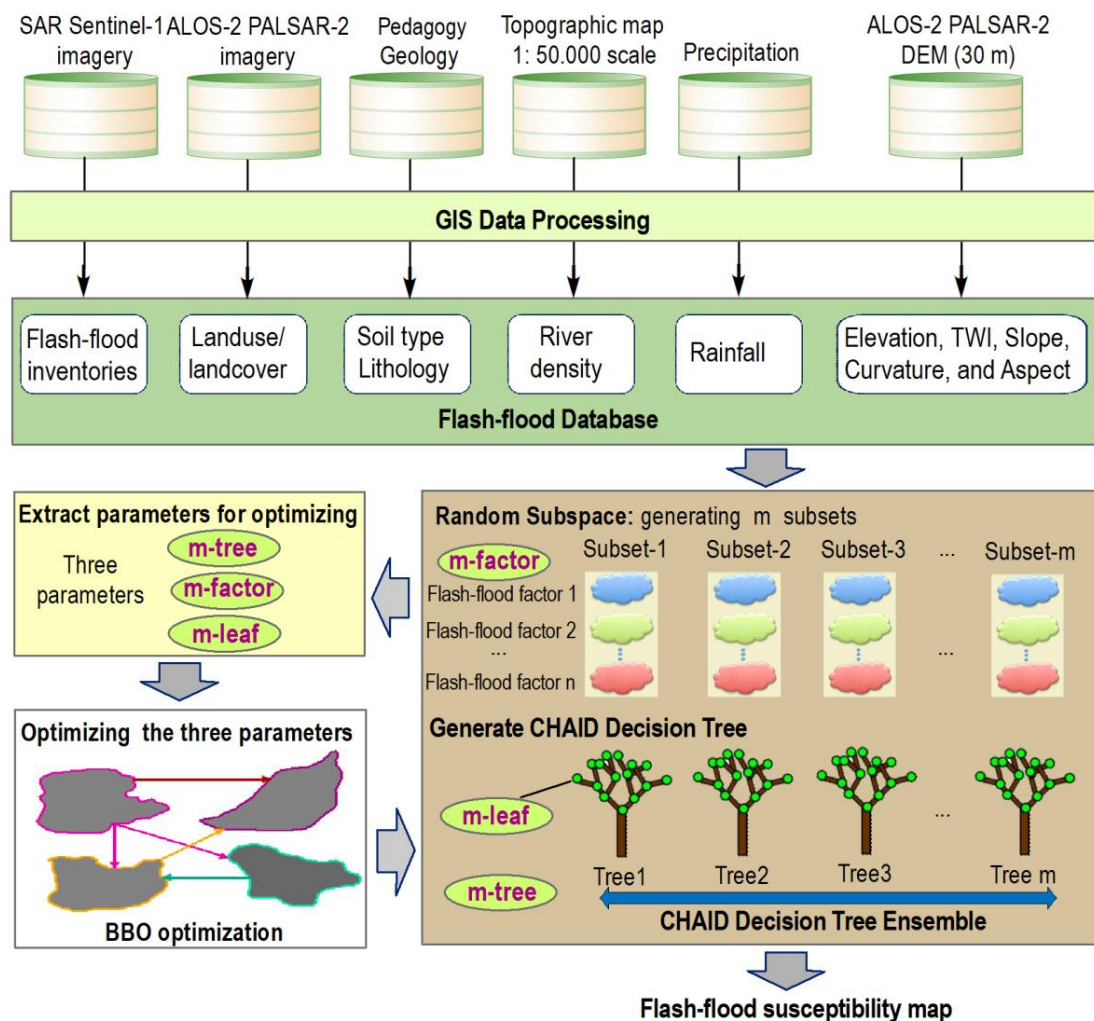


Figure 4. The flowchart with the CHAID-RS-BBO model for predicting flash flood susceptibility.

##### 4.1. Flash-Flood Database Establishment, Coding and Checking

In this step, the flash flood database for the Luc Yen, which consists of 1866 polygons, was constructed using Sentinel-1 SAR images and field investigations with handheld GPS and the ten selected influencing factors. The database associated with the geodatabase model in the ESRI ArcCatalog function was employed due to the ability to optimize its performance [88].

Because the CHAID model cannot read and understand the flash-flood-influencing factors directly, a coding process is required to convert all values in the factor maps into the range 0–1. In our research, values in six continuous factors (river density, rainfall, topographic wetness index, elevation, slope, curvature) were rescaled into the above range, whereas the other categorical factors (LULC, soil type, lithology, and aspect) were coded using the method described in [58].

Subsequently, a total number of 1866 points representing flash flood locations were divided into two datasets: 70% of the locations were randomly selected and used as the training set, and the remaining 30% of locations were used as the testing set to validate the model accuracy, as suggested in [56,89–91]. Finally, a sampling process was performed to generate values of the ten influencing factors.

#### 4.2. Establishing the CHAID-RS and the Cost Function

To generate the CHAID Decision Tree Ensemble using the Random Subspace framework (CHAID-RS), we determine three important parameters that are required to optimize: (1) number of CHAID trees used in the ensemble (m-tree); (2) number of the influencing factors used for the CHAID trees (m-factor); and (3) the minimum number of samples per leaf in the CHAID trees (m-leaf). The other parameters of the CHAID-RS model are used as the default values [92]. The three parameters were searched and optimized using the BBO algorithm.

Before optimizing these three parameters, it is necessary to design a cost function for the model. In this research, the cost function (CoF) (Equation (5)) proposed in [54] was adopted:

$$\text{CoF} = \sum_{i=1}^n \frac{(\text{FLPR}_i - \text{FLIV}_i)^2}{n} \quad (5)$$

where  $\text{FLPR}_i$  is the predicted output of the flash flood model;  $\text{FLIV}_i$  is the flood inventory value;  $n$  is the total samples used.

#### 4.3. Optimizing the CHAID-RS Using the BBO Algorithm

To search and optimize the three parameters, m-tree, m-factor, and m-leaf, for the CHAID-RS model, a three-dimensional searching space was established: m-tree = [1–500]; m-factor = [2–10]; and m-leaf = [2–20]. The three parameters were then transferred into a BBO matrix for optimizing. The other parameters of the BBO are as follows: the population size was 50; the maximum immigration and emigration values were 1.0; the mutation and crossover values were 0.25 and 0.95, respectively; and the total number of iterations used was 1000 [42]. Each individual of the population has three characteristics, which are the three parameters of the CHAID-RS model. The CoF was used to measure the suitability of the habitat. Herein, the smaller the CoF value, the better the habitat. Finally, the combination with the lowest CoF value was determined, and the best m-tree, m-factor, and m-leaf were derived. The best model was called the CHAID-RS-BBO model.

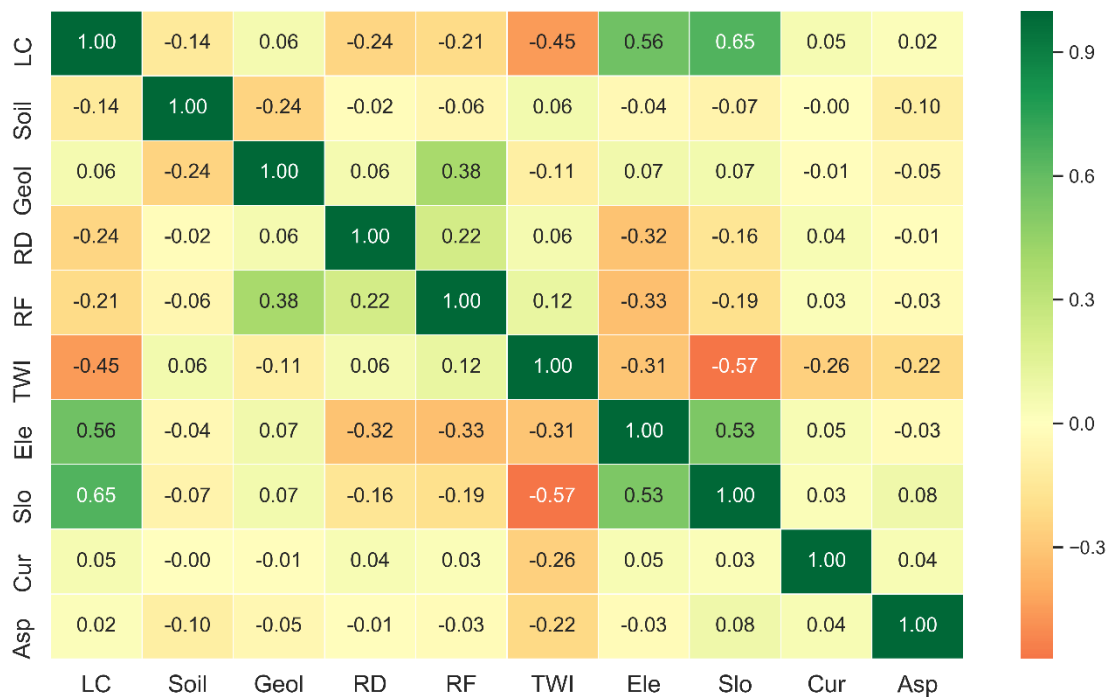
#### 4.4. Final CHAID-RS-BBO Model and Flash Flood Susceptibility

Once the CHAID-RS-BBO model was obtained, the performance of the model on both the training dataset and the validation dataset was checked. In this research, positive predictive value (PPV), and negative predictive value (NPV), sensitivity, specificity, accuracy, kappa, ROC curve and area under the curve (AUC) were used. Since explanations of these metrics for measuring the quality of spatial models are common in the literature, e.g., [93–95], we do not repeat these explanations here. In the final step, the CHAID-RS-BBO model was used to estimate the flash flood susceptibility index for each pixel of the Luc Yen district and generate the flash flood susceptibility map.

## 5. Results

### 5.1. Correlation of the Predictors of Flash Floods

The results of the Pearson's correlation among ten influencing factors (LULC, soil type, lithology, river density, rainfall, topographic wetness index, elevation, slope, curvature, and aspect) is presented in Figure 5. As can be seen from Figure 5, the highest positive correlation value (0.65) was observed between the LULC and the slope factors, whereas the largest negative correlation value of  $-0.57$  was observed between the TWI and the slope factors in the study area. However, these correlation values are less than those of 0.7, which is the threshold value of the collinearity problem [96]. Therefore, it is concluded that there is no correlation problem among the considered affecting factors.



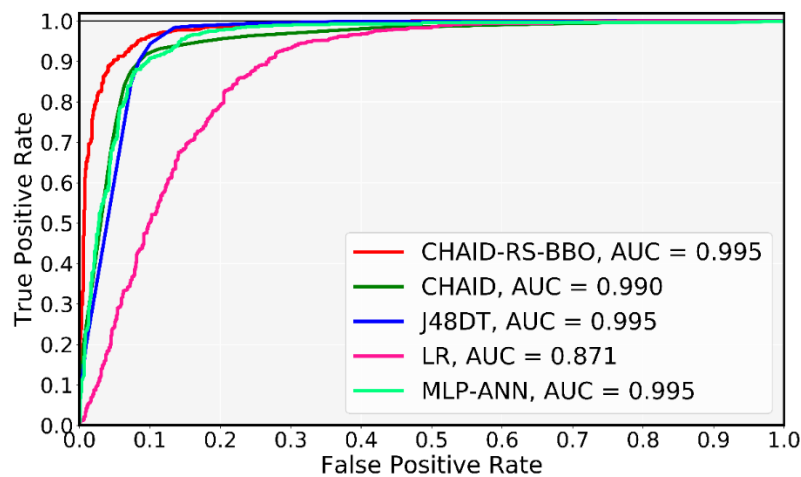
**Figure 5.** Pearson correlation of the flash flood predictors. LC: landuse/landcover; Soil: Soil type; Geol: Lithology; RD: River density; RF: Rainfall; TWI: Topographic Wetness Index; Ele: Elevation; Slo: Slope; Cur: Curvature; Asp: Aspect.

### 5.2. Training the Flash Flood Models

The training set accounts for 70% of the total dataset; the results in the training phase for the flash flooding occurrence using machine learning models are shown in Table 3 and Figure 6. It can be clearly observed that the CHAID-RS-BBO, the CHAID, the J48DT, the logistic regression, and the MLP-NN models had very good overall accuracies in the training dataset. The values of the AUC ranged from 0.871 to 0.979 (CHAID-RS-BBO = 0.979, CHAID = 0.949, J48DT = 0.955, logistic regression = 0.871, MLP-NN = 0.942). Besides, these corresponding numbers showed high predictive performances in terms of accuracy and kappa coefficient. The accuracies of the five ML models ranged from 81.36 to 91.00, whereas the kappa values were observed between 0.634 and 0.867.

**Table 3.** Performance of the flash flood models in the training phase.

Metrics	CHAID-RS-BBO	CHAID	J48DT	Logistic Regression	MLP-NN
True positive	867	832	893	835	868
True negative	828	823	786	654	774
False positive	41	76	15	73	40
False negative	80	85	122	254	134
PPV (%)	95.48	91.63	98.35	91.96	95.59
NPV (%)	91.19	90.64	86.56	72.03	85.24
Sensitivity (%)	91.55	90.73	87.98	76.68	86.63
Specificity (%)	95.28	91.55	98.13	89.96	95.09
Accuracy (%)	93.34	91.13	92.46	81.99	90.42
Kappa	0.867	0.823	0.849	0.634	0.808
AUC	0.979	0.949	0.955	0.871	0.953



**Figure 6.** The ROC curve of the five flash flood models in the training phase.

Overall, the CHAID-RS-BBO model had the highest performance in the training phase (AUC = 0.979, accuracy = 93.34, kappa = 0.867), followed by the J48-DT (AUC = 0.955, accuracy = 92.46, kappa = 0.849) and the CHAID model (AUC = 0.949, accuracy = 91.13, kappa = 0.823).

In contrast to the ensemble-based models, the logistic regression model produced the lowest performance (AUC = 0.871, accuracy = 81.99, kappa = 0.634). Figure 6 shows the predictive performance of the models in the training phase using the AUC indicator. It can also be clearly seen from the graph that the proposed model performed well and produced the best predictive performance for flash flood susceptibility in the training dataset.

### 5.3. Validating the Flash Flood Models

The results in the testing phase, using 30% of the total datasets for predicting flash flooding occurrence models, are shown in Table 3 and Figure 6. As can be observed from Table 4, the proposed ensemble-based model yielded the highest prediction performances with AUC = 0.960, accuracy = 91.00 and kappa = 0.820, followed by the MLP-NN, the CHAID, and the J48DT model. Conversely, the logistic regression model had the lowest performance in terms of the AUC, the accuracy, and the kappa coefficients (AUC = 0.880, accuracy = 81.36, kappa = 0.627). Generally, the results showed that the ensemble-based models archived high accuracy and satisfactory predictive performance for flash flooding accident, and this outcome can be clearly seen in Figure 7.

**Table 4.** Performance of the flash flood models in the validation phase.

Metrics	CHAID-RS-BBO	CHAID	J48DT	Logistic Regression	MLP-NN
True positive	364	338	363	350	367
True negative	344	337	315	283	323
False positive	25	51	26	39	22
False negative	45	52	74	106	66
PPV (%)	93.57	86.89	93.32	89.97	94.34
NPV (%)	88.43	86.63	80.98	72.75	83.03
Sensitivity (%)	89.00	86.67	83.07	76.75	84.76
Specificity (%)	93.22	86.86	92.38	87.89	93.62
Accuracy (%)	91.00	86.76	87.15	81.36	88.69
Kappa	0.820	0.735	0.743	0.627	0.774
AUC	0.960	0.899	0.893	0.880	0.942

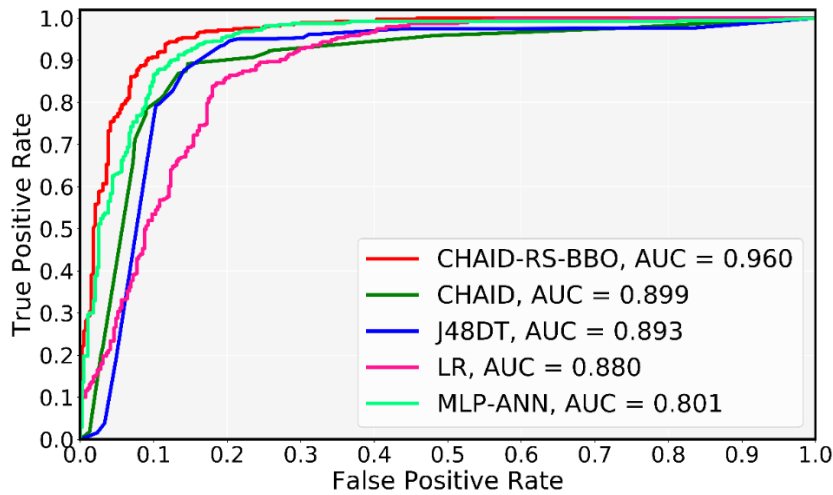


Figure 7. The ROC curve of the five flash flood models in the validating phase.

5.4. Flash Flood Susceptibility Maps

Since the CHAID-RS-BBO had the highest predictive performance in both the training and the testing phases and outperformed the benchmark models, we employed this model to map the flash flooding susceptibility in the study area. Accordingly, the CHAID-RS-BBO model was also used to calculate the flash flood susceptibility for all the pixels in the map of the case study. The predictive results of flash flood capacity were converted into a raster format and presented in the ArcGIS environment. Figure 8 illustrates the spatial prediction of the flash flood in the study area ranging from 0.022 to 0.9101.

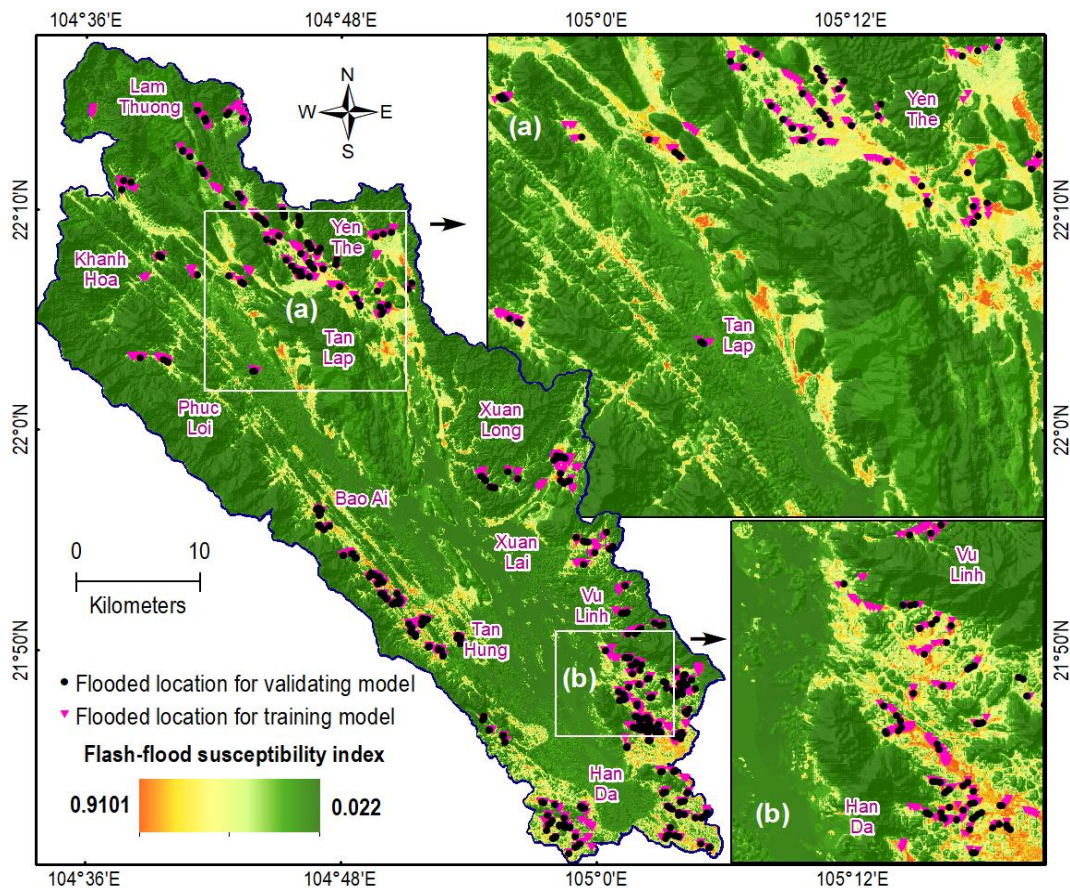


Figure 8. Flash flood map for the Luc Yen area using the CHAID-RS-BBO model.

As can be seen from Figure 8, the highest flash-flood susceptibility index was observed in the steep mountainous highland areas, where the flash floods often occur largely during the storm season associated with tropical typhoons. In contrast, the lowest rate was presented in the lowland area closed to rivers and streams.

## 6. Discussion

This study proposed a novel framework based on Sentinel-1 SAR images and field investigations combined with a new ensemble-based model for spatial prediction of flash flood hazards. Ten flood flash predictors were selected based on a review of the literature and the interpretations of the correlations of them with flash floods in the study area. As suggested in previous work [54,97], correlations among these predictors should be checked before going ahead to the modeling process. In this work, Pearson correlation analysis confirmed that these predictors are valid for modeling where all correlation values are less than 0.7. Consequently, the high performance of the CHAID-RS-BBO model indicates that these predictors were selected, processed, and coded successfully.

Regarding the final flood model, this is a hybrid of three components, CHAID, RS, and BBO, in which the CHAID plays the classifier in a tree-like structure manner, whereas the RS with the feature sub-spacing framework helps to reduce error rates of the flood model by generating various sub-datasets for the forest of the CHAID classifiers. Additionally, the BBO was integrated to optimize the three parameters (m-tree, m-factor, m-leaf) of the hybrid model. In our work, the merit of the BBO is that, with 1000 iterations run, a total of 50,000 possible combinations of m-tree, m-factor, and m-leaf for the CHAID-RS model were checked and compared, in order to select the best combination. The high prediction capability of the CHAID-RS-BBO model indicates that the three parameters were globally searched and optimized.

The validity of the hybrid CHAID-RS-BBO for flash flood modeling was confirmed through comparison with those of five benchmark machine learning algorithms. The proposed model was the most accurate in predicting the flash flood events and outperformed the benchmarks, indicating the CHAID-RS-BBO is promising for flash flood studies.

## 7. Concluding Remarks

This research presents a novel modeling approach for flash flood modeling with a new hybrid of machine learning, geospatial data, and available remote sensing data. Based on the findings, some conclusions can be drawn, as follows:

- With the flash flood inventories and six predictors, the remote sensing data, Sentinel-1 SAR, Sentinel-2 and ALOS-PALSAR DEM, are important sources for flash flood modeling.
- With its high performance, it can be concluded that CHAID-RS-BBO is a new tool for flash flood modeling.
- The susceptibility map, which reveals the flash flood hotspots in Luc Yen, might help the local government and decision-makers to minimize the flash flood impacts in the selection and collection of the water of the flash floods for life requirements and development projects.
- The current study recommends the creation of precise and updated meteorology, morphometry, hydrology, geology, topography, and socioeconomic studies. Early warning systems (EWS) have to be developed to predict flash floods and consequently minimize losses and reduce damage. Last, but not least, a national plan for flash-flood disaster management and risk reduction has to be enabled.

**Author Contributions:** Formal analysis: V.-N.N., A.D.T., M.P.D., P.T.T.N., V.-H.N., N.Q.L., D.T.B. Methodology: V.-N.N., P.Y., M.A., A.D.T., T.D.P., P.T.T.N., N.Q.L., V.-H.N. and D.T.B. Writing—original draft, V.-N.N., P.Y., M.A., A.D.T., T.D.P., D.T.B. Writing—review & editing: V.-H.N., T.D.P. and D.T.B. All authors have read and agreed to the published version of the manuscript.



**Funding:** This research was supported by Project “Research to building the map of partition and flash flood warnings with high resolution for some Northwestern provinces in order to enhance the ability to cope with natural disasters of the community to serving new rural area building”. The national target programme on new rural area building, stage 2016–2020. Ministry of Agriculture and Rural Development. (Project No-03/HD-KHCN-NTM).

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

## References

1. Fernandes, O.; Murphy, R.; Adams, J.; Merrick, D. Quantitative Data Analysis: CRASAR Small Unmanned Aerial Systems at Hurricane Harvey. In Proceedings of the 2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Philadelphia, PA, USA, 6–8 August 2018.
2. Marchi, L.; Borga, M.; Preciso, E.; Gaume, E. Characterisation of selected extreme flash floods in Europe and implications for flood risk management. *J. Hydrol.* **2010**, *394*, 118–133. [[CrossRef](#)]
3. Kjeldsen, T.R. Modelling the impact of urbanization on flood frequency relationships in the UK. *Hydrol. Res.* **2010**, *41*, 391–405. [[CrossRef](#)]
4. Alexander, L.V. Global observed long-term changes in temperature and precipitation extremes: A review of progress and limitations in IPCC assessments and beyond. *Weather Clim. Extrem.* **2016**, *11*, 4–16. [[CrossRef](#)]
5. Shafapour Tehrany, M.; Kumar, L.; Shabani, F. A novel GIS-based ensemble technique for flood susceptibility mapping using evidential belief function and support vector machine: Brisbane, Australia. *PeerJ* **2019**, *7*, e7653. [[CrossRef](#)]
6. Lyu, H.M.; Sun, W.J.; Shen, S.L.; Arulrajah, A. Flood risk assessment in metro systems of mega-cities using a GIS-based modeling approach. *Sci. Total Environ.* **2018**, *626*, 1012–1025. [[CrossRef](#)]
7. Yu, J.J.; Qin, X.S.; Larsen, O. Joint Monte Carlo and possibilistic simulation for flood damage assessment. *Stoch. Environ. Res. Risk Assess.* **2012**, *27*, 725–735. [[CrossRef](#)]
8. Merz, B.; Kreibich, H.; Schwarze, R.; Thielen, A. Review article “Assessment of economic flood damage”. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 1697–1724. [[CrossRef](#)]
9. Ogie, R.I.; Adam, C.; Perez, P. A review of structural approach to flood management in coastal megacities of developing nations: Current research and future directions. *J. Environ. Plan. Manag.* **2019**, *63*, 127–147. [[CrossRef](#)]
10. Gourley, J.J.; Flamig, Z.L.; Vergara, H.; Kirstetter, P.E.; Clark, R.A.; Argyle, E.; Arthur, A.; Martinaitis, S.; Terti, G.; Erlingis, J.M.; et al. The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 361–372. [[CrossRef](#)]
11. Archer, D.R.; Fowler, H.J. Characterising flash flood response to intense rainfall and impacts using historical information and gauged data in Britain. *J. Flood Risk Manag.* **2015**, *11*, S121–S133. [[CrossRef](#)]
12. Chang, H.; Franczyk, J. Climate change, land-use change, and floods: Toward an integrated assessment. *Geogr. Compass* **2008**, *2*, 1549–1579. [[CrossRef](#)]
13. Rahmati, O.; Darabi, H.; Haghighi, A.T.; Stefanidis, S.; Kornejady, A.; Nalivan, O.A.; Tien Bui, D. Urban Flood Hazard Modeling Using Self-Organizing Map Neural Network. *Water* **2019**, *11*, 2370. [[CrossRef](#)]
14. Mansur, A.V.; Brondizio, E.S.; Roy, S.; de Miranda Araújo Soares, P.P.; Newton, A. Adapting to urban challenges in the Amazon: Flood risk and infrastructure deficiencies in Belém, Brazil. *Reg. Environ. Chang.* **2017**, *18*, 1411–1426. [[CrossRef](#)]
15. Zhou, Z.; Liu, S.; Zhong, G.; Cai, Y. Flood Disaster and Flood Control Measurements in Shanghai. *Nat. Hazards Rev.* **2017**, *18*. [[CrossRef](#)]
16. Papaioannou, G.; Efstratiadis, A.; Vasiliades, L.; Loukas, A.; Papalexiou, S.; Koukouvinos, A.; Tsoukalas, I.; Kossieris, P. An Operational Method for Flood Directive Implementation in Ungauged Urban Areas. *Hydrology* **2018**, *5*, 24. [[CrossRef](#)]
17. Barredo, J.I.; Engelen, G. Land Use Scenario Modeling for Flood Risk Mitigation. *Sustainability* **2010**, *2*, 1327–1344. [[CrossRef](#)]
18. Winsemius, H.C.; Van Beek, L.P.H.; Jongman, B.; Ward, P.J.; Bouwman, A. A framework for global river flood risk assessments. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 1871–1892. [[CrossRef](#)]
19. Tsakiris, G. Flood risk assessment: Concepts, modelling, applications. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 1361–1369. [[CrossRef](#)]

20. Tien Bui, D.; Pradhan, B.; Nampak, H.; Bui, Q.T.; Tran, Q.A.; Nguyen, Q.P. Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. *J. Hydrol.* **2016**, *540*, 317–330. [[CrossRef](#)]
21. Lee, B.J.; Kim, S. Gridded Flash Flood Risk Index Coupling Statistical Approaches and TOPLATS Land Surface Model for Mountainous Areas. *Water* **2019**, *11*, 504. [[CrossRef](#)]
22. Giustarini, L.; Chini, M.; Hostache, R.; Pappenberger, F.; Matgen, P. Flood Hazard Mapping Combining Hydrodynamic Modeling and Multi Annual Remote Sensing data. *Remote Sens.* **2015**, *7*, 14200–14226. [[CrossRef](#)]
23. Li, C.; Cheng, X.; Li, N.; Du, X.; Yu, Q.; Kan, G. A Framework for Flood Risk Analysis and Benefit Assessment of Flood Control Measures in Urban Areas. *Int. J. Environ. Res. Public Health* **2016**, *13*, 787. [[CrossRef](#)]
24. Komi, K.; Neal, J.; Trigg, M.A.; Diekkrüger, B. Modelling of flood hazard extent in data sparse areas: A case study of the Oti River basin, West Africa. *J. Hydrol.* **2017**, *10*, 122–132. [[CrossRef](#)]
25. Seejata, K.; Yodying, A.; Wongthadam, T.; Mahavik, N.; Tantanee, S. Assessment of flood hazard areas using Analytical Hierarchy Process over the Lower Yom Basin, Sukhothai Province. *Procedia Eng.* **2018**, *212*, 340–347. [[CrossRef](#)]
26. Wang, Y.; Hong, H.; Chen, W.; Li, S.; Pamučar, D.; Gigović, L.; Drobniak, S.; Bui, D.T.; Duan, H. A Hybrid GIS Multi-Criteria Decision-Making Method for Flood Susceptibility Mapping at Shangyou, China. *Remote Sens.* **2018**, *11*, 62. [[CrossRef](#)]
27. Pham, T.D.; Xia, J.; Ha, N.T.; Bui, D.T.; Le, N.N.; Takeuchi, W. A Review of Remote Sensing Approaches for Monitoring Blue Carbon Ecosystems: Mangroves, Seagrasses and Salt Marshes during 2010–2018. *Sensors* **2019**, *19*, 1933. [[CrossRef](#)]
28. Schlaffer, S.; Matgen, P.; Hollaus, M.; Wagner, W. Flood detection from multi-temporal SAR data using harmonic analysis and change detection. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *38*, 15–24. [[CrossRef](#)]
29. Twele, A.; Cao, W.; Plank, S.; Martinis, S. Sentinel-1-based flood mapping: A fully automated processing chain. *Int. J. Remote Sens.* **2016**, *37*, 2990–3004. [[CrossRef](#)]
30. Chatziantoniou, A.; Psomiadis, E.; Petropoulos, G. Co-Orbital Sentinel 1 and 2 for LULC Mapping with Emphasis on Wetlands in a Mediterranean Setting Based on Machine Learning. *Remote Sens.* **2017**, *9*, 1259. [[CrossRef](#)]
31. Samanta, S.; Pal, D.K.; Palsamanta, B. Flood susceptibility analysis through remote sensing, GIS and frequency ratio model. *Appl. Water Sci.* **2018**, *8*, 66. [[CrossRef](#)]
32. Arora, A.; Pandey, M.; Siddiqui, M.A.; Hong, H.; Mishra, V.N. Spatial flood susceptibility prediction in Middle Ganga Plain: Comparison of frequency ratio and Shannon’s entropy models. *Geocarto Int.* **2019**, 1–32. [[CrossRef](#)]
33. Ngo, P.T.; Hoang, N.D.; Pradhan, B.; Nguyen, Q.; Tran, X.; Nguyen, Q.; Nguyen, V.; Samui, P.; Tien Bui, D. A Novel Hybrid Swarm Optimized Multilayer Neural Network for Spatial Prediction of Flash Floods in Tropical Areas Using Sentinel-1 SAR Imagery and Geospatial Data. *Sensors* **2018**, *18*, 3704. [[CrossRef](#)] [[PubMed](#)]
34. Chang, L.C.; Amin, M.; Yang, S.N.; Chang, F.J. Building ANN-Based Regional Multi-Step-Ahead Flood Inundation Forecast Models. *Water* **2018**, *10*, 1283. [[CrossRef](#)]
35. Jahangir, M.H.; Mousavi Reineh, S.M.; Abolghasemi, M. Spatial prediction of flood zonation mapping in Kan River Basin, Iran, using artificial neural network algorithm. *Weather Clim. Extrem.* **2019**, *25*, 100215. [[CrossRef](#)]
36. Pham, B.T.; Prakash, I.; Bui, D.T. Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* **2018**, *303*, 256–270. [[CrossRef](#)]
37. Chen, W.; Hong, H.; Li, S.; Shahabi, H.; Wang, Y.; Wang, X.; Ahmad, B.B. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *J. Hydrol.* **2019**, *575*, 864–873. [[CrossRef](#)]
38. Jiang, M.; Fang, Y.; Su, Y.; Cai, G.; Han, G. Random Subspace Ensemble With Enhanced Feature for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2019**. [[CrossRef](#)]
39. Atieh, M.A.; Pang, J.K.; Lian, K.; Wong, S.; Tawse-Smith, A.; Ma, S.; Duncan, W.J. Predicting peri-implant disease: Chi-square automatic interaction detection (CHAID) decision tree analysis of risk indicators. *J. Periodontol.* **2019**, *90*, 834–846. [[CrossRef](#)]

40. Althuwaynee, O.F.; Pradhan, B.; Park, H.J.; Lee, J.H. A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. *Landslides* **2014**, *11*, 1063–1078. [[CrossRef](#)]
41. Díaz-Pérez, F.M.; Bethencourt-Cejas, M. CHAID algorithm as an appropriate analytical method for tourism market segmentation. *J. Destin. Mark. Manag.* **2016**, *5*, 275–282. [[CrossRef](#)]
42. Pham, B.T.; Nguyen, M.D.; Bui, K.T.T.; Prakash, I.; Chapi, K.; Bui, D.T. A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil. *Catena* **2019**, *173*, 302–311. [[CrossRef](#)]
43. Kaveh, M.; Mesgari, M.S. Improved biogeography-based optimization using migration process adjustment: An approach for location-allocation of ambulances. *Comput. Ind. Eng.* **2019**, *135*, 800–813. [[CrossRef](#)]
44. Jaafari, A.; Panahi, M.; Pham, B.T.; Shahabi, H.; Bui, D.T.; Rezaie, F.; Lee, S. Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *Catena* **2019**, *175*, 430–445. [[CrossRef](#)]
45. SYB. *Yen Bai Statistical Year Book 2017*; Statistical Publishing House: Hanoi, Vietnam, 2018; p. 470.
46. Tien Bui, D.; Hoang, N.D. A Bayesian framework based on a Gaussian mixture model and radial-basis-function Fisher discriminant analysis (BayGmmKda V1.1) for spatial prediction of floods. *Geosci. Model Dev.* **2017**, *10*, 3391–3409. [[CrossRef](#)]
47. Viet Nghia, N. *Study to Build Flash Flood Prediction and Zoning Maps with High Resolution for Some Northwestern Provinces of Vietnam to Enhance Community's Ability to Respond to Natural Disasters and New Rural Development Strategies*; The Ministry of Agriculture and Rural Development of Vietnam: Hanoi, Vietnam, 2020.
48. Costache, R.; Popa, M.C.; Bui, D.T.; Diaconu, D.C.; Ciubotaru, N.; Minea, G.; Pham, Q.B. Spatial predicting of flood potential areas using novel hybridizations of fuzzy decision-making, bivariate statistics, and machine learning. *J. Hydrol.* **2020**, 124808. [[CrossRef](#)]
49. Tehrany, M.S.; Lee, M.J.; Pradhan, B.; Jebur, M.N.; Lee, S. Flood susceptibility mapping using integrated bivariate and multivariate statistical models. *Environ. Earth Sci.* **2014**, *72*, 4001–4015. [[CrossRef](#)]
50. Duong, P.C.; Trung, T.H.; Nasahara, K.N.; Tadono, T. JAXA High-Resolution Land Use/Land Cover Map for Central Vietnam in 2007 and 2017. *Remote Sens.* **2018**, *10*, 1406. [[CrossRef](#)]
51. Armenakis, C.; Du, E.; Natesan, S.; Persad, R.; Zhang, Y. Flood Risk Assessment in Urban Areas Based on Spatial Analytics and Social Factors. *Geosciences* **2017**, *7*, 123. [[CrossRef](#)]
52. Youssef, A.M.; Pradhan, B.; Sefry, S.A. Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. *Environ. Earth Sci.* **2015**, *75*. [[CrossRef](#)]
53. Chen, Y.; Liu, R.; Barrett, D.; Gao, L.; Zhou, M.; Renzullo, L.; Emelyanova, I. A spatial assessment framework for evaluating flood risk under extreme climates. *Sci. Total Environ.* **2015**, *538*, 512–523. [[CrossRef](#)]
54. Bui, D.T.; Ngo, P.-T.T.; Pham, T.D.; Jaafari, A.; Minh, N.Q.; Hoa, P.V.; Samui, P. A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *CATENA* **2019**, *179*, 184–196. [[CrossRef](#)]
55. Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhaug, I.; Prakash, I.; Tien Bui, D. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci. Total Environ.* **2018**, *627*, 744–755. [[CrossRef](#)] [[PubMed](#)]
56. Tien Bui, D.; Hoang, N.D.; Pham, T.D.; Ngo, P.T.T.; Hoa, P.V.; Minh, N.Q.; Tran, X.T.; Samui, P. A new intelligence approach based on GIS-based Multivariate Adaptive Regression Splines and metaheuristic optimization for predicting flash flood susceptible areas at high-frequency tropical typhoon area. *J. Hydrol.* **2019**, *575*, 314–326. [[CrossRef](#)]
57. Japan Aerospace Exploration Agency ALOS Global Digital Surface Model ALOS World 3D—30m. Available online: <https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm> (accessed on 8 July 2019).
58. Tien Bui, D.; Hoang, N.D.; Martínez-Álvarez, F.; Ngo, P.T.T.; Hoa, P.V.; Pham, T.D.; Samui, P.; Costache, R. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Sci. Total Environ.* **2020**, *701*, 134413. [[CrossRef](#)] [[PubMed](#)]
59. Hölting, B.; Coldewey, W.G. Hydrogeology. In *Springer Textbooks in Earth Sciences, Geography and Environment*; Springer: Berlin/Heidelberg, Germany, 2019.
60. Cosby, B.J.; Hornberger, G.M.; Clapp, R.B.; Ginn, T.R. A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils. *Water Resour. Res.* **1984**, *20*, 682–690. [[CrossRef](#)]

61. Roger, F.; Leloup, P.H.; Jolivet, M.; Lacassin, R.; Trinh, P.T.; Brunel, M.; Seward, D. Long and complex thermal history of the Song Chay metamorphic dome (Northern Vietnam) by multi-system geochronology. *Tectonophysics* **2000**, *321*, 449–466. [[CrossRef](#)]
62. Clift, P.D.; Sun, Z. The sedimentary and tectonic evolution of the Yinggehai–Song Hong basin and the southern Hainan margin, South China Sea: Implications for Tibetan uplift and monsoon intensification. *J. Geophys. Res. Solid Earth* **2006**, *111*. [[CrossRef](#)]
63. Polyakov, G.V.; Shelepaev, R.A.; Hoa, T.T.; Izokh, A.E.; Balykin, P.A.; Phuong, N.T.; Hung, T.Q.; Nien, B.A. The Nui Chua layered peridotite-gabbro complex as manifestation of Permo-Triassic mantle plume in northern Vietnam. *Russ. Geol. Geophys.* **2009**, *50*, 501–516. [[CrossRef](#)]
64. Lepvrier, C.; Faure, M.; Van, V.N.; Vu, T.V.; Lin, W.; Trong, T.T.; Hoa, P.T. North-directed Triassic nappes in Northeastern Vietnam (East Bac Bo). *J. Asian Earth Sci.* **2011**, *41*, 56–68. [[CrossRef](#)]
65. Pham, B.T.; Bui, D.T.; Dholakia, M.B.; Prakash, I.; Pham, H.V.; Mehmood, K.; Le, H.Q. A novel ensemble classifier of rotation forest and Naïve Bayer for landslide susceptibility assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. *Geomat. Nat. Hazards Risk* **2017**, *8*, 649–671. [[CrossRef](#)]
66. Quan, V.T.H.; Giao, P.H. Geochemical evaluation of shale formations in the northern Song Hong basin, Vietnam. *J. Pet. Explor. Prod. Technol.* **2019**, *9*, 1839–1853. [[CrossRef](#)]
67. Glenn, E.P.; Morino, K.; Nagler, P.L.; Murray, R.S.; Pearlstein, S.; Hultine, K.R. Roles of saltcedar (*Tamarix* spp.) and capillary rise in salinizing a non-flooding terrace on a flow-regulated desert river. *J. Arid Environ.* **2012**, *79*, 56–65. [[CrossRef](#)]
68. Jeong, H.G.; Ahn, J.B.; Lee, J.; Shim, K.M.; Jung, M.P. Improvement of daily precipitation estimations using PRISM with inverse-distance weighting. *Theor. Appl. Climatol.* **2020**, *139*, 923–934. [[CrossRef](#)]
69. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *J. Hydrol.* **2014**, *512*, 332–343. [[CrossRef](#)]
70. Chen, C.Y.; Yu, F.C. Morphometric analysis of debris flows and their source areas using GIS. *Geomorphology* **2011**, *129*, 387–397. [[CrossRef](#)]
71. Grabs, T.; Seibert, J.; Bishop, K.; Laudon, H. Modeling spatial patterns of saturated areas: A comparison of the topographic wetness index and a dynamic distributed model. *J. Hydrol.* **2009**, *373*, 15–23. [[CrossRef](#)]
72. Rahmati, O.; Pourghasemi, H.R.; Zeinivand, H. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto Int.* **2015**, *31*, 42–70. [[CrossRef](#)]
73. Mojaddadi, H.; Pradhan, B.; Nampak, H.; Ahmad, N.; Ghazali, A.H.B. Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1080–1102. [[CrossRef](#)]
74. Althuwaynee, O.F.; Pradhan, B.; Ahmad, N. Landslide susceptibility mapping using decision-tree based CHi-squared automatic interaction detection (CHAID) and Logistic regression (LR) integration. *IOP Conf. Ser.* **2014**, *20*, 012032. [[CrossRef](#)]
75. Amalita, N.; Kurniawati, Y.; Fitria, D. Characteristics of bidikmisi’s scholarship awardee in FMIPA UNP using chi-squared automatic interaction detection. *J. Phys.* **2019**, *1317*, 012012. [[CrossRef](#)]
76. Kass, G.V. Significance Testing in Automatic Interaction Detection (A.I.D.). *Appl. Stat.* **1975**, *24*, 178. [[CrossRef](#)]
77. Yeon, Y.K.; Han, J.G.; Ryu, K.H. Landslide susceptibility mapping in Injae, Korea, using a decision tree. *Eng. Geol.* **2010**, *116*, 274–283. [[CrossRef](#)]
78. Park, S.J.; Lee, C.W.; Lee, S.; Lee, M.J. Landslide Susceptibility Mapping and Comparison Using Decision Tree Models: A Case Study of Jumunjin Area, Korea. *Remote Sens.* **2018**, *10*, 1545. [[CrossRef](#)]
79. Baker, S.; Cousins, R.D. Clarification of the use of CHI-square and likelihood functions in fits to histograms. *Nucl. Instrum. Methods Phys. Res.* **1984**, *221*, 437–442. [[CrossRef](#)]
80. Tin Kam, H. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844. [[CrossRef](#)]
81. Pham, B.T.; Tien Bui, D.; Prakash, I.; Dholakia, M.B. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *CATENA* **2017**, *149*, 52–63. [[CrossRef](#)]
82. Simon, D. Biogeography-Based Optimization. *IEEE Trans. Evol. Comput.* **2008**, *12*, 702–713. [[CrossRef](#)]
83. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Let a biogeography-based optimizer train your Multi-Layer Perceptron. *Inf. Sci.* **2014**, *269*, 188–209. [[CrossRef](#)]

84. Roy, B.; Singh, M.P.; Singh, A. A novel approach for rainfall-runoff modelling using a biogeography-based optimization technique. *Int. J. River Basin Manag.* **2019**, *1*–14. [[CrossRef](#)]
85. Moayedi, H.; Osouli, A.; Tien Bui, D.; Foong, L.K. Spatial Landslide Susceptibility Assessment Based on Novel Neural-Metaheuristic Geographic Information System Based Ensembles. *Sensors* **2019**, *19*, 4698. [[CrossRef](#)]
86. Kaur, H.; Bhatia, R. Medical Image Segmentation using Penalized FCM and Pollination based Optimization Approach. *Int. J. Comput. Appl.* **2015**, *118*, 32–35. [[CrossRef](#)]
87. Hadidi, A. A robust approach for optimal design of plate fin heat exchangers using biogeography based optimization (BBO) algorithm. *Appl. Energy* **2015**, *150*, 196–210. [[CrossRef](#)]
88. Zeiler, M.; Murphy, J. *Modeling Our World: The ESRI Guide to Geodatabase Concep*; ESRI Press: Redlands, CA, USA, 2010; p. 297.
89. Gnecco, G.; Morisi, R.; Roth, G.; Sanguineti, M.; Taramasso, A.C. Supervised and semi-supervised classifiers for the detection of flood-prone areas. *Soft Computing* **2017**, *21*, 3673–3685. [[CrossRef](#)]
90. Costache, R.; Bui, D.T. Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania. *Sci. Total Environ.* **2019**, *691*, 1098–1118. [[CrossRef](#)]
91. Rahmati, O.; Yousefi, S.; Kalantari, Z.; Uuemaa, E.; Teimurian, T.; Keesstra, S.; Pham, T.D.; Tien Bui, D. Multi-hazard exposure mapping using machine learning techniques: A case study from Iran. *Remote Sens.* **2019**, *11*, 1943. [[CrossRef](#)]
92. Ibarguren, I.; Lasarguren, A.; Pérez, J.M.; Muguerza, J.; Gurrutxaga, I.; Arbelaitz, O. BFPART: Best-First PART. *Inf. Sci.* **2016**, *367*, 927–952. [[CrossRef](#)]
93. Costache, R.; Tien Bui, D. Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. *Sci. Total Environ.* **2020**, *712*, 136492. [[CrossRef](#)]
94. Tehrany, M.S.; Jones, S.; Shabani, F.; Martínez-Álvarez, F.; Bui, D.T. A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using logitboost machine learning classifier and multi-source geospatial data. *Theor. Appl. Climatol.* **2019**, *137*, 637–653. [[CrossRef](#)]
95. Rahmati, O.; Kornejady, A.; Samadi, M.; Deo, R.C.; Conoscenti, C.; Lombardo, L.; Dayal, K.; Taghizadeh-Mehrjardi, R.; Pourghasemi, H.R.; Kumar, S. PMT: New analytical framework for automated evaluation of geo-environmental modelling approaches. *Sci. Total Environ.* **2019**, *664*, 296–311. [[CrossRef](#)]
96. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46. [[CrossRef](#)]
97. Avand, M.; Janizadeh, S.; Tien Bui, D.; Pham, V.H.; Ngo, P.T.T.; Nhu, V.H. A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. *Int. J. Digit. Earth* **2020**, *1*–22. [[CrossRef](#)]

