A New Hybrid Firefly–PSO Optimized Random Subspace Tree Intelligence for Torrential Rainfall-Induced Flash Flood Susceptible Mapping

Viet-Ha Nhu [1,2] , **Phuong-Thao Thi Ngo** [3,*] , **Tien Dat Pham** [4] , **Jie Dou** [5,6] , **Xuan Song** [7] , **Nhat-Duc Hoang** [8] , **Dang An Tran** [9] , **Duong Phan Cao** [10,11] , **İbrahim Berkan Aydilek** [12] , **Mahdis Amiri** [13] , **Romulus Costache** [14] , **Pham Viet Hoa** [15] and **Dieu Tien Bui** [16]

1   Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam; nhuvietha@tdtu.edu.vn
2   Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam
3   Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
4   Center for Agricultural Research and Ecological Studies (CARES), Vietnam National University of Agriculture (VNUA), Trau Quy, Gia Lam, Hanoi 100000, Vietnam; dat6784@gmail.com
5   Three Gorges Research Center for Geo-Hazards, Ministry of Education, China University of Geosciences, Wuhan 430074, China; douj888@vos.nagaokaut.ac.jp
6   Department of Civil and Environmental Engineering, Nagaoka University of Technology, Nagaoka 1603-1, Japan
7   SUSTech-UTokyo Joint Research Center on Super Smart City, Department of Computer Science and Engineering, Southern University of Science and Technology (SUSTech), Shenzhen 518055, China; songx@sustech.edu.cn
8   Faculty of Civil Engineering, Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam; hoangnhatduc@dtu.edu.vn
9   Faculty of Water Resources Engineering, Thuyloi University, 175 Tay Son, Dong Da district, Hanoi 100000, Vietnam; antd@tlu.edu.vn
10   Graduate School of Life and Environmental Sciences, University of Tsukuba, Tennoudai 1-1-1, Tsukuba 305-8572, Japan; s1936042@s.tsukuba.ac.jp
11   Hydraulic Construction Institute—Vietnam Academy for Water Resources, No. 3, Alley 95, Chua Boc Street, Dong Da District, Hanoi 116765, Vietnam
12   Department of Computer Engineering, Faculty of Engineering, Harran University, 63050 Şanlıurfa, Turkey; BerkanAydilek@harran.edu.tr
13   Department of Watershed & Arid Zone Management, Gorgan University of Agricultural Sciences & Natural Resources, Gorgan 49138-15739, Iran; m.amiri@gau.ac.ir
14   Research Institute of the University of Bucharest, 90-92 Sos. Panduri, 5th District, 050663 Bucharest, Romania; romulus.costache@icub.unibuc.ro
15   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 700000, Vietnam; pvhoa@hcmig.vast.vn
16   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: ngotphuongthao5@duytan.edu.vn

check for updates

Abstract: Flash flood is one of the most dangerous natural phenomena because of its high magnitudes and sudden occurrence, resulting in huge damages for people and properties. Our work aims to propose a state-of-the-art model for susceptibility mapping of the flash flood using the decision tree random subspace ensemble optimized by hybrid firefly–particle swarm optimization (HFPS), namely the HFPS-RSTree model. In this work, we used data from a flood inventory map consisting of 1866 polygons derived from Sentinel-1 C-band synthetic aperture radar (SAR) data and a field survey conducted in the northwest mountainous area of the Van Ban district, Lao Cai Province in Vietnam. A total of eleven flooding conditioning factors (soil type, geology, rainfall, river density,

elevation, slope, aspect, topographic wetness index (TWI), normalized difference vegetation index (NDVI), plant curvature, and profile curvature) were used as explanatory variables. These indicators were compiled from a geological and mineral resources map, soil type map, and topographic map, ALOS PALSAR DEM 30 m, and Landsat-8 imagery. The HFPS-RSTree model was trained and verified using the inventory map and the eleven conditioning variables and then compared with four machine learning algorithms, i.e., the support vector machine (SVM), the random forests (RF), the C4.5 decision trees (C4.5 DT), and the logistic model trees (LMT) models. We employed a range of statistical standard metrics to assess the predictive performance of the proposed model. The results show that the HFPS-RSTree model had the best predictive performance and achieved better results than those of other benchmarks with the ability to predict flash flood, reaching an overall accuracy of over 90%. It can be concluded that the proposed approach provides new insights into flash flood prediction in mountainous regions.

Keywords: flood modeling; firefly; Particle Swarm Optimization; flash flood; GIS; remote sensing; rainfall

---

1. Introduction

Flash floods that occurr in tropical and semi-tropical areas, caused by extraordinary rainfall, are one of the most dangerous natural phenomena due to the significant socio-economic damage and loss of human lives, particularly in the frequent cyclone regions in Southeast Asia [1,2]. Floods are often classified into different types, i.e., city flooding, river flooding, coastal flooding, and flash flood [3], of which flash floods are more vulnerable and severe because of their speed in short timescales [4,5]. Prior studies suggest that most of the areas are exposed to offensive and destructive flooding, resulting in an increase in the huge damages, casualties, and financial losses during flooding events. Thus, in order to prevent and control the floods, the susceptibility areas, where the potential flood risks are high, should be identified and mapped [6]. On the other hand, human factors, i.e., deforestation and unplanned land-use changes, also considerably contribute to the occurrence of sudden flooding [7] because forests play an important role in reducing surface runoff and transferring excess water to groundwater. Moreover, population growth causes land conversion from forested areas to new settlements built in flood-prone areas. This situation becomes more severe because of the impacts under a changing climate along with land-use changes, which is anticipated to exceed 1 trillion US$ in damage by 2050 [8]. However, accurate and timely prediction of the flash flood still remains challenging because of the complex nature of this phenomenon [9]. Thus, the development of a cost-effective, reliable, and precise accuracy model for predicting and mapping the occurrence of flash floods in areas with high and frequently-induced rainfall is essential in order to support sustainable land-use planning [10].

A previous literature review shows that a large number of studies have been conducted to predict the probability of flooding based on the three main categories, ranging from the traditional analysis, the rainfall-runoff approach to the pattern classification [11]. In recent years, the rapid development of innovative technologies involved earth observations (EO), geographic information system (GIS)-based approaches, and machine learning techniques, which have been proven as promising tools to account for the complexity of spatial flood modeling [12]. Importantly, the integration of satellite remotely sensed imagery and GIS data had been proven as an effective way to map and evaluate flash flood damages [13,14]. For instance, Klemas [14] reported the use of satellite imagery and modeling techniques to predict flood vulnerability, whereas Lee et al. [15] reported the usability of the random forest methods for mapping flood vulnerability in the metropolis. Recently, Khosravi, et al. [16] used GIS-based frequency and weight ratio statistical bivariate statistical models for mapping flash flooding susceptibility. A wide range of attempts have been made to map flash flooding using various

artificial intelligence techniques optimized by metaheuristic algorithms for flooding capacity [17,18]. More recent studies used different machine learning algorithms in predicting and zoning the flash flooding areas [19–21]. However, only a few studies integrated remotely sensed data and spatial data in machine learning techniques for improving the accuracy of spatial prediction of flash floods, despite the fact that air-borne remote sensing data provide a number of benefits such as easier repeatability, low cost, and wider area coverage [21,22], resulting in a lack of cost-effective, precise, and timely models for the susceptibility mapping of flash floods. Thus, this study aims at developing a state-of-the-art model incorporating Sentinel-1 C band free-of-charge data and an advanced machine learning algorithm using the decision tree-based random subspace optimized by hybrid swarm intelligence, namely the HFPS-RSTree model, for the spatial prediction of flash floods in a mountainous area in Northwestern Vietnam.

2. The Employed Algorithms

2.1. Decision Tree Algorithm

The decision tree (DT) algorithm is a simple supervised learning classifier [23]. While other supervised learning algorithms collate all available features together to determine each individual label, the DT operates multiple steps based on decision rules to decide the outcome of a label class. It creates a tree-like structure—that is, nodes represent tests on attributes while branches and leaves represent the consequences of the tests and a category label, respectively. At each node, tests can be applied to one or more of the attributes, namely univariate and multivariate applications. The univariate application analyzes a single attribute while the multivariate application simultaneously tests for one or more attributes. For instance, the Gini index can be applied for single-attribute splits or univariate applications, whereas the support vector machine [24] can be used for the multivariate approaches.

The benefit of this approach is that it is simple and flexible. It can not only be applied to both categorical and continuous data, but it also performs rapidly due to its requirement of simple mathematical calculations. Its structure can be easily visualized for interpretation. However, it sometimes provides a non-optimal result or overfitting. Overfitting can be solved by removing branches.

2.2. Random Subspace Ensemble

Ensemble-based learning techniques are well-known methods for multiple classifiers [25,26], of which the random subspace (RS), proposed first by Ho [23], has proven to be one of the most potent techniques. To handle the deficiency of a single decision tree method, RS takes full advantage of multiple classifiers that are decision trees. To predict flash flood output, each decision tree votes for the flood class and the majority vote of all decision trees is an aid to decide a final outcome. This can reduce overfitting and non-optimal resolution, which may be major issues of single-classifier approaches. More significant improvements have been reported by training each classifier with a random subset of the reference data as opposed to using only a subset of input attributes for that classifier. Although this reduction process may reduce the performance of individual classifiers, it can deal with too strong correlations causing unreliable solutions. It also reduces the amount of calculation time, while the remaining reference data could be adopted for the independent accuracy assessment of the random subspace algorithm.

2.3. Hybrid Firefly–Particle Swarm Algorithm (HFPS)

One difficulty in using ensemble-based learning for predicting flash floods is in optimizing the model parameters, and in this context, metaheuristic optimization algorithms have demonstrated their superiority [27,28]. In this research, a hybrid optimization algorithm (HFPS) proposed by Aydilek [29], which is an integration of the firefly algorithm (FA) and particle swarm optimization (PSO), was used. The reason for this selection is that the HFPS algorithm inherits the fast computation of PSO and the

robustness of FA to form a new powerful algorithm. Consequently, the HFPS algorithm outperforms benchmark algorithms in various engineering problems [29]. Overall, the procedure of the HFPS can be briefly described as follows.

(1) Determine the population of the swarm, the position and the velocity for each particle, and the total number of iterations used.
(2) Establish a cost function to measure the fitness of each particle, called particle best (*pbest*), and then compare all *pbests* to obtain the global best (*gbest*).
(3) For each iteration, calculate and update the position (*Pos*) and the velocity (*Vel*) for all particles in the swarm using Equations (1) and (2) and then compute *pbest* and *gbest*. If the fitness is not improved, *Pos* and *Vel* for each practice will be updated using Equations (3) and (4).

$$Pos_i(t+1) = Pos_i(t) + B_0 e^{-\gamma r_{ij}^2}\left(Pos_i(t) - gbest^t\right) + a \tag{1}$$

$$Vel_i(t+1) = Pos_i(t+1) - Pos_i(tp) \tag{2}$$

$$Vel_i(t+1) = wVel_i(t) + C_1 r_1(pbest_i(t) - Pos_i(t)) + C_2 r_2(gbest(t) - Pos_i(t)) \tag{3}$$

$$Pos_i(t+1) = Pos_i(t) + Vel_i(t+1) \tag{4}$$

where Pos_i and Vel_i are the position and the velocity of *i*-th particle or *i*-th firefly, *a* is the random parameter from 0 to 1, r_{ij} is the distance between the two fireflies, γ is the light absorption coefficient of FA, B_0 is the attractiveness value, *w* is the *inertia* weight of PSO, *tp* is the temporal position, r_1 and r_2 are random parameters $\in [0, 1]$, C_1 and C_2 are the acceleration coefficients, and *t* is the current iteration.

(4) Compute the best *gbest* in all iterations, and then extract the coordinates of the particle with this *gbest*. The coordinate values are called the optimized parameters for the flash flood ensemble model.

3. Study Area and Spatial Data

3.1. Descriptions of the Study Site

We conducted the current work in the Van Ban district, located in the mountainous the Lao Cai Province (approximately 263 km from Northwest Hanoi), Vietnam. The total area is approximately 1435 km^2, accounting for around 22.5% of Lao Cai Province. It lies between latitudes of 21°57′32″N and 22°17′12″N and between longitudes of 103°57′18″E and 104°30′38″E (Figure 1).

The study area has a complex terrain condition, lying between two large mountain ranges, the Hoang Lien Son in the northwest and the Con Voi in the southeast. The topography contains approximately 90% mountainous area and around 10% low land area. The former area consists of various hills and mountainous ranges, with altitude ranging from 700 m to 1500 m and an average slope from 25 to 35°, exceeding 50° in some areas. The remaining areas are valleys at an altitude of 400–700 m. The highest place is located in the Nam Chay commune, at a height of approximately 2875 m, while the lowest is along the Nam Chan stream area, with an altitude of 85 m. In the study area, there are small streams and springs starting from the Hoang Lien Son and the Nui Voi mountainous areas and discharging into the Hong River in the northeast. The study area is highly vulnerable to flash floods due to the complex terrain and dense drain network; they particularly occur when rapid runoff from hilly and mountainous areas discharges quickly into Ngoi Nhu, Nam Tha, and Ngoi Chan streams in a short time before reaching the Hong River [30].
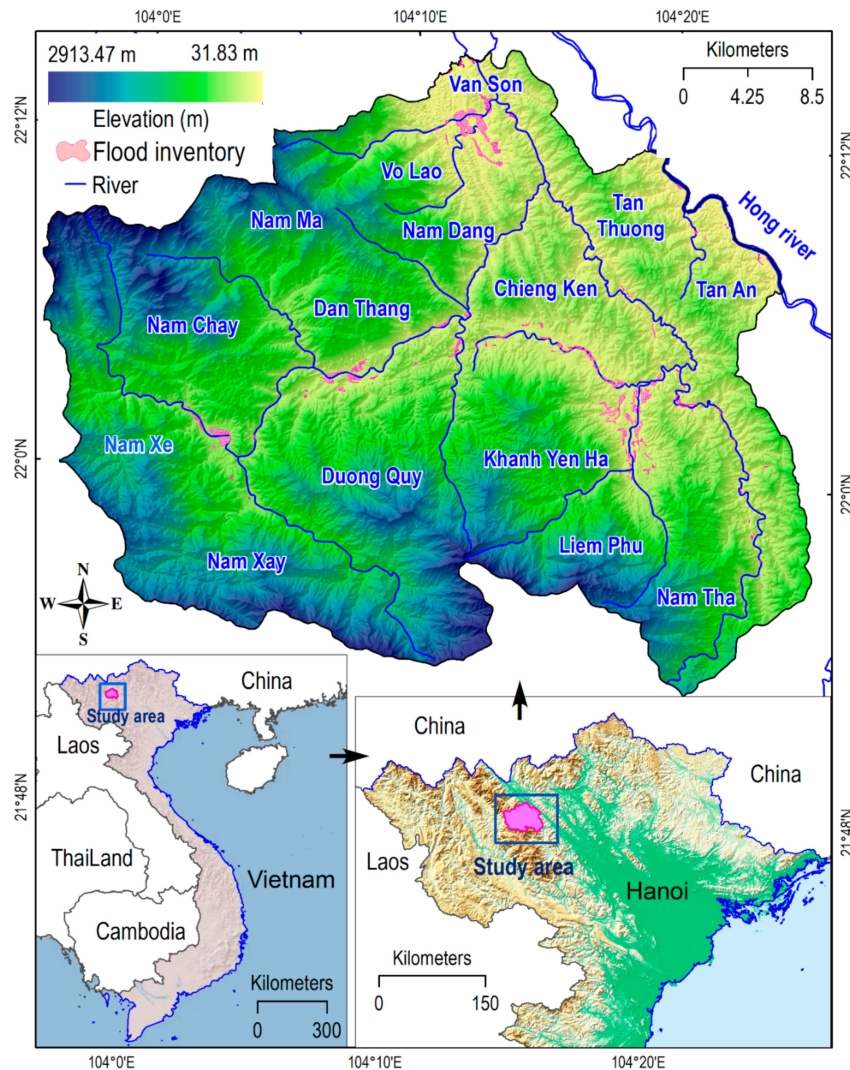
Figure 1. Location map of the Van Ban district and flooded inventories.

Geologically, the study site has a total of fifteen formations and a complex outcrop. Seven formations occupy over 89% of the total study site: Ye Yen Sun (26.98%), TLNT (21.60%), Sinh Quyen (11.66%), Bac Ha (11.46%), PS complex (5.98%), Cam Duong (6.01%), and Suoi Bang (4.87%) (Figure 2a). The dominant lithology area consists of biotite granite, marble, motley limestone, clay shale, quartz-plagioclase-biotite, clay sericite shale schist, and crystalline schist (See Table 1).

The study site has a subtropical monsoon climate with two seasons, of which the rainy season lasts from April to September and the dry season starts in October and ends in March. Average annual precipitation rates range from 1500 mm to 2500 mm and are mainly allocated to the rainy season, making up 70.74–89.25% of the total annual precipitation. Noticeably, very high precipitation intensity events often occur in the rainy season within a short period, observed in steep slopes, leading to the frequent occurrence of flash floods along with landslides in the case study.

Table 1. Geological characteristics in the study area.

No.	Formation	Symbol	Main Lithology
1	Ye Yen Sun	γ/E1ys	Biotite granite, biotite-amphibol granite, granite biotite, and granite biotite-amphibol granite pegmatite
2	TL-NT	K_1ntl-τλK_1nk	Tufogen conglomerate, tufogen sandstone, shale, and black coal shale-quartz orthophyry
3	Sin Quyen	PR2sq	Feldspar-biotite schist, biotite interlaced with quartz mica, mica schist-graphite, biotite, feldspar-mica schist, and tremolite marble
4	Bao Ha	νPR2bh	Gabrodiabas, diabase, gabbro amphybolit, and amphybolit
5	Po Sen Complex	γ/PZ1ps	Aplite, banded plagio-granite, diorite, granodiorite, and pegmatite veins
6	Cam Duong	ε1cd	Sandstone, quartz-carbonate schist, actinolite schist, quartzite, conglomerate, quartz-mica schist, and black schist
7	Suoi Bang	T3n-rsb	Sandstone, siltstone, claystone, claystone mixed coal, and coaly lenses
8	Da Dinh	PR3đđ	Marble, dolomite, dolomite, and tremolite marble
9	Xom Giau	γPR2xg	Granit microcline, granite aplit, and granite pegmatite
10	Phu Sa Phin	ξεγK2pp	Syenite porphyry, granosyenite porphyry, syenite porphyry, granite porphyry, and granite felspar
11	Nam Thep	J1nt	Sandstone, siltstone, thin layer interbedded claystone, and black shale lens
12	Chang Pung	∈2cp	Clay shale, marl, and oolitic limestone
13	YYS Complex	YYS	Granit microcline, granit aplit, and granit pegmatite
14	Tram Tau		Tuffogenic shale, siltstone, clay shale, tufogen conglomerate, tufogen sandstone, coal-bearing shale, and tuffaceous rhyolite
15	Quaternary	Qa	Granule, breccia, boulder, sand, grit, clay, and silt

3.2. Data Collection

3.2.1. Flash Flood Inventory Map

Flood inventory maps are often required in order to investigate the relationship between flood and causative agents [27,31,32]. To prepare the flood-prone map, the initial step is to acquire the relevant data and to construct a spatial geodatabase. In the present study, a total of 200 flash flood locations was interpreted using Sentinel-1 C band free-of-charge data to generate an inventory map, as suggested by Nguyen et al. [21]. Accordingly, 2653 flash flooding polygons occurring in the rainfall season during the year 2018 were identified. These flooded polygons were utilized to construct an inventory map for the study site, of which a total of 1858 (Figure 1) polygons were employed for the training phase to predict flash floods occurring in the study site, and the remaining 796 polygons were utilized for the testing phase to validate the predictive performance [31].

3.2.2. Flash Flood Indicators

It is widely recognized that flash flooding events often take place on a local scale, depending mainly on rainfall, land use, topography, and soil features of the region [16,31,33,34]. Therefore, identification of the conditioning factors associated with flash floods is often required in predicting the possibility of flash floods occurring in a specific region. However, the conditioning or influencing factors in each flash flood event are complicated and depend on a number of factors involved. In the current work, we selected eleven influencing factors for flash flood modeling, including geology, soil type, river density, rainfall, slope, elevation, aspect, plan curvature, profile curvature, TWI, and NDVI, based on the literature reviews [19,35,36]. Herein, geology, soil type, and river density indicators were derived from the geology map with a scale of 1:50,000, acquired from the geological and mineral resources map of the Van Ban district [37] and the soil type and the topographic maps of Vietnam at a scale of 1:50,000. The rainfall indicator was obtained from the stations in the study area, whereas the slope, elevation, plan curvature, aspect, profile curvature, and TWI conditioning factors were computed from the ALOS-PALSAR DEM 30 m [38]. The NDVI factor was computed from Landsat-8 imagery (acquired on 20 December 2017) at a spatial resolution of 30 m. All indicators were transformed in a raster format at 30 m spatial resolution to construct a flash flood susceptibility map.

Geology: Geology is an important factor related to flash flooding occurrence [39,40]. This is because different geological terrains have various capacities of water absorption and therefore can be susceptible to rapid runoff generation during high rainfall events, exacerbating the potentiality of severe flooding to downstream regions [27,40]. The geology characteristics were classified into fifteen formations (Figure 2a) consisting of different mother-rocks such as sedimentary, igneous, and metamorphic components. The geological characteristics of the study site are presented and summarized in Table 2.

Soil type: Hydrologically, soil type is among the crucial factors influencing infiltration, runoff generation, and soil erosion processes affecting flash flood characteristics [41]. It is widely recognized that different soil types have different properties (soil moisture, soil texture, and soil profile). Thus, the type of soil directly influences the formation of flash flood flow and its components (i.e., water, muds, and alluvial) [42]. In this study, thirteen soil types were observed in the study area, of which the Fa, Ha, and Fs soil types exceed 90% of the total area, followed by the Fj type and other soils (Figure 2b).

River density: River density plays a vital role in carrying flash flood flow out of the watershed [43]. Although the characteristics of the flash flood may vary according to different topographical conditions, a higher river density likely has a more significant impact on flood flow expansion [16]. For instance, a dense river system in flat areas could lead to rapidly expanding flood flow, while it has a reverse trend in steep areas. Thus, we considered the river density factor, which was generated from the digital elevation model (DEM) (Figure 2c).
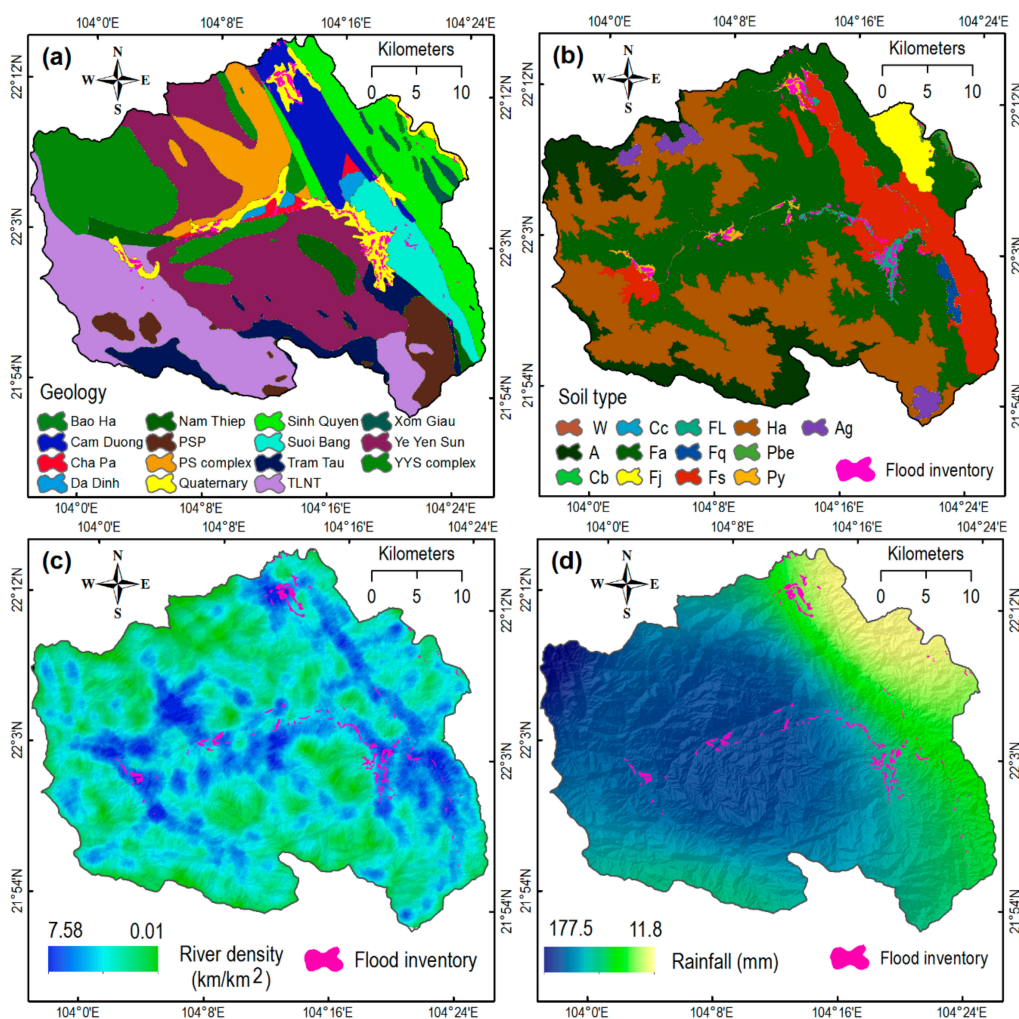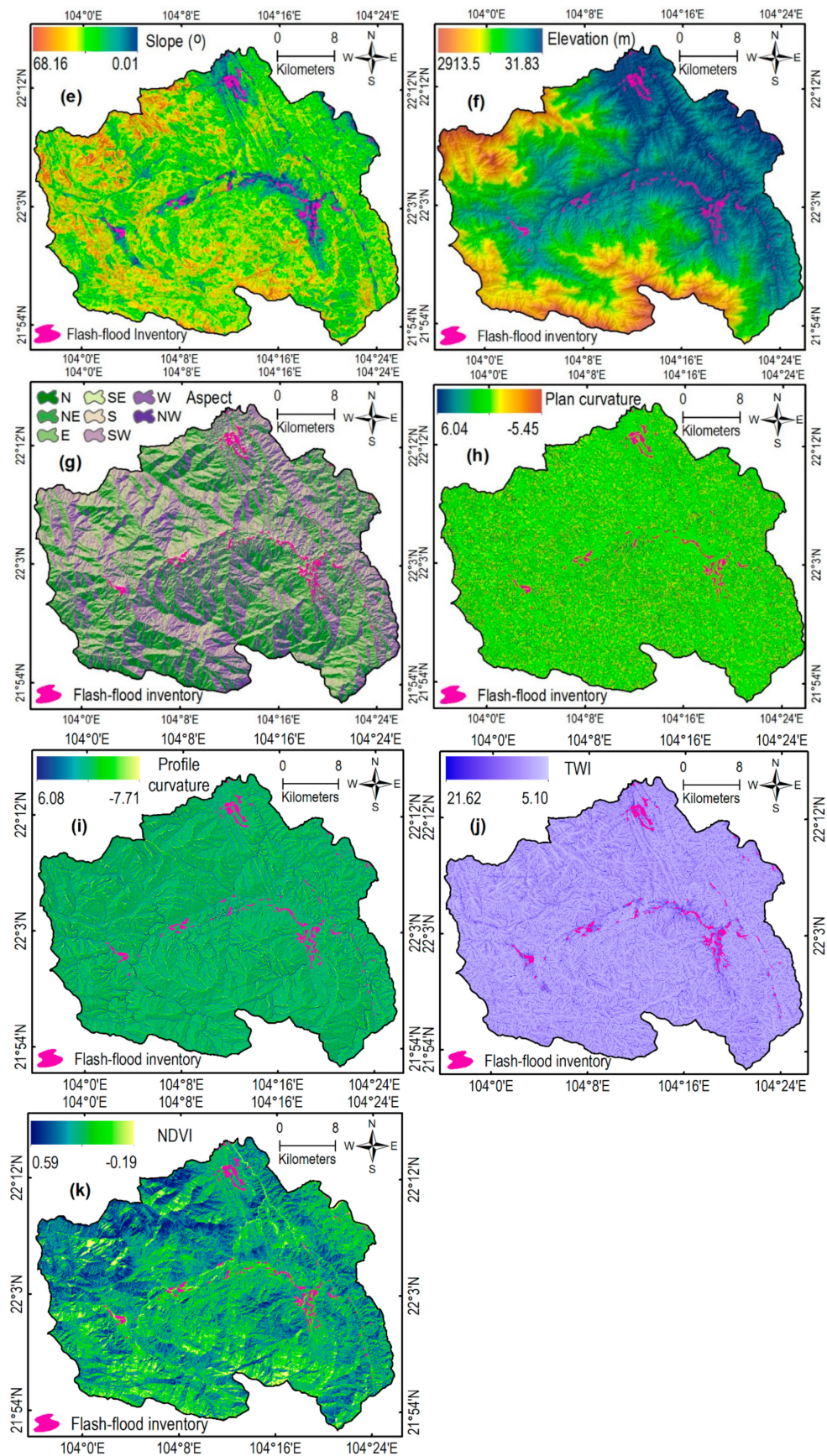

Figure 2. *Cont.*

Figure 2. Flood influencing factors: (**a**) geology; (**b**) soil type; (**c**) river density; (**d**) rainfall; (**e**) slope; (**f**) elevation; (**g**) aspect; (**h**) plan curvature; (**i**) profile curvature; (**j**) TWI; (**k**) NDVI.

Rainfall: Rainfall, which is characterized by rainfall intensity, duration, and frequency, is the most important influencing factor of any hydrological processes within a watershed. Although flash flood flow likely depends on a number of factors involved, high rainfall intensity tends to contribute to forming high energy and fast mass transferring of the flash flood in a specific area. As the study site was located in a tropical monsoon region with high rainfall intensity occurring in steep areas, the site is highly vulnerable to flash floods and landslides [19]. In this work, the highest cumulative rainfall events which occurred for 72 h during the last three years were employed to compute the rainfall map using the inverse distance weight (IDW) method [31]. The rainfall level recorded within three days was around 177.5 mm in the west and approximately 11.8 mm in southeastern areas, interpolated in the ArcGIS software (Figure 2d).

Slope and elevation: Slope and elevation are two main drivers of topographical conditions indicating the speed of flash flood flow within a watershed. For example, high altitude coupled with steep slopes has a greater probability of contributing to extreme flash floods, even in the event of low rainfall intensity [19,31]. Meanwhile, high rainfall intensity in flat areas with small slopes will probably result in less flash flood occurrences. It was noted that most of the high slope angle is located in the high elevation areas generating runoff, forming and speeding up flash flood flow in the study area. In contrast, the low slopes along streams (Ngoi Nhu, Nam Tha, and Ngoi Chan) may decrease the capacity of carrying flash flood flow out of the watershed. The slope layer revealed a large amount of variation, ranging from 0.01 to 68.16 degrees (Figure 2e), whereas the elevation map was generated from a DEM with 30 m spatial resolution, showing elevation ranging between approximately 32 m and 3000 m (Figure 2f). The DEM was generated using the national topographic map at a scale of 1:50,000, acquired from the Vietnam Institute of Geosciences and Mineral Resources.

Aspect: The aspect factor is another component of topography presenting the potential stream flow direction and sensitivity processes that regulate the components of a flash flood [44,45]. The aspect map created for the study area was categorized into eight classes [16,46], as shown in Figure 2g. The positions of flash floods occurring in the case study were corresponded to the aspect map, showing the influencing level of this indicator on the probability of flash flood occurrences.

Plan curvature: Plan curvature reveals the morphometrical characteristics and indicates the change in a slope's inclination or aspect [45]. Plan curvature may largely affect the acceleration and deceleration of water and muds/sediment during downslope flow and, therefore, likely influences the velocity of the flash flood [45].

More importantly, the plan curvature influences the divergence of flow, thus deeply affecting the flash flood energy and mass transfer from upstream to downstream in a specific watershed. Plan curvature values are generally defined as concavity (positive), convexity (negative), and flat (zero), which are largely affected by the runoff processes [47,48]. We generated the plan curvature map using the DEM with a pixel size of 30 m × 30 m. Figure 2h shows that roughly 75% of the study area is covered by the concave zones.

Profile curvature: Profile curvature corresponds to the direction of the maximum slope, thus indicating the convergence and the divergence of a surface flow [49]. A negative value in the top of the mountains indicates that the surface is upwardly convex, while a positive value reveals that the surface is upwardly concave at that location (Figure 2i). A zero value of the profile curvature shows that the surface is linear. The profile curvature often influences the acceleration or deceleration of flash flow across the surface area. We used the DEM with a grid of 30 m × 30 m to generate the profile curvature map in the current work. Figure 2i shows that approximately 80% of the study area is occupied by concave zones.

Topographic wetness index (TWI): The TWI is considered the most critical parameter measuring topographic controls of basic hydrological processes [50]. The TWI map was created using the altitude map by applying Equation (5) [51,52].

$$TWI = In\frac{(A_s)}{(\tan\beta)} \tag{5}$$

where A_s is an upslope area, and β is the slope angle at one pixel.

Figure 2j shows that the TWI ranges from 5.10 to 21.62, in which high TWI values indicate the greater capability for water accumulation in the study area.

Normalized difference vegetation index (NDVI): The NDVI is a crucial indicator, showing the degree of vegetation coverage, which largely influences flood processes [31]. Greater NDVI values demonstrate higher vegetation coverage, while lower values indicate less vegetation. Previous studies show that low vegetation coverage indicates high probability of flash flood occurrence [27,36].

The NDVI map (Figure 2k) for the study area was calculated and computed using Landsat-8 Operational Land Imager (OLI) multispectral imagery with a pixel size of 30 m × 30 m for predicting flash flood susceptibility (Equation (6)) [31].

$$NDVI = (NIR - RED)/(NIR + RED) \tag{6}$$

where RED and NIR are the surface reflectance of the red and the near-infrared wavelengths derived from Landsat-8 OLI, respectively.

The NDVI values range between −0.19 and 0.59, indicating the different impact levels of vegetation coverage on flash flood processes.

4. Proposed HFPS-RSTree for Flash Flood Susceptibility Modeling

In this work, the flash flood indicators and inventories were processed using ArcMap 10.6 (See Figure 3). The HFPS-RSTree model was computed and constructed by the authors in the Matlab environment. The RSTree is available in the API Python Weka Wrapper [53], whereas the HFPS code in Matlab was introduced by Aydilek [29].
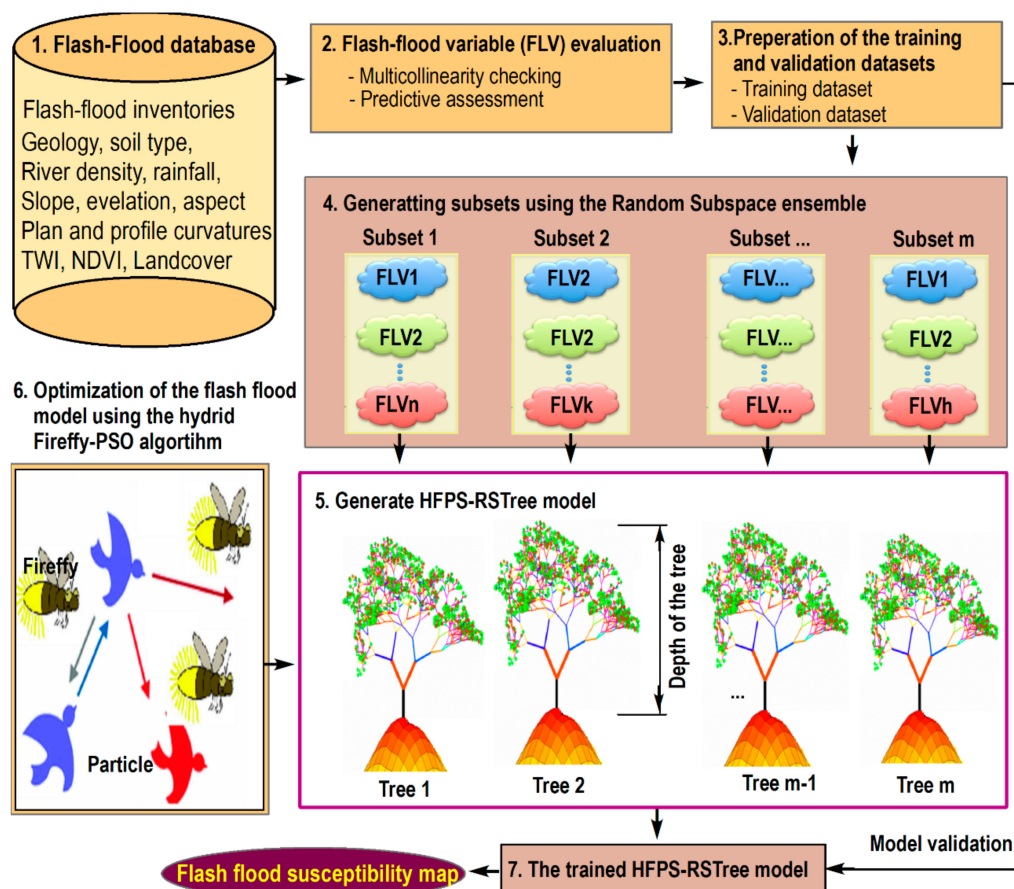

Figure 3. The proposed HFPS-RSTree model for flash flood susceptibility mapping.

4.1. Database Establishment

The geospatial database for flash floods in the Van Ban district was constructed using ArcCatalog. The flash flood inventory map and the eleven influencing factors were converted into a raster format with a spatial resolution of 30 m. Note that in the proposed model, a number of factors (slope, elevation, plan curvature, profile curvature, TWI, NDVI, river density, and rainfall) were represented as continuous values, while the remaining categorical indicators, including the aspect, soil type, and lithology factors, were converted into numeric values using the method suggested in [31].

The flash flood inventories were randomly split into two subsets for the flash flood modeling in the next phase, of which the training dataset consisted of 1858 polygons and the validating dataset contained 796 polygons.

4.2. Configuration of the HFPS-RSTree Model

The structure of the proposed HFPS-RSTree model consists of three algorithms: the RS ensemble, the decision tree algorithm, and the HFPS optimization. Using the training dataset, the RS ensemble will generate *m* subsets (*m-ss*), and each subset will have *p* flash flood indicators (*p-ffi*). Each subset will be used to generate a tree using the decision tree algorithm, where the maximum depth of the tree (*d-max*) must be defined. Therefore, the HFPS-RSTree model was configured using values determined by the three above parameters, *m-ss*, *p-ffi*, and *d-max*.

Herein, the HFPS algorithm was integrated in order to search for and optimize the best combination of them autonomously. A number of parameters used for the HFPS algorithm were suggested by Aydilek [29]. Accordingly, the acceleration coefficient was set to 1.49445 for both C_1 and C_2. The swarm population was 30, whereas the total number of iterations was 1000. The searching space was as follows: *m-ss* ∈ [10–500], *p-ffi* ∈ [1–11], and *d-max* ∈ [1–30]. It should be noted that the default maximum depth of the tree was computed using an integer value.

4.3. The Objective Function and Training the HFPS-RSTree Model

To quantitatively measure the best combination of the three parameters (*m-ss*, *p-ffi*, and *d-max*), an objective function (ObjF) must be established, and in this work, the ObjF suggested in [27] was used, as shown below.

$$\text{ObjF} = \frac{1}{n} \sum_{i=1}^{n} \left(Predict_i - Target_i \right)^2 \tag{7}$$

where $Predict_i$ is the estimated value of the HFPS-RSTree model; $Target_i$ is the flash flood inventory value; *n* is the total number of samples.

4.4. Model Performance Assessment

The model's performance was assessed using a number of statistical measures, such as the receiver operating characteristic (ROC) curve and area under the curve (AUC), the overall accuracy, and the kappa coefficient, because these metrics have been widely used for checking the performance of flash flood modeling in the literature. Detailed formulas of these statistical measures can be found in [27,54,55].

5. Results and Analysis

5.1. Variable Importance Ranking

In this study, variable importance was assessed using the random forest algorithm. The results in Table 2 show that the slope, aspect, and elevation factors had higher importance for assessing flood risk, thus minimizing the impact on the occurrence of flash floods in the case study. Remarkably, the slope factor is likely the most important factor for predicting the spatial distribution of the flash flood in this study. Other remaining factors, such as the aspect, elevation, plan curvature, profile curvature, TWI,

NDVI, river density, lithology, rainfall pattern, and soil type, were ranked from 2 to 10, respectively, in the occurrence of floods in the study area.

Table 2. The relative importance of flash flood indicators using the random forest model.

Indicators	Average Impurity Decreased	Number of Nodes Used	Ranking
Slope	0.42	15,290	1
Aspect	0.41	7763	2
Elevation	0.36	17,282	3
Plan curvature	0.35	10,012	4
Profile curvature	0.32	9738	5
TWI	0.29	9960	6
NDVI	0.27	8955	7
River density	0.26	10,551	8
Lithology	0.25	2668	9
Rainfall pattern	0.23	9042	10
Soil type	0.21	3370	11

5.2. Model Performance and Comparison

Figure 4 and Table 3 show the predictive performance of the HFPS-RSTree, the RF, the C4.5-DT, the LMT, and the SVM algorithms in the training and the testing phases. The AUC for the prediction-rate curve demonstrates how well the model predicts the flash flood. The results in Figure 4 and Table 3 show that the proposed algorithms performed very well in both the training and the validation datasets. It could be observed that the AUC values of the HFPS-RSTree, the RF, the C4.5-DT, the LMT, and the SVM models were 0.973, 0.970, 0.920, 0.945, and 0.964, respectively, in the training data, whereas these corresponding values were 0.967, 0.965, 0.914, 0.927, and 0.951, respectively, in the testing data, showing satisfactory results for the spatial prediction of flash floods in the study area.

Table 3. Comparison of the HFPS-RSTree, the RF, the C4.5-DT, the LMT, and the SVM for the flash flood modeling.

| Metrics	HFPS-RSTree	RF	C4.5-DT	LMT	SVM
Training Phase					
True positive	1811	1817	1766	1799	1654
True negative	1626	1612	1613	1581	1753
False positive	37	31	82	49	194
False negative	222	236	235	267	95
Positive predictive values (PPV) (%)	98.00	98.32	95.56	97.35	89.50
Negative predictive values (NPV) (%)	87.99	87.23	87.28	85.55	94.86
Sensitivity (%)	89.08	88.50	88.26	87.08	94.57
Specificity (%)	97.78	98.11	95.16	96.99	90.04
Overall Accuracy (%)	92.99	92.78	91.42	91.45	92.18
Kappa	0.860	0.856	0.823	0.829	0.844
AUC	0.973	0.970	0.920	0.945	0.964
Validation Phase					
True positive	782	783	763	779	681
True negative	677	656	665	644	740
False positive	12	11	31	15	113
False negative	117	138	129	150	54
PPV (%)	98.49	98.61	96.10	98.11	85.77
NPV (%)	85.26	82.62	83.75	81.11	93.20
Sensitivity (%)	86.99	85.02	85.54	83.85	92.65
Specificity (%)	98.26	98.35	95.55	97.72	86.75
Overall Accuracy (%)	91.88	90.62	89.92	89.61	89.48
Kappa	0.838	0.812	0.799	0.792	0.790
AUC	0.967	0.965	0.914	0.927	0.951
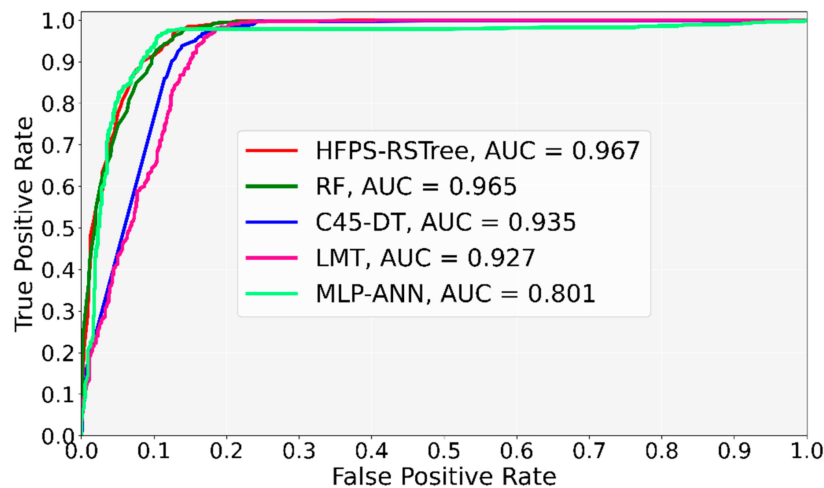
Figure 4. The AUC for the flash flood models using the validation dataset.

Overall, the HFPS-RSTree model yielded the highest predictive performance both in the training phase (kappa = 0.860, overall accuracy = 92.99) and in the testing phase (kappa = 0.838, overall accuracy = 91.88), followed by the RF algorithm. In contrast, the SVM algorithm produced the lowest performance (kappa = 0.844, overall accuracy = 92.18 in the training set and kappa = 0.790, overall accuracy = 89.48 in the testing set). The results showed that the ensemble-based methods using the decision tree learning algorithm yielded better predictive performance than those of well-known machine learning (ML) algorithms in this study. Our results are in agreement with the recent studies reported by [21,56]. We conclude that, among the five ML algorithms, the proposed model using a combination of the decision tree ensemble-based algorithm and an advanced optimization technique produced the most precise results for the spatial prediction of flash floods in the study area.

Table 4 shows the Wilcoxon rank-sum test results for five ML models. It can be clearly seen that all pairwise comparisons were statistically significant except the RF vs. C4.5-DT (*p*-value = 0.099) and the RF vs. LMT (*p*-value = 0.055).

Table 4. Wilcoxon rank-sum test for the five flash flood models.

No.	Pairwise Comparison	Z Statistics Value	*p*-Value	Statistical Significance
1	HFPS-RSTree vs. RF	3.577	0.0003	Yes
2	HFPS-RSTree vs. C4.5-DT	2.598	0.0094	Yes
3	HFPS-RSTree vs. LMT	4.274	<0.0001	Yes
4	HFPS-RSTree vs. SVM	−10.404	<0.0001	Yes
5	RF vs. C4.5-DT	−1.647	0.0996	No
6	RF vs. LMT	1.921	0.0548	No
7	RF vs. SVM	−11.037	<0.0001	Yes
8	C4.5-DT vs. LMT	6.870	<0.0001	Yes
9	C4.5-DT vs. SVM	−13.251	<0.0001	Yes
10	LMT vs. SVM	−10.846	<0.0001	Yes

5.3. Flash Flood Susceptibility Map

Since the HFPS-RSTree produced the best predictive performance regarding the AUC, overall accuracy, and kappa index among the five ML models for predicting flash flood risk, we employed this model to compute the flash flood susceptibility map in the study area. The final model results were transformed into a raster format and interpreted in the ArcGIS environment. The flash flood susceptibility map was generated and visualized, as shown in Figure 5. The susceptibility index was varied from 0.01–1.00, of which the darker blue color in the map represents the high-frequency occurrences of flash floods. In contrast, the brighter yellow color shows the low probabilities of flash flood risk.
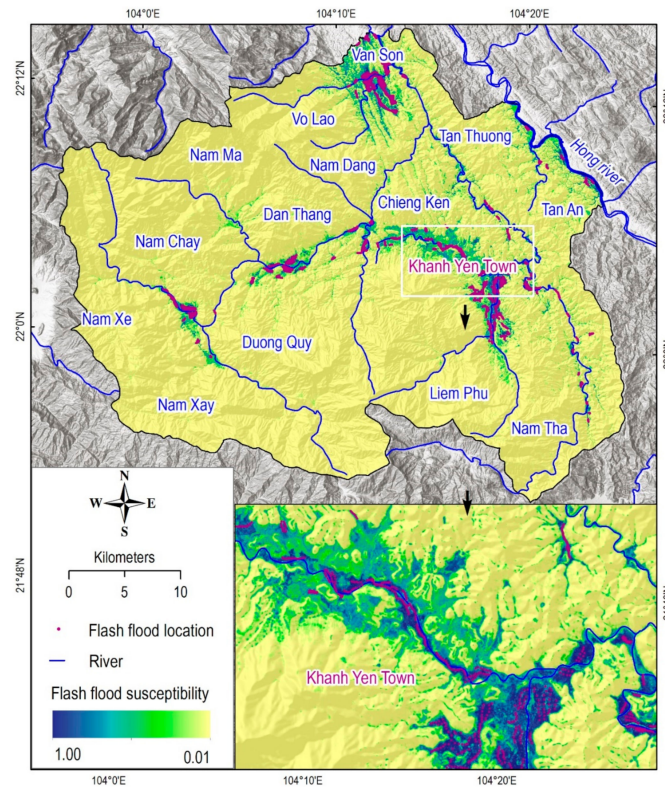
Figure 5. Flash flood susceptibility map of the Van Ban district using the proposed HFPS-RSTree model.

The visual interpretation in Figure 5 shows that the highest possibility of flash floods likely occurred in Khanh Yen town, followed by the Van Son, the Dan Thang, and the Nam Chay communes. These areas are flat and are located closer to the rivers that were likely the most affected by the flash flood risk during the last five years. Therefore, the policymakers or local authorities should pay more attention to these areas when prioritizing the development of flood risk measures. In contrast, the other areas have a lower probability of flash flood. This is possibly due to the terrain slope of these areas being steep, which may prevent water accumulation.

6. Discussion

In the last decade, the adverse effects of global warming have resulted in a higher frequency of floods in various regions around the globe [57–59]; therefore, new studies to develop better tools for flood prediction are highly necessary. In this research, we proposed a new modeling approach, named the HFPS-RSTree, for the spatial prediction of flash flood susceptibility, with a case study of a high-frequency torrential rainfall area. The proposed HFPS-RSTree is a new machine learning ensemble consisting of three components: the decision tree (Tree), the random subspace (RS), and the HFPS technique. Herein, the flood ensemble model was created using the RS and Tree, while the HFPS was integrated in order to optimize the model.

As a result, the precise accuracy of the HFPS-RSTree model for the spatial prediction of flash floods indicates that a combination of the HFPS, the RS, and the Tree techniques is efficient in predicting flash flood potential areas. This is due to the mechanism of ensemble-based learning, in which the RS plays a vital role in generating flood subsets to ensure the diversity of the final ensemble model. Thus, during the last ten years, decision tree ensemble-based learning methods have confirmed their high prediction power in various domains [60–62] in which flood studies have been conducted [15,26]. The results of this HFPS-RSTree model in this regard confirm the above statement.

The success of building the HFPS-RSTree model is also strongly dependent on three parameters, namely the number of subsets (*m-ss*), the number of indicators used in these subsets (*p-ffi*), and the

maximum depth of the tree (*d-max*); therefore, these parameters should be carefully determined. The highest performance of the HFPS-RSTree model, compared to the RF, C4.5-DT, LMT, and SVM, shows that these parameters have been searched and optimized successfully by the HFPS algorithm. This is a reasonable result because the HFPS has proven its capacity in searching and optimizing parameters in various engineering domains recently [29].

In this research, eleven indicators were considered for flash flood modeling, and the superior performance of the HFPS-RSTree model demonstrates that these indicators were selected and processed properly. Among these indicators, the slope degree and slope direction are likely the most important factors for mapping and predicting flash floods in the present study. This result is consistent with the results reported by Tehrany et al. [63], showing that flood-prone areas are often located in flat areas and low altitudes. On the other hand, as the slope increases, the rate of water infiltration decreases, and the water velocity increases [16].

7. Concluding Remarks

We proposed a new ensemble machine learning model, namely the HFPS-RSTree model, to map the spatial prediction of flash floods in the present work. The Van Ban district, located in the northern mountainous region of Vietnam, was selected as a case study. The predictive performance results of the HFPS-RSTree were compared with the four machine learning techniques, namely the RF, C4.5 DT, LMT, and SVM models. The conclusions which can be drawn from the results of the current study are the following:

- The integration of HFPS, RS, and Tree, which results in a new ensemble model, is capable of predicting flash floods accurately. HFPS is a useful tool for optimizing the RSTree model.
- The HFPS-RSTree model yielded higher predictive performance than those of other benchmarks such as the RF, C4.5-DT, LMT, and SVM models, which was confirmed by the Wilcoxon rank-sum test. This denotes that the HFPS-RSTree model is a promising tool to be considered for flash flood studies.
- Regarding the 11 conditioning flash flood indicators, the slope and the aspect factors are the most important features.
- Finally, the flash flood susceptibility map may assist local authorities and policymakers with watershed management and sustainable development in the district.

Author Contributions: Conceptualization, data curation, and investigation: V.-H.N., P.-T.T.N., P.V.H., and D.C.P.; methodology: V.-H.N., P.-T.T.N., T.D.P., J.D., X.S., N.-D.H., D.T.B.; Software: İ.B.A. and D.T.B.; writing—original draft preparation, P.-T.T.N., T.D.P., D.A.T., D.P.C., M.A., R.C.; writing—review and editing, T.D.P., D.J., X.S., and D.T.B. All authors have read and agreed to the published version of the manuscript.

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

References

1. Peduzzi, P. Prioritizing protection? *Nat. Clim. Chang.* **2017**, *7*, 625–626. [CrossRef]
2. Hu, P.; Zhang, Q.; Shi, P.; Chen, B.; Fang, J. Flood-induced mortality across the globe: Spatiotemporal pattern and influencing factors. *Sci. Total Environ.* **2018**, *643*, 171–182. [CrossRef] [PubMed]
3. Kundzewicz, Z.W.; Kanae, S.; Seneviratne, S.I.; Handmer, J.; Nicholls, N.; Peduzzi, P.; Mechler, R.; Bouwer, L.M.; Arnell, N.; Mach, K.; et al. Flood risk and climate change: Global and regional perspectives. *Hydrol. Sci. J.* **2013**, *59*, 1–28. [CrossRef]
4. Schmittner, K.-E.; Giresse, P. Modelling and application of the geomorphic and environmental controls on flash flood flow. *Geomorphology* **1996**, *16*, 337–347. [CrossRef]

5. Yussouf, N.; Knopfmeier, K.H. Application of the warn-on-forecast system for flash-flood-producing heavy convective rainfall events. *Q. J. R. Meteorol. Soc.* **2019**, *145*, 2385–2403. [CrossRef]

6. Postek, K.; Den Hertog, D.; Kind, J.; Pustjens, C. Adjustable robust strategies for flood protection. *Omega* **2019**, *82*, 142–154. [CrossRef]

7. Peptenatu, D.; Grecu, A.; Simion, A.G.; Gruia, K.A.; Andronache, I.; Draghici, C.C.; Diaconu, D.C. Deforestation and frequency of floods in Romania. In *Water Resources Management in Romania*; Negm, A., Romanescu, G., Zeleňáková, M., Eds.; Springer Water; Springer: Cham, Switzerland, 2020; pp. 279–306. [CrossRef]

8. Alfieri, L.; Cohen, S.; Galantowicz, J.; Schumann, G.J.P.; Trigg, M.A.; Zsoter, E.; Prudhomme, C.; Kruczkiewicz, A.; Coughlan de Perez, E.; Flamig, Z.; et al. A global network for operational flood risk reduction. *Environ. Sci. Policy* **2018**, *84*, 149–158. [CrossRef]

9. Edouard, S.; Vincendon, B.; Ducrocq, V. Ensemble-based flash-flood modelling: Taking into account hydrodynamic parameters and initial soil moisture uncertainties. *J. Hydrol.* **2018**, *560*, 480–494. [CrossRef]

10. Costache, R.; Bui, D.T. 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]

11. 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]

12. Tzavella, K.; Fekete, A.; Fiedrich, F. Opportunities provided by geographic information systems and volunteered geographic information for a timely emergency response during flood events in Cologne, Germany. *Nat. Hazards* **2017**, *91*. [CrossRef]

13. Klemas, V. Remote sensing of floods and flood-prone areas: An overview. *J. Coast. Res.* **2015**, *31*, 1005–1013. [CrossRef]

14. Rahmati, O.; Darabi, H.; Panahi, M.; Kalantari, Z.; Naghibi, S.A.; Ferreira, C.S.S.; Kornejady, A.; Karimidastenaei, Z.; Mohammadi, F.; Stefanidis, S.; et al. Development of novel hybridized models for urban flood susceptibility mapping. *Sci. Rep.* **2020**, *10*, 1–19. [CrossRef]

15. Lee, S.; Kim, J.-C.; Jung, H.-S.; Lee, M.J.; Lee, S. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in seoul metropolitan city, korea. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1185–1203. [CrossRef]

16. 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]

17. 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 susceptibilitgy modeling in a high-frequency tropical cyclone area using gis. *J. Hydrol.* **2016**, *540*, 317–330. [CrossRef]

18. Termeh, S.V.R.; Kornejady, A.; Pourghasemi, H.R.; Keesstra, S. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Sci. Total Environ.* **2018**, *615*, 438–451. [CrossRef]

19. 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]

20. 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]

21. Nguyen, V.-N.; Yariyan, P.; Amiri, M.; Dang Tran, A.; Pham, T.D.; Do, M.P.; Thi Ngo, P.T.; Nhu, V.-H.; Quoc Long, N.; Tien Bui, D. A new modeling approach for spatial prediction of flash flood with biogeography optimized chaid tree ensemble and remote sensing data. *Remote Sens.* **2020**, *12*, 1373. [CrossRef]

22. 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, seagrassesand salt marshes during 2010–2018. *Sensors* **2019**, *19*, 1933. [CrossRef] [PubMed]

23. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844.

24. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: New York, NY, USA, 2013.

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

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

27. 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]

28. Ngo, P.-T.T.; Hoang, N.-D.; Pradhan, B.; Nguyen, Q.K.; Tran, X.T.; Nguyen, Q.M.; Nguyen, V.N.; 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]

29. Aydilek, İ.B. A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems. *Appl. Soft Comput.* **2018**, *66*, 232–249. [CrossRef]

30. SYB. *Yen Bai Statistical Year Book 2017*; Statistical Publishing House: Hanoi, Vietnam, 2018; p. 470.

31. 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]

32. Rahmati, O.; Yousefi, S.; Kalantari, Z.; Uuemaa, E.; Teimurian, T.; Keesstra, S.; Pham, T.D.; Bui, D.T. Multi-hazard exposure mapping using machine learning techniques: A case study from Iran. *Remote Sens.* **2019**, *11*, 1943. [CrossRef]

33. Naulin, J.P.; Payrastre, O.; Gaume, E. Spatially distributed flood forecasting in flash flood prone areas: Application to road network supervision in southern france. *J. Hydrol.* **2013**, *486*, 88–99. [CrossRef]

34. 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]

35. Costache, R.; Hong, H.; Pham, Q.B. Comparative assessment of the flash-flood potential within small mountain catchments using bivariate statistics and their novel hybrid integration with machine learning models. *Sci. Total Environ.* **2020**, *711*, 134514. [CrossRef] [PubMed]

36. Hosseini, F.S.; Choubin, B.; Mosavi, A.; Nabipour, N.; Shamshirband, S.; Darabi, H.; Haghighi, A.T. Flash-flood hazard assessment using ensembles and bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Sci. Total Environ.* **2020**, *711*, 135161. [CrossRef] [PubMed]

37. Truong, L.X.; Mitamura, M.; Kono, Y.; Raghavan, V.; Yonezawa, G.; Truong, Q.X.; Do, H.T.; Tien Bui, D.; Lee, S. Enhancing prediction performance of landslide susceptibility model using hybrid machine learning approach of bagging ensemble and logistic model tree. *Appl. Sci.* **2018**, *8*, 1046. [CrossRef]

38. Japan Aerospace Exploration Agency Alos Global Digital Surface Model Alos World 3d—30 m. Available online: https://www.Eorc.Jaxa.Jp/alos/en/aw3d30/index.htm (accessed on 5 July 2019).

39. Skias, S.G. The effectiveness of engineering geology in coping with flash floods: A systems approach. In *Coping with Flash Floods*; Gruntfest, E., Handmer, J., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 2001; pp. 115–122.

40. Vannier, O.; Anquetin, S.; Braud, I. Investigating the role of geology in the hydrological response of mediterranean catchments prone to flash-floods: Regional modelling study and process understanding. *J. Hydrol.* **2016**, *541*, 158–172. [CrossRef]

41. Sangati, M.; Borga, M.; Rabuffetti, D.; Bechini, R. Influence of rainfall and soil properties spatial aggregation on extreme flash flood response modelling: An evaluation based on the sesia river basin, north western italy. *Adv. Water Resour.* **2009**, *32*, 1090–1106. [CrossRef]

42. Lovat, A.; Vincendon, B.; Ducrocq, V. Assessing the impact of resolution and soil datasets on flash-flood modelling. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 1801–1818. [CrossRef]

43. Pallard, B.; Castellarin, A.; Montanari, A. A look at the links between drainage density and flood statistics. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 1019–1029. [CrossRef]

44. Bisht, S.; Chaudhry, S.; Sharma, S.; Soni, S. Assessment of flash flood vulnerability zonation through geospatial technique in high altitude himalayan watershed, himachal pradesh India. *Remote Sens. Appl. Soc. Environ.* **2018**, *12*, 35–47. [CrossRef]
45. Arabameri, A.; Pourghasemi, H.R. 13—Spatial modeling of gully erosion using linear and quadratic discriminant analyses in gis and r. In *Spatial Modeling in Gis and R for Earth and Environmental Sciences*; Pourghasemi, H.R., Gokceoglu, C., Eds.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 299–321.
46. Saha, S.; Roy, J.; Arabameri, A.; Blaschke, T.; Tien Bui, D. Machine learning-based gully erosion susceptibility mapping: A case study of eastern India. *Sensors* **2020**, *20*, 1313. [CrossRef]
47. 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]
48. Arabameri, A.; Pradhan, B.; Pourghasemi, R.H.; Rezaei, K.; Kerle, N. Spatial modelling of gully erosion using gis and r programing: A comparison among three data mining algorithms. *Appl. Sci.* **2018**, *8*, 1369. [CrossRef]
49. Florinsky, I.V. Topographic surface and its characterization. In *Digital Terrain Analysis in Soil Science and Geology*, 2nd ed.; Florinsky, I.V., Ed.; Academic Press: Cambridge, UK, 2016; Chapter 2, pp. 7–76.
50. Sørensen, R.; Zinko, U.; Seibert, J. On the calculation of the topographic wetness index: Evaluation of different methods based on field observations. *Hydrol. Earth Syst. Sci.* **2006**, *10*, 101–112. [CrossRef]
51. 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]
52. Jiang, L.; Ling, D.; Zhao, M.; Wang, C.; Liang, Q.; Liu, K. Effective identification of terrain positions from gridded dem data using multimodal classification integration. *ISPRS Int. J. Geo Inf.* **2018**, *7*, 443. [CrossRef]
53. Reutermann, P. Python3 Wrapper for the Weka Machine Learning Workbench. Available online: https://pypi.Org/project/python-weka-wrapper3/ (accessed on 15 January 2019).
54. Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Bui, D.T.; Pham, B.T.; Khosravi, K. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environ. Model. Softw.* **2017**, *95*, 229–245. [CrossRef]
55. 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]
56. Ha, N.T.; Manley-Harris, M.; Pham, T.D.; Hawes, I. A comparative assessment of ensemble-based machine learning and maximum likelihood methods for mapping seagrass using sentinel-2 imagery in tauranga harbor, New Zealand. *Remote Sens.* **2020**, *12*, 355. [CrossRef]
57. Bubeck, P.; Dillenardt, L.; Alfieri, L.; Feyen, L.; Thieken, A.H.; Kellermann, P. Global warming to increase flood risk on european railways. *Clim. Chang.* **2019**, *155*, 19–36. [CrossRef]
58. Alfieri, L.; Bisselink, B.; Dottori, F.; Naumann, G.; de Roo, A.; Salamon, P.; Wyser, K.; Feyen, L. Global projections of river flood risk in a warmer world. *Earth Future* **2017**, *5*, 171–182. [CrossRef]
59. Carvalho, K.S.; Wang, S. Characterizing the indian ocean sea level changes and potential coastal flooding impacts under global warming. *J. Hydrol.* **2019**, *569*, 373–386. [CrossRef]
60. Kocev, D.; Vens, C.; Struyf, J.; Džeroski, S. Tree ensembles for predicting structured outputs. *Pattern Recognit.* **2013**, *46*, 817–833. [CrossRef]
61. Schnier, S.; Cai, X. Prediction of regional streamflow frequency using model tree ensembles. *J. Hydrol.* **2014**, *517*, 298–309. [CrossRef]
62. Torres-Barrán, A.; Alonso, Á.; Dorronsoro, J.R. Regression tree ensembles for wind energy and solar radiation prediction. *Neurocomputing* **2019**, *326*, 151–160. [CrossRef]
63. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. Spatial prediction of flood susceptible areas using rule based decision tree (dt) and a novel ensemble bivariate and multivariate statistical models in gis. *J. Hydrol.* **2013**, *504*, 69–79. [CrossRef]