







Article

A Comparative Study of Kernel Logistic Regression, Radial Basis Function Classifier, Multinomial Naïve Bayes, and Logistic Model Tree for Flash Flood Susceptibility Mapping

Binh Thai Pham ¹, Tran Van Phong ², Huu Duy Nguyen ³, Chongchong Qi ⁴,
Nadhir Al-Ansari ^{5,*}, Ata Amini ⁶, Lanh Si Ho ^{7,*}, Tran Thi Tuyen ⁸, Hoang Phan Hai Yen ^{9,*},
Hai-Bang Ly ¹, Indra Prakash ¹⁰ and Dieu Tien Bui ¹¹

- ¹ University of Transport Technology, Hanoi 100000, Viet Nam; binhpt@utt.edu.vn (B.T.P.); banglh@utt.edu.vn (H.-B.L.)
 - ² Institute of Geological Sciences, Vietnam Academy of Sciences and Technology, 84 Chua Lang Street, Dong da, Hanoi 100000, Viet Nam; tphong1617@gmail.com
 - ³ Faculty of Geography, VNU University of Science, 334 Nguyen Trai, Hanoi 100000, Viet Nam; huuduy151189@gmail.com
 - ⁴ School of Resources and Safety Engineering, Central South University, Changsha 410083, China; chongchong.qi@gmail.com
 - ⁵ Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden
 - ⁶ Kurdistan Agricultural and Natural Resources Research and Education Center, AREEO, Sanandaj 66177-15175, Iran; a.amini@areeo.ac.ir
 - ⁷ Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
 - ⁸ Department of Resource and Environment Management, School of Agriculture and Resources, Vinh University, Nghe An 470000, Vietnam; ttt.dhv@gmail.com
 - ⁹ Department of Geography, School of Social Education, Vinh University, Nghe An 470000, Vietnam
 - ¹⁰ Department of Science & Technology, Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Government of Gujarat, Gandhinagar 382002, India; indra52prakash@gmail.com
 - ¹¹ Geographic Information System group, Department of Business and IT, University of South-Eastern Norway, 3674 Notodden, Norway; dieu.t.bui@usn.no
- * Correspondence: nadhir.alansari@ltu.se (N.A.-A.); hosilanh@duytan.edu.vn (L.S.H.); yenhphdhv@gmail.com (H.P.H.Y.)

Received: 30 September 2019; Accepted: 10 January 2020; Published: 15 January 2020



Abstract: Risk of flash floods is currently an important problem in many parts of Vietnam. In this study, we used four machine-learning methods, namely Kernel Logistic Regression (KLR), Radial Basis Function Classifier (RBFC), Multinomial Naïve Bayes (NBM), and Logistic Model Tree (LMT) to generate flash flood susceptibility maps at the minor part of Nghe An province of the Center region (Vietnam) where recurrent flood problems are being experienced. Performance of these four methods was evaluated to select the best method for flash flood susceptibility mapping. In the model studies, ten flash flood conditioning factors, namely soil, slope, curvature, river density, flow direction, distance from rivers, elevation, aspect, land use, and geology, were chosen based on topography and geo-environmental conditions of the site. For the validation of models, the area under Receiver Operating Characteristic (ROC), Area Under Curve (AUC), and various statistical indices were used. The results indicated that performance of all the models is good for generating flash flood susceptibility maps (AUC = 0.983–0.988). However, performance of LMT model is the best among the four methods (LMT: AUC = 0.988; KLR: AUC = 0.985; RBFC: AUC = 0.984; and NBM: AUC = 0.983). The present study would be useful for the construction of accurate flash flood susceptibility maps with the objectives of identifying flood-susceptible areas/zones for proper flash flood risk management.

Keywords: flash flood; kernel logistic regression; radial basis function network; multinomial naïve Bayes; logistic model tree; machine learning; Vietnam

1. Introduction

Flooding is considered to be one of the most dangerous natural disasters, associated with damage to properties, infrastructure, and people around the world [1,2]. Approximately 90% of human losses occur from flooding in Asia, especially in tropical cyclone regions such as Southeast Asia [3,4]. There are many types of floods including pluvial (surface), fluvial (riverine), and coastal (surge). The main difference between pluvial and fluvial flood is that pluvial flood caused by heavy rainfall creates a flood event independent of an overflowing water body, whereas fluvial flood is caused by excessive rainfall over an extended period of time which is dependent on overflowing water bodies. Floods also occur due to excessive amounts of snow melt and sudden breaking of natural and manmade dams. Pluvial floods can also occur at higher elevation areas that lie above coastal and river floodplains. Flash flooding is characterized by intense, high-velocity torrential rainfall within a short period. Flash floods can occur on the ground surface as well as on the riverbed. Much environmental research has indicated that human activities affect the water cycle, such as deforestation. Forests play a critical role in the fight against natural disasters. However, there is an increasing trend towards deforestation in recent years regarding development [5]. Erratic rainfall due to climate change, in conjunction with deforestation and un-planned city development, has resulted in the occurrence of more flash floods with disastrous consequences, which require greater attention from government and other organizations. Although it is impossible to prevent flash floods, their accurate prediction by appropriate model studies may help in reducing damage [6].

The determination of flash flood susceptibility zones is essential for risk management strategies and is helpful for the decision-makers to manage land-use planning [7,8]. A flood susceptibility map will show areas where floods are likely to occur. Flood susceptibility is defined as a quantitative or qualitative assessment of an area with spatial distribution of flood, where probability of flood occurrence is likely [9]. This is a measure of the probability of future floods likely to occur depending on meteorological conditions [10]. However, there is a limit to the temporal frequency of floods. Flood hazard is a phenomenon that may cause loss of life, injury or other health impacts, property damage, loss of livelihoods and services, social and economic disruption, or environmental damage (<http://www.charim.net/methodology/31>). It is a combination of extent, depth, and flow velocity [11]. The information needed depends on the hazard interpretation (evacuation, building damage, early warning etc.). It depends on the intensity of the phenomenon within specified time and area [11]. However, flood risk is a measure of the damage anticipated to occur in an area [12]. Risk is often expressed as a combination of exposure, vulnerability, and flood hazard [13,14]. A hazard map is not a risk map. The risk is dependent on the hazard and potential damage [12]. A risk analysis includes the impact of one or more hazards, taking into account the vulnerability and resilience of the elements at risk [15]. In general, a flash flood susceptibility map is a critical tool for flood risk management [16]. However, it is difficult to accurately predict specific areas which would be affected most, because of the nature and dynamics of meteorological (climatic) conditions [16].

In recent years, different statistical methods have been developed and applied effectively in flood susceptibility mapping. Presently, Machine Learning (ML) or Artificial Intelligence (AI) methods, which are advanced soft computing approaches for natural hazard prediction and assessment, are mostly used for the flood study [17]. These methods are based on effective and objective mathematical algorithms for analysis and prediction [18–21]. Some popular ML methods used for flood susceptibility assessment are Artificial Neural Networks (ANN) [22,23], Logistic Model Trees (LMT) [24], Support Vector Machines (SVM), Logistic Regression (LR) [25,26], Adaptive Neuro-Fuzzy Inference Systems (ANFIS) [27], and Neural-Fuzzy (NF) approach [28,29]. So far, there is no existing

model that can be applied in all regions for flood susceptibility assessment and mapping accurately [30]. There is a need for ongoing research to explore the possibility of the selection of appropriate models for accurate identification and mapping of flash flood-susceptible areas. With this objective, we have experimented with the four ML models, namely Kernel Logistic Regression (KLR), Radial Basis Function Classifier (RBFC), Multinomial NB (MNB), and LMT, which were not applied and compared earlier in flash flood studies. These models were applied in the Nghe An province, which is one of the flash flood-prone areas of Vietnam. All these models use supervised learning algorithms to solve classification problems with high prediction accuracy. Receiver Operating Characteristic (ROC) and various statistical measures were used to validate and compare the performance of the models. Results were compared to select the best method among these four models for flash flood susceptibility mapping. Arc Map 10.2 and Weka 3.7.12 software were used to process data and generating flash flood susceptibility maps.

2. Description of Study Area

Vietnam in general and Nghe An in particular has been affected by different natural hazards such as flood, arsenic pollution [31], radiation hazard [32], erosion [33–35], sea level rise [36,37], earthquakes [38–42], volcanos [43,44], and landslides [45]. Nghe An province is in the North Central Coast region of Vietnam (Figure 1). The morphology of the region consists of mountains, midlands, plains, and coastal areas. The topography of the area is very complicated, with very steep slopes, narrow valleys, and deep gorges. In the study area, the highest peak is Pulaileng peak (2711 m) in the Ky Son district, and the lowest area is the plain in Quynh Luu, Dien Chau, and Yen Thanh districts, which is only 0.2 m above the sea level. Mountains and hills account for 83% of the province's natural land.

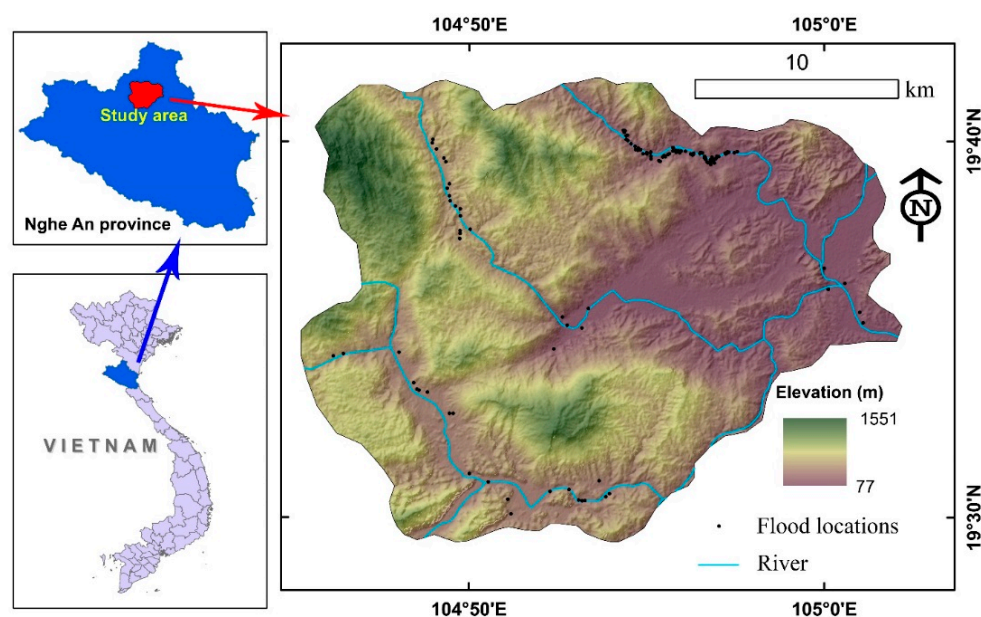


Figure 1. Location of the study area and flash floods.

In Nghe An province, rainfall is concentrated in the coastal zone and the eastern slopes of the Truong Son mountain range. The rainy season, lasting until December, has most rain between September and November. These maximums are associated with atmospheric disturbances that develop in the inter-tropical convergence zone, and with tropical cyclones. Agricultural area increase and dam filling are some of anthropogenic causes of deforestation [46,47]. Loss of watershed forest makes flood prevention difficult.

Nghe An province has seven river basins with a total length of rivers and streams in the region of 9828 km, giving an average density of 0.7 km/km². The steep upstream slopes are associated with

dense hydrological networks that add to the complexity of flash floods in the event of a rain episode of increasing intensity. In this study, a minor part of Nghe An province (Longitudes: 104.7544° N to 105.0364° N and Latitudes: 19.4890° E to 19.6947° E) is selected for flash flood mapping (Figure 1).

3. Data Used

3.1. Flash Flood Inventory

In the modelling, a knowledge of historical flash floods is important [24,48]. Thus, a flash flooding inventory map is essential. Every year, there are 10–15 flash floods in Vietnam due to extreme weather conditions causing heavy rainfall within a short period. A large part of Nghe An's surface is covered by forests, which play an essential role in the fight against flash floods and landslides. However, in recent years, forested areas have decreased because of agricultural activity and other anthropogenic activities of development. Therefore, flash floods have become increasingly hazardous in this area. Typhoons in this area also cause flash flood. In 2018 in Nghe An flash flood caused severe damage to properties and material: 6 houses collapsed, 5 schools were affected, more than 19,000 hectares of rice and vegetables damaged, and more than 15,000 m of road was affected besides loss of lives.

In this research, an inventory map with 126 flash flood events (locations) obtained from the Department of Natural Resources and Environment, Nghe An province (Vietnam) and verified from aerial photographs, satellite images, and field surveys were used for the construction of a flash flood inventory map (Figure 1).

3.2. Flash Flood Influencing Parameters

For flash flood modelling, it is crucial to select the appropriate influencing factors adapted for flash flood assessment. In our research, the choice of factors is based on the nature of flash flood observation related to different conditions of study area such as physical, hydrologic, climatic conditions, and human activity. A total of 10 factors, including soil, slope, curvature, river density, flow direction, distance from rivers, elevation, aspect, land use, and geology (Figure 2), were selected and used for analysis and modelling. In this research, a digital elevation model (DEM) with a resolution of 20 m were constructed from topographic maps at a scale of 1:50,000. DEM was used to extract the geomorphology factors (slope, aspect, curvature, and elevation) and hydrology factors (river density and distance from the river). This data was verified from the data of the Department of Natural Resources and Environment, Nghe An province (Vietnam).

Slope is an essential factor for studying flash flood susceptibility because it controls the speed of water flow from high to low altitude [49]. In this study, five main classes are used for the slope map (Figure 2a). Aspect is related to the directions of water flow affecting flash flood occurrence [50] and aspect map was built with eight classes: flat, north, northeast, east, southeast, south, southwest, and northwest (Figure 2b). Curvature is a conditioning factor in flash flood modelling that influences accumulation and runoff on the slope. In addition, flash flood zones are linked to convergence of topographic height [51]. Curvature classes used in this research are concave, flat, and convex (Figure 2c). River density is related to surface runoff, which can promote flash flooding. Areas closer to the river are more prone to experience flooding. Density of rivers and distance from rivers are considered the main factors affecting the occurrence of a flash flood [52]. Maps of river density and distance from rivers were constructed with various classes (Figure 2d,f). Flow direction, which is the direction in which water travels, is considered to be a conditioning factor of flash flood. Flow direction of this area was grouped into eight classes: 1, 2, 4, 8, 16, 32, 64, and 128 (Figure 2e). Elevation is a conditioning factor due to the weathering of rocks and soil on the slope [53,54]. An elevation map was constructed with five groups: 77–297.3, 297.3–487.4, 487.4–695.5, 695.5–961.4, and 961.4–1551.1 m (Figure 2g).

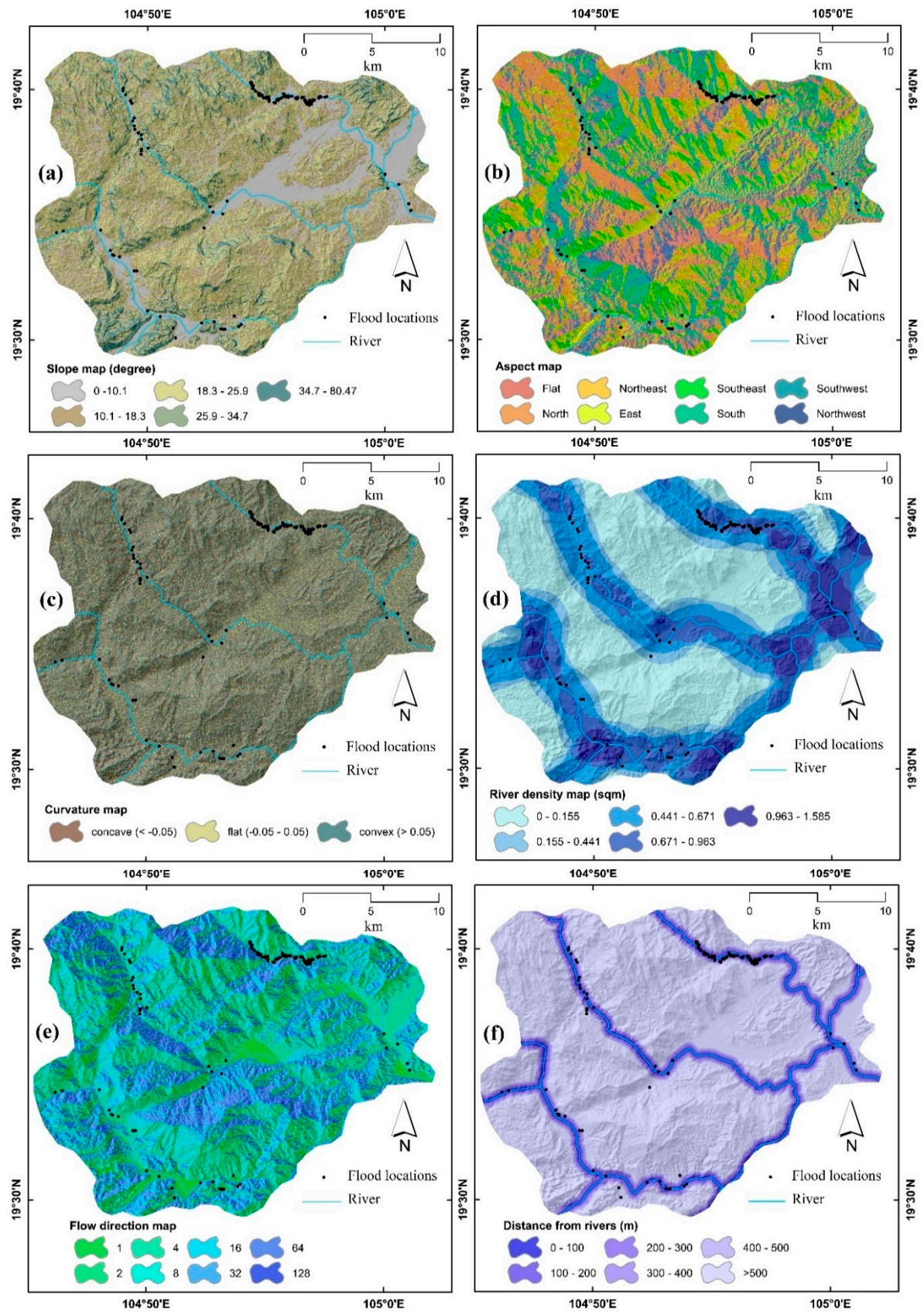


Figure 2. Cont.

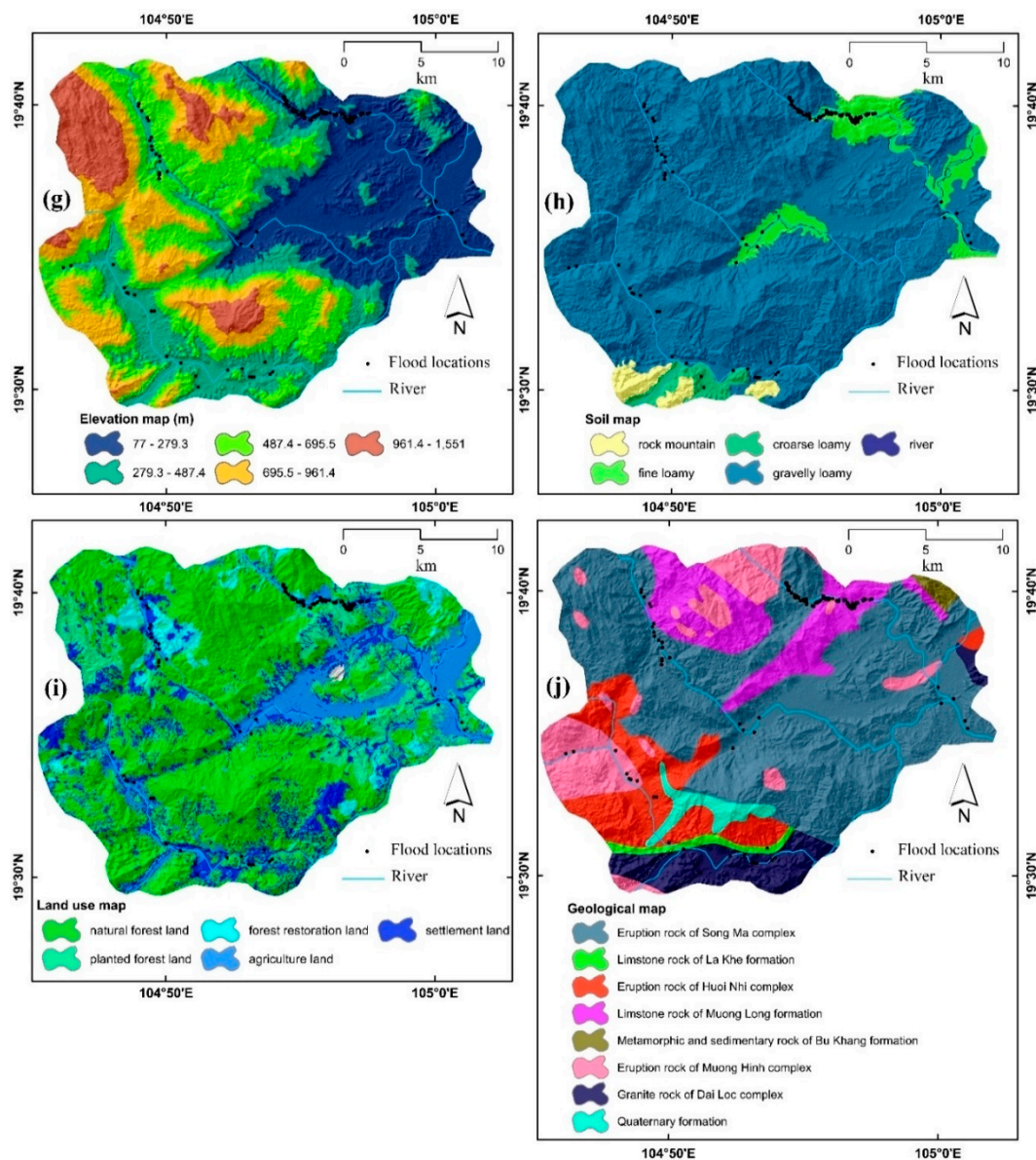


Figure 2. Maps of flash flood conditioning factors: (a) slope, (b) aspect, (c) curvature, (d) river density, (e) flow direction, (f) distance from rivers, (g) elevation, (h) soil, (i) land use, and (j) geology.

Soil type is considered an essential factor that is strongly related to rainfall runoff mechanisms affecting flash flood occurrence [55]. In this study, soil type was divided in five categories. The soil map was extracted from the MONRE geologic map at a scale of 1:100,000 (Figure 2h). Land use is an essential conditioning factor in flash flood research as it affects surface runoff. Runoff often occurs differently on agricultural and settlement lands. In addition, forests play an important role in reducing runoff speed and reducing the possibility of flash floods. A land use map (1:100,000 scale) of this area was extracted from the Landsat 7 satellite and classified into five types: natural forest land, planted forest land, forest restoration land, agriculture land, and settlement land (Figure 2i). Geology is an essential factor related to the process of runoff and infiltration, thus affecting flash flood occurrence. In this area, a geology map was compiled based on four tiles of the Geoscience and Mineral Resources Map of Vietnam at a scale of 1: 100,000 and constructed with eight classes: eruption rock of Song Ma complex, limestone rock of La Khe formation, eruption rock of Huoi Nhi complex, limestone rock of Muong Long formation, metamorphic and sedimentary rock of Bu Khang formation, eruption rock of Muong Hinh complex, granite rock of Dai Loc complex, and Quaternary formation.

Muong Hinh complex, granite rock of Dai Loc complex, and sedimentary and metamorphic of Song Ca formation, quaternary formation (Figure 2j).

4. Methods Used

In this study, selection of ML model depends on the type of data and nature of the problem. In the present study our data is of labeled type. Therefore, we have selected supervised algorithm-based models, namely LMT, KLR, NBM, and RBFC. The reason for the selection of these four ML models is that, as per the literature review, performance and prediction capabilities of these models are good but they were not applied and compared earlier for flash flood studies.

4.1. Logistic Model Tree (LMT)

LMT is a method that integrates two algorithms: C4.5 and LR. In LMT, the gain ratio information of C4.5 is used to split the tree into node and leaves, whereas the LogitBoost algorithm is applied to adapt the LR functions occurring at a tree node [56]. Out of these algorithms, C4.5 is considered to be a standard algorithm for creating classification rules in the form of decision tree. C4.5 is often referred to as a statistical classifier, which is an extension part of ID3. The information gain ratio is the default criteria of choosing to split attributes in C4.5. Instead of using the information gain as ID3, the information gain avoids the bias of selecting attributes with different values. In the LMT model, the overfitting problem is significant. To solve this challenge, the Classification and Regression Trees (CART) algorithm is used for the pruning the tree during training [57]. CART is one of the important machine-learning algorithms presenting information in a way that is intuitive and easy to visualize. CART encloses a nonparametric regression algorithm that “grows” a decision tree based on a technical binary hesitation. In LMT, let c be the sum of flash flood and non-flash flood layers and $x = x_i$ ($i = 1 - n$) be defined as flash flood conditioning factors (n is the number of the factors used). The probabilities at the leaf nodes are measured using the linear LR model as follows [56]:

$$p(c|x) = \frac{\exp(L_c(x))}{\sum_{c'=1}^c \exp(L_{c'}(x))} \quad (1)$$

where while $L_c(x)$ is the least-squares fit that is changed using following equation:

$$\sum_{c'=1}^c L_c(x) = 0 \quad (2)$$

4.2. Kernel Logistic Regression (KLR)

KLR is considered to be one of the best known machine-learning techniques for classification using nonlinear LR and probabilistic current [58]. To learn the parameters, this model estimates the class-posterior probabilities with the kernel’s log-linear function combination by applying the penalized maximum likelihood method [59]. In this model, the kernel function is used to look at a discriminant function with a goal of dealing with the classification problem by transforming the original input space into a high-dimensional feature space. Considering the predisposing factors of the flash flood as the input vector x , and the kernel function is used to complete the nonlinear transformation of x . As a result, the nonlinear form of the LR can be formulated as follows:

$$\text{logit}\{p\} = \omega \cdot \varphi(x) + b \quad (3)$$

where w and b are the optimal model parameters obtained by minimizing a cost function, which represents the regularized negative-log likelihood of the data [60], and p presents the probability of flash flood that occurs in an area.

4.3. Multinomial Naïve Bayes (NBM)

NBM relies on a probabilistic method with separated training and testing processes [61]. For the training process, suppose $t = t_i$ represents the flash flood and non-flash flood classes and $c = c_i$ ($i = 1 - n$) is defined as flash flood conditioning factors (n is the number of the factors used). The probability of each event in a class of can be measured using the following formula:

$$P(t|c) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'}}, \quad (4)$$

where T_{ct} is the sum of times t emerges in the training information of factor c , and $\sum_{t' \in V} T_{ct'}$ is the sum of attributes in factor c . To avoid problems that occur when T_{ct} is zero or some events are not present in the training data, smoothing of the square is performed by adding one to each equation:

$$P(t|c) = \frac{T_{ct} + 1}{\sum_{t' \in V} (T_{ct'} + 1)} = \frac{T_{ct} + 1}{(\sum_{t' \in V} T_{ct'}) + B'} \quad (5)$$

For the best class, the maximum a posteriori (MAP) formula is applied to avoid underflow of the test process:

$$C_{map} = \operatorname{argmax}_{c \in C} [\log P(c) + \sum_{1 \leq k \leq nr} \log P(t/c)] \quad (6)$$

where $p(c)$ is given by $P(C) = \frac{N_c}{N}$, N_c is the sum of data in layer c , and N is the sum of information in the dataset.

4.4. Radial Basis Function Classifier (RBFC)

RBFC is a supervised neural network considering an approximation problem in poly-dimensional space which is used to answer questions such as interpolation and recognition [62]. In this learning process, the network is looking for a surface in multidimensional space, which allows for a better comparison of the training dataset. Correspondingly, the test data can be interpolated using the multidimensional surface [62]. The network is composed of three layers: the first is the input layer, the second is the masked layer, and the last is the output layer. Each layer is grouped by the elements that make up the inputs and outputs. The elements of each layer are linked to transmit the information (the elements of each layer are not related).

In the process of transmitting information, a Gaussian function is used as the following radial basis function:

$$h_j(x) = \exp\left(-\frac{\|x - c_j\|^2}{r^2}\right) \quad (7)$$

where $h_j(x)$ is output data defined as flash flood or non-flash flood classes from j . The element in the hidden layer where the activation function is applied to analyze the relationship between input and output variables, $x = (x_1, \dots, x_n)$ is the input data vector of flash flood conditioning factors linked to the element in the hidden layer, c_j is inferred as the centrepoint of the basis function and r is radius of the basis function.

4.5. Validation Methods

Validation methods such as Area Under the ROC Curve (AUC) and various statistical measures were used to validate and compare the models in this study. ROC curve is a popular measure to evaluate the accuracy of the model and can be used to determine the accuracy of natural hazard susceptibility mapping [63–68]. Two values are used to build the ROC curve: sensitivity and 100-specificity [69–74]. Performance of the models is analyzed quantitatively using the area under the

curve (AUC) [75–80]. An AUC value of 1 indicates the best classification, while 0.5 corresponds to non-accurate models [81–85]. AUC values are calculated according to the equation:

$$AUC = \sum TP + \sum \frac{TN}{P} + N \quad (8)$$

where TP and TN are considered the rate of pixels classified correctly as flood and non-flood, P and N are the total number of flash floods and non-flash floods, respectively.

Various statistical measures such as accuracy (ACC), sensitivity (SST), specificity (SPF), root mean squared errors ($RMSE$), kappa (K) positive predictive value (PPV), and negative predictive value (NPV) were also selected to validate flood flash modelling [86]. PPV and NPV are the values of pixel probabilities classified correctly as “flood” occurrence and “non-flood” occurrence [87]. The proportion of flash flood pixels is represented by SST value and proportion of non-flash flood pixels is represented by SPF . K is used to analyze the accuracy of modelling [88]. K value varies between -1 and 1. Values of K close to 1 represent better reliability [8]. ACC is the ratio of the rate number of correct predictions and the total number of predictions [88]. $RMSE$ represents the difference between data observations and data estimates [89–103]. Equations for the different measures are given below:

$$SST = \frac{TP}{TP + FN} \quad (9)$$

$$SPF = \frac{TN}{TN + FP} \quad (10)$$

$$PPV = \frac{TP}{FP + TP} \quad (11)$$

$$NPV = \frac{TN}{FN + TN} \quad (12)$$

$$K = \frac{P_p - P_{exp}}{1 - P_{exp}} \quad (13)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (X_{predicted} - X_{actual})^2} \quad (15)$$

where FP and FN are the rate of pixels classified incorrectly as the flood and non-flood. P_p is the rate of pixels classified correctly for flood or non-flood. Expected agreements is defined by P_{exp} . $X_{predicted}$ and X_{actual} are the predicted and real values in the training samples or the testing samples of the models, and n is the total number of samples in the training samples or testing samples.

5. Modelling Methodology

Methodology used for constructing the flash flood susceptibility map of study area includes five steps (Figure 3): (1) Collection of data: Various thematic maps of factors were constructed using ArcGIS software in raster format with 20 m pixel size. These maps were sampled with flash inventory to generate the sampling data for further processing; (2) Dataset preparation: In this study, the sampling data has been randomly shared by two parts: the training data (70%) used for constructing the models and maps, and the validation data (30%) used for validation of the models and maps; (3) Model configuration and implementation. Four models, namely KLR, RFBC, NBM, and LMT, were constructed using training data. Out of these models, RBFC was constructed with batch size, number of functions, number of threads, ridge, and seed of 100, 2, 1, 0.01, and 1, respectively; NBM was built with batch size of 100; LMT was built with batch size, minimum number of instances, and number of boosting iterations of 100, 15, and 1, respectively; KLR was built with batch size,

lambda, number of threads, and seed of 100, 0.01, 1, and 1, respectively; (4) Model validation: In this step, validation of the flash flood susceptibility models was conducted by using PPV, NPV, SST, SPE, ACC, RMSE, K, and AUC values; (5) Development of flash flood susceptibility maps: In this step, flash flood susceptibility was evaluated using flood flash susceptibility indices that were produced from the model construction processes. These indices were then transferred to all the pixels of the flash flood zone in the study space and classified to determine susceptibility levels using natural breaks classification method in ArcGIS application—a popular method for classifying the natural hazard susceptibility classes [104].

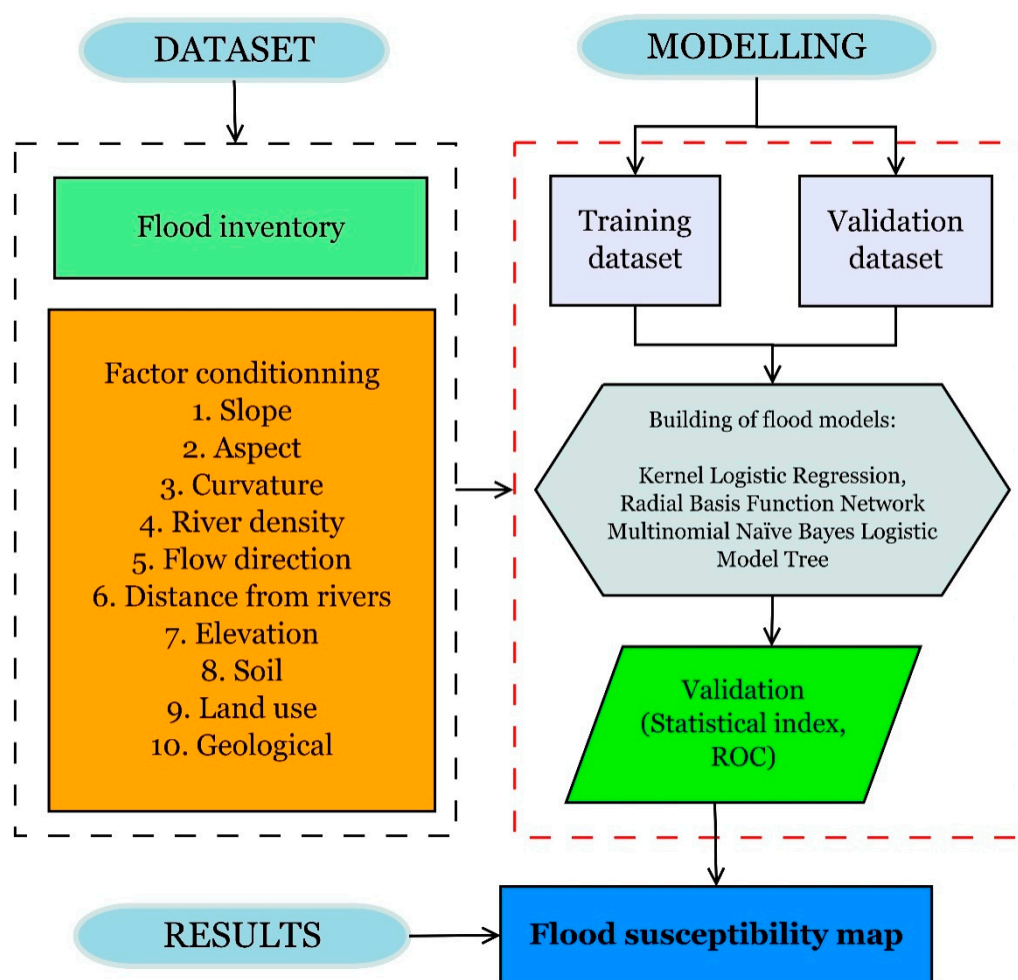


Figure 3. Methodological flow chart of this study.

6. Results and Analysis

6.1. Models Validation and Comparison

Performance of the models (RBFC, NBM, LMT, and KLR) is shown in Figures 4–6 and summarized in Table 1, which is based on both the training and validation datasets. For the training data, the results show that KLR and RBFC have the highest values of PPV (94.32%), KLR has the highest values of NPV (95.45%), SST (95.4%), SPF (94.38%), and ACC (94.89%) compared with those of other models. In the case of the validation data, LMT and NBM achieve the highest values of PPV (94.74%), LMT, KLR, and RBFC have the highest values of NPV (97.37%), LMT has the highest value of SST (97.3%), SPF (94.38%), and ACC (96.05%) (Figure 4). In terms of K value, KLR has the highest value of K (0.8977) with training data whereas LMT has the highest value of K (0.9211) with validation data (Figure 5). Regarding the RMSE value, KLR has the highest value of RMSE (0.215) with training data whereas

LMT has the highest value of RMSE (0.184) with validation data (Table 1). Based on these results, it can be stated that performance of KLR is better than other models in the training dataset; however, LMT has the best predictive capability compared to other models in terms of validation dataset.

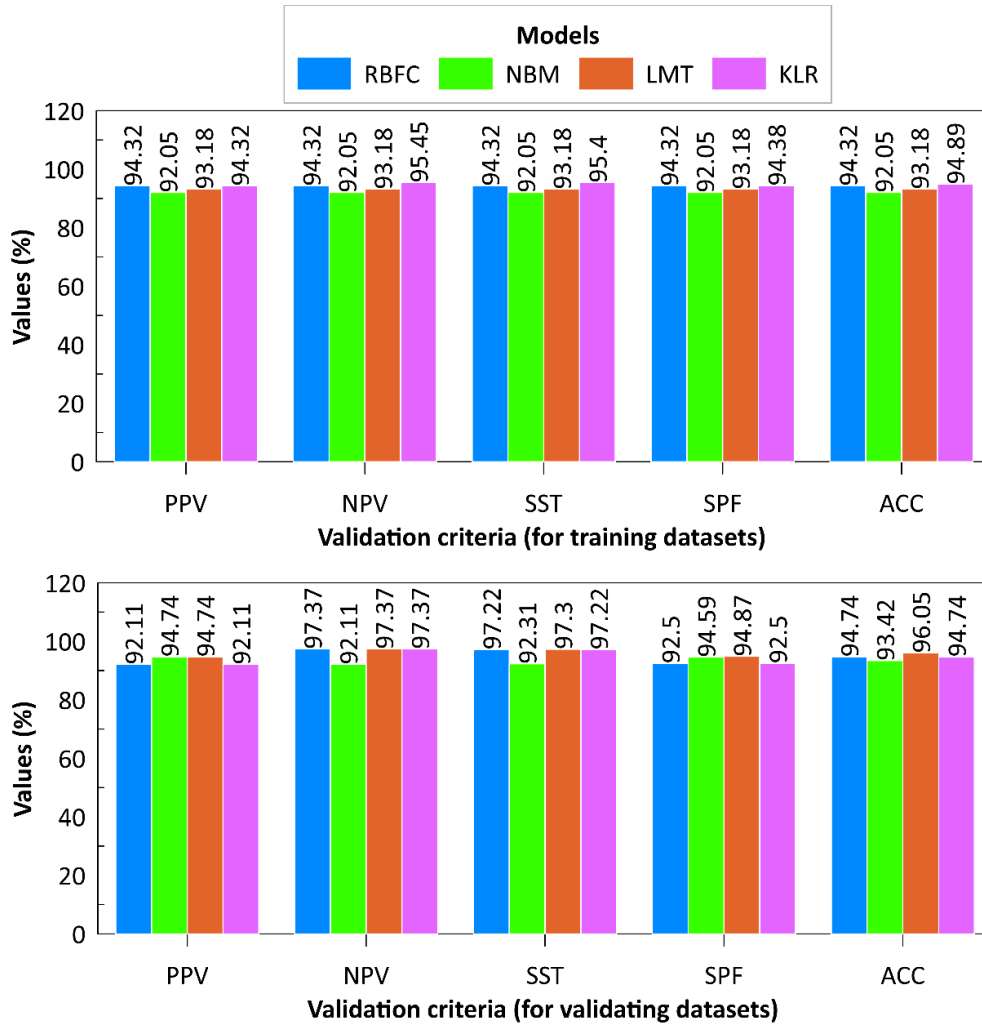


Figure 4. Value of statistical measures of the models.

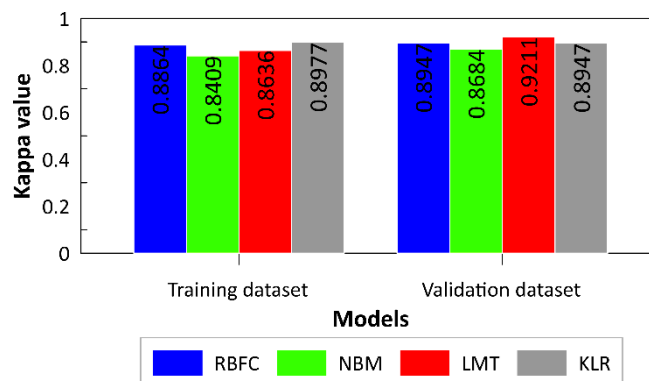


Figure 5. Kappa values of the models.

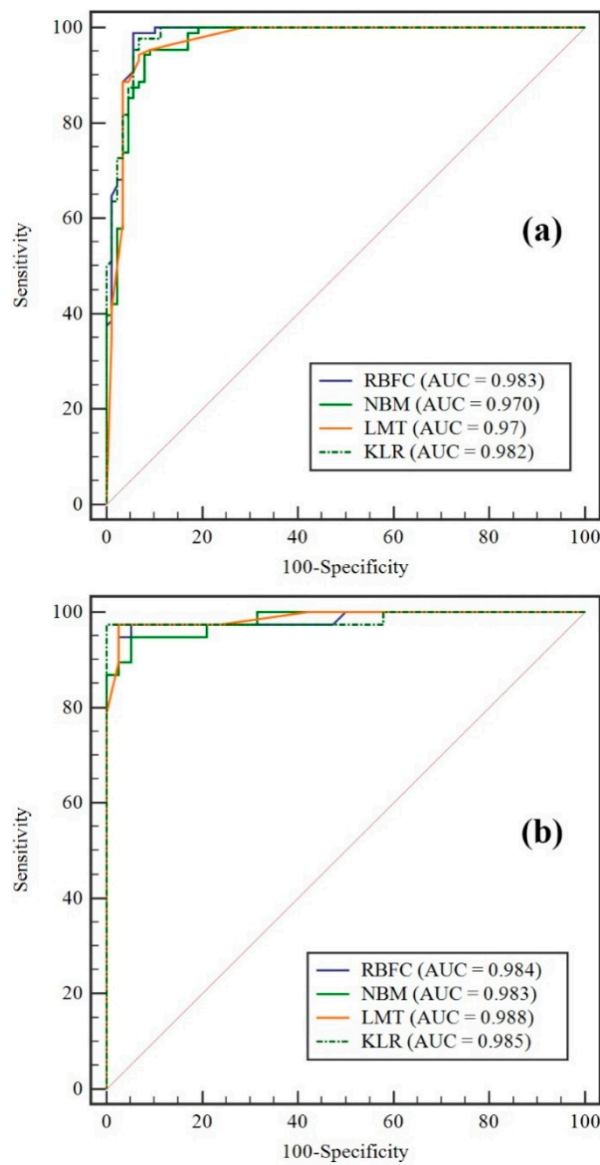


Figure 6. ROC analysis of the models: (a) training dataset; and (b) testing dataset.

Table 1. Summary of validation results of the models.

Statistical Measures	Models							
	Training Dataset				Validation Dataset			
	KLR	RBFC	NBM	LMT	KLR	RBFC	NBM	LMT
PPV	94.32	94.32	92.05	93.18	92.11	92.11	94.74	94.74
NPV	95.45	94.32	92.05	93.18	97.37	97.37	92.11	97.37
SST	95.4	94.32	92.05	93.18	97.22	97.22	92.31	97.3
SPF	94.38	94.32	92.05	93.18	92.5	92.5	94.59	94.87
ACC (%)	94.98	94.32	92.05	93.18	94.47	94.74	93.42	96.05
RMSE	0.215	0.222	0.254	0.241	0.205	0.207	0.217	0.241
K	0.8977	0.8864	0.8409	0.8636	0.8947	0.8947	0.8684	0.9211
AUC	0.982	0.983	0.970	0.97	0.985	0.984	0.983	0.988

ROC curve results indicate that RBFC model (AUC = 0.983) outperforms three other models in terms of the training prediction rate (KLR:AUC = 0.982; NBM:AUC = 0.970; and LMT:AUC = 0.970). In terms of validation, LMT is more accurate in comparison to the other models with the AUC of

0.988, followed by KLR with AUC of 0.985, RBFC with AUC of 0.984 and NBM with AUC of 0.983, respectively (Figure 6).

6.2. Flash Flood Susceptibility Map

Flash flood susceptibility maps were constructed using four ML models (KLR, RBFC, NBM, and LMT) with five classes: very low, low, moderate, high, and very high (Figure 7). The distribution of each susceptibility class on the maps obtained with different methods is shown in Figure 8. A map generated by KLR model indicates that 61.84% of the pixels are in the very low class, 6.372% in the moderate class and 13.18 in the very high. In the map constructed by RBFC model, 47.63% of the study area is in the very low level, 11.33% in the moderate level, and 12.94% in the very high level. The map built by NBM model shows 62.59% of the study area as very low level, 6.641% as moderate level, and 11.96% as very high level. Finally, the map constructed by LMT model shows that 40.06% of the area is in the very low level, 6.163% in the moderate level and 9.589% in the very high level (Figure 8). Validation of the maps using frequency ratio, which is a ratio of percentage of flash flood pixels observed on each susceptibility class, and percentage of all pixels of susceptibility class, was also done as shown in Figure 8. Validation results show that most of the flash flood pixels were observed in high and very high levels. However, the frequency ratio of flash flood observed in high and very high classes of the map produced by LMT is higher than those of other maps produced by other models (KLR, RBFC, and NBM). Thus, it can be stated that the map produced by LMT is more reliable than those of other models.

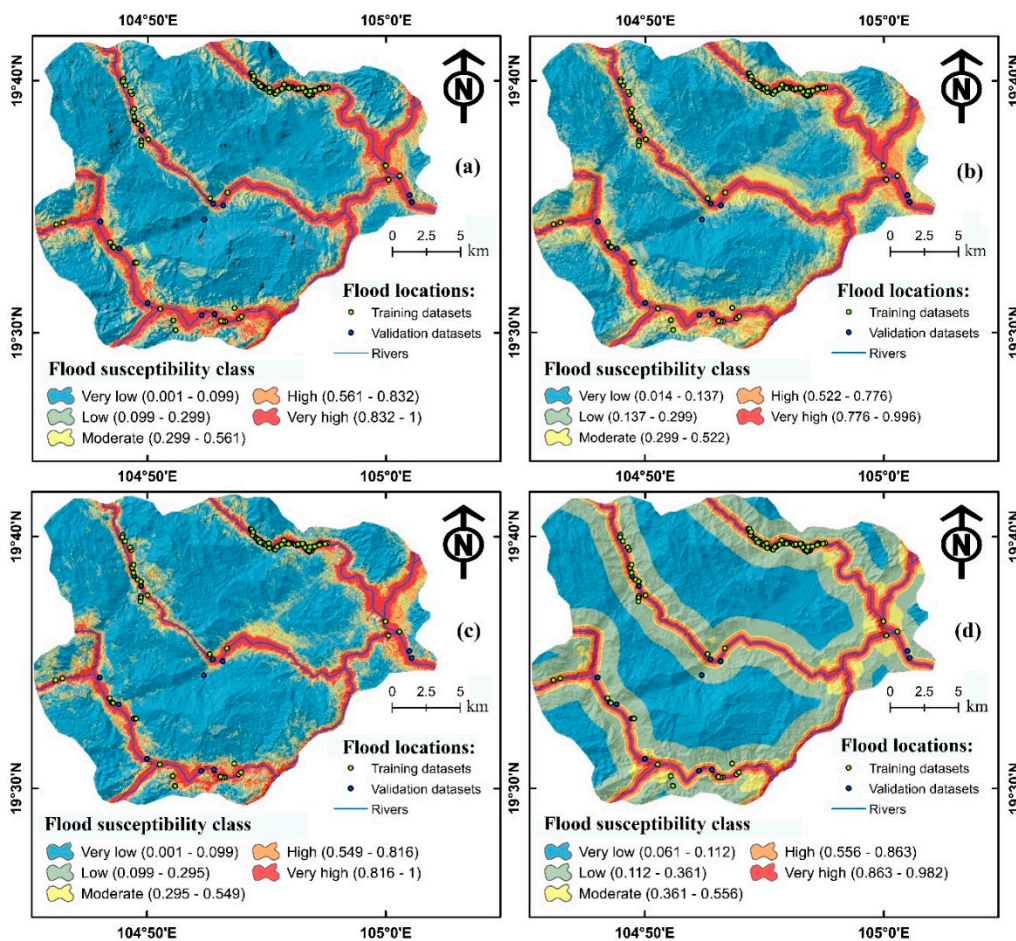


Figure 7. Flood susceptibility maps using various models: (a) KLR, (b) RBFC, (c) NBM, (d) LMT.

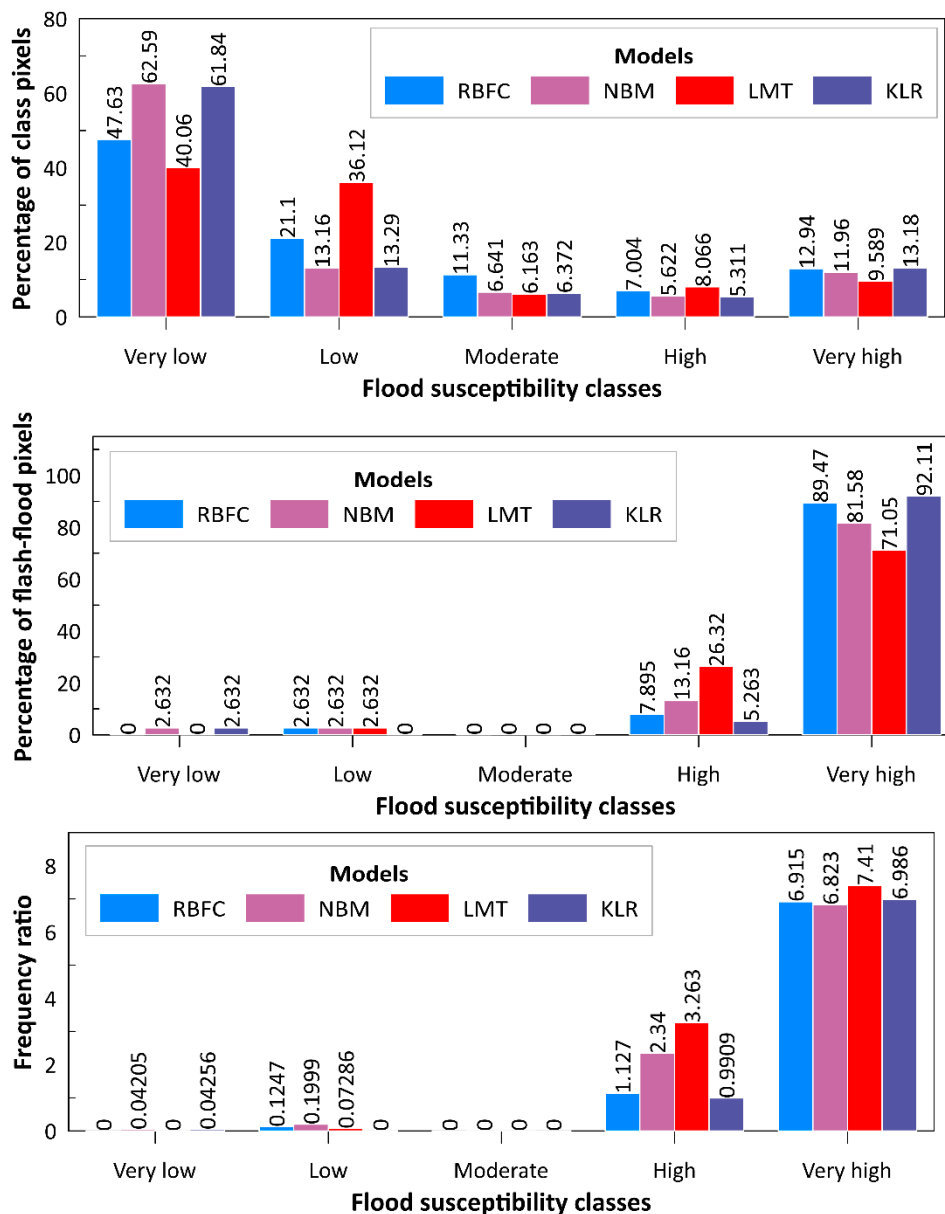


Figure 8. Analysis of the frequency of flash floods on the susceptibility maps (class pixels represents the total number of pixels in whole susceptibility class and flash flood pixels is the total number of flash flood pixels observed in the susceptibility class).

7. Discussion

Determining the areas that are most susceptible to flash floods is considered to be the most critical issue for risk management and land-use planning. Although there are several different methods developed and applied for the flash flood zone prediction around the world, generation of a flash flood susceptibility map using suitable methods for a specific area remains a topic of concern among researchers. In this study, the main purpose is to assess and compare various methods to choose the best for generating an accurate flash flood susceptibility map of the mountain area of the Nghe An province, which is one of the most affected flash flood disaster area in Vietnam. For flash flood modelling, four methods, namely KLR, RBFC, NBM, and LMT, were selected as these are advanced and effective ML models for natural hazard prediction and assessment [105–107]. Conditioning factors may change depending on the local geo-environmental conditions of the study area [108]. In general, flash flooding occurs mainly on watersheds, especially in hilly areas, where the topography is favorable to rapid flow

(runoff) in the event of heavy rainfall within a short time. Loss of vegetation accentuates the flooding process. Topography and river density affect the occurrence of flash flood [109]. Considering this, ten factors, namely soil, slope, curvature, river density, flow direction, distance from rivers, elevation, aspect, land use, and geology, were used to construct the flood database for modelling.

In the context of spatial planning, selection of suitable models for the generation of accurate flood susceptibility map is desirable to avoid damage to property and human losses [110]. Out of the four models proposed in this paper, KLR is the best compared with other models using training data. However, LMT achieves a higher predictive capability during the validation process. This model is more reliable than the other models for flash flood susceptibility mapping. Performance of LMT is related to its robustness, noise reduction, and variance, as well as the reduction of overfitting. Thus, LMT is better compared to other models because of its reduced overfitting and variance. In addition, KLR uses the fractal dimension for input data, and thus performed well in the training dataset. Results also indicate that NBM has less accuracy compared to the other three models, as it rests on the independent hypothesis of the conditioning factors that could influence its accuracy. Overall, the four flash flooding models have an acceptable performance for assessing flash flood susceptibility but LMT is the best compared with other models.

Even though flash flood prediction ability may decrease when a low proportion of training samples were used, in the present case, models demonstrated robustness. With the complexity of flash floods and the interaction of several factors, a comparison of more modelling methods are required and different sets of characteristics and factors can be determined using various techniques that would make it possible to give different points of view regarding feature selection and improvement of performance of machine-learning models.

8. Conclusions

In this study, four ML models, namely LMT, KLR, RBFC, and NBM, were used to generate flash flood susceptibility maps of Nghe An province in Vietnam. For this purpose, 126 flash flood historic events and ten conditioning factors (soil, slope, curvature, river density, flow direction, distance from rivers, elevation, aspect, land use, and geology) were used for the construction the flash flood database for modelling. Various methods such as area under ROC curve (AUC), and several statistical measures were used for the validation and comparison of the models.

Validation results show that LMT had the best performance (AUC = 0.988), followed by KLM (0.985), RBFC (0.984), and NBM (0.983), respectively. LMT model also achieved the highest PPV (94.74%), NPV (97.37%), SST (97.3%), SPF (94.38%), and ACC (96.05%) in comparison to other models. Therefore, this method can be used for flash flood susceptibility mapping of other areas also. There is always scope for improvement in the performance of methods adopted in this study by using different combinations of ML models considering greater numbers of flash flood events and influencing factors depending on the physical, hydrological, and meteorological conditions of the area.

Author Contributions: Conceptualization, B.T.P., N.A.-A., H.D.N., L.S.H., H.-B.L., I.P., A.A., and D.T.B.; Data curation, L.S.H., H.D.N., T.T.T. and H.P.H.Y.; Formal analysis, T.V.P., H.D.N., C.Q., N.A.-A., L.S.H., T.T.T., H.P.H.Y. and H.-B.L.; Funding acquisition, N.A.-A.; Methodology, B.T.P., T.V.P., and D.T.B.; Project administration, B.T.P., N.A.-A., and I.P.; Supervision, B.T.P., H.-B.L., I.P. and D.T.B.; Validation, H.P.H.Y., H.-B.L., A.A., and I.P.; Visualization, H.D.N., A.A., T.T.T. and H.P.H.Y.; Writing—original draft, B.T.P., T.V.P., H.D.N., A.A., C.Q., N.A.-A., L.S.H., T.T.T., H.P.H.Y. and H.-B.L.; Writing—review and editing, A.A., B.T.P., N.A.-A., and I.P. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the research fund of Vinh University, Vietnam in Nghe An Province, Vietnam.

Acknowledgments: We thank to the Department of Natural Resources and Environment, Nghe An province (Vietnam) for providing us the data used in this research.

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

References

1. Peduzzi, P. Flooding: Prioritizing protection? *Nat. Clim. Chang.* **2017**, *7*. [[CrossRef](#)]
2. Bubeck, P.; Thielen, A. What helps people recover from floods? Insights from a survey among flood-affected residents in Germany. *Reg. Environ. Chang.* **2018**, *18*, 287–296. [[CrossRef](#)]
3. Dutta, D.; Herath, S. Trend of Floods in Asia and Flood Risk Management with Integrated River Basin Approach. In Proceedings of the 2nd International Conference of Asia-Pacific Hydrology and Water Resources Association, Singapore, 5–9 July 2004.
4. Smith, K. *Environmental Hazards: Assessing Risk and Reducing Disaster*; Routledge: Abingdon-on-Thames, UK, 2003.
5. Roche, Y.; De Koninck, R. Les enjeux de la déforestation au Vietnam. *VertigO* **2002**, *3*. [[CrossRef](#)]
6. Cloke, H.L.; Pappenberger, F. Ensemble flood forecasting: A review. *J. Hydrol.* **2009**, *375*, 613–626. [[CrossRef](#)]
7. 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.* **2016**, *75*, 12. [[CrossRef](#)]
8. Janizadeh, S.; Avand, M.; Jaafari, A.; Phong, T.V.; Bayat, M.; Ahmadisharaf, E.; Prakash, I.; Pham, B.T.; Lee, S. Prediction Success of Machine Learning Methods for Flash Flood Susceptibility Mapping in the Tafresh Watershed, Iran. *Sustainability* **2019**, *11*, 5426. [[CrossRef](#)]
9. Rahman, M.; Ningsheng, C.; Islam, M.M.; Dewan, A.; Iqbal, J.; Washakh, R.M.A.; Shufeng, T. Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis. *Earth Syst. Environ.* **2019**, *3*, 585–601. [[CrossRef](#)]
10. Quinn, P.; Hutchinson, D.; Diederichs, M.; Rowe, R.K. Characteristics of large landslides in sensitive clay in relation to susceptibility, hazard, and risk. *Can. Geotech. J.* **2011**, *48*, 1212–1232. [[CrossRef](#)]
11. Islam, M.M.; Sado, K. Flood hazard assessment in Bangladesh using NOAA AVHRR data with geographical information system. *Hydrol. Process.* **2000**, *14*, 605–620. [[CrossRef](#)]
12. Zhou, Q.; Mikkelsen, P.S.; Halsnaes, K.; Arnbjerg-Nielsen, K. Framework for economic pluvial flood risk assessment considering climate change effects and adaptation benefits. *J. Hydrol.* **2012**, *414*, 539–549. [[CrossRef](#)]
13. Apel, H.; Thielen, A.H.; Merz, B.; Blöschl, G. Flood risk assessment and associated uncertainty. *Nat. Hazards Earth Syst. Sci.* **2004**, *4*, 295–308. [[CrossRef](#)]
14. De Risi, R.; Jalayer, F.; De Paola, F.; Carozza, S.; Yonas, N.; Giugni, M.; Gasparini, P. From flood risk mapping toward reducing vulnerability: The case of Addis Ababa. *Nat. Hazards* **2019**, 1–29. [[CrossRef](#)]
15. Zou, Q.; Zhou, J.; Zhou, C.; Song, L.; Guo, J. Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP. *Stoch. Environ. Res. Risk Assess.* **2013**, *27*, 525–546. [[CrossRef](#)]
16. Kubal, C.; Haase, D.; Meyer, V.; Scheuer, S. Integrated urban flood risk assessment—adapting a multicriteria approach to a city. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 1881–1895. [[CrossRef](#)]
17. Bui, D.T.; Tsangaratos, P.; Ngo, P.-T.T.; Pham, T.D.; Pham, B.T. Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. *Sci. Total Environ.* **2019**, *668*, 1038–1054. [[CrossRef](#)]
18. Jaafari, A.; Zenner, E.K.; Pham, B.T. Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. *Ecol. Inform.* **2018**, *43*, 200–211. [[CrossRef](#)]
19. Shirzadi, A.; Soliamani, K.; Habibnejhad, M.; Kaviani, A.; Chapi, K.; Shahabi, H.; Chen, W.; Khosravi, K.; Thai Pham, B.; Pradhan, B.; et al. Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. *Sensors* **2018**, *18*, 3777. [[CrossRef](#)]
20. Khosravi, K.; Sartaj, M.; Tsai, F.T.-C.; Singh, V.P.; Kazakis, N.; Melesse, A.M.; Prakash, I.; Bui, D.T.; Pham, B.T. A comparison study of DRASTIC methods with various objective methods for groundwater vulnerability assessment. *Sci. Total. Environ.* **2018**, *642*, 1032–1049. [[CrossRef](#)]
21. Dou, J.; Yunus, A.P.; Tien Bui, D.; Sahana, M.; Chen, C.-W.; Zhu, Z.; Wang, W.; Pham, B.T. Evaluating GIS-Based Multiple Statistical Models and Data Mining for Earthquake and Rainfall-Induced Landslide Susceptibility Using the LiDAR DEM. *Remote Sens.* **2019**, *11*, 638. [[CrossRef](#)]
22. Radmehr, A.; Araghinejad, S. Developing Strategies for Urban Flood Management of Tehran City Using SMCDM and ANN. *J. Comput. Civ. Eng.* **2014**, *28*, 05014006. [[CrossRef](#)]

23. Falah, F.; Rahmati, O.; Rostami, M.; Ahmadisharaf, E.; Daliakopoulos, I.N.; Pourghasemi, H.R. Artificial Neural Networks for Flood Susceptibility Mapping in Data-Scarce Urban Areas. In *Spatial Modeling in GIS and R for Earth and Environmental Sciences*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 323–336.
24. 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](#)]
25. Nandi, A.; Mandal, A.; Wilson, M.; Smith, D. Flood hazard mapping in Jamaica using principal component analysis and logistic regression. *Environ. Earth Sci.* **2016**, *75*. [[CrossRef](#)]
26. Pradhan, B. Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. *J. Spat. Hydrol.* **2009**, *9*, 1–18.
27. Ahmadi, M.; Karimi, M.; Alizadeh, S.; Shirzadi, A.; Parvinnejhad, D.; Shahabi, H.; Panahi, M. Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). *Geocarto Int.* **2019**, *34*, 1252–1272. [[CrossRef](#)]
28. Mukerji, A.; Chatterjee, C.; Raghuvanshi, N. Flood forecasting using ANN, Neuro-Fuzzy, and Neuro-GA models. *J. Hydrol. Eng.* **2009**, *14*, 647–652. [[CrossRef](#)]
29. Hong, H.; Panahi, M.; Shirzadi, A.; Ma, T.; Liu, J.; Zhu, A.-X.; Chen, W.; Kougiyas, I.; Kazakis, N. Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Sci. Total Environ.* **2018**, *621*, 1124–1141. [[CrossRef](#)] [[PubMed](#)]
30. Bui, Q.-T.; Nguyen, Q.-H.; Nguyen, X.L.; Pham, V.D.; Nguyen, H.D.; Pham, V.-M. Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *J. Hydrol.* **2019**, *581*, 124379. [[CrossRef](#)]
31. Nga, D.V.; Trang, P.T.K.; Duyen, V.T.; Mai, T.T.; Lan, V.T.M.; Viet, P.H.; Postma, D.; Jakobsen, R. Spatial variations of arsenic in groundwater from a transect in the Northwestern Hanoi. *Vietnam J. Earth Sci.* **2018**, *40*, 70–77. [[CrossRef](#)]
32. Nguyet, N.T.A.; Duong, N.T.; Schimmelmann, A.; Huong, N. Human exposure to radon radiation geohazard in Rong Cave, Dong Van Karst Plateau Geopark, Vietnam. *Vietnam J. Earth Sci.* **2018**, *40*, 117–125. [[CrossRef](#)]
33. Thai, T.H.; Thao, N.P.; Dieu, B.T. Assessment and simulation of impacts of climate change on erosion and water flow by using the soil and water assessment tool and GIS: Case Study in Upper Cau River basin in Vietnam. *J. Earth Sci.* **2017**, *39*, 376–392. [[CrossRef](#)]
34. Van Hung, P.; Quan, N.C. The characteristics of active faults and the erosion hazard in coastal-river mouth zones of North Central Vietnam. *Vietnam J. Earth Sci.* **2016**, *38*, 46–58.
35. Son, P.Q.; Anh, N.D. Evolution of the coastal zone in Hai Hau district (Nam Dinh province) and nearest region over the last 100 years based on analysis topographic maps and multi-temporal remote sensing data. *Vietnam J. Earth Sci.* **2016**, *38*, 118–130.
36. Van Thanh, N.; Le, D.T.; Tinh, N.A.; Lan, T.D.; Hens, L. Shifting challenges for coastal green cities. *Vietnam J. Earth Sci.* **2017**, *39*, 109–129. [[CrossRef](#)]
37. Hens, L.; Tinh, N.A.; Hanh, T.H.; Cuong, N.S.; Lan, T.D.; Van Thanh, N.; Le, D.T. Sea-level rise and resilience in Vietnam and the Asia-Pacific: A synthesis. *Vietnam J. Earth Sci.* **2018**, *40*, 126–152. [[CrossRef](#)]
38. Hoan, V.T.; Lu, N.T.; Rodkin, M.; Quang, N.; Huong, P.T. Seismic activity characteristics in the East Sea area. *Vietnam J. Earth Sci.* **2018**, *40*, 240–252. [[CrossRef](#)]
39. Lu, N.T.; Burmin, V.Y.; Hang, P.T.T.; Hoan, V.T.; Giang, H.T. Estimation of errors in determination of main parameters of earthquake hypocenter, recorded by the national seismic network of Vietnam. *J. Volcanol. Seismol.* **2018**, *40*, 1–16. [[CrossRef](#)]
40. Nhung, B.T.; Phuong, N.H.; Nam, N.T. Assessment of earthquake-induced liquefaction hazard in urban areas of Hanoi city using LPI-based method. *Vietnam J. Earth Sci.* **2018**, *40*, 78–96. [[CrossRef](#)]
41. Van Duan, B.; Duong, N.A. The relation between fault movement potential and seismic activity of major faults in Northwestern Vietnam. *Vietnam J. Earth Sci.* **2017**, *39*, 240–255. [[CrossRef](#)]
42. Nguyen-Van, H.; Van Phong, T.; Trinh, P.T.; Van Liem, N.; Thanh, B.N.; Pham, B.T.; Bui, D.T.; Bieu, N.; Vinh, H.Q.; Xuyen, N.Q.; et al. Recent tectonics, geodynamics and seismotectonics in the Ninh Thuan Nuclear Power plants and surrounding regions, South Vietnam. *J. Asian Earth Sci.* **2020**, *187*, 104080. [[CrossRef](#)]

43. Hoang, N.; Shakirov, R.B.; Huong, T.T. Geochemistry of late miocene-pleistocene basalts in the Phu Quy island area (East Vietnam Sea): Implication for mantle source feature and melt generation. *J. Earth Sci.* **2017**, *39*, 270–288.
44. Tachihara, H.; Honda, T.; Tuat, L.T.; Van Thom, B.; Hoang, N.; Chikano, Y.; Yoshida, K.; Tung, N.T.; Danh, P.N.; Hung, N.B.; et al. Geological values of lava caves in Krongno Volcano Geopark, Dak Nong, Vietnam. *J. Earth Sci.* **2018**, *40*, 299–319.
45. Van Tu, T.; Duc, D.M.; Tung, N.M.; Cong, V.D. Preliminary assessments of debris flow hazard in relation to geological environment changes in mountainous regions, North Vietnam. *J. Earth Sci.* **2016**, *38*, 277–286.
46. Boissau, S.; Castella, J.-C.; Thanh, N. La distribution des terres de forêt au Nord Viêt Nam: Droit d’usage et gestion des ressources. *Cah. Agric.* **2003**, *12*, 307–320.
47. Castella, J.-C.; Boissau, S.; Hai Thanh, N.; Novosad, P. Impact of forestland allocation on land use in a mountainous province of Vietnam. *Land Use Policy* **2006**, *23*, 147–160. [[CrossRef](#)]
48. 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*, 1–19. [[CrossRef](#)]
49. Pham, B.T.; Jaafari, A.; Prakash, I.; Singh, S.K.; Quoc, N.K.; Bui, D.T. Hybrid computational intelligence models for groundwater potential mapping. *Catena* **2019**, *182*, 104101. [[CrossRef](#)]
50. Aryal, S.; Mein, R.; O’Loughlin, E. The Concept of Effective Length in Hillslopes: Assessing the Influence of Climate and Topography on the Contributing Areas of Catchments. *Hydrol. Process.* **2003**, *17*, 131–151. [[CrossRef](#)]
51. Manfreda, S.; Nardi, F.; Samela, C.; Grimaldi, S.; Taramasso, A.; Roth, G.; Sole, A. Investigation on the Use of Geomorphic Approaches for the Delineation of Flood Prone Areas. *J. Hydrol.* **2014**. [[CrossRef](#)]
52. Vojtek, M.; Vojteková, J. Flood Susceptibility Mapping on a National Scale in Slovakia Using the Analytical Hierarchy Process. *Water* **2019**, *11*, 364. [[CrossRef](#)]
53. Nguyen, V.V.; Pham, B.T.; Vu, B.T.; Prakash, I.; Jha, S.; Shahabi, H.; Shirzadi, A.; Ba, D.N.; Kumar, R.; Chatterjee, J.M. Hybrid machine learning approaches for landslide susceptibility modeling. *Forests* **2019**, *10*, 157. [[CrossRef](#)]
54. Yilmaz, I. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: Conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ. Earth Sci.* **2009**, *61*, 821–836. [[CrossRef](#)]
55. Geris, J.; Tetzlaff, D.; McDonnell, J. The relative role of soil type and tree cover on water storage and transmission in northern headwater catchments. *Hydrol. Process.* **2015**, *29*, 1844–1860. [[CrossRef](#)]
56. Landwehr, N.; Hall, M.; Frank, E. Logistic model trees. *Mach. Learn.* **2005**, *59*, 161–205. [[CrossRef](#)]
57. Breiman, L. *Classification and Regression Trees*; Routledge: Abingdon-on-Thames, UK, 2017.
58. Cawley, G.; Talbot, N. Efficient approximate leave-one-out cross-validation for kernel logistic regression. *Mach. Learn.* **2008**, *71*, 243–264. [[CrossRef](#)]
59. Tien Bui, D.; Tuan, T.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2015**, *13*, 361–378. [[CrossRef](#)]
60. Cawley, G.C.; Talbot, N.L. Efficient model selection for kernel logistic regression. In Proceedings of the 17th International Conference on Pattern Recognition (ICPR), Cambridge, UK, 26–26 August 2004; pp. 439–442.
61. Isabelle, G.; Maharani, W.; Asror, I. Analysis on Opinion Mining Using Combining Lexicon-Based Method and Multinomial Naïve Bayes. In Proceedings of the 2018 International Conference on Industrial Enterprise and System Engineering (IcoIESE 2018), Yogyakarta, Indonesia, 21–22 November 2018.
62. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1994.
63. Pham, B.T.; Prakash, I.; Khosravi, K.; Chapi, K.; Trinh, P.T.; Ngo, T.Q.; Hosseini, S.V.; Bui, D. A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. *Geocarto Int.* **2019**, *34*, 1385–1407. [[CrossRef](#)]
64. Miraki, S.; Zanganeh, S.H.; Chapi, K.; Singh, V.P.; Shirzadi, A.; Shahabi, H.; Pham, B.T. Mapping groundwater potential using a novel hybrid intelligence approach. *Water Resour. Manag.* **2019**, *33*, 281–302. [[CrossRef](#)]
65. Pham, B.T.; Prakash, I.; Jaafari, A.; Bui, D.T. Spatial prediction of rainfall-induced landslides using aggregating one-dependence estimators classifier. *J. Indian Soc. Remote Sens.* **2018**, *46*, 1457–1470. [[CrossRef](#)]

66. Abedini, M.; Ghasemian, B.; Shirzadi, A.; Shahabi, H.; Chapi, K.; Pham, B.T.; Bin Ahmad, B.; Tien Bui, D. A novel hybrid approach of bayesian logistic regression and its ensembles for landslide susceptibility assessment. *Geocarto Int.* **2019**, *34*, 1427–1457. [[CrossRef](#)]
67. Pham, B.T.; Bui, D.T.; Pham, H.V.; Le, H.Q.; Prakash, I.; Dholakia, M. Landslide hazard assessment using random subspace fuzzy rules based classifier ensemble and probability analysis of rainfall data: A case study at Mu Cang Chai District, Yen Bai Province (Viet Nam). *J. Indian Soc. Remote Sens.* **2017**, *45*, 673–683. [[CrossRef](#)]
68. Pham, B.T. A novel classifier based on composite hyper-cubes on iterated random projections for assessment of landslide susceptibility. *J. Geol. Soc. India* **2018**, *91*, 355–362. [[CrossRef](#)]
69. Pradhan, A.; Kim, Y.-T. Relative effect method of landslide susceptibility zonation in weathered granite soil: A case study in Deokjeok-ri Creek, South Korea. *Nat. Hazards* **2014**, *72*, 1189–1217. [[CrossRef](#)]
70. Termeh, S.V.R.; Khosravi, K.; Sartaj, M.; Keesstra, S.D.; Tsai, F.T.-C.; Dijkma, R.; Pham, B.T. Optimization of an adaptive neuro-fuzzy inference system for groundwater potential mapping. *Hydrogeol. J.* **2019**, *27*, 2511–2534. [[CrossRef](#)]
71. Pham, B.T.; Prakash, I.; Dou, J.; Singh, S.K.; Trinh, P.T.; Tran, H.T.; Le, T.M.; Van Phong, T.; Khoi, D.K.; Shirzadi, A.; et al. A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto Int.* **2019**, 1–25. [[CrossRef](#)]
72. Pham, B.T.; Prakash, I. Machine learning methods of kernel logistic regression and classification and regression trees for landslide susceptibility assessment at part of Himalayan area, India. *Indian J. Sci. Technol.* **2018**, *11*, 1–10. [[CrossRef](#)]
73. Thai Pham, B.; Shirzadi, A.; Shahabi, H.; Omidvar, E.; Singh, S.K.; Sahana, M.; Talebpour Asl, D.; Bin Ahmad, B.; Kim Quoc, N.; Lee, S.; et al. Landslide susceptibility assessment by novel hybrid machine learning algorithms. *Sustainability* **2019**, *11*, 4386. [[CrossRef](#)]
74. Dou, J.; Yunus, A.P.; Bui, D.T.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.-W.; Khosravi, K.; Yang, Y.; Pham, B.T. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Sci. Total Environ.* **2019**, *662*, 332–346. [[CrossRef](#)]
75. Dou, J.; Yunus, A.P.; Bui, D.T.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.-W.; Han, Z.; Pham, B.T. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides* **2019**, 1–18. [[CrossRef](#)]
76. Tien Bui, D.; Shirzadi, A.; Chapi, K.; Shahabi, H.; Pradhan, B.; Pham, B.T.; Singh, V.P.; Chen, W.; Khosravi, K.; Bin Ahmad, B.J.W. A Hybrid Computational Intelligence Approach to Groundwater Spring Potential Mapping. *Water* **2019**, *11*, 2013. [[CrossRef](#)]
77. Phong, T.V.; Phan, T.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Chapi, K.; Ly, H.-B.; Ho, L.S.; Quoc, N.K.; Pham, B.T.; et al. Landslide susceptibility modeling using different artificial intelligence methods: A case study at Muong Lay district, Vietnam. *Geocarto Int.* **2019**, 1–24. [[CrossRef](#)]
78. Nohani, E.; Moharrami, M.; Sharafi, S.; Khosravi, K.; Pradhan, B.; Pham, B.T.; Lee, S.; Melesse, A.M. Landslide susceptibility mapping using different GIS-based bivariate models. *Water* **2019**, *11*, 1402. [[CrossRef](#)]
79. Dou, J.; Yunus, A.P.; Xu, Y.; Zhu, Z.; Chen, C.-W.; Sahana, M.; Khosravi, K.; Yang, Y.; Pham, B.T. Torrential rainfall-triggered shallow landslide characteristics and susceptibility assessment using ensemble data-driven models in the Dongjiang Reservoir Watershed, China. *Nat. Hazards* **2019**, *97*, 579–609. [[CrossRef](#)]
80. Pham, B.T.; Nguyen, V.-T.; Ngo, V.-L.; Trinh, P.T.; Ngo, H.T.T.; Bui, D.T. A novel hybrid model of rotation forest based functional trees for landslide susceptibility mapping: A case study at Kon Tum Province, Vietnam. In Proceedings of the International Conference on Geo-Spatial Technologies and Earth Resources, Hanoi, Vietnam, 5–6 October 2017; pp. 186–201.
81. Walter, S. The partial area under the summary ROC curve. *Stat. Med.* **2005**, *24*, 2025–2040. [[CrossRef](#)]
82. Tien Bui, D.; Shirzadi, A.; Shahabi, H.; Geertsema, M.; Omidvar, E.; Clague, J.J.; Thai Pham, B.; Dou, J.; Talebpour Asl, D.; Bin Ahmad, B.; et al. New Ensemble Models for Shallow Landslide Susceptibility Modeling in a Semi-Arid Watershed. *Forests* **2019**, *10*, 743. [[CrossRef](#)]
83. Chang, K.-T.; Merghadi, A.; Yunus, A.P.; Pham, B.T.; Dou, J. Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. *Sci. Rep.* **2019**, *9*, 1–21. [[CrossRef](#)] [[PubMed](#)]

84. Thai Pham, B.; Tien Bui, D.; Prakash, I.J.C.E.; Systems, E. Landslide susceptibility modelling using different advanced decision trees methods. *Civ. Eng. Environ. Syst.* **2018**, *35*, 139–157. [[CrossRef](#)]
85. Nguyen, P.T.; Tuyen, T.T.; Shirzadi, A.; Pham, B.T.; Shahabi, H.; Omidvar, E.; Amini, A.; Entezami, H.; Prakash, I.; Phong, T.V. Development of a novel hybrid intelligence approach for landslide spatial prediction. *Appl. Sci.* **2019**, *9*, 2824. [[CrossRef](#)]
86. Pham, B.T.; Pradhan, B.; Tien Bui, D.; Prakash, I.; Dholakia, M.B. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Model. Softw.* **2016**, *84*, 240–250. [[CrossRef](#)]
87. Tien Bui, D.; Pradhan, B.; Lofman, O.; Revhaug, I. Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naive Bayes Models. *Math. Probl. Eng.* **2012**, *2012*. [[CrossRef](#)]
88. Bennett, N.D.; Croke, B.F.W.; Guariso, G.; Guillaume, J.H.A.; Hamilton, S.H.; Jakeman, A.J.; Marsili-Libelli, S.; Newham, L.T.H.; Norton, J.P.; Perrin, C.; et al. Characterising performance of environmental models. *Environ. Model. Softw.* **2013**, *40*, 1–20. [[CrossRef](#)]
89. Khosravi, K.; Shahabi, H.; Pham, B.T.; Adamowski, J.; Shirzadi, A.; Pradhan, B.; Dou, J.; Ly, H.-B.; Gróf, G.; Ho, H.L.; et al. A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. *J. Hydrol.* **2019**, *573*, 311–323. [[CrossRef](#)]
90. Qi, C.; Fourie, A. Cemented paste backfill for mineral tailings management: Review and future perspectives. *Miner. Eng.* **2019**, *144*, 106025. [[CrossRef](#)]
91. Qi, C.; Ly, H.-B.; Chen, Q.; Le, T.-T.; Le, V.M.; Pham, B.T.J.C. Flocculation-dewatering prediction of fine mineral tailings using a hybrid machine learning approach. *Chemosphere* **2019**, *244*, 125450. [[CrossRef](#)]
92. Khosravi, K.; Daggupati, P.; Alami, M.T.; Awadh, S.M.; Ghareb, M.I.; Panahi, M.; Pham, B.T.; Rezaie, F.; Qi, C.; Yaseen, Z.M. Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: A case study in Iraq. *Comput. Electron. Agric.* **2019**, *167*, 105041. [[CrossRef](#)]
93. Khosravi, K.; Barzegar, R.; Miraki, S.; Adamowski, J.; Daggupati, P.; Alizadeh, M.R.; Pham, B.T.; Alami, M.T. Stochastic Modeling of Groundwater Fluoride Contamination: Introducing Lazy Learners. *Ground Water* **2019**. [[CrossRef](#)] [[PubMed](#)]
94. Bayat, M.; Ghorbanpour, M.; Zare, R.; Jaafari, A.; Pham, B.T. Application of artificial neural networks for predicting tree survival and mortality in the Hyrcanian forest of Iran. *Comput. Electron. Agric.* **2019**, *164*, 104929. [[CrossRef](#)]
95. Nguyen, M.D.; Pham, B.T.; Tuyen, T.T.; Yen, H.P.H.; Prakash, I.; Vu, T.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Dou, J.; et al. Development of an Artificial Intelligence Approach for Prediction of Consolidation Coefficient of Soft Soil: A Sensitivity Analysis. *Open Constr. Build. Technol. J.* **2019**, *13*, 178–188. [[CrossRef](#)]
96. Pham, B.T.; Nguyen, M.D.; Van Dao, D.; Prakash, I.; Ly, H.-B.; Le, T.-T.; Ho, L.S.; Nguyen, K.T.; Ngo, T.Q.; Hoang, V.; et al. Development of artificial intelligence models for the prediction of Compression Coefficient of soil: An application of Monte Carlo sensitivity analysis. *Sci. Total. Environ.* **2019**, *679*, 172–184. [[CrossRef](#)]
97. Pham, B.T.; Nguyen, M.D.; Bui, K.-T.T.; Prakash, I.; Chapi, K.; Bui, D. 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](#)]
98. Le, L.M.; Ly, H.-B.; Pham, B.T.; Le, V.M.; Pham, T.A.; Nguyen, D.-H.; Tran, X.-T.; Le, T.-T. Hybrid Artificial Intelligence Approaches for Predicting Buckling Damage of Steel Columns Under Axial Compression. *Materials* **2019**, *12*, 1670. [[CrossRef](#)]
99. Ly, H.-B.; Pham, B.T.; Dao, D.V.; Le, V.M.; Le, L.M.; Le, T.-T. Improvement of ANFIS Model for Prediction of Compressive Strength of Manufactured Sand Concrete. *Appl. Sci.* **2019**, *9*, 3841. [[CrossRef](#)]
100. Nguyen, H.-L.; Pham, B.T.; Son, L.H.; Thang, N.T.; Ly, H.-B.; Le, T.-T.; Ho, L.S.; Le, T.-H.; Bui, D.T. Adaptive Network Based Fuzzy Inference System with Meta-Heuristic Optimizations for International Roughness Index Prediction. *Appl. Sci.* **2019**, *9*, 4715. [[CrossRef](#)]
101. Pham, B.T.; Le, L.M.; Le, T.-T.; Bui, K.-T.T.; Le, V.M.; Ly, H.-B.; Prakash, I. Development of advanced artificial intelligence models for daily rainfall prediction. *Atmos. Res.* **2020**, *237*, 104845. [[CrossRef](#)]
102. Nguyen, H.-L.; Le, T.-H.; Pham, C.-T.; Le, T.-T.; Ho, L.S.; Le, V.M.; Pham, B.T.; Ly, H.-B. Development of Hybrid Artificial Intelligence Approaches and a Support Vector Machine Algorithm for Predicting the Marshall Parameters of Stone Matrix Asphalt. *Appl. Sci.* **2019**, *9*, 3172. [[CrossRef](#)]

103. Ly, H.-B.; Le, L.M.; Duong, H.T.; Nguyen, T.C.; Pham, T.A.; Le, T.-T.; Le, V.M.; Nguyen-Ngoc, L.; Pham, B.T. Hybrid Artificial Intelligence Approaches for Predicting Critical Buckling Load of Structural Members under Compression Considering the Influence of Initial Geometric Imperfections. *Appl. Sci.* **2019**, *9*, 2258. [[CrossRef](#)]
104. Stefanidis, S.; Stathis, D. Assessment of flood hazard based on natural and anthropogenic factors using analytic hierarchy process (AHP). *Nat. Hazards* **2013**, *68*, 569–585. [[CrossRef](#)]
105. Chen, W.; Shahabi, H.; Shirzadi, A.; Hong, H.; Akgun, A.; Tian, Y.; Liu, J.; Zhu, A.-X.; Li, S. Novel hybrid artificial intelligence approach of bivariate statistical-methods-based kernel logistic regression classifier for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* **2019**, *78*, 4397–4419. [[CrossRef](#)]
106. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T.; Duan, Z.; Ma, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* **2017**, *151*, 147–160. [[CrossRef](#)]
107. Pham, B.T.; Bui, D.T.; Pourghasemi, H.R.; Indra, P.; Dholakia, M.B. susceptibility assessment in the Uttarakhand area (India) using GIS: A comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theor. Appl. Climatol.* **2017**, *128*, 255–273. [[CrossRef](#)]
108. Wang, Q.; Li, W.; Wu, Y.; Pei, Y.; Xie, P. Application of statistical index and index of entropy methods to landslide susceptibility assessment in Gongliu (Xinjiang, China). *Environ. Earth Sci.* **2016**, *75*. [[CrossRef](#)]
109. Pandey, V.K.; Sharma, M.C. Probabilistic landslide susceptibility mapping along Tipri to Ghuttu highway corridor, Garhwal Himalaya (India). *Remote Sens. Appl. Soc. Environ.* **2017**, *8*, 1–11. [[CrossRef](#)]
110. Zhou, C.; Yin, K.; Cao, Y.; Ahmed, B.; Li, Y.; Catani, F.; Pourghasemi, H.R. Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China. *Comput. Geosci.* **2018**, *112*, 23–37. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).