Vietnam Journal of Earth Sciences, 44(1), 55-72, https://doi.org/10.15625/2615-9783/16626

Vietnam Academy of Science and Technology
Vietnam Journal of Earth Sciences
http://www.vjs.ac.vn/index.php/jse

Flood susceptibility modeling using Radial Basis Function Classifier and Fisher's linear discriminant function

Chinh Luu[1], Duc Dam Nguyen[2], Mahdis Amiri[3], Tran Van Phong[4], Quynh Duy Bui[5], Indra Prakash[6], Binh Thai Pham[2*]

[1]*Faculty of Hydraulic Engineering, National University of Civil Engineering, Hanoi, Vietnam*
[2]*University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Hanoi, Vietnam*
[3]*Department of Watershed and Arid Zone Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, 49189-434, Iran*
[4]*Institute of Geological Sciences, VAST, Hanoi, Vietnam*
[5]*Department of Geodesy, National University of Civil Engineering, Hanoi, Vietnam*
[6]*DDG(R) Geological Survey of India, Gandhinagar 382015, India*

Received 23 March 2021; Received in revised form 11 June 2021; Accepted 08 October 2021

ABSTRACT

Floods are among the most frequent highly disastrous hazards affecting life, property, and the environment worldwide. While various models are available to predict flood susceptibility, no model is accurate enough to be used for all flood-prone areas. Model development using different algorithms is a continuous process to improve the prediction accuracy of flood susceptibility. In the study, we used the Radial Basis Function and Fisher's linear discriminant function to develop a flood susceptibility map for a case study of Quang Binh Province. The model development used ten variables (elevation, slope, curvature, river density, distance from river, geomorphology, land use, flow accumulation, flow direction, and rainfall). For model training and validation, input data was split into a 70:30 ratio according to flood locations. Statistical indexes were used to evaluate model performance such as Receiver Operating Characteristic, the Area Under the ROC Curve, Root Mean Square Error, Accuracy, Sensitivity, Specificity, and Kappa index. Results indicated that the radial basis function classifier model had better performance in predicting flood susceptible areas based on the statistical measures (PPV = 92.00%, NPV = 87.00%, SST = 87.62%, SPF = 91.58%, ACC = 89.50%, Kappa = 0.790, MAE = 0.204, RMSE = 0.292 and AUC = 0.957. Therefore, the radial basis function classifier algorithm model is appropriate for predicting flood susceptibility in Quang Binh Province.

Keywords: Radial Basis Function Classifier, Fisher's linear discriminant function, Important variables, floods, Quang Binh.

1. Introduction

Floods are the most destructive natural hazards that frequently occur worldwide as a result of heavy rainfall, storm surge on land, tropical cyclone and tsunami in coastal areas. In recent years, more than 89% of the world's natural hazards, such as floods, droughts,

*Corresponding author, Email: binhpt@utt.edu.vn

hurricanes, have been proven to be related to climate change (Convertino et al., 2019). The consequences of flooding have either short- or long-term impacts on human societies. Floods also cause severe damage to the environment and economic losses to individuals, communities, and government (Choubin et al., 2019). Floods pose a challenge to decision-makers and planners to better manage this hazard to prevent damages (Whan et al., 2020).

Several factors exacerbate flooding in an area. High and monsoon rains, tropical cyclones, rapid snowmelt, and insufficient drainage systems are all examples of these complex and interconnected variables. Different spatial and temporal dimensions of rainfall are the primary and severe causes of floods. In addition to the above factors, socio-economic and ecological conditions can significantly increase the impact of floods (Halgamuge & Nirmalathas, 2017).

Complete flood risk prevention is unreasonable and impossible, but predicting flood susceptibility can significantly reduce flood damage. Recently, floods in residential areas and other human-made structures are creating many challenges for flood management (Lyu et al., 2019). Therefore, flood-affected areas must be identified, and the destructive effects of floods should be reduced. For this purpose, it is required to identify flood affecting parameters and flood susceptible areas using modern technology (Qiao et al., 2017). Flood susceptibility assessments are usually done by statistical methods or by physical models of precipitation and run-off (McCuen, 2016). The statistical methods include weights of evidence, frequency ratio, certainty factor, statistical index, and evidential belief function (Romulus Costache & Zaharia, 2017; Khosravi et al., 2016; Siahkamari et al., 2018). In addition, multi-criteria decision-making techniques have yielded successful results in flood assessment such as Analytical Hierarchy Process (AHP), Analytical Network Process (ANP), Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), and VIKOR (Dano et al., 2019; Khosravi et al., 2019; W. Yang et al., 2018). Recently, remote sensing technology and GIS methods have been increasingly used in flood prediction and management (Das, 2019). Several programs are used for run-off rainfall modeling, including HEC-RAS and MIKE (Brunner, 1995; Zhou et al., 2012). However, these hydrological models require long monitoring of meteorological and geoenvironmental data (Kim et al., 2015).

Nowadays, spatial modeling is done using Machine Learning (ML) approach for flood susceptibility modeling due to their high efficiency, accuracy, and predictability (Ahmadlou et al., 2019). ML algorithms often used for flood susceptibility modeling include Support Vector Machine (Tehrany et al., 2014), Logistic Regression (Pham et al., 2020), Artificial Neural Network (R. Costache et al., 2020), Bayesian Logistic Regression (Vogel et al., 2014), Decision Tree, Random Forest, Alternating Decision Tree, Logistic Model Tree, Naïve Bayes tree (Pham et al., 2020), Reduced Error Pruning Tree (W. Chen et al., 2019), k-nearest neighbor, and Deep Learning Neural Networks (Panahi et al., 2021).

In the present study, we used Radial Basis Function Classifier and Fisher's linear discriminant function to develop a flood susceptibility map for the case study of Quang Binh Province. The models were evaluated using standard statistical methods. The data analysis and visualization were carried out using Weka and ArcGIS software.

2. Methods used

2.1. Radial basis function classifier

Radial Based Function Classifier (RBFC) is a type of Artificial Neural Network with a much faster training process than Multi-Layer Perceptron (MLP) for the function approximation and classification problems (Yue Wu et al., 2012). The task of the RBF Classifier algorithm is to model classification or grouping problems. In general, this algorithm is used for different applications in various fields of science (Savitha et al., 2012). These systems generally use local-based functions, such as Gaussians or inverted quadrants, to solve the grouping problems (Ince et al., 2012). RBF classifiers include the compatibility of kernel-based classifiers that utilize a constant variance for each RBF kernel and select RBF centers from the input dataset (He et al., 2019). RBF is a function of actual value, whether it belongs to the distance from the source. Accordingly, the relation $\emptyset(x) = \emptyset(\|x\|)$ can be expressed. If it belongs on another distance, such as spot c, we will have a relationship $\emptyset(x, c) = \emptyset(\|x - c\|)$. It should be noted that any function named "ϕ" that creates the attribute $\emptyset(x) = \emptyset(\|x\|)$ is a radial function (Li et al., 2002). One of the main tasks of the RBF is to solve real multivariate interpolation (MI) problems (Frank, 2014; Shastry et al., 2017). Amongst the different types of RBF methods, one of the most important and most used is the "Gaussian" subordinate, which is expressed as the following equation (Young-Sup & Sung-Yang, 1997):

$$\emptyset(x) = \exp(-|x-c|^2 / 2\partial^2),\ \partial > 0 \tag{1}$$

where ∂ is the Gaussian function scatter. The value ∂ defines the Gaussian function expansion and controls interpolation.

2.2. Fisher's linear discriminant function

The Fisher Linear Discriminant Function Algorithm (FLDA) is a technique for dimensionality reduction before classification that involves finding a linear combination of characteristics to describe or distinguish two or more classes of objects or events. In this analysis, the idea is to maximize a function to provide a significant separation between the projected class means to minimize the class overlap for better classification with the slight variance within each category.

Fisher's linear discriminant algorithm is widely used for 2-instances of natural multivariate data. The central concept for its credibility and validity is that the variances matrices/ matrixes of 2-innate groups are equivalent. It is crucial to study the attributes of FLDA in cases where the variance matrices are not identical (Roh et al., 2019). FLDA is used in attribute selection to decrease dimensions and to differentiate various categories. Data subsets are located in the N by M matrix. Thus, the total scattering matrix is defined as below (Arahal & Berenguel, 1998; Saastamoinen et al., 1998):

$$S_t = \sum (X_n - X_{mean})(X_n - X_{mean})^T \tag{2}$$

In the above formula, X_{mean} is the total vector, X_j represents the results of classes for rows of X_n vectors. S_j illustrates the scatter (S_j) within the X matrix for class j obtained from the following equation:

$$S_j = \sum (X_n - X_{j.mean})(X_n - X_{j.mean})^T \tag{3}$$

where X_j shows the average vector for the susceptibility results category j; T is mentioned to the relocation. If J is the total number of categories outstanding for the data as a total, the within-category distribute (S_w) to the matrix is:

$$S_w = \sum S_j \tag{4}$$

The among-category distribute (S_b) for matrix X is:

$$S_b = \sum n_j (X_{j.mean} - X_{mean})(X_{j.mean} - X_{mean})^T \tag{5}$$

where nj is the whole observations within the category j; the whole -distribute (S_t) matrix as:

$$S_t = S_w + S_b \tag{6}$$

Finally, the first FLDA vector is shown below:

$$\text{Max } w^t_1 S_b w_1 / w^t_1 S_w w_1$$

The second FLDA vector (w_k) can also be created that maximizes. The diffuse among the J categories but together minimizes the diffuse within each category. The FLDA vectors demonstrate the special vectors w_k of the universal special value presentation:

$$S_b\, w_k = \lambda_j\, S_w w_k \tag{7}$$

λ_j demonstrates the measure of throughout segregation between the j categories.

2.3. Validation methods

2.3.1. Receiver Operating Characteristic (ROC) curve

One of the most common and well-known evaluation statistical indicators is the Receiver Operating Characteristics (ROC) curve. The specific results of ROC can be generalized and trusted. The ROC technique is a probabilistic approach for evaluating algorithms over a wide range of thresholds (Avand et al., 2020). The ROC graph consists of two axes x and y. Each of these represents the characteristics of specificity and sensitivity, respectively (Jaafari, 2018). The Area Under the ROC curve (AUC) is usually used as a criterion for measuring the predictive accuracy of algorithms. The AUC value varies between "0.5–1". A value of 0.5 indicates the poor performance of algorithms in predicting susceptibility to natural hazards such as floods and landslides. A value of AUC 1 indicates the strong performance of algorithms in predicting flood and landslide susceptibility. In general, the closer the AUC value is to 1, the higher the accuracy of the algorithm, whereas the closer to 0.5, the lower the algorithm's accuracy (Wei Chen et al., 2019). The equation of the evaluation method using the AUC method is as follows:

$$AUC = \frac{\sum E + \sum I}{A + B} \tag{8}$$

where the "E" illustrates the numeral of the exactly categorized flood happens, "I" defines the numeral of the false categorized flood happens, "A" defines the whole numeral of non-flood samples, and "B" defines the whole numeral of flood samples.

2.3.2. Statistical Indexes

Statistical indicators are used to evaluate the performance of the models, including ROC, AUC, Positive Predictive Value (PPV), Negative Predictive Value (NPV), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) (Pham et al., 2021; Tran and Prakash, 2020; Van Phong et al., 2020), Accuracy, Sensitivity, Specificity, and Kappa index (K). The PPV and NPV criteria are the probabilities of correctly grouped pixels as "flood" and "non-flood". The Sensitivity indicator shows the ratio of flood pixels, while the Specificity shows the proportion of non-flood pixels. The K indicator is a useful statistic that may be used to assess random agreement among categorization variables. K varies between 1 and -1. If K values are close to 1, it indicates the high credibility and reliability of the algorithm in predicting flood susceptibility. The Accuracy criterion estimates the ratio of accurate flood forecast to total flood forecast (De Rosa et al., 2019; Prăvălie & Costache, 2014). The RMSE denotes the difference between observed and calculated data. The MAE is an extent of errors among binary observations. The higher values of SPF, PPV, NPV, ACC, SST, and K and lower RMSE and MAE denote the higher performance of the model in predicting flood susceptibility. The evaluation criteria' equations are given in published works (Singh et al., 2018; Yanli Wu et al., 2020).

3. Study area

Quang Binh is a North Central coastal province, located at latitude 16°56'20" to 18°5'12" North and longitude 105°36'55" to 106°59'37" East (Fig. 1). This province was chosen for the research because it is one of Vietnam's most flood-prone regions (Luu et al., 2019).

Quang Binh has an area of 800 km^2 and a coast length of 116.04 km. The population of the province is 887,600 in 2018. The topography of the area is adulatory with steep hills and mountains. Rivers and streams in this area are of short length. Thus, floods occur with high intensity, raising the floodwater quickly in the river valleys and low-lying areas. Floods and storms occur mainly from August to November. Severe storms and floods have recently occurred in this region in 1995, 1999, 2007, 2008, 2010, 2016, and 2020.

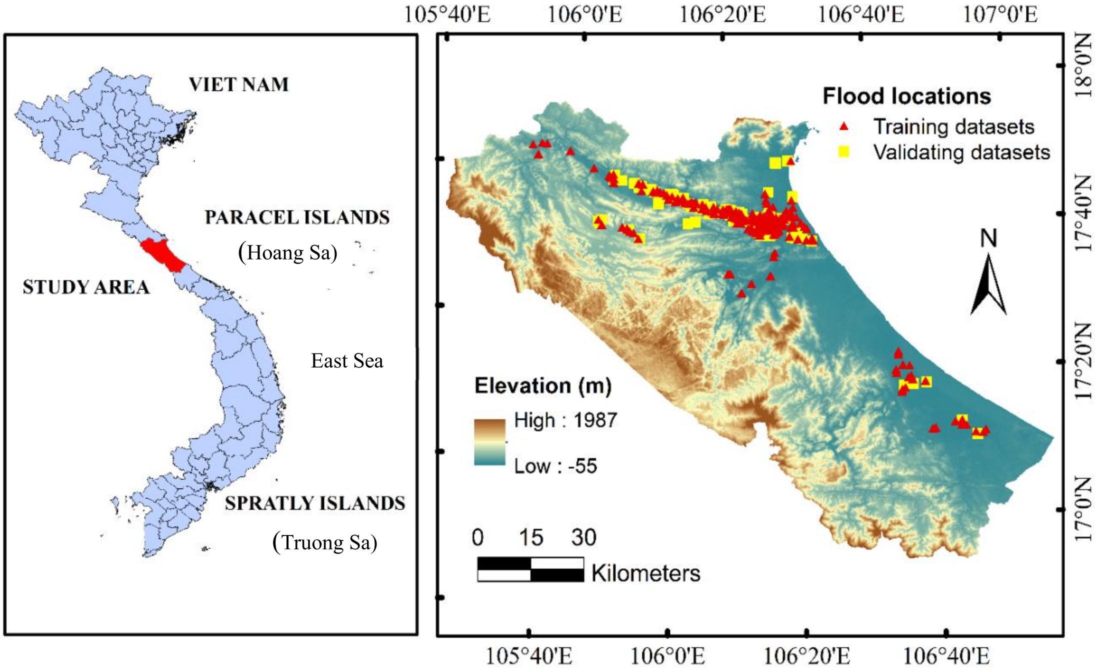

Figure 1. The research area's location and historical flood points

4. Data used

4.1. Flood inventory

The flood inventory database was developed from the historical flood record, remote sensing, and field survey. Flood inventory is required for developing flood susceptibility maps. The flood inventory database used in this research contains 321 historical flooded sites from 2007, 2010, and 2016 (Figs. 1, 2). This data was collected from the Quang Binh Center for Hydro-Meteorological Forecasting department and used to develop flood susceptibility maps in the present study.

Figure 2. Flooding photos of the study area (Source: https://www.quangbinh.gov.vn)

4.2. Flood influencing factors

This study selected ten flood influencing factors: elevation, slope, curvature, river density, distance from the river, geology, flow accumulation, flow direction, land use, and rainfall (Fig. 3 and Table 1). The selection of these factors is based on the geo-environmental conditions of the study area and literature review of many studies relating to flood susceptibility modeling (Dottori et al., 2018; Gonzalez-Arqueros et al., 2018). The data used in this study were also used in other works (Luu et al., 2021). Figure 4 shows the frequency ratio analysis of the affecting factors with flood inventory of the study area.

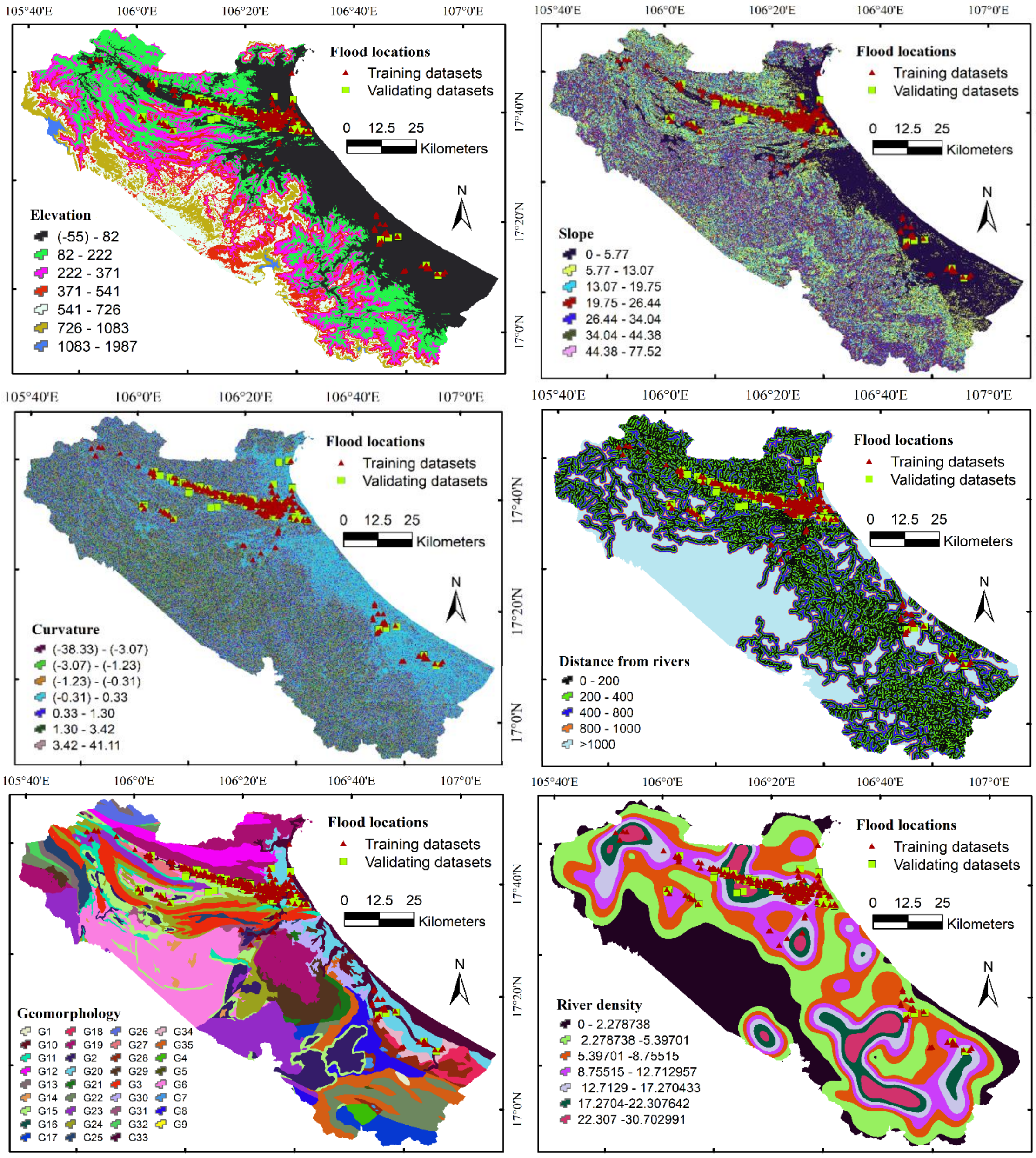

Figure 3. Flood affecting variables are shown in thematic maps (Luu et al., 2021)

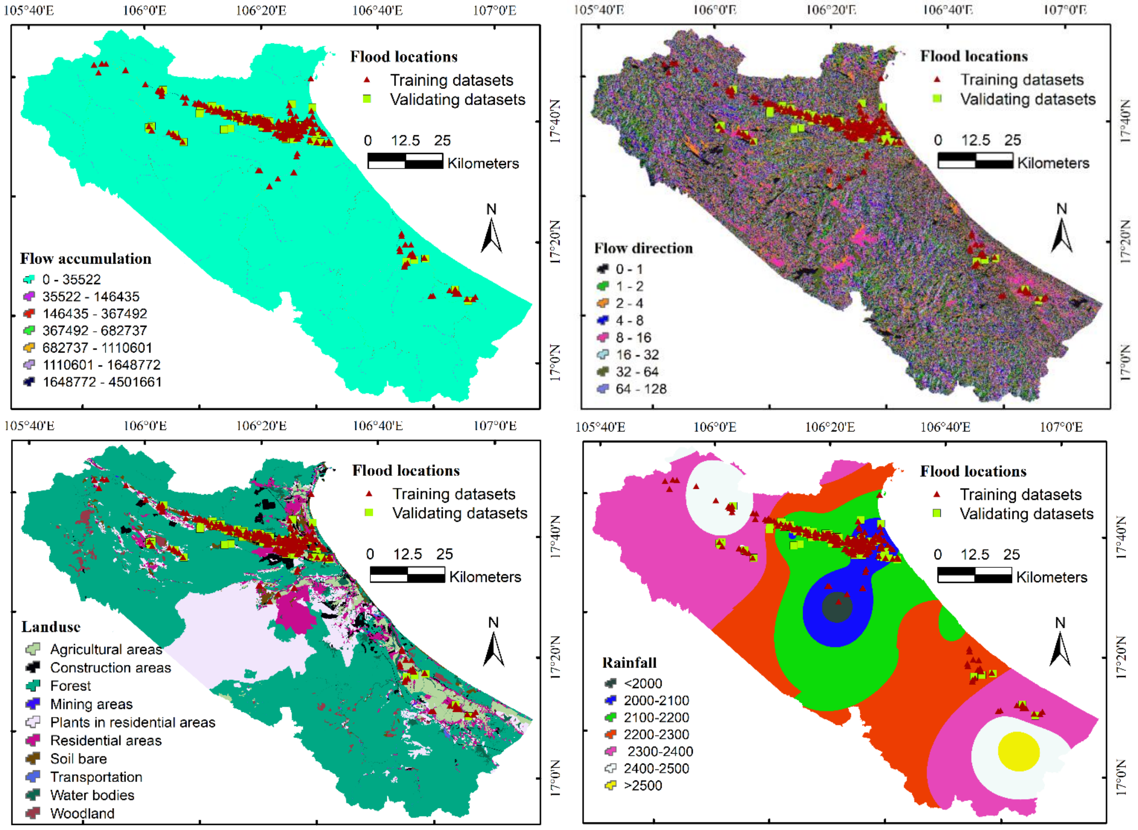

Figure 3. Continue

Table 1. Flood susceptibility modeling spatial variables and datasets (Luu et al., 2021)

Factors	Sources	Spatial resolution	Years
Elevation	EARTHDATA SEARCH	30 m	2013
Slope	EARTHDATA SEARCH	30 m	2013
Curvature	EARTHDATA SEARCH	30 m	2013
Flow direction	EARTHDATA SEARCH	30 m	2013
Flow accumulation	EARTHDATA SEARCH	30 m	2013
Rainfall	16 rainfall stations	30 m	Yearly, 1986-2016
River density	River network map	1:50,000	2015
Distance from river	River network map	1:50,000	2015
Land-use	Land-use categories map	1:50,000	2015
Geology	Geological map	1:50,000	2015

4.2.1. Elevation

The elevation has a significant impact on the flood situation (Razavi Termeh et al., 2018). Water flows from higher altitudes and collects at lower elevations in flat areas, producing floods in river valleys, flood plains, and rvier banks. The elevation map was generated from the DEM of the research region and categorized into seven class intervals using the natural breaks classification technique.

4.2.2. Slope

Steep slopes result in higher surface run-off and soil erosion, whereas gentle slopes cause accumulation of the water (Agassi et al., 1990). The DEM was used to create the slope angle map, which was split into seven groups. The values of the slope angle vary from 0 to 77.52°.

4.2.3. Curvature

Another topographic element derived from the DEM map is curvature. Curvature splits the surface into three types of slopes: convex, concave, and flat (Shafizadeh-Moghadam et al., 2018). The run-off will be more on the convex surface in comparison to the concave surface. Accumulation of the water would be on the flat surfaces. The natural break classification technique was used to divide the raster map into seven groups.

4.2.4. River density

Most Quang Binh rivers and streams originate in the province's territory and then flow into the sea. Rainfall in the river basins determines the flood regime in rivers. The higher density of the river creates more run-off, but it also results in flooding of the larger areas of territory occupied by several river basins. The river density map was derived from a river network map using the Dissolve tool in ArcGIS Pro.

4.2.5. Distance from the river

The Euclidean Distance tool in ArcGIS Pro is used to determine distance from the river using river network data. Due to an excess of water from river banks, inundation zones are often seen around rivers (Santangelo et al., 2011). The distance from the river was divided into five class intervals in this research using the manual categorization technique.

4.2.6. Geomorphology

Geomorphology is considered in flood susceptibility modeling since it affects the water run-off and infiltration process (Bui et al., 2019). Geomorphology is related to landforms that are dependent on sub-surface geology and erosional and depositional processes. There are 35 geological formation types in the study, and all data was obtained from the Quang Binh Department of Natural Resources and Environment.

4.2.7. Lan duse

Land use is a significant variable in research focusing on flood susceptibility modeling (Panahi et al., 2021). Land use categories affect the run-off and infiltration, thus flooding. The land use map of Quang Binh province is provided by the Department of Natural Resources and Environment. It consists of ten categories: agricultural areas, residential areas, construction areas, plants in residential areas, transportation, forest, woodland, soil bare, water bodies, and mining areas.

Land use is a major variable in research focusing on flood vulnerability modeling (Panahi et al., 2021). Land usage in the region has an impact on run-off and infiltration, resulting in floods. The Department of Natural Resources and Environment provided the land use map for Quang Binh province. Agricultural areas, residential areas, construction areas, plants in residential areas, transportation, forest, woodland, soil bare, water bodies, and mining areas are the 10 categories on the land use map.

4.2.8. Flow accumulation

The flow accumulation is a metric that shows how fast the river network is moving (R. Costache et al., 2020). Using ArcGIS software, the flow accumulation map was created from the flow direction map by calculating accumulated flow based on the cumulative weight of all cells flowing from upstream into every downward cell. The map was created with seven class intervals Using the natural break categorization technique.

4.2.9. Flow direction

Flow direction is a hydrological element that indicates the flow's direction. Flow direction maps show the direction of surface flow and represent the contribution of run-off towards flooding. Flow direction maps depict the direction of surface flow and the amount of run-off that contributes to floods. The Arc Hydro tool in ArcGIS Pro was used to generate the flow direction raster. The flow direction map was created with five class intervals using the natural break classification technique.

4.2.10. Rainfall

Rainfall is a critical factor in modeling flood susceptibility (Y. Wang et al., 2019). There are ten rainfall stations in Quang Binh province: Kien Giang, Le Thuy, Mai Hoa, Ba Don, Viet Trung, Trooc, Dong Tam, Tuyen Hoa, Dong Hoi, and Minh Hoa, with the availability of continuous data from the year 1965-2016. The *Inverse Distance Weighted* technique was used to create a rainfall map for the study region. The natural break categorization technique was used to divide a rainfall map into seven groups.

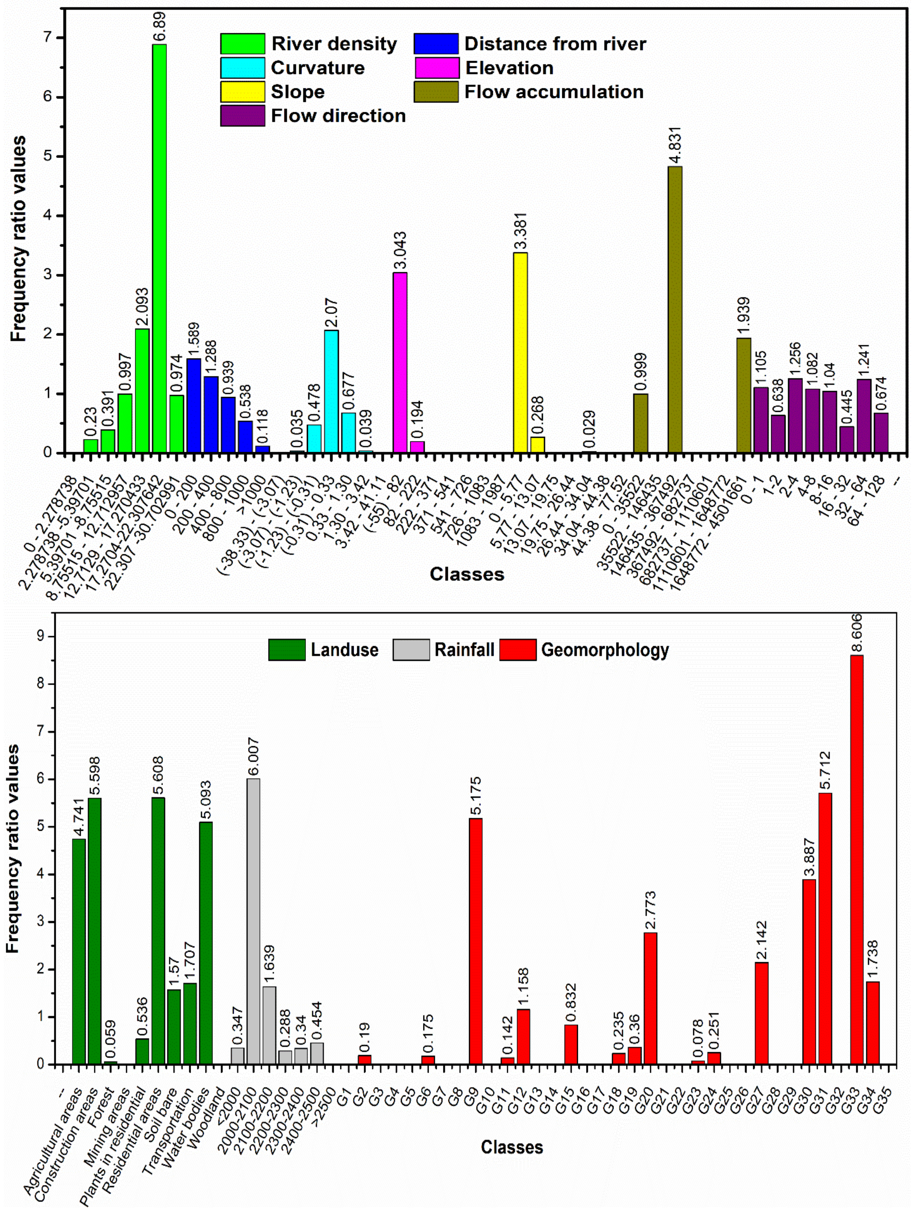

Figure 4. Frequency ratio of the flood influencing factors

5. Methodology flowchart

The methodology adopted for modeling and mapping flood susceptibility is shown in Figure 5 and summarized in the following steps:

Step 1: Selecting flood influencing variables from different sources.

Step 2: Recording flood events in the study area and dividing them into two groups for model training (70% of data set) and model validation (30% of data set).

Step 3: Predictive modeling of flood susceptibility RBFC and FLDA. Table 2 shows the hyper-parameters used for each model in this study.

Step 4: Evaluate the accuracy of the algorithms' performance mentioned in the previous step using SPF, PPV, NPV, ACC, MAE, RMSE, AUC, SST, and K indices.

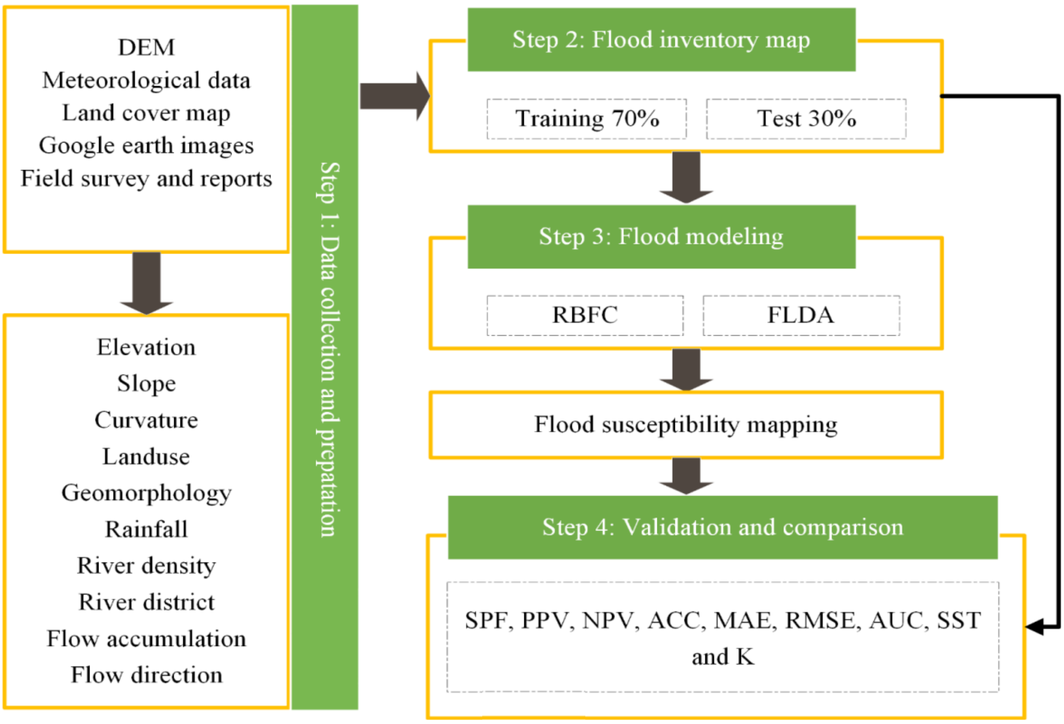

Figure 5. Methodological framework of flood susceptibility modeling

Table 2. Hyper-parameters used for each model in this study

No	Hyper-parameters	Models	
		FLDA	RBFC
1	Batch size	100	100
2	Debug	False	False
3	Do not check capabilities	False	False
4	Num decimal places	2	2
5	Ridge	1.0E-6	0.02
6	Num function	-	2
7	Num threads	-	1
8	Pool size	-	1
9	Scale optimization option	-	Use scale per unit
10	Seed	-	1
11	Tolerance	-	1.0E-6
12	Use attribute weights		False
13	Use CGD	-	False

6. Results

6.1. Validation of the models

Table 3 shows the performance of both RBFC and FLDA models overtraining and validation phases. In the training phase, the RBFC algorithm has higher performance according to the PPV (97.37%), SPF (89.66%), ACC (86.93%), SST (84.55%), and NPV (83.49%). FLDA model has higher performance in terms of PPV (88.99%), SPF (88.52%), ACC (86.93%), SST (85.46%), and NPV (84.86%). On the other hand, the FLDA model has higher performance for K (0.739), and RBFC has higher accuracy in MAE (0.208). In the testing phase, the RBFC algorithm has higher performance in terms of PPV (92.00%), SPF (91.58%), ACC (89.50%), SST (87.62%), and NPV (87.00%). Accordingly, it has the highest accuracy in predicting flood sensitivity. On the other hand, the K indicator is higher in the RBFC algorithm (0.790), and the MAE index is higher in the RBFC algorithm (0.204). Figure 6 shows the analysis of RMSE for 400 samples of training data. The index of RMSE for the RBFC algorithm is 0.300 and for the FLDA algorithm is 0.318, which indicates that the RBFC algorithm is more accurate in forecasting flood susceptibility in the study area. Accordingly, Fig. 7 shows the RMSE for test data with 200 randomly selected samples of flood locations. The result shows that the RBFC algorithm with a value of RMSE 0.292 is more accurate than the FLDA algorithm with a value of RMSE 0.313 on predicting flood susceptibility. Therefore, the RBFC model is more effective.

Table 3. Performance analysis of the models

No	Parameter	Traning		Validation	
		RBFC	FLDA	RBFC	FLDA
1	PPV (%)	90.37	88.99	92.00	90.00
2	NPV (%)	83.49	84.86	87.00	88.00
3	SST (%)	84.55	85.46	87.62	88.24
4	SPF (%)	89.66	88.52	91.58	89.80
5	ACC (%)	86.93	86.93	89.50	89.00
6	Kappa (k)	0.379	0.739	0.790	0.780
7	MAE	0.208	0.246	0.204	0.243

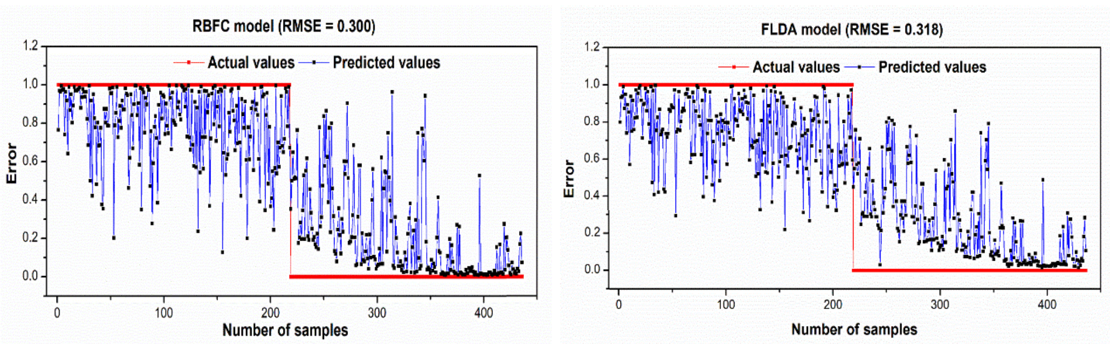

Figure 6. RMSE of the models in training phase

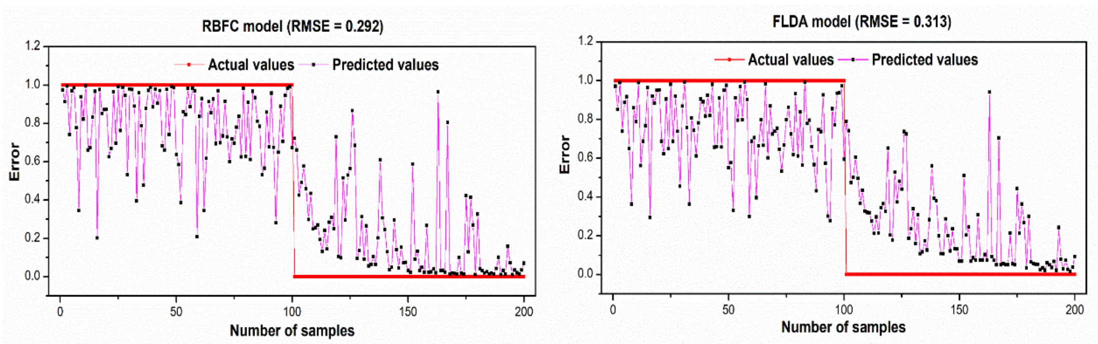

Figure 7. RMSE of the models in validation phase

Figure 8, on the other hand, shows the accuracy and precision of the algorithms used by the AUC standard for training and testing data. The results indicate that in the training phase, the accuracy of the RBFC algorithm (AUC = 0.951) is higher than the accuracy of the FLDA algorithm (AUC = 0.945). Also, in the validating stage, the precision of the RBFC algorithm (AUC = 0.957) is higher than the FLDA algorithm (AUC = 0.948).

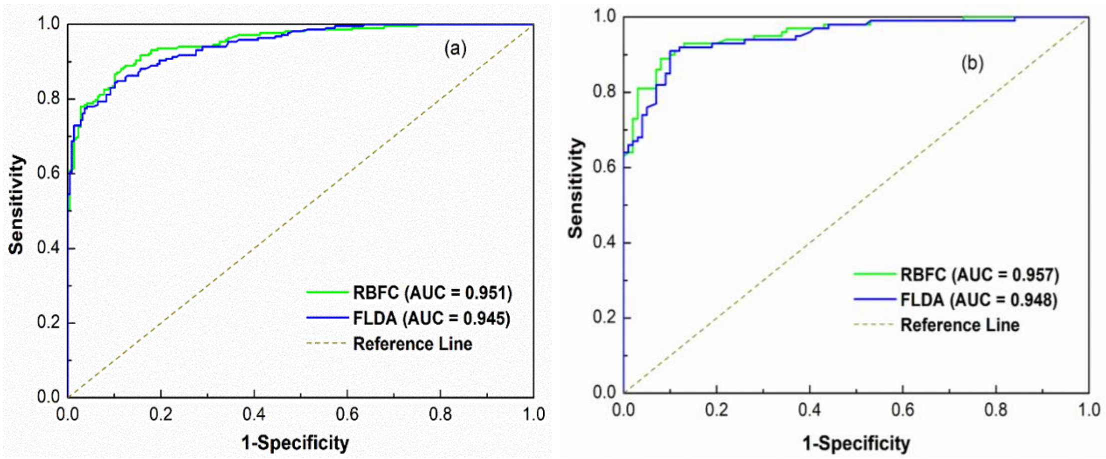

Figure 8. AUC in the training (a) and validation (b) stages of the models

6.2. Construction of flood susceptibility maps

The flood susceptibility maps generated by RBFC and FLDA algorithms are shown in Figure 9. These maps were divided into 5 classes using Reclassify tool and Natural Break method in ArcGIS software. Flood susceptibility map using RBFC model include classes: Very low (0.001-0.12), low (0.12-0.287), moderate (0.287-0.513), high (0.513-0.751), and very high (0.751-0.993), whereas map developed using the FLDA model include classes: very low (0-0.126), low (0.126-0.302), moderate (0.302-0.507), high (0.507-0.726), and very high (0.726-0.998).

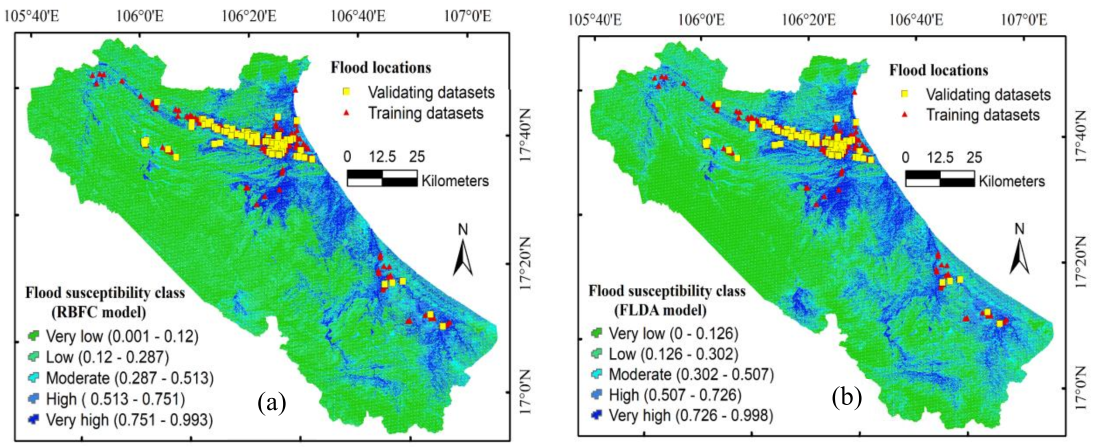

Figure 9. Flood susceptibility maps developed using the RBFC (a) and FLDA (b) models

Figure 10 shows the analysis of the density of flood occurrence points in 5 classes of flood susceptibility using RBFC and FLDA models. It can be seen that the percentage of flood pixels in very low, low, moderate, high, and very high classes is 45.08%, 22.77%, 11.68%, 9.96%, and 10.76%, respectively, according to the RBFC algorithm. Also, according to the FLDA algorithm, the percentage of flood occurrence pixels is 39.67%, 21%, 15.3%, 13.28%, and 10.5%, respectively, in very low, low, moderate, high, and very high classes of the map.

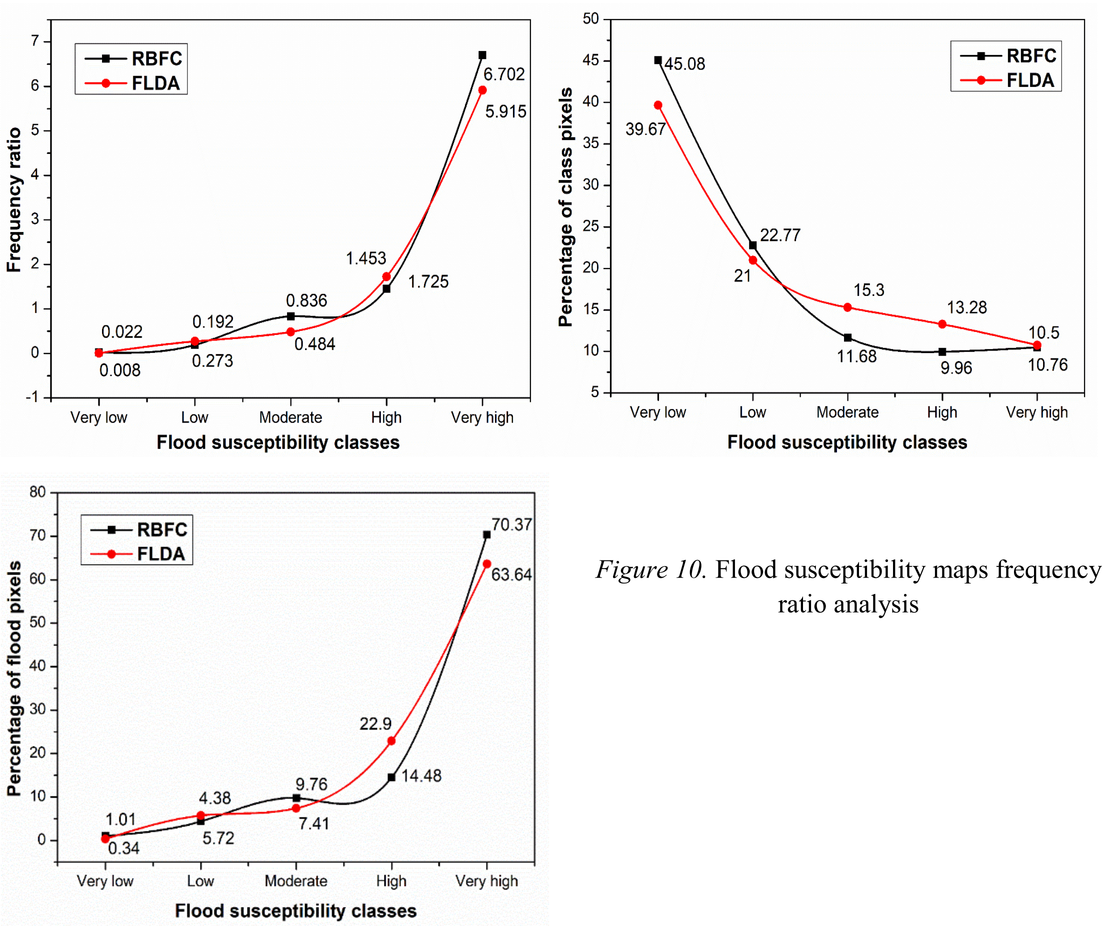

Figure 10. Flood susceptibility maps frequency ratio analysis

7. Discussions

Flood damage can be exacerbated by the aggravating factors of population growth, climate change, and land use. It is necessary to identify flood-prone areas for risk management and planning. Thus, flood sensitivity zoning is a good option to solve this problem (Shahabi et al., 2021). On the other hand, more and more studies on the extent and intensity of flood hazards worldwide use modeling methods (Khosravi et al., 2016). Accordingly, various models have been developed to predict flood sensitivity. However, accurate prediction of flood sensitivity areas is still under discussion, and different models should be used and tested for this purpose. Some models, such as physical and hydrological models, can not be applied in areas where data is scarce or lacks long-term monitoring data (Tien Bui et al., 2020). Instead, new emerged models and algorithms such as ML algorithms can deal with more popular data. In ML models, all variables affecting flood susceptibility can be used for spatial modeling. Therefore, it can predict flood-prone areas more accurately (Pham et al., 2017).

The selection of affecting variables is essential, and it depends on the environmental conditions and a specific region (Z. Yang et al., 2020). For example, during heavy rainfall, flash floods can occur in mountainous areas due to the topography. On the other hand, bare lands without vegetation can increase the speed of floodwater and creating flash-flood. In general, it can be said that the topography, land use, and course of rivers affect the occurrence and intensification of flood susceptibility (Pandey et al., 2020; Q. Wang et al., 2016). Due to the complexity of predicting flood sensitivity using different spatial modeling approaches, it is necessary to use different algorithms and compare them, as seen in many studies (Fu et al., 2020; Zenggang et al., 2019; J. Zhu et al., 2018). In this study, two of the practical ML models, RBFC and FLDA, were selected for flood susceptibility modeling. The advantage of these models is to reduce the dimensions of the problem and thus simplify and reach the answer faster (Y. Zhu et al., 2018). The results showed that the RBFC model had higher quality and efficiency according to the AUC criteria.

The results indicate that the two ML models of RBFC and FLDA, with AUCs of 0.957 and 0.948, respectively, perform well for flood susceptibility mapping in the study region. In another research, Pham et al. (2020) used RBFC algorism for flood susceptibility modeling and had an AUC of 0.984. According to the current study's model performance evaluation, the FBFC tree has the greatest prediction ability, with an accuracy of 89.5%. Pham et al. (2018) also used RBFC and combined it with other ML models to estimate the flood susceptibility. Thus, the flood susceptibility maps constructed by the RBFC and FLDA techniques can provide helpful information for disaster risk reduction activities, especially in localities where real-time simulation models have not been developed yet.

8. Conclusions

In the present study, two critical models, namely RBFC and FLDA, were used for modeling the flood susceptibility for a case study of Quang Binh province. The statistical analysis results indicated that the RBFC algorithm has better performance and can develop an accurate flood susceptible map of the research area. The RBFC is a type of Artificial Neural Network has shown good prediction result of flood susceptibility model; thus it can also be applied in other flood-prone areas. In addition, the RFBC algorithm can be combined with other ML algorithms to develop more accurate mapping and prediction models. The proposed approach can be extended to other regions that have available topography, hydrology, geology, and land use. The limitation of this study is that we applied single models for flood susceptibility modeling; hybrid models should be developed for mapping flood susceptibility in the area for further studies. These validated models could be used for other flood-prone regions, taking into account the local geo-environmental features.

Acknowledgments

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.08-2019.319.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

Agassi M., Morin J., Shainberg I., 1990. Slope, Aspect, and Phosphogypsum Effects on Runoff and Erosion. Soil Science Society of America Journal, 54(4), 1102-1106. https://doi.org/10.2136/sssaj1990.03615995005400040030x.

Ahmadlou M., Karimi M., Alizadeh S., Shirzadi A., Parvinnejhad D., Shahabi H., Panahi M., 2019. Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). Geocarto International, 34(11), 1252-1272. Doi: 10.1080/10106049.2018.1474276.

Arahal M.R., Berenguel M., 1998. Constructive Radial Basis Function Networks for Mobile Robot Positioning. IFAC Proceedings Volumes, 31(2), 181-186. https://doi.org/10.1016/S1474-6670(17)44193-0.

Avand M., Janizadeh S., Tien Bui D., Pham V.H., Ngo P.T.T., Nhu V.-H., 2020. A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. International Journal of Digital Earth, 13(12), 1408-1429. Doi: 10.1080/17538947.2020.1718785.

Brunner G.W., 1995. HEC-RAS River Analysis System. Hydraulic Reference Manual. Version 1.0.

Bui D.T., Ngo P.-T.T., Pham T.D., Jaafari A., Minh N.Q., Hoa P.V., Samui P., 2019. A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. CATENA, 179, 184-196. https://doi.org/10.1016/j.catena.2019.04.009.

Chen W., et al., 2019. Spatial Prediction of Landslide Susceptibility Using GIS-Based Data Mining Techniques of ANFIS with Whale Optimization Algorithm (WOA) and Grey Wolf Optimizer (GWO). Applied Sciences, 9(18), 3755. Retrieved from https://www.mdpi.com/2076-3417/9/18/3755.

Chen W., Hong H., Li S., Shahabi H., Wang Y., Wang X., Ahmad B.B., 2019. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. J. of Hydrology, 575, 864-873. https://doi.org/10.1016/j.jhydrol.2019.05.089.

Choubin B., Moradi E., Golshan M., Adamowski J., Sajedi-Hosseini F., Mosavi A., 2019. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Sci. Total Environ, 651(Pt2), 2087-2096. https://doi.org/10.1016/j.scitotenv.2018.10.064.

Convertino M., Annis A., Nardi F., 2019. Information-theoretic portfolio decision model for optimal flood management. Environmental Modelling & Software, 119, 258-274. https://doi.org/10.1016/j.envsoft.2019.06.013.

Costache R., et al., 2020. Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment. J Environ Manage, 265, 110485. https://doi.org/10.1016/j.jenvman.2020.110485.

Costache R., Zaharia L., 2017. Flash-flood potential assessment and mapping by integrating the weights-of-evidence and frequency ratio statistical methods in GIS environment - case study: Bâsca Chiojdului River catchment (Romania). J. of Earth System Sci., 126(4), 59. Doi: 10.1007/s12040-017-0828-9.

Dano U.L., et al., 2019. Flood Susceptibility Mapping Using GIS-Based Analytic Network Process: A Case Study of Perlis, Malaysia. Water, 11(3), 615. Retrieved from https://www.mdpi.com/2073-4441/11/3/615.

Das S., 2019. Geospatial mapping of flood susceptibility and hydro-geomorphic response to the floods in Ulhas basin, India. Remote Sensing Applications: Society and Environment, 14, 60-74. https://doi.org/10.1016/j.rsase.2019.02.006.

De Rosa P., Fredduzzi A., Cencetti C., 2019. Stream Power Determination in GIS: An Index to Evaluate the Most 'Sensitive'Points of a River. Water, 11(6), 1145. Retrieved from https://www.mdpi.com/2073-4441/11/6/1145.

Dottori F., Martina M.L.V., Figueiredo R., 2018. A methodology for flood susceptibility and vulnerability analysis in complex flood scenarios. Journal of Flood Risk Management, 11, S632-S645. https://doi.org/10.1111/jfr3.12234.

Frank E., 2014. Fully supervised training of Gaussian radial basis function networks in WEKA.

Fu X., Pace P., Aloi G., Yang L., Fortino G., 2020. Topology optimization against cascading failures on wireless sensor networks using a memetic algorithm. Computer Networks, 177, 107327. https://doi.org/10.1016/j.comnet.2020.107327.

Gonzalez-Arqueros M.L., Mendoza M.E., Bocco G., Solis Castillo B., 2018. Flood susceptibility in rural settlements in remote zones: The case of a mountainous basin in the Sierra-Costa region of Michoacan, Mexico. J Environ Manage, 223, 685-693. https://doi.org/10.1016/j.jenvman.2018.06.075.

Halgamuge M.N., Nirmalathas A., 2017. Analysis of large flood events: Based on flood data during 1985-2016 in Australia and India. International Journal of Disaster Risk Reduction, 24, 1-11. https://doi.org/10.1016/j.ijdrr.2017.05.011.

He Q., et al., 2019. Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. Science of The Total Environment, 663, 1-15. https://doi.org/10.1016/j.scitotenv.2019.01.329

Ince T., Kiranyaz S., Gabbouj M., 2012. Evolutionary RBF classifier for polarimetric SAR images. Expert Systems with Applications, 39(5), 4710-4717. https://doi.org/10.1016/j.eswa.2011.09.082.

Jaafari A., 2018. LiDAR-supported prediction of slope failures using an integrated ensemble weights-of-evidence and analytical hierarchy process. Environmental Earth Sciences, 77(2), 42. Doi: 10.1007/s12665-017-7207-3.

Khosravi K., et al., 2019. A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. Journal of Hydrology, 573, 311-323. https://doi.org/10.1016/j.jhydrol.2019.03.073.

Khosravi K., Nohani E., Maroufinia E., Pourghasemi H.R., 2016. A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. Natural Hazards, 83(2), 947-987. Doi: 10.1007/s11069-016-2357-2.

Kim B., Sanders B.F., Famiglietti J., Guinot V., 2015. Urban flood modeling with porous shallow-water equations: a case study of model errors in the presence of anisotropic porosity. Journal of Environmental Hydrology, 523, 680-692. Doi: 10.1016/j.jhydrol.2015.01.059.

Li Y., Pont M.J., Barrie Jones N., 2002. Improving the performance of radial basis function classifiers in condition monitoring and fault diagnosis applications where `unknown’ faults may occur. Pattern Recognition Letters, 23(5), 569-577. https://doi.org/10.1016/S0167-8655(01)00133-7.

Luu C., et al., 2021. GIS-Based Ensemble Computational models for Flood Susceptibility Prediction in the Quang Binh Province, Vietnam. Journal of Hydrology, 126500. https://doi.org/10.1016/j.jhydrol.2021.126500.

Luu C., von Meding J., Mojtahedi M., 2019. Analyzing Vietnam’s national disaster loss database for flood risk assessment using multiple linear regression-TOPSIS. International Journal of Disaster Risk Reduction, 40, 101153. https://doi.org/10.1016/j.ijdrr.2019.101153.

Lyu H.-M., Shen S.-L., Zhou A.-N., Zhou W.-H., 2019. Flood risk assessment of metro systems in a subsiding environment using the interval FAHP-FCA approach. Sustainable Cities and Society, 50, 101682. https://doi.org/10.1016/j.scs.2019.101682.

McCuen R.H., 2016. Modeling hydrologic change: statistical methods: CRC press.

Panahi M., et al., 2021. Deep learning neural networks for spatially explicit prediction of flash flood probability. Geoscience Frontiers, 12(3). Doi: 10.1016/j.gsf.2020.09.007.

Pandey V.K., Pourghasemi H.R., Sharma M.C., 2020. Landslide susceptibility mapping using maximum entropy and support vector machine models along the highway corridor, Garhwal Himalaya. Geocarto International, 35(2), 168-187. Doi: 10.1080/10106049.2018.1510038.

Pham B.T., Bui D.T., Dholakia M.B., Prakash I., Pham H.V., Mehmood K., Le H.Q., 2017. A novel ensemble classifier of rotation forest and Naïve Bayer for landslide susceptibility assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. Geomatics, Natural Hazards and Risk, 8(2), 649-671. Doi: 10.1080/19475705.2016.1255667.

Pham B.T., et al., 2020. A Comparative Study of Kernel Logistic Regression, Radial Basis Function

Classifier, Multinomial Naïve Bayes, and Logistic Model Tree for Flash Flood Susceptibility Mapping. Water, 12(1). Doi: 10.3390/w12010239.

Pham B.T., et al., 2021. Estimation of shear strength parameters of soil using Optimized Inference Intelligence System. Vietnam Journal of Earth Sciences, 43, 189-198.

Pham B.T., Shirzadi A., Tien Bui D., Prakash I., Dholakia M.B., 2018. A hybrid machine learning ensemble approach based on a Radial Basis Function neural network and Rotation Forest for landslide susceptibility modeling: A case study in the Himalayan area, India. International Journal of Sediment Research, 33(2), 157-170. https://doi.org/10.1016/j.ijsrc.2017.09.008.

Prăvălie R., Costache R.. 2014. The analysis of the susceptibility of the flash-floods' genesis in the area of the hydrographical basin of Bâsca Chiojdului river. Forum geografic, XIII(1), 39-49. Doi: 10.5775/fg.2067-4635.2014.071.i.

Qiao Y.-K., Peng F.-L., Wang Y., 2017. Monetary valuation of urban underground space: A critical issue for the decision-making of urban underground space development. Land Use Policy, 69, 12-24. https://doi.org/10.1016/j.landusepol.2017.08.037.

Razavi Termeh S.V., Kornejady A., Pourghasemi H.R., Keesstra S., 2018. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. Sci. Total Environ, 615, 438-451. https://doi.org/10.1016/j.scitotenv.2017.09.262.

Roh S.-B., Oh S.-K., Pedrycz W., Seo K., Fu Z., 2019. Design methodology for Radial Basis Function Neural Networks classifier based on locally linear reconstruction and Conditional Fuzzy C-Means clustering. International Journal of Approximate Reasoning, 106, 228-243. https://doi.org/10.1016/j.ijar.2019.01.008.

Saastamoinen A., Pietilä T., Värri A., Lehtokangas M., Saarinen J., 1998. Waveform detection with RBF network - Application to automated EEG analysis. Neurocomputing, 20(1), 1-13. https://doi.org/10.1016/S0925-2312(98)00005-8.

Santangelo N., Santo A., Di Crescenzo G., Foscari G., Liuzza V., Sciarrotta S., Scorpio V., 2011. Flood susceptibility assessment in a highly urbanized alluvial fan: the case study of Sala Consilina (southern Italy). Natural Hazards and Earth System Sciences, 11(10), 2765-2780. Doi: 10.5194/nhess-11-2765-2011.

Savitha R., Suresh S., Sundararajan N., Kim H.J., 2012. A fully complex-valued radial basis function classifier for real-valued classification problems. Neurocomputing, 78(1), 104-110. https://doi.org/10.1016/j.neucom.2011.05.036.

Shafizadeh-Moghadam H., Valavi R., Shahabi H., Chapi K., Shirzadi A., 2018. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. J. Environ Manage, 217, 1-11. Doi: 10.1016/j.jenvman.2018.03.089.

Shahabi H., et al., 2021. Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. Geoscience Frontiers, 12(3), 101100. https://doi.org/10.1016/j.gsf.2020.10.007.

Shastry K.A., Sanjay H.A., Deexith G., 2017. Quadratic-radial-basis-function-kernel for classifying multi-class agricultural datasets with continuous attributes. Applied Soft Computing, 58, 65-74. https://doi.org/10.1016/j.asoc.2017.04.049.

Siahkamari S., Haghizadeh A., Zeinivand H., Tahmasebipour N., Rahmati O., 2018. Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models. Geocarto International, 33(9), 927-941. Doi: 10.1080/10106049.2017.1316780.

Singh C., Walia E., Kaur K.P., 2018. Enhancing color image retrieval performance with feature fusion and non-linear support vector machine classifier. Optik, 158, 127-141. https://doi.org/10.1016/j.ijleo.2017.11.202.

Tehrany M.S., Pradhan B., Jebur M.N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. Journal of Hydrology, 512, 332-343. https://doi.org/10.1016/j.jhydrol.2014.03.008.

Tien Bui D., et al., 2020. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. Science of The Total Environment, 701, 134413. https://doi.org/10.1016/j.scitotenv.2019.134413.

Tran V.Q., Prakash I., 2020. Prediction of soil loss due to erosion using support vector machine model. Vietnam Journal of Earth Sciences, 42(3), 247-254. https://doi.org/10.15625/0866-7187/42/3/15050.

Van Phong T., Ly H.-B., Trinh P.T., Prakash I., Btjvjoes P., 2020. Landslide susceptibility mapping using Forest by Penalizing Attributes (FPA) algorithm based machine learning approach. Vietnam Journal of Earth Sciences, 42(3), 237-246. https://doi.org/10.15625/0866-7187/42/3/15047.

Vogel K., Riggelsen C., Korup O., Scherbaum F., 2014. Bayesian network learning for natural hazard analyses. Natural Hazards and Earth System Science, 14(9), 2605-2626. https://doi.org/10.5194/nhess-14-2605-2014.

Wang Q., Li W., Wu Y., Pei Y., Xie P., 2016. Application of statistical index and index of entropy methods to landslide susceptibility assessment in Gongliu (Xinjiang, China). Environmental Earth Sciences, 75(7), 599. Doi: 10.1007/s12665-016-5400-4.

Wang Y., et al., 2019. Flood susceptibility mapping in Dingnan County (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. J Environ Manage, 247, 712-729. Doi: 10.1016/j.jenvman.2019.06.102.

Whan K., Sillmann J., Schaller N., Haarsma R., 2020. Future changes in atmospheric rivers and extreme precipitation in Norway. Climate Dynamics, 54(3), 2071-2084.

Wu Y., Ke Y., Chen Z., Liang S., Zhao H., Hong H.. 2020. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. CATENA, 187, 104396. https://doi.org/10.1016/j.catena.2019.104396.

Wu Y., Wang H., Zhang B., Du K.L., 2012. Using Radial Basis Function Networks for Function Approximation and Classification. ISRN Applied Mathematics, 2012, 1-34. Doi: 10.5402/2012/324194.

Yang W., Xu K., Lian J., Ma C., Bin L., 2018. Integrated flood vulnerability assessment approach based on TOPSIS and Shannon entropy methods. Ecological Indicators, 89, 269-280. https://doi.org/10.1016/j.ecolind.2018.02.015.

Yang Z., Mourshed M., Liu K., Xu X., Feng S., 2020. A novel competitive swarm optimized RBF neural network model for short-term solar power generation forecasting. Neurocomputing, 397, 415-421. https://doi.org/10.1016/j.neucom.2019.09.110.

Young-Sup H., Sung-Yang B., 1997. Recognition of unconstrained handwritten numerals by a radial basis function neural network classifier. Pattern Recognition Letters, 18(7), 657-664. https://doi.org/10.1016/S0167-8655(97)00056-1.

Zenggang X., Zhiwen T., Xiaowen C., Xue-min Z., Kaibin Z., Conghuan Y., 2019. Research on Image Retrieval Algorithm Based on Combination of Color and Shape Features. Journal of Signal Processing Systems. Doi: 10.1007/s11265-019-01508-y.

Zhou Q., Mikkelsen P.S., Halsnæs K., Arnbjerg-Nielsen K., 2012. Framework for economic pluvial flood risk assessment considering climate change effects and adaptation benefits. Journal of Hydrology, 414-415, 539-549. https://doi.org/10.1016/j.jhydrol.2011.11.031.

Zhu J., Shi Q., Wu P., Sheng Z., Wang X., 2018. Complexity Analysis of Prefabrication Contractors' Dynamic Price Competition in Mega Projects with Different Competition Strategies. Complexity, 2018, 5928235. Doi: 10.1155/2018/5928235.

Zhu Y., Wang Z., Cao C., Gao D., 2018. Regularized fisher linear discriminant through two threshold variation strategies for imbalanced problems. Knowledge-Based Systems, 150, 57-73. https://doi.org/10.1016/j.knosys.2018.02.035.