

Flood Susceptibility Mapping Using Machine Learning Algorithms: A Case Study in Huong Khe District, Ha Tinh Province, Vietnam

Nguyen, D. L.,^{1,3*} Chou, T. Y.,² Hoang, T. V.² and Chen, M. H.²

¹Ph.D. Program for Infrastructure Planning and Engineering, College of Construction and Development Feng Chia University, Taichung 40724, Taiwan, E-mail: st_loc@gis.tw*

²Geographic Information Systems Research Center, Feng Chia University, Taichung 40724, Taiwan, E-mail: jimmy@gis.tw, van@gis.tw, ivy@gis.tw

³Faculty of Natural Resources and Environment, Vietnam National University of Agriculture, Trau Quy Gia Lam, Hanoi, Vietnam

*Corresponding Author

DOI: <https://doi.org/10.52939/ijg.v19i7.2739>

Abstract

A flood is a natural catastrophe that causes heavy damage not only to people but also to properties. To prevent and mitigate flood damage, an accurate flood susceptibility map that reveals highly potential flood-prone areas is essential. This study aims to construct flood susceptibility maps in the Huong Khe district using three machine learning algorithms, namely the K - Nearest Neighbour (KNN), the Support Vector Machine (SVM) and Artificial Neural Network (ANN). Training and testing datasets were extracted from Sentinel-1 SAR images. Seven causative factors were selected as input for predictive models after removing high-correlation factors and unimportant factors through a rigorous screening process by analyzing the Pearson correlation coefficient (PCC) and calculating the information gain ratio (InGR). The model's hyperparameters were found by grid search algorithm integrated 5-fold cross-validation. The three optimal flood susceptibility models showed excellent performance, with very high accuracy indices in the training and testing phases, over 90% of overall accuracy and UAC values. High and very high susceptibility classes on flood susceptibility maps accounted for around 18% of the total study area and were mainly located in residential and agricultural areas. Thus, there is a need to make proper land use planning for these areas to reduce damage in flood seasons.

Keywords: ANN, Flood Susceptibility Mapping, Grid Search, Information Gain Ratio, KNN, Sentinel-1 SAR, SVM

1. Introduction

Asia-Pacific is one of the most disaster-prone areas in the world. According to the United Nations Economic and Social Commission for Asia and the Pacific, from 1970 to 2020, this region recorded 57% of global fatalities from disasters and 87% of the global population was influenced by natural hazards [1]. More detail, natural hazards influenced 6.9 billion people and killed over 2 million people. Located in the Western Pacific, Vietnam frequently suffered intense storms, especially in the Central provinces. These storms resulted in severe floods, destroyed houses and infrastructures, caused crop failure, and killed many residents. For example, in the year 2020, from October 6 to November 15, this region was struck by nine consecutive storms. It caused severe and widespread flooding in eight provinces, including Nghe An, Ha Tinh, Quang

Binh, Quang Tri, Thua Thien Hue, Da Nang, Quang Nam, and Quang Ngai. According to the Viet Nam Red Cross report, these provinces suffered heavy losses of people, shelters, and properties. 357 residents were killed, 876 residents were injured, 511,172 houses were submerged, over 360 schools were flooded, and 30,000 hectares of crops were devastated [2]. It was recorded as the worst disaster which hit Central Viet Nam in the past 100 years. To avoid severe flood damages, flood prevention and mitigation strategies must be pre-planned, whereby a flood susceptibility map is an essential document. This kind of map reveals flood-prone areas under specific conditions of topography, hydrology, land cover, rainfall, and artificial constructions.

There were various methods applied in flood susceptibility assessment (FSA), including hydrological and hydraulic models [3] [4] and [5], Analytic Hierarchy Process (AHP) [6] [7] and [8], the integration of Fuzzy Logic and Analytic Hierarchy Process (F-AHP) [9] [10] and [11], frequency ratio (FR) [12] [13] and [14], the weight of evidence (WoE) [12] [14] and [15], logistic regression (LR) [16] and [17]. Physically-based models, such as MIKE FLOOD and HEC-RAS, have proven their efficacies in developing flood susceptibility models. These models require complex data that is not easy to meet for large areas, such as river cross-section, meteorological and hydrological data in long duration [18] and [19]. To develop a flood susceptibility map using an expert-based approach (such as AHP or F-AHP), experts evaluate the contribution of various influencing factors to the flood susceptibility index. However, these subjective judgments lead to predictive errors [20]. Flood susceptibility models constructed by statistical approaches, including FR, WoE, and LR, were rated as comprehensible and reliable. However, they require high-quality data on historical floods and influencing factors [21]. Furthermore, they do not examine the correlation between influencing factors. So redundant data may exist in the influencing factor database.

Machine learning algorithms (MLAs) have brought high accuracy in natural disaster prediction fields. Compare to expert-based and statistical approaches, MLAs have produced better results [22] [23] and [24]. It is because the input data for MLAs

typically has the non-existence of highly correlated factors and low contribution factors by the feature selection process. Besides, the performance of machine learning models depends not only on input data and learning algorithms but also on hyperparameters. Previous studies on flood susceptibility assessment using MLAs used the "trial and error" method to find the best hyperparameters [25] and [26]. This method is unreasonable and needs to be improved by a better search algorithm.

In conclusion, the main objective of this study was to use MLAs for constructing FSMs in Huong Khe district, the most frequently flooding district of Ha Tinh province, Vietnam. To achieve this objective, sub-objectives were implemented: (1) building a historical flood database using Sentinel-1 SAR satellite images, (2) generating influencing factors database from original geospatial data, (3) eliminating highly correlated and low contribution influencing factors through the feature selection process, (4) tuning hyperparameters by Grid Search, (5) developing FSMs using selected MLAs, and (6) Assessing the accuracy and comparing the performance of the developed FSMs.

2. Methodology

The overall methodology of the study is represented in Figure 1. It includes developing a historical flood database, causative factor preparation, constructing optimal flood susceptibility models, and models' performance assessment, flood susceptibility mapping.

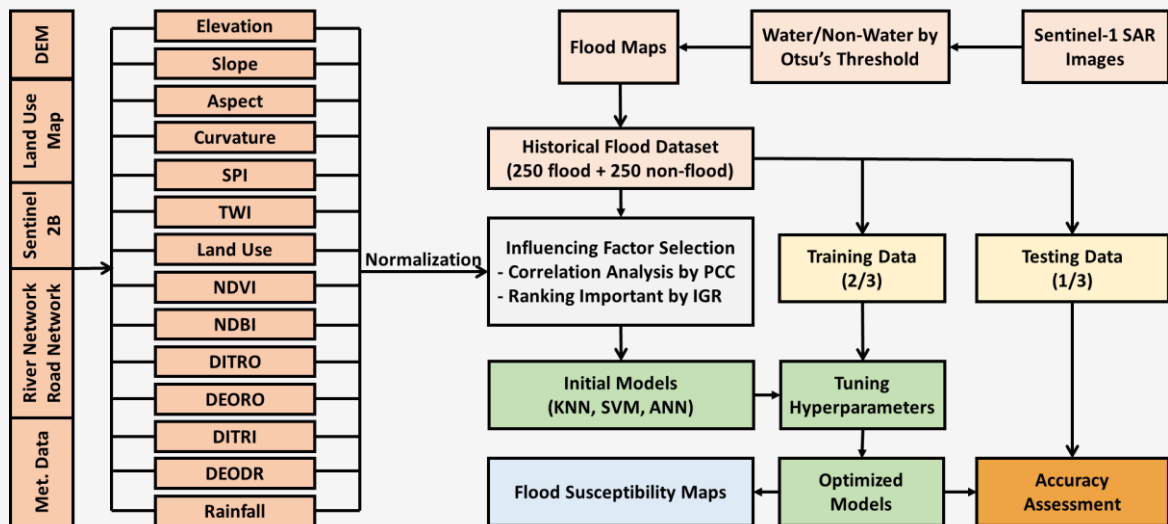


Figure 1: Methodology of the study

2.1 Study Area

Huong Khe is a border district of Ha Tinh province. It shares about 50km borderline with Lao People's Democratic Republic to the West and borders Tuyen Hoa district (Quang Binh province) to the South. The North and the East are bordered by Vu Quang, Thach Ha, and Can Loc districts of Ha Tinh province, respectively. This mountainous district ranges in height from 3m to 1440m. It is covered by forest land, agricultural land, specially used land, and residential land with a corresponding percentage of 78.58%, 14.22%, 2.65%, and 0.75%, respectively [27]. The three latter land-use types are distributed in low lands and are often seriously flooded.

Every year, this district receives high rainfall amounts. According to the statistical data from 2010 to 2020 at Huong Khe meteorological station, the average annual rainfall was 2560mm. Among months of the rainy season, September and October were the wettest months in terms of rainfall amount, with 641mm and 586mm. The corresponding figures in the three remaining months of the rainy season (including July, August, and November) were 268mm, 223mm, and 247mm, respectively. Because of the high rainfall amount, this district has frequently suffered natural disasters, such as floods and landslides. The location of the study area is represented in Figure 2.

2.2 Historical Flood Database

In Ha Tinh province, historical flood data has been not recorded as geospatial data, such as points or polygons. In practice, the flood situation is stored in the reports with simple descriptive information about where the flood occurred, how large it extended, and how deep it was. Therefore, collecting historical flood data from flood control departments for training and testing flood susceptibility models is not feasible. Researchers and scientists have used synthetic aperture radar (SAR) images to extract geo-information about past floods to solve this difficulty. In the period 2006-2011, ALOS PALSAR provided L-band SAR images that were useful for flood mapping [28] [29] and [30]. Recently, free Sentinel-1 SAR images have been effectively utilized in flood detecting and monitoring [31] [32] and [33]. Flood information extracted from SAR images has been used as training and testing data for flood susceptibility mapping models [34] [35] and [36]. Hence, SAR satellite images are a crucial data source for developing flood susceptibility models. To prepare training and testing data for flood susceptibility models, three Sentinel-1 SAR data images were used to extract historical flood data. These images included two images acquired in May and October 2020, and one image acquired in September 2019, and were preprocessed using the flowchart proposed by Filipponi [37]. Table 1 displays the detailed information of the Sentinel-1 images.

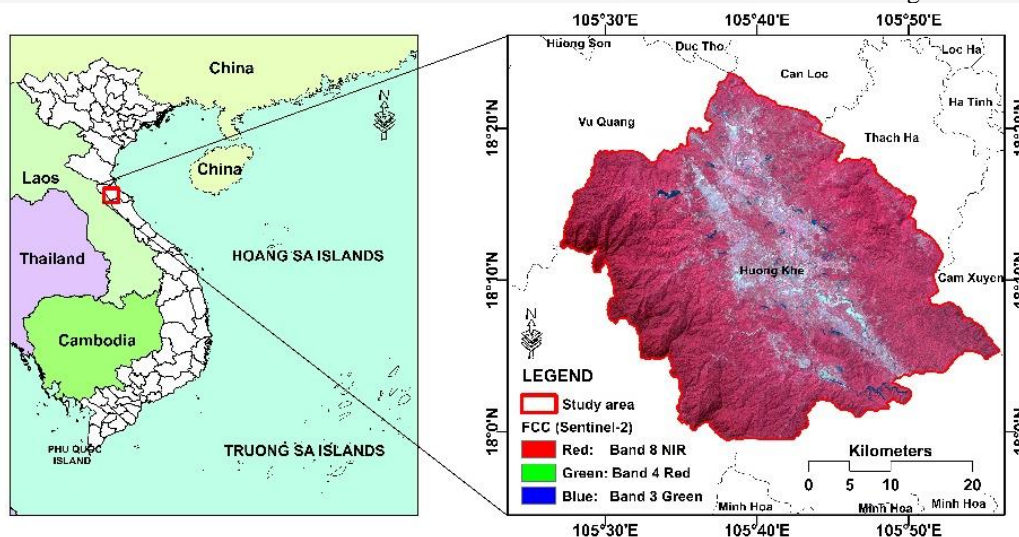


Figure 2: Location of study area

Table 1: Sentinel-1 images used to construct flood database

ID	Satellite	Date of Acquisition	Pass Direction	Condition
1	S1-A	06-Sep-2019	Ascending	Flood Event
2	S1-B	18-May-2020	Descending	Dry Season
3	S1-A	18-Oct-2020	Ascending	Flood Event

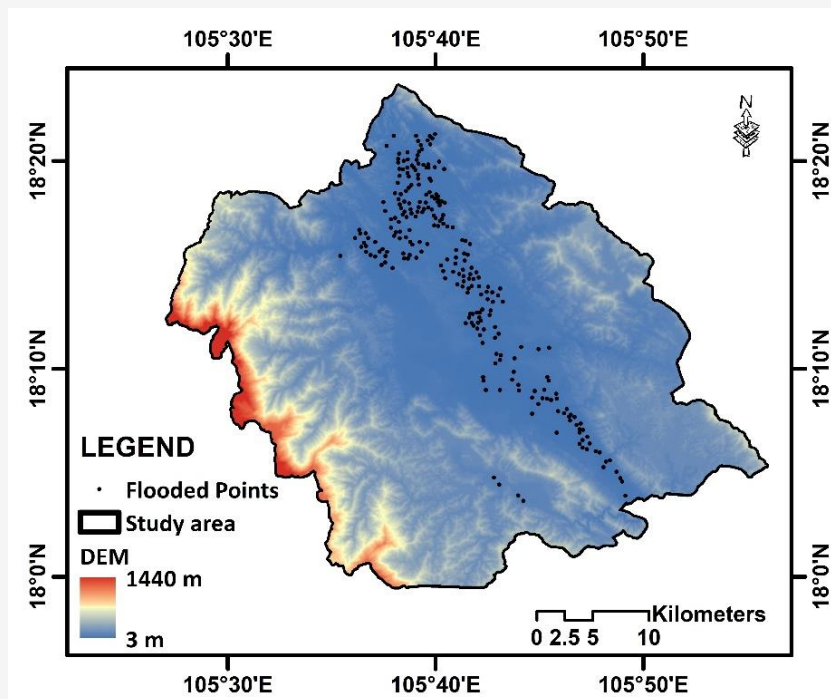


Figure 3: Historical flooded points

Among the three images, the image taken in May 2020, before the rainy and storm season, was used as the pre-flood event image, while the remaining images collected at flood times were used as flood event images. To detect and delineate floodwater, the image ratioing method was applied, resulting in ratio images by calculating the ratio of digital number value of corresponding pixels on pre-flood event images and flood event images. Floodwater was delineated by an Otsu's threshold method. 250 flood points were randomly created from inundated flood areas over the two years 2019 and 2020 and 250 non-flood points were also randomly generated from non-flood sites. Finally, the entire dataset of 500 points was divided into two subsets, namely training and testing, in a ratio of 2 to 1. The location of flooded points is represented in Figure 3.

2.3 Causative Factors and Data

Through a careful review of previous studies on flood susceptibility assessment and the own characteristics of the study site, fourteen causative factors were chosen for developing flood susceptibility models in the Huong Khe district. They consist Elevation (ELE), Slope (SLO), Aspect (ASP), Curvature (CUR), Stream Power Index (SPI), Topographic Wetness Index (TWI), Land Use (LU), Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), Rainfall (RAI), Distance to River (DITRI), Density of Drainage (DEODR), Distance to Road

(DITRO), Density of Road (DEORO). ELE, SLO, CUR, ASP, SPI, and TWI were derivatives of DEM with a spatial resolution of 10m. This DEM was created from topographic maps provided by the Center of Survey and Mapping Data (Department of Natural Resources and the Environment of Vietnam). NDVI and NDBI were calculated from Sentinel-2B satellite images freely provided by European Space Agency. LU map was generalized from land use maps of all communes of Huong Khe district. The average rainfall amount of eleven flood events from 2010-2020 of 10 rainfall stations was used to interpolate the rainfall map by the Kriging method. DITRI and DEODR were established based on river networks extracted from topographic maps. DITRO and DEORO were made from road networks collected from OpenStreetMap. All causative factors maps were converted to raster with a spatial resolution of 10m and normalized to the range [0,1] by the min-max method (Equation 1). ArcGIS 10.8 software was utilized to create and normalize all causative factor maps (Figure 4).

$$V_{Norm} = \frac{V_i - V_{min}}{V_{max} - V_{min}}$$

Equation 1

Where:

- V_{Norm} is normalized value
- V_{min} is minimum value
- V_{max} is maximum value
- V_i is input value

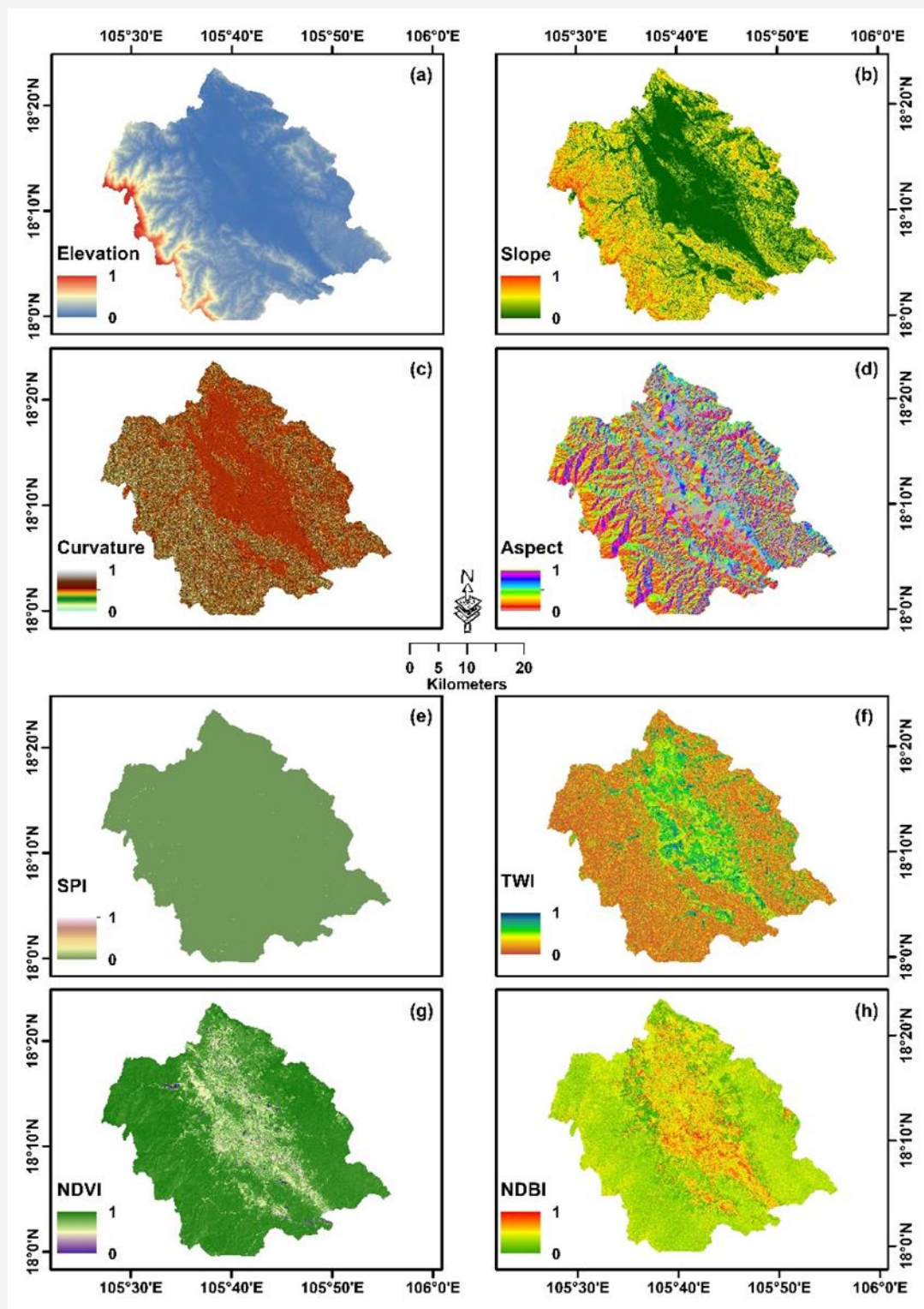


Figure 4: Causative factors: (a) Elevation; (b) Slope; (c) Curvature; (d) Aspect; (e) Stream power index; (f) Topographic wetness index; (g) Normalized difference vegetation index; (h) Normalized difference built-up index

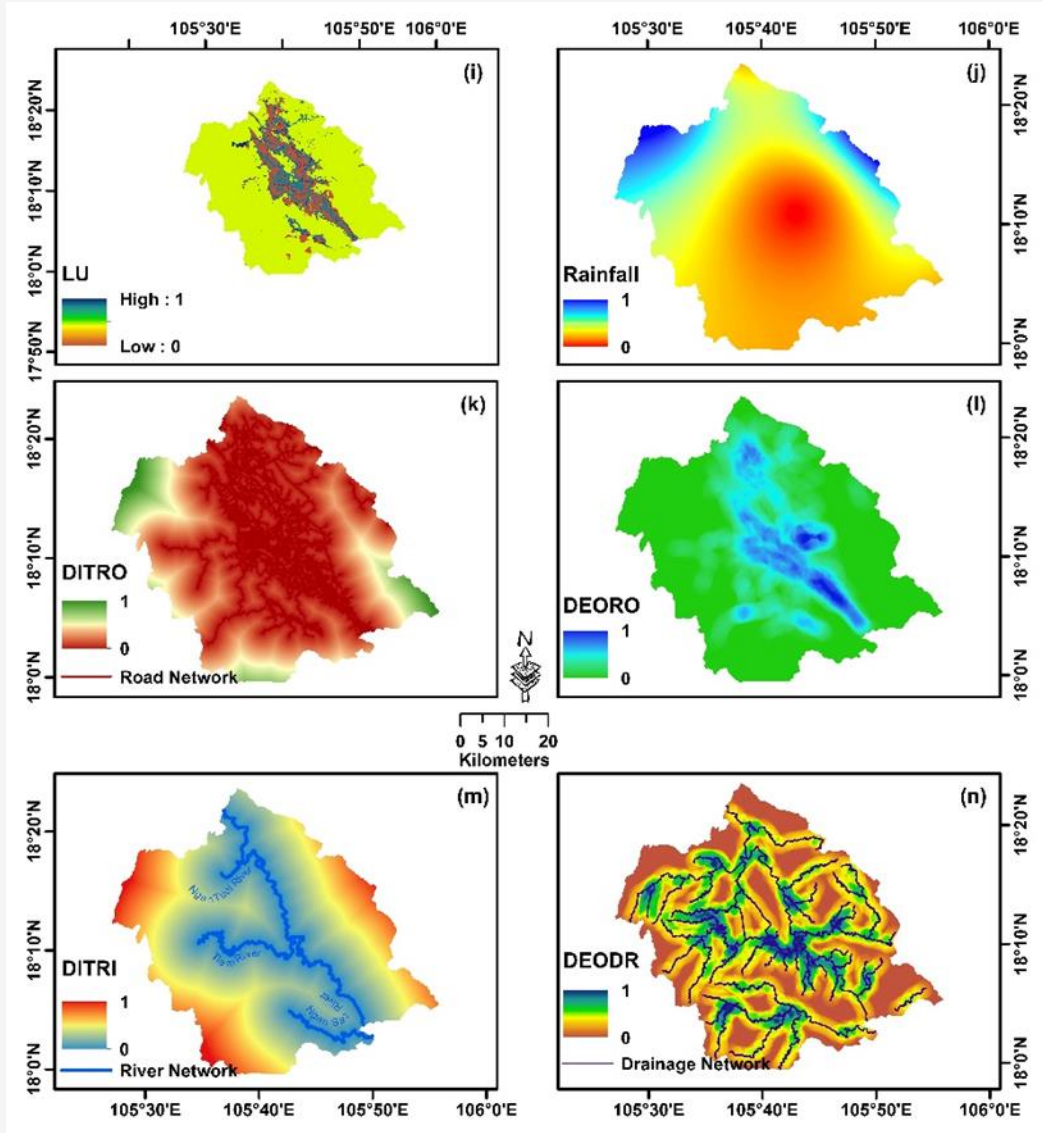


Figure 4: Causative factors: (i) Land Use; (j) Rainfall; (k) Distance to road; (l) Density of road; (m) Distance to river; (n) Density of river

2.4 Causative Factor Selection

2.4.1 Correlation analysis

Studies on disaster susceptibility assessment typically used Pearson's correlation coefficient (PCC) to analyze the correlation between causative factors [38] and [39]. Assuming that X and Y are two causative factors, the Pearson's correlation coefficient (r) is computed by following equation:

$$r = \frac{\sum_1^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{\sum_1^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_1^n (Y_i - \bar{Y})^2}} \quad \text{Equation 2}$$

Where:

X_i and Y_i are samples of causative factors
 \bar{X} and \bar{Y} are means of X and Y
 n is number of samples

The absolute value of r ranges from 0 to 1. Two factors that have an absolute value of r greater than 0.6 are considered to have a strong correlation [40]. Then one factor will be eliminated.

2.4.2 Information Gain Ratio

Information Gain Ratio (InGR) is utilized to measure the information contribution level of causative factors to prediction models. It has been widely applied in disaster forecasting models [41] and [42]. Assuming that training dataset T consists of n samples and m factors (F_i). Output class C consists of two classes: flood and non-flood. The GR of causative factor F_i is defined as following equation:

$$InGR(T, F) = \frac{Infor(T) - Infor(T, F)}{SplitInfor(T, F)} \quad \text{Equation 3}$$

Where:

$Infor(T)$ is information entropy for the dataset

$Infor(T, F)$ is amount of information ($T1, T2, \dots, Tm$) split from T regarding the causal factor F

$SplitInfor(T, F)$ is the potential information generated by dividing the training data T into m subsets.

2.5 Machine Learning Algorithms

2.5.1 Support Vector Machine

The Support Vector Machine (SVM) was firstly introduced in 1992 by Boser et al., [43]. It is a robust supervised learning algorithm that can be applied for both classification and regression. It has been effectively used for natural disaster prediction, including landslides [44] and [45] and floods [46] and [47]. The powerful algorithm aims to find an optimal hyperplane to correctly distinguish data points, with the hyperplane being considered optimal when it has the widest margin. For the SVM algorithm, kernel function, penalty (C), and gamma (γ) are hyperparameters that directly affect its performance. The penalty parameter (C) controls the trade-off between achieving a low training error and maintaining a wider margin. It determines the level of misclassification that the SVM classifier is willing to tolerate. The gamma (γ) is a parameter of the kernel function, it defines the influence of each training example and affects the shape of the decision boundary. In this study, the default kernel, Radial Basis Function (RBF), was used because it produced higher accuracy compared to other kernels.

2.5.2 K-Nearest Neighbour (KNN)

To be considered the most straightforward algorithm in the machine learning field [48], KNN calculates the distance from the candidate point to K neighbor points. K is an integer and should be an odd number [49]. The distance can be calculated by Euclidean distance, Manhattan distance, or Minkowski distance. And then, the candidate point will be assigned to the class for which the number of neighbors is maximum. KNN is highly recommended for real-time applications because it is speedy [50]. However, it does not work well with large datasets and high dimension data.

2.5.3 Artificial Neural Network (ANN)

Artificial neural network (ANN) has been a powerful model to predict disaster susceptibility. It can solve the complex relationships between causative factors and natural disasters [51]. ANN simulates the behavior of the human brain to process input information [52]. This research used the multi-layer perceptron (MLP) model, which has been proven to be effective for landslide susceptibility mapping and flood susceptibility mapping. MLP model contains three layers: one input layer, one or more hidden layers, and one output layer. The input layer is responsible for preparing the data for the model, and it consists of nodes for the input variables. The function of the hidden layer is to process the data, and the number of hidden layers needed depends on the complexity of the problem being solved. Finally, the output layer consists of nodes that represent the output results.

The ANN has several hyperparameters, including the activation function, solver, learning rate, and learning rate initialization. The activation function is a crucial component of the MLP as it introduces non-linearity into the network. It is applied to the output of each neuron, enabling the modeling of complex relationships between inputs and outputs. The solver determines the algorithm used to optimize the weights and biases of the MLP during training. It controls how the network adjusts its parameters to minimize the error between predicted and actual outputs. The learning rate governs the step size taken during each iteration of the optimization process, influencing the extent to which the weights and biases are adjusted based on the calculated gradient. The learning rate initialization parameter establishes the initial value of the learning rate. The three selected algorithms have showcased their impressive abilities in modeling landslide susceptibility mapping and flood susceptibility mapping in different regions over the world. In this study, we aim to evaluate their abilities in developing flood susceptibility models specifically for Huong Khe district, the mountainous region of central Vietnam.

2.6 Tuning Hyperparameters

The performance of machine learning algorithms mainly depends on data quality and hyperparameter configuration. The optimal hyperparameters are typically found through the hyperparameter tuning process. This process can be done by a manually or automatically. In the first method, different hyperparameters are experimented with through a "trial and error" approach [53] and [54]. In the second method, the optimal set of hyperparameters can be come out by algorithms, such as Grid Search

[55] and [56], Bayesian Optimization [57], Genetic Algorithm [58], or Whale Optimization [59]. Grid Search is the simplest among automatic searching methods but typically gives reliable results. It generates combinations of hyperparameters' values and calculates the model's performance corresponding to each variety. The results are tracked and the optimal combination is released at the end of the calculating process. This study used the Grid Search algorithm and 5-fold cross-validation to find the best hyperparameters to optimize the models' accuracy.

2.7 Accuracy Assessment

There are a lot of metrics to assess the performance of machine learning models. This study used some primary statistical metrics, containing Precision (Equation 4), Recall (Equation 5), and Overall Accuracy (Equation 6) are extracted from the confusion matrix. Furthermore, the receiver operating characteristics curve (ROC) and area under the ROC curve (AUC) were also applied to indicate the algorithms' performance.

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Equation 4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{Equation 5}$$

$$\text{Overall Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \quad \text{Equation 6}$$

Where:

TP is short for True Positive

TN is short for True Negative

FP is short for False Positive

FN is short for False Negative

3. Results and Discussion

3.1 Causative Factor Selection

Figure 5 describes the InGR values of the causative factors and shows that there is a significant difference in their importance. ELE and LU are the most critical factors, with the highest InGR values of 0.577 and 0.449, respectively. NDVI is also an essential factor with an InGR value of 0.409. DITRO and TWI are equally important, as indicated by their identical InGR values of 0.292 and 0.291, respectively. The InGR values of seven other factors (DEORO, DITRI, SLO, NDBI, CUR, RAI, and DEODR) decrease from 0.239 to 0.078. Notably, the InGR values of SPI and ASP are both equal to 0, indicating that these two factors do not contribute any valuable information to the prediction model. Consequently, these factors are removed from the flood susceptibility models. Figure 6 displays the PCC between the causative factors, revealing six pairs of factors that have a strong correlation, with absolute PCC values higher than 0.6. Among these pairs, the DITRI-DITRO and TWI-SLO pairs have the highest PCC values of 0.73 and -0.70, respectively. The remaining four pairs, DITRI-ELE, NDBI-NDVI, DITRO-ELE, and DEORO-ELE, have PCC values of 0.68, 0.65, 0.61, and -0.61, respectively. To create an optimal input dataset, the factors with an InGR value of 0 are eliminated. Additionally, when there are two factors with a PCC greater than 0.6, the factor with the lower InGR is removed. As a result, seven factors were removed, including ASP, SPI, NDBI, SLO, DITRI, DEORO, and DITRO. The seven remaining factors used for developing the FSMs are ELE, LU, NDVI, CUR, TWI, RAI, and DEODR.

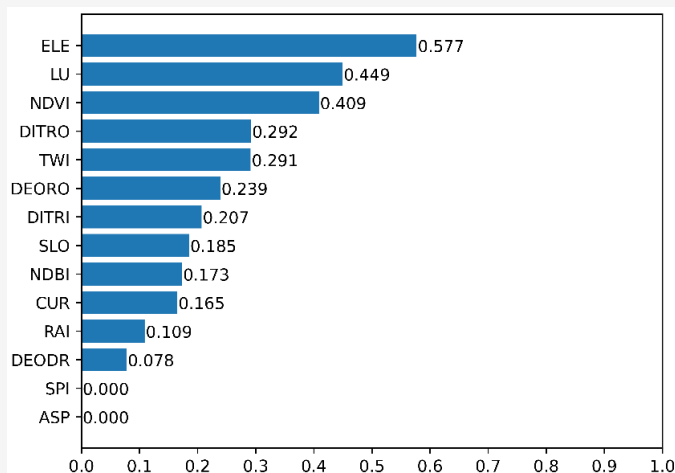


Figure 5: Information Gain Ratio

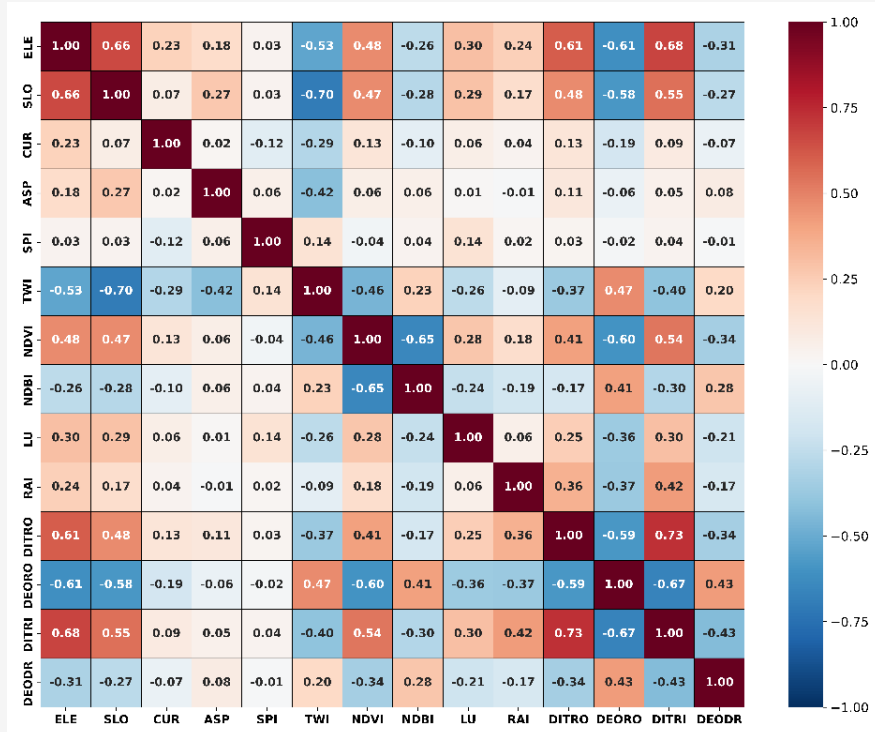


Figure 6: Pearson correlation matrix between 14 causative factors

3.2 Tuning Hyperparameters

Table 2 shows default values and tuned values of hyperparameters for the three chosen machine learning models. The results reveal that tuned hyperparameters slightly improved models' performance. The overall accuracy (OA) of KNN, SVM, and ANN increased from 0.91642, 0.91045, and 0.91940 to 0.92239, 0.91642, and 0.92239, respectively.

3.3 Assessing FSMs' Accuracy

Table 3 and Figure 7 reveal information on the accuracy of the developed models. From an overall perspective, all accuracy metrics are very high. In the training phase, the ANN model performed the best with overall accuracy and AUC values of 0.92239 and 0.92727, respectively, followed by the KNN model with corresponding figures of 0.92239 and 0.90428. The SVM model performed the worst with an overall accuracy of 0.91642 and an AUC value of 0.90873. On the other hand, in the testing phase, the SVM model was the best performer with overall accuracy and AUC values of 0.92727 and 0.94312, respectively. The KNN model, by contrast, performed the worst with overall accuracy and AUC values of 0.91515 and 0.91329, respectively. The corresponding figures for the ANN model were 0.92727 and 0.93416.

3.4 Flood Susceptibility Mapping

The flood susceptibility maps for the Huong Khe district were generated based on the three optimal flood susceptibility models (FSMs). Each pixel of these raster maps is assigned a digital number that represents the flood probability value, ranging from 0 to 1. The susceptible level was then classified into 5 categories: very low, low, moderate, high, and very high. The corresponding range of values for each category are 0-0.2, 0.2-0.4, 0.4-0.6, 0.6-0.8, and 0.8-1, respectively. These maps are shown in Figure 8. As shown in Figure 8, high and very high susceptibility areas are mainly distributed in low-lying agricultural and residential areas. On the other hand, low susceptibility areas are located in higher areas covered by forest. The total percentage of high and very high susceptibility areas calculated by the KNN, SVM, and ANN models are 18.9%, 17.3%, and 17.6%, respectively, while the corresponding figures for low and very low susceptibility areas are 80.1%, 80.7%, and 81.0%. The moderate susceptibility class accounts for a small percentage of the area, only 1.1%, 2.0%, and 4.5%, respectively. Thus, there is a significant difference in the proportion of area under various susceptibility levels.

Table 2: Default and optimal hyperparameters of algorithms

K - Nearest Neighbour			
ID	Hyperparameter	Default value	Optimal value
1	n_neighbors	5	7
2	weights	uniform	uniform
Model's accuracy		0.91642	0.92239
Support Vector Machine Algorithm			
ID	Hyperparameter	Default value	Optimal value
1	C	1	100
2	γ	scale	1
Model's accuracy		0.91045	0.91642
Artificial Neural Network			
ID	Hyperparameter	Default value	Optimal value
1	activation	relu	'logistic'
2	solver	adam	sgd
3	learning rate	constant	constant
4	Learning rate init	0.001	0.1
Model's accuracy		0.91940	0.92239

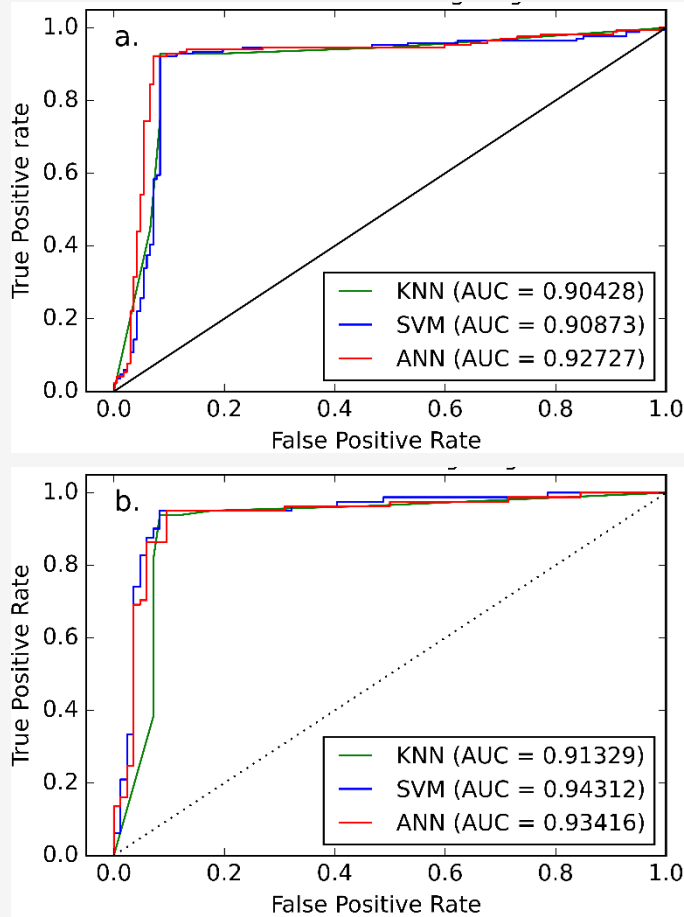
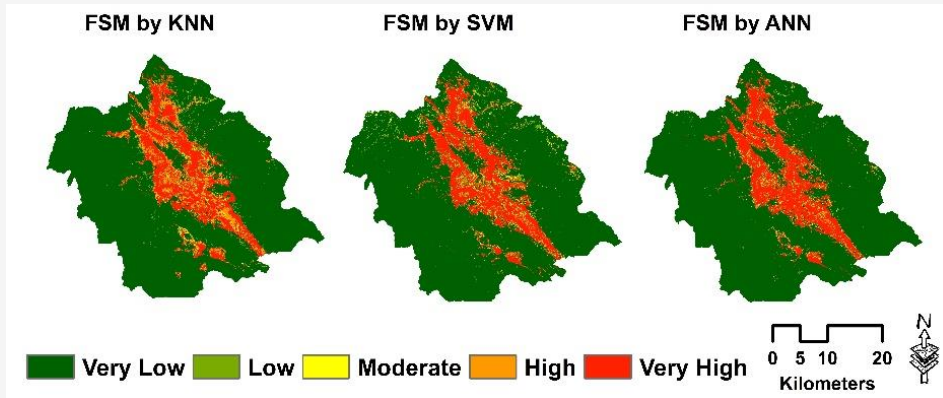
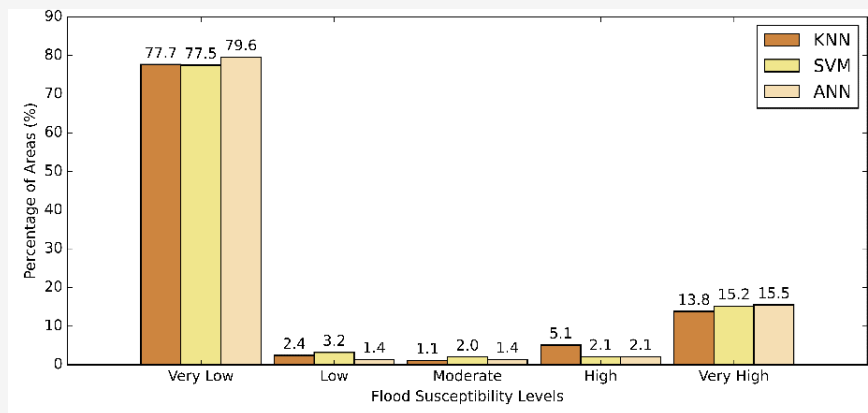
**Figure 7:** ROC curve of: (a) Training phase, (b) Testing phase

Table 3: Accuracy metrics of three models in training and testing phases

ID	Metric	Training			Testing		
		KNN	SVM	ANN	KNN	SVM	ANN
1	Accuracy	0.92239	0.91642	0.92239	0.91515	0.92727	0.92727
2	Precision	0.91828	0.91759	0.92564	0.89412	0.91566	0.90588
3	Recall	0.92906	0.91729	0.92299	0.93827	0.93827	0.95062
4	AUC	0.90428	0.90873	0.92727	0.91329	0.94312	0.93416

**Figure 8:** Flood susceptibility maps developed by three models**Figure 9:** Percentage of susceptibility classes by two models

3.5 Discussion

The InGR values of the causative factors indicate their importance in the flood susceptibility model. The study found that ELE and LU were the most critical factors, followed by NDVI. These findings suggest that topography, land use, and vegetation cover significantly influence flood susceptibility. On the other hand, factors such as SPI and ASP did not contribute valuable information and were therefore removed from the models. Strong correlations were observed between certain factors, such as DITRI-DITRO and TWI-SLO. These findings indicate that some factors provide redundant or overlapping information. To optimize the input dataset, factors with lower InGR values were eliminated when there were strong correlations between two factors. This process helped refine the selection of factors used in developing the flood susceptibility models. The hyper-parameter tuning process showed that tuned

hyper-parameters slightly improved the models' performance in terms of overall accuracy. This finding highlights the importance of optimizing model parameters to achieve better results. The accuracy assessment results indicated high overall accuracy for all three models. In the training phase, the ANN model performed the best, followed by the KNN model, while the SVM model had the lowest accuracy. However, in the testing phase, the SVM model outperformed the other models. These findings suggest that different models may exhibit varying performance in different phases, highlighting the importance of evaluating models on independent datasets. The flood susceptibility maps generated based on the three optimal FSMs provided valuable insights into the areas at risk of flooding.

The distribution of high and very high susceptibility areas was predominantly observed in low-lying agricultural and residential areas, while low susceptibility areas were mostly located in higher areas covered by forests. These findings demonstrate the ability of the models to effectively identify areas prone to flooding and provide useful information for flood risk management and mitigation.

4. Conclusion

This study aimed to develop flood susceptibility models using machine learning techniques and assess their accuracy. The results demonstrated that elevation, land use, and vegetation cover were significant factors influencing flood susceptibility. Through the selection and elimination of causative factors based on their importance and correlation, an optimal input dataset was constructed. Hyperparameter tuning was performed to enhance the models' performance, leading to slight improvements in overall accuracy. The accuracy assessment revealed that all three models achieved high overall accuracy, with variations observed between the training and testing phases. The SVM model performed the best in the testing phase, indicating its potential for accurate flood susceptibility prediction. The generated flood susceptibility maps provided valuable information for identifying areas at high risk of flooding. The maps highlighted the importance of factors such as topography and land use in determining flood-prone areas. These findings contribute to the understanding of flood vulnerability in the study area and can assist in implementing effective flood risk management strategies.

References

- [1] ESCAP, (2021). Asia-Pacific Disaster Report 2021: Resilience in a Riskier World: Managing Systemic Risks from biological and other Natural Hazards, Bangkok. 1-114. https://www.unescap.org/sites/default/d8files/knowledge-products/Asia-Pacific%20Disaster%20Report%202021_full%20version_0.pdf.
- [2] VNRC, (2021). *Final Report: Viet Nam: Floods*, VNRC, Vietnam.
- [3] Mai, D. T. and De Smedt, F., (2017). A Combined Hydrological and Hydraulic Model for Flood Prediction in Vietnam Applied to the Huong River Basin as a Test Case Study. *Water*, Vol. 9(11). <https://doi.org/10.3390/w9110879>.
- [4] Nogherotto, R., Fantini, A., Raffaele, F., Di Sante, F., Dottori, F., Coppola, E. and Giorgi, F., (2022). A Combined Hydrological and Hydraulic Modelling Approach for the Flood Hazard Mapping of the Po River Basin. *Journal of Flood Risk Management*, Vol. 15. <https://doi.org/10.1111/jfr3.12755>.
- [5] Zhang, K., Shalehy, M. H., Ezaz, G. T., Chakraborty, A., Mohib, K. M. and Liu, L., (2022). An Integrated Flood Risk Assessment Approach Based on Coupled Hydrological-Hydraulic Modeling and Bottom-Up Hazard Vulnerability Analysis. *Environmental Modelling & Software*, Vol. 148. <https://doi.org/10.1016/j.envsoft.2021.105279>.
- [6] Aydin, M. C. and Sevgi Birincioğlu, E., (2022). Flood Risk Analysis Using GIS-Based Analytical Hierarchy Process: A Case Study of Bitlis Province. *Applied Water Science*, Vol. 12. <https://doi.org/10.1007/s13201-022-01655-x>.
- [7] Seejata, K., Yodying, A., Wongthadam, T., Mahavik, N., and Tantanee, S., (2018). Assessment of Flood Hazard Areas using Analytical Hierarchy Process over the Lower Yom Basin, Sukhothai Province. *Procedia Engineering*, Vol. 212. 340-347. <https://doi.org/10.1016/j.proeng.2018.01.044>.
- [8] Swain, K. C., Singha, C. and Nayak, L., (2020). Flood Susceptibility Mapping through the GIS-AHP Technique Using the Cloud. *ISPRS International Journal of Geo-Information*, Vol. 9(12). <https://doi.org/10.3390/ijgi9120720>.
- [9] Ekmekcioğlu, Ö., Koc, K. and Özger, M., (2021). District Based Flood Risk Assessment in Istanbul Using Fuzzy Analytical Hierarchy Process. *Stochastic Environmental Research and Risk Assessment*, Vol. 35, 617-637. <https://doi.org/10.1007/s00477-020-01924-8>.
- [10] Parsian, S., Amani, M., Moghimi, A., Ghorbanian, A. and Mahdavi, S., (2021). Flood Hazard Mapping Using Fuzzy Logic, Analytical Hierarchy Process, and Multi-Source Geospatial Datasets. *Remote Sensing*, Vol. 13(23). <https://doi.org/10.3390/rs13234761>.
- [11] Zou, Q., Zhou, J., Zhou, C., Song, L. and Guo, J., (2013). Comprehensive Flood Risk Assessment Based on Set Pair Analysis-Variable Fuzzy Sets Model and Fuzzy AHP. *Stochastic Environmental Research and Risk Assessment*, Vol. 27, 525-546. <https://doi.org/10.1007/s00477-012-0598-5>.

- [12] Rahmati, O., Pourghasemi, H. R. and Zeinivand, H., (2016). Flood Susceptibility Mapping Using Frequency Ratio and Weights-of-Evidence Models in the Golastan Province, Iran. *Geocarto International*, Vol. 31, 42-70. <https://doi.org/10.1080/10106049.2015.1041559>.
- [13] Sarkar, D. and Mondal, P., (2019). Flood Vulnerability Mapping Using Frequency Ratio (FR) Model: A Case Study on Kulik River Basin, Indo-Bangladesh Barind Region. *Applied Water Science*, Vol. 10(17). <https://doi.org/10.1007/s13201-019-1102-x>.
- [14] Shafapour Tehrani, M., Shabani, F., Neamah Jebur, M., Hong, H., Chen, W. and Xie, X., (2017). GIS-Based Spatial Prediction of Flood Prone Areas Using Standalone Frequency Ratio, Logistic Regression, Weight of Evidence and their Ensemble Techniques. *Geomatics, Natural Hazards and Risk*, Vol. 8, 1538-1561. <https://doi.org/10.1080/19475705.2017.1362038>.
- [15] Khosravi, K., Nohani, E., Maroufinia, E. and 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*, Vol. 83, 947-987. <https://doi.org/10.1007/s11069-016-2357-2>.
- [16] Ali, S. A., Parvin, F., Pham, Q. B., Vojtek, M., Vojteková, J., Costache, R., Linh, N. T. T., Nguyen, H. Q., Ahmad, A. and Ghorbani, M. A., (2020). GIS-based Comparative Assessment of Flood Susceptibility Mapping Using Hybrid Multi-Criteria Decision-Making Approach, Naïve Bayes Tree, Bivariate Statistics and Logistic Regression: A Case of Topľa Basin, Slovakia. *Ecological Indicators*, Vol. 117. <https://doi.org/10.1016/j.ecolind.2020.106620>.
- [17] Al-Juaidi, A. E. M., Nassar, A. M. and Al-Juaidi, O. E. M., (2018). Evaluation of Flood Susceptibility Mapping Using Logistic Regression and GIS Conditioning Factors. *Arabian Journal of Geosciences*, Vol. 11. <https://doi.org/10.1007/s12517-018-4095-0>.
- [18] Ongdas, N., Akiyanova, F., Karakulov, Y., Muratbayeva, A. and Zinabdin, N., (2020). Application of HEC-RAS (2D) for Flood Hazard Maps Generation for Yesil (Ishim) River in Kazakhstan. *Water*, Vol. 12(10). <https://doi.org/10.3390/w12102672>.
- [19] Patro, S., Chatterjee, C., Mohanty, S., Singh, R. and Raghuvanshi, N. S., (2009). Flood Inundation Modeling Using MIKE FLOOD and Remote Sensing Data. *Journal of the Indian Society of Remote Sensing*, vol. 37, 107-118. <https://doi.org/10.1007/s12524-009-0002-1>.
- [20] El Jazouli, A., Barakat, A. and Khellouk, R., (2019). GIS-multicriteria Evaluation Using AHP for Landslide Susceptibility Mapping in Oum Er Rbia High Basin (Morocco). *Geoenvironmental Disasters*, Vol. 6. <https://doi.org/10.1186/s40677-019-0119-7>.
- [21] Wubalem, A., Tesfaw, G., Dawit, Z., Getahun, B., Mekuria, T. and Jothimani, M., (2021). Comparison of Statistical and Analytical Hierarchy Process Methods on Flood Susceptibility Mapping: In a Case Study of the Lake Tana Sub-Basin in Northwestern Ethiopia. Vol. 13, 1668-1688. <https://doi.org/10.5194/nhess-2020-332>.
- [22] Akinci, H. and Zeybek, M., (2021). Comparing Classical Statistic and Machine Learning Models in Landslide Susceptibility Mapping in Ardanuc (Artvin), Turkey. *Natural Hazards*, Vol. 108, 1515-1543. <https://doi.org/10.1007/s11069-021-04743-4>.
- [23] Liang, Z., Wang, C. M., Zhang, Z. M. and Khan, K. U. J., (2020). A Comparison of Statistical and Machine Learning Methods for Debris Flow Susceptibility Mapping. *Stochastic Environmental Research and Risk Assessment*, Vol. 34, 1887-1907. <https://doi.org/10.1007/s00477-020-01851-8>.
- [24] Vojtek, M., Vojteková, J., Costache, R., Pham, Q. B., Lee, S., Arshad, A., Sahoo, S., Linh, N. T. T. and Anh, D. T., (2021). Comparison of Multi-Criteria-Analytical Hierarchy Process and Machine Learning-Boosted Tree Models for Regional Flood Susceptibility Mapping: A Case Study from Slovakia. *Geomatics, Natural Hazards and Risk*, Vol. 12, 1153-1180. <https://doi.org/10.1080/19475705.2021.1912835>.
- [25] Pham, B. T., Jaafari, A., Prakash, I. and Bui, D. T., (2019). A Novel Hybrid Intelligent Model of Support Vector Machines and the MultiBoost Ensemble for Landslide Susceptibility Modeling. *Bulletin of Engineering Geology and the Environment*, Vol. 78, 2865-2886. <https://doi.org/10.1007/s10064-018-1281-y>.
- [26] Yaseen, A., Lu, J. and Chen, X., (2022). Flood Susceptibility Mapping in an Arid Region of Pakistan through Ensemble Machine Learning Model. *Stochastic Environmental Research and Risk Assessment*, Vol. 36, 3041-3061. <https://doi.org/10.1007/s00477-022-02179-1>.

- [27] HTSO, (2021). *Ha Tinh Statistical Yearbook 2020*, Ha Tinh.
- [28] Arnesen, A. S., Silva, T. S. F., Hess, L. L., Novo, E. M. L. M., Rudorff, C. M., Chapman, B. D. and McDonald, K. C., (2013). Monitoring Flood Extent in the Lower Amazon River Floodplain Using ALOS/PALSAR ScanSAR Images. *Remote Sensing of Environment*, Vol. 130, 51-61. <https://doi.org/10.1016/j.rse.2012.10.035>.
- [29] Dimitrios, D. A., Diofantos, G. H., Athos, A., Kyriacos, T., Adrianos, R., Silas, M., Stelios, P. and Tymvios, F., (2012). *Flood Mapping of Yialias River Catchment Area in Cyprus Using ALOS PALSAR Radar Images*, Proceedings of SPIE - The International Society for Optical Engineering, Vol. 8531, <https://doi.org/10.1117/12.974581>.
- [30] Yulianto, F., Sofan, P., Zubaidah, A., Sukowati, K. A. D., Pasaribu, J. M. and Khomarudin, M. R., (2015). Detecting Areas Affected by Flood Using Multi-Temporal ALOS PALSAR Remotely Sensed Data in Karawang, West Java, Indonesia. *Natural Hazards*, Vol. 77, 959-985. <https://doi.org/10.1007/s11069-015-1633-x>.
- [31] David, C. M., Sarah, L. D. and Hannah, L. C., (2021). Floodwater Detection in Urban Areas Using Sentinel-1 and WorldDEM Data. *Journal of Applied Remote Sensing*, Vol. 15(3). <https://doi.org/10.1117/1.JRS.15.032003>.
- [32] Moharrami, M., Javanbakht, M. and Attarchi, S., (2021). Automatic Flood Detection Using Sentinel-1 Images on the Google Earth Engine. *Environmental Monitoring and Assessment*, Vol. 193. <https://doi.org/10.1007/s10661-021-09037-7>.
- [33] Sipelgas, L., Aavaste, A. and Uiboupin, R., (2021). Mapping Flood Extent and Frequency from Sentinel-1 Imagery during the Extremely Warm Winter of 2020 in Boreal Floodplains and Forests. *Remote Sensing*, Vol. 13(23). <https://doi.org/10.3390/rs13234949>.
- [34] Mohammadi, A., Kamran, K. V., Karimzadeh, S., Shahabi, H. and Al-Ansari, N., (2020). Flood Detection and Susceptibility Mapping Using Sentinel-1 Time Series, Alternating Decision Trees, and Bag-ADTree Models. *Complexity*, Vol. 2020. <https://doi.org/10.1155/2020/4271376>.
- [35] Ngo, P. T. T., Hoang, N. D., Pradhan, B., Nguyen, Q. K., Tran, X. T., Nguyen, Q. M., Nguyen, V. N., Samui, P. and Tien Bui, D., (2018). A Novel Hybrid Swarm Optimized Multilayer Neural Network for Spatial Prediction of Flash Floods in Tropical Areas Using Sentinel-1 SAR Imagery and Geospatial Data. *Sensors*, Vol. 18(11). <https://doi.org/10.3390/s18113704>.
- [36] Samarasinghea, S. M., Nandalalb, H. K., Weliwitiyac, D. P., Fowzed, J. S. M., Hazarikad, M. K. and Samarakoond, L., (2010). *Application of Remote Sensing and GIS for Flood Risk Analysis: A Case Study at Kalu-Ganga River, Sri Lanka*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, 110-115.
- [37] Filipponi, F., (2019). Sentinel-1 GRD Preprocessing Workflow. The 3rd International Electronic Conference on Remote Sensing (ECRS 2019), Sciforum Electronic Conference Series, Vol. 3, 1-5. <https://sciforum.net/manuscripts/6201/manuscript.pdf>.
- [38] Tehrany, M. S., Jones, S. and Shabani, F., (2019). Identifying the Essential Flood Conditioning Factors for Flood Prone Area Mapping Using Machine Learning Techniques. *CATENA*, Vol. 175, 174-192. <https://doi.org/10.1016/j.catena.2018.12.011>.
- [39] Kalantar, B., Ueda, N., Saeidi, V., Ahmadi, K., Halin, A. A. and Shabani, F., (2020). Landslide Susceptibility Mapping: Machine and Ensemble Learning Based on Remote Sensing Big Data. *Remote Sensing*, Vol. 12(11). <https://doi.org/10.3390/rs12111737>.
- [40] Wang, C. N., Le, T. M., Nguyen, H. K. and Ngoc-Nguyen, H., (2019). Using the Optimization Algorithm to Evaluate the Energetic Industry: A Case Study in Thailand. *Processes*, Vol. 7(2). <https://doi.org/10.3390/pr7020087>.
- [41] Yu, L., Cao, Y., Zhou, C., Wang, Y. and Huo, Z., (2019). Landslide Susceptibility Mapping Combining Information Gain Ratio and Support Vector Machines: A Case Study from Wushan Segment in the Three Gorges Reservoir Area, China. *Applied Sciences*, Vol. 9(22). <https://doi.org/10.3390/app9224756>.
- [42] Rane, P. R. and Vincent, S., (2022). Landslide Susceptibility Mapping Using Machine Learning Algorithms for Nainital, India. *Engineered Science*, Vol. 17, 142-155. <https://doi.org/10.30919/es8d600>.
- [43] Boser, B. E., Guyon, I. M. and Vapnik, V. N., (1992). *A Training Algorithm for Optimal Margin Classifiers*, 144-152. <https://doi.org/10.1145/130385.130401>.
- [44] Kamran, K. V., Feizizadeh, B., Khorrami, B. and Ebadi, Y., (2021). A Comparative Approach of Support Vector Machine Kernel

- Functions for GIS-Based Landslide Susceptibility Mapping. *Applied Geomatics*, Vol. 13, 837-851. <https://doi.org/10.1007/s12518-021-00393-0>.
- [45] Saha, S., Saha, A., Hembram, T. K., Kundu, B. and Sarkar, R., (2022). Novel Ensemble of Deep Learning Neural Network and Support Vector Machine for Landslide Susceptibility Mapping in Tehri Region, Garhwal Himalaya. *Geocarto International*, 1-26. <https://doi.org/10.1080/10106049.2022.2120638>.
- [46] Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F. and Mosavi, A., (2019). An Ensemble Prediction of Flood Susceptibility Using Multivariate Discriminant Analysis, Classification and Regression Trees, and Support Vector Machines. *Science of the Total Environment*, Vol. 651, 2087-2096. <https://doi.org/10.1016/j.scitotenv.2018.10.064>.
- [47] Sahana, M., Rehman, S., Sajjad, H. and Hong, H., (2020). Exploring Effectiveness of Frequency Ratio and Support Vector Machine Models in Storm Surge Flood Susceptibility Assessment: A Study of Sundarban Biosphere Reserve, India. *CATENA*, Vol. 189. <https://doi.org/10.1016/j.catena.2019.104450>.
- [48] Nouh, R. M., Lee, H. H., Lee, W. J. and Lee, J. D., (2019). A Smart Recommender Based on Hybrid Learning Methods for Personal Well-Being Services. *Sensors*, Vol. 19(2). <https://doi.org/10.3390/s19020431>.
- [49] Ernest, Y. B., Joseph, O. and Daniel, A. A., (2020). Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review. *Journal of Data Analysis and Information Processing*, Vol. 8(4), 341-357. <https://doi.org/10.4236/jdaip.2020.84020>.
- [50] Rattanasak, A., Uthansakul, P., Uthansakul, M., Jumphoo, T., Phapatanaburi, K., Sindhupakorn, B. and Rooppakhun, S., (2022). Real-Time Gait Phase Detection Using Wearable Sensors for Transtibial Prosthesis Based on a kNN Algorithm. *Sensors*, Vol. 22(11). <https://doi.org/10.3390/s22114242>.
- [51] Sarkar, T. and Mishra, M., (2018). Soil Erosion Susceptibility Mapping with the Application of Logistic Regression and Artificial Neural Network. *Journal of Geovisualization and Spatial Analysis*, Vol. 2(8). <https://doi.org/10.1007/s41651-0180015-9>.
- [52] Kim, D. H., Kim, Y. J. and Hur, D. S., (2014). Artificial Neural Network-Based Breakwater Damage Estimation Considering Tidal Level Variation. *Ocean Engineering*, Vol. 87, 185-190. <https://doi.org/10.1016/j.oceaneng.2014.06.001>.
- [53] Sajadi, P., Sang, Y. F., Gholamnia, M., Bonafoni, S. and Mukherjee, S., (2022). Evaluation of the Landslide Susceptibility and its Spatial Difference in the whole Qinghai-Tibetan Plateau Region by Five Learning Algorithms. *Geoscience Letters*, Vol. 9(9). <https://doi.org/10.1186/s40562-022-00218-x>.
- [54] Ullah, K., Wang, Y., Fang, Z., Wang, L. and Rahman, M., (2022). Multi-Hazard Susceptibility Mapping Based on Convolutional Neural Networks. *Geoscience Frontiers*, Vol. 13 (5). <https://doi.org/10.16/j.gsf.2022.101425>.
- [55] Zhao, P., Masoumi, Z., Kalantari, M., Aflaki, M. and Mansourian, A., (2022). A GIS-Based Landslide Susceptibility Mapping and Variable Importance Analysis Using Artificial Intelligent Training-Based Methods. *Remote Sensing*, Vol. 14(1). <https://doi.org/10.3390/rs14010211>.
- [56] Zhou, S., and Fang, L., (2015). Support Vector Machine Modeling of Earthquake-Induced Landslides Susceptibility in Central Part of Sichuan Province, China. *Geoenvironmental Disasters*, Vol. 2(2). <https://doi.org/10.1186/s40677-014-0006-1>.
- [57] Sameen, M. I., Pradhan, B. and Lee, S., (2020). Application of Convolutional Neural Networks Featuring Bayesian Optimization for Landslide Susceptibility Assessment. *CATENA*, Vol. 186. <https://doi.org/10.1016/j.catena.2019.104249>.
- [58] Shahabi, H., Shirzadi, A., Ronoud, S., Asadi, S., Pham, B. T., Mansouripour, F., Geertsema, M., Clague, J. J., and Bui, D. T., (2021). Flash Flood Susceptibility Mapping Using a Novel Deep Learning Model Based on Deep Belief Network, Back Propagation and Genetic Algorithm. *Geoscience Frontiers*, Vol. 12(3). <https://doi.org/10.1016/j.gsf.2020.10.007>.
- [59] Tien Bui, D., Abdullahi, M. A. M., Ghareh, S., Moayedi, H. and Nguyen, H., (2021). Fine-Tuning of Neural Computing Using Whale Optimization Algorithm for Predicting Compressive Strength of Concrete. *Engineering with Computers*, Vol. 37, 701-712. <https://doi.org/10.1007/s00366-0190085-0-w>.