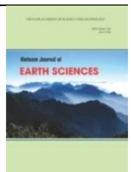
Vietnam Journal of Earth Sciences, 48(3), 356−378, https://doi.org/10.15625/2615-9783/24463

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse



Flood susceptibility prediction and adaptive capacity of community-based machine learning and socioeconomic data: Case study in Da Nang city, Vietnam

Huu Duy Nguyen[1], Thi Sen Tran[1], Quang-Hai Truong[2*]

[1] *Faculty of Geography, University of Science, Vietnam National University, Ha Noi, Hanoi, Vietnam*
[2] *Institute of Vietnamese Studies & Development Sciences, Vietnam National University (VNU), Hanoi 10000, Vietnam*

Received 25 February 2026; Received in revised form 31 March 2026; Accepted 08 April 2026

ABSTRACT

This study aims to predict flood susceptibility and community adaptation capacity based on machine learning and socioeconomic data. Da Nang City was selected as the case study in this study. Five machine learning algorithms, Random Forest (RF), Adaboost (ADB), Bagging (BA), Gradient Boosting (GB), and XGBoost (XGB) were used to predict flood susceptibility, and 80 households were selected to assess the community's adaptation capacity. The findings indicate that the RF model performed better than the other models, with an AUC score of 0.989, followed by ADB (0.987), BA (0.985), XGB (0.984), and GB (0.983). The eastern regions are affected by very high and high flooding, including Hai Chau, Thanh Khe, Ngu Hanh Son districts, and part of Son Tra. These regions have low elevations and high construction density. However, western mountainous areas, such as the Hoa Vang district and part of the Lien Chieu district, are in very low- and low-flood areas. The adaptive capacity of communities in Da Nang City is shaped by natural, physical, human, social, and financial resources.

Keyword: Flood susceptibility, adaptation capacity, Da Nang, Vietnam.

1. Introduction

Floods are among the most dangerous natural disasters, causing significant damage to people and the socioeconomic development of many countries worldwide (Khosravi, Shahabi et al. 2019, Nachappa, Piralilou et al., 2020). According to the World Disaster Report (CRED, 2023), in the period 2000–2022, floods accounted for 43% of all climate-related disasters, affecting about 2.3 billion people and causing damage of approximately 650 billion dollars (Hendrawan, Rahardjo et al., 2025). In recent years, driven by climate change and urban development, floods have become more intense and frequent. Specifically, the average number of floods worldwide per year has increased from 120–150 in the period 1980–2000 to 200–250 in the period 2000–2022. Asia is one of the regions most affected by floods worldwide, accounting for more than 60% of global flood events (Mirza, 2011). During the period 2010–2022, floods caused almost 300 billion dollars in economic losses in this region, killed nearly 45,000 people, and affected more than 1.2 billion

*Corresponding author, Email: truongquanghai.ivides@gmail.com

people (Nguyen, Nguyen et al., 2024). Vietnam is among the countries most affected by floods, due to its long coastline, dense river network, and frequent exposure to storms. Each year, floods cause significant damage to human and socioeconomic conditions; specifically, the 2020 floods caused more than 240 deaths and economic losses exceeding \$1.3 billion. Therefore, predicting flood susceptibility and evaluating community adaptive capacity are considered effective solutions to support local authorities and planners in providing appropriate solutions, including structural and nonstructural measures to minimize the impact of floods.

From the literature review, it is evident that many methods have been deployed to assess flood susceptibility, depending on the conditions, objectives, and characteristics of the research. These methods are divided into four main groups: physical models, remote sensing - GIS, statistical methods, and machine learning. Methods based on physical models, typically MIKE FLOOD, HEC-RAS (Namara, Damisse et al., 2022), or SWMM (Babaei, Ghazavi et al., 2018), use the Saint-Venant and Navier-Stokes equations to simulate hydrological and hydraulic processes (Pappenberger, Dutra et al., 2012). However, this method can simulate the flow in detail and can integrate different scenarios into the simulation process. However, physically based models require very detailed input data, such as topographic, meteorological, and hydrological data, and require a long time to calibrate. This poses significant challenges for application in areas with limited data. To overcome this limitation, many recent studies have applied remote sensing methods combined with GIS in flood assessment and simulation. This method uses satellite images (Sentinel, Landsat) (Perivolioti, Zachopoulos et al., 2024), radar (SAR) (Amitrano, Di Martino et al., 2024), or unmanned aerial vehicles (UAVs) (Dimitrov, Borisova et al., 2024, Panjavarnam, S.S. et al., 2024) to identify flooded areas in space and time. Although this method can be applied over a large area, especially in difficult-to-access areas, its accuracy can be affected by spatial and temporal resolution. Furthermore, this method is limited in its ability to simulate flood intensity.

In recent years, many studies have applied statistical methods such as linear regression, time series analysis (ARIMA) (Gegenleithner, Pirker et al., 2025), or correlation analysis to exploit the relationship between natural conditions, hydrometeorology, and flood areas. Although this method is easy to implement, it consumes little computational resources and has been proven effective. However, the method has limitations: it does not describe the physical nature of floods, accuracy is significantly reduced under climate change conditions, and it is not flexible to unusual factors such as rapid changes in terrain or land use. More recently, with the development of artificial intelligence, many studies have used machine learning methods for flood assessment and forecasting. These models include Random Forest (Tan, Li et al., 2024), Support Vector Machine (SVM) (Youssef, Pradhan et al., 2022), Xgboost (Ren, Pang et al., 2024), Adaboost (Jahanbani, Vahidnia et al., 2024), artificial neural networks (ANN) (Elsafi, 2014), which are capable of learning and simulating linear and nonlinear multidimensional relationships between input factors such as terrain, rainfall, water level, humidity, and land use,... This method allows processing large volumes of data, provides fast and accurate forecasts, and can be integrated with real-time data (Amiri, Soltani et al., 2024, Wahba, Sharaan et al., 2024). However, machine learning models are often considered “black box” models, difficult to explain their operation mechanisms, require large training data, and can suffer from overfitting if not properly calibrated (Gao, Liao et al., 2024).

An overview of the studies shows that each method has certain advantages and limitations. In the context of increasing demands for accuracy and practical application, the trend toward developing integrated models that combine the advantages of multiple methods is becoming the main approach in flood risk research and management. From this perspective, many studies have focused on the role of community resilience as a key component in managing community flood risk. The adaptive capacity of the community to floods is defined as the ability to adjust behavior, organize, and mobilize resources to minimize the negative impacts of floods, while taking advantage of opportunities to recover and rebuild after disasters (Haase, 2013). In the context of climate change, which has increased the frequency and intensity of extreme weather events, adaptive capacity becomes a key factor in protecting the livelihoods, lives, and assets of communities, especially in vulnerable areas. Adaptive capacity depends on natural, physical, human, social, and economic capital (Nguyen, Nguyen et al., 2023). According to Smit and Wandel (2006), adaptive capacity is not simply short-term responses to disasters but also includes long-term strategies to enhance resilience in the context of increasingly complex climate change. The assessment of adaptive capacity not only provides a comprehensive view of the community's readiness to respond to floods but also helps clarify internal strengths and weaknesses, thereby enabling appropriate recommendations to improve resilience. Furthermore, assessing the community's adaptive capacity plays a fundamental role in policy-making and the development of appropriate flood-prevention plans. Finally, this assessment helps to improve the effectiveness of interventions by identifying vulnerable population groups and designing appropriate solutions for each group.

This study aims to predict flood susceptibility and evaluate community adaptive capacity using machine learning models and socioeconomic data in Da Nang city, where floods occur frequently. This study proposes an integrated approach between machine learning models and multisource data (GIS, remote sensing, sociological surveys) to analyze and simulate flood space in a coastal urban area. The novelty of this study lies in integrating flood-susceptibility forecasting and community adaptation assessment, thereby providing a more comprehensive view of flood risks across physical and socioeconomic dimensions. This study provides valuable support for local authorities and planners in developing appropriate strategies to mitigate flood impacts, particularly in the context of climate change and sea-level rise.

2. Study and material

2.1. Study Area

The study area is located between 15°15’ and 16°40’N and 107°17’ and 108°20’E. This area has diverse terrain, clearly differentiated between coastal, plain, and mountainous areas. Mountainous terrain accounts for a large proportion, with a common altitude of 700–1500 m and a large slope (> 400 m), concentrated mainly in the West and Northwest regions. The coastal area is predominantly a low-lying plain formed by alluvial deposits from rivers such as the Han and Cu De, with relatively flat terrain and elevations typically ranging from 0 to 20 meters above sea level. Da Nang City is situated within a tropical monsoon climate zone, characterized by two distinct seasons: a dry season from January to August and a rainy season from September to December. Total rainfall in the study area ranges from 2150.4 mm to 3947.1 mm, with approximately 80% concentrated in the rainy season (Fig. 1).

The hydrological system of Da Nang City mainly consists of short rivers with steep slopes and strong seasonal flow fluctuations. The main river systems include the Han River, the Cu De River, the Cau Do - Tuy Loan River, and the Vu Gia - Thu Bon River, of which the Han and Cu De Rivers play an important role in urban flood drainage.

Da Nang City currently has a total agricultural land area of approximately 3,146 ha and a forest land area of approximately 61,956 ha. However, in recent years, rapid urbanization and industrialization have led to significant reductions in agricultural and forest land, especially in suburban areas. Conversion of land use purposes not only alters the natural flow but also increases the frequency and intensity of floods in urban areas. In recent years, the study area has been affected by many severe floods, including the historic flood in October 2020, which killed 43 people, flooded more than 100,000 houses, caused an estimated economic loss of 3,000 billion VND, and caused serious landslides in mountainous districts. Floods in 2022 caused damage to 1,200 houses, with water depths of 0.5–1.5 m; 50 houses were severely damaged, requiring the urgent evacuation of 12,000 people; 1,500 hectares of rice and crops were flooded, and 200 hectares of aquaculture were damaged.

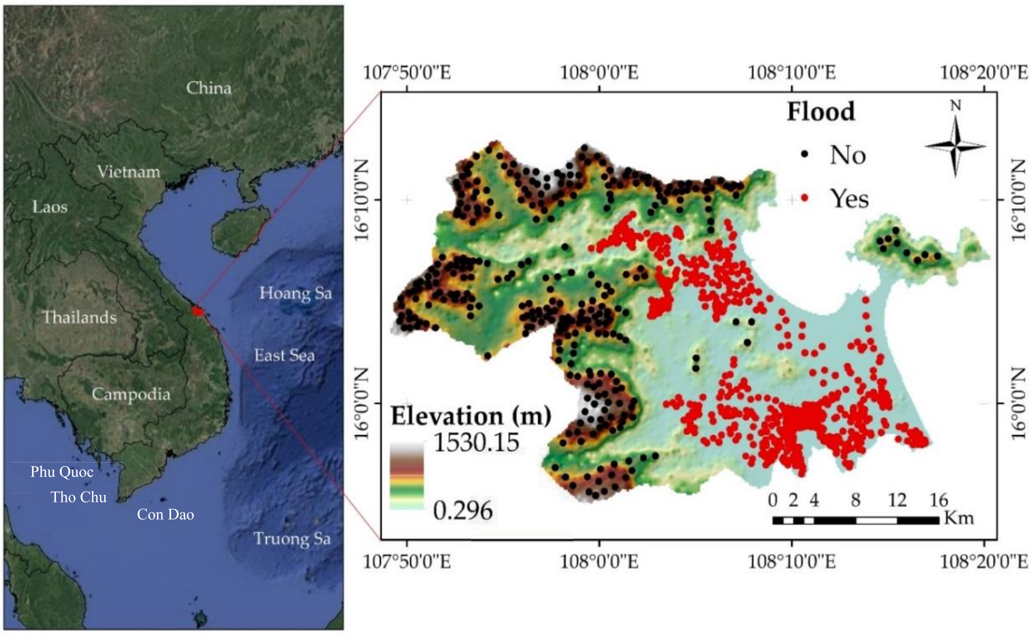


Figure 1. Location of the study area in this study

2.1.1. Flood Inventory

Flood inventory maps play a crucial role in machine-learning-based flood-susceptibility modeling by highlighting locations where flooding has occurred and the factors that cause it. In this study, flood inventory maps were extracted from different sources: the first is the flood marks collected during the field mission in 2025. To improve the flood inventory, we used Sentinel 1A images to detect flood zones in 2020 and 2022. This process involves the following steps:

First, the Sentinel 1A image is orbitally corrected by applying the correct orbital files to improve spatial accuracy. Then, after cropping to the boundary, the Sentinel image is radiometrically corrected to convert the raw data into backscatter coefficients. When processing the Sentinel 1A image, this study used the Lee filter to reduce speckle noise and improve image quality.

The Sentinel 1A image is then geometrically corrected to minimize distortions caused by terrain and reference frames, which is essential to ensure the accuracy of the spatial analysis. This study selected the VH channel for analysis because it performs better at detecting flooded areas than other channels.

The third step is the classification of aqueous and nonaqueous layers; this is performed using an automatic thresholding method based on the Otsu algorithm. This algorithm uses image histogram analysis to determine thresholds separating the aqueous and nonaqueous layers. Areas with lower backscatter coefficients are classified as aquatic, whereas those with higher backscatter coefficients are considered nonaqueous. Finally, flooded areas are calculated by subtracting the pre-flood area from the post-flood area. Finally, 650 flood points were collected to build the machine learning model.

This study uses the binary model, so identifying non-flood areas is essential. No-flood areas that were defined as areas never affected by floods. Normally, areas with steep slopes and high elevations have a very low probability of flooding. Therefore, based on the elevation and slope map, we selected 460 flood-free points. These points were combined with 650 flood points to build the flood susceptibility model.

2.1.2. Conditioning factor

Selecting conditioning factors is a critical step in applying machine learning to assess flood susceptibility in any region. This is because the machine learning model learns the relationship between past flood locations and their causes to predict areas with a probability of future flooding (Al-Juaidi 2023). In this study, the conditioning factors were classified into four main groups: environmental factors, hydrology, climate and human activities, namely elevation, curvature, slope, aspect, rainfall, distance from the river, distance to the road, geology, soil, Land Use/Land Cover (LULC), Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), Normalized Difference Water Index (NDWI), and Normalized Difference Moisture Index (NDMI) (Fig. 1A).

Elevation, curvature, slope, and aspect were derived from a 10 m resolution digital elevation model (DEM). Distances to rivers and roads were calculated from a 1:50,000 topographic map using the Euclidean distance method. Geological and soil data were obtained from the Ministry of Agriculture and Environment. LULC is available at https://www.landcovermapping.org/en/landcover/#. Annual rainfall is available at https://chrsdata.eng.uci.edu/. NDVI, NDBI, NDWI, and NDMI were calculated using Sentinel 2A images at https://browser.dataspace.copernicus.eu/.

Elevation plays an important role in identifying flood-prone areas because flooding tends to occur in low-lying areas (Vojtek and Vojteková 2019). In the study area, the elevation value ranges from 0 to 1530. The western regions tend to be higher, while the eastern regions are lower. Therefore, eastern regions are often affected by increasingly severe flooding. The slope is an essential factor in predicting flood susceptibility because it directly influences the speed of water flow and surface water accumulation (Liuzzo, Sammartano et al., 2019; Vojtek and Vojteková, 2019). Areas with low slopes promote water stagnation and increase the likelihood of flooding. In the study area, the slope varies from 0 to 55 degrees.

The curvature was selected to assess flood susceptibility because it reflects the land's shape and influences water flow behavior (Al-Juaidi, 2023). While the aspect determines the slope orientation, it directly influences solar radiation exposure, evaporation, and runoff dynamics (Vojtek and Vojteková, 2019; Kaya and Derin, 2023).

Rainfall is considered an essential factor in assessing the probability of flooding because it acts as a trigger (Mahmoud and Gan, 2018). In fact, prolonged heavy rain can overload the infiltration capacity, leading to the accumulation of surface water. In fact, continuous rainfall increases runoff and reduces the water's ability to infiltrate the soil. All of this causes flooding.

The distance from the river plays an important role in the probability of flooding in any region. Areas near rivers are more prone to flooding. This is because when the river level rises and overflows, the water flows quickly, flooding nearby areas. Therefore, the distance to the river is proportional to the probability of flooding (Vojtek and Vojteková, 2019).

Distance to roads directly influences flood probability by affecting both infiltration capacity and flow velocity. Furthermore, areas near roads are often more affected by flooding because concentrated rainfall flows quickly along roads (Versini, Gaume et al., 2010).

Geology plays an important role in the likelihood of flooding because it directly affects soil permeability and water storage. A bedrock that is poorly permeable or impermeable limits the infiltration of rainwater into the lower layers. All of this increases the likelihood of flooding. On the contrary, areas with loose or highly fractured bedrock allow rainwater to infiltrate more quickly, reducing the amount of water remaining on the surface. However, soil type influences an area's ability to absorb, retain, and drain rainwater, thereby directly affecting surface runoff (Mahmoud and Gan, 2018).

LULC directly influences the probability of flooding in any region by affecting soil permeability and flow velocity. For example, concreted areas during urban development reduce soil permeability and increase surface runoff, thus increasing the risk of flooding. On the contrary, vegetated areas have good water permeability, reducing surface runoff and, consequently, flood risk. In the study area, rapid urbanization over recent years has increased the concrete surface area, thereby increasing surface runoff and exacerbating flooding (Rashidiyan and Rahimzadegan, 2024).

NDVI represents the density of vegetation in an area. Areas with high vegetation density (high NDVI) often have a low probability of flooding because vegetation helps reduce surface runoff. On the contrary, in areas with low vegetation density (low NDVI), rainwater is poorly absorbed, leading to increased surface runoff and flooding (Khosravi, Shahabi et al., 2019).

The NDBI is an index used to assess built-up density. Areas with high built-up density are often characterized by the concentration of structures, such as houses, roads, and concrete surfaces, which reduces the soil's water-absorption capacity and increases surface runoff. This exacerbates the flood (Hoang and Liou 2024).

The NDWI index indicates the presence of bodies of water. High NDWI values often reflect the presence of bodies of water or wetlands such as rivers, lakes, and swamps. These areas often have a high risk of flooding due to their low relief and drainage capacities. The NDMI index represents soil moisture, an important index for flood analysis. Since this index is closely related to soil saturation, areas with high NDMI values often no longer have the capacity to absorb additional rainfall. In the event of heavy rainfall, water cannot penetrate the soil but runs off the surface, increasing runoff and increasing the risk of flooding (Sharma, Mishra et al., 2014).

2.2. Methodology

The methodology adopted in this study to assess flood susceptibility and community adaptive capacity consists of four main steps: (i) the first is data collection; (ii) construction of the machine learning model; (iii) construction of the flood susceptibility map; and iv) evaluation of community adaptive capacity (Fig. 2).

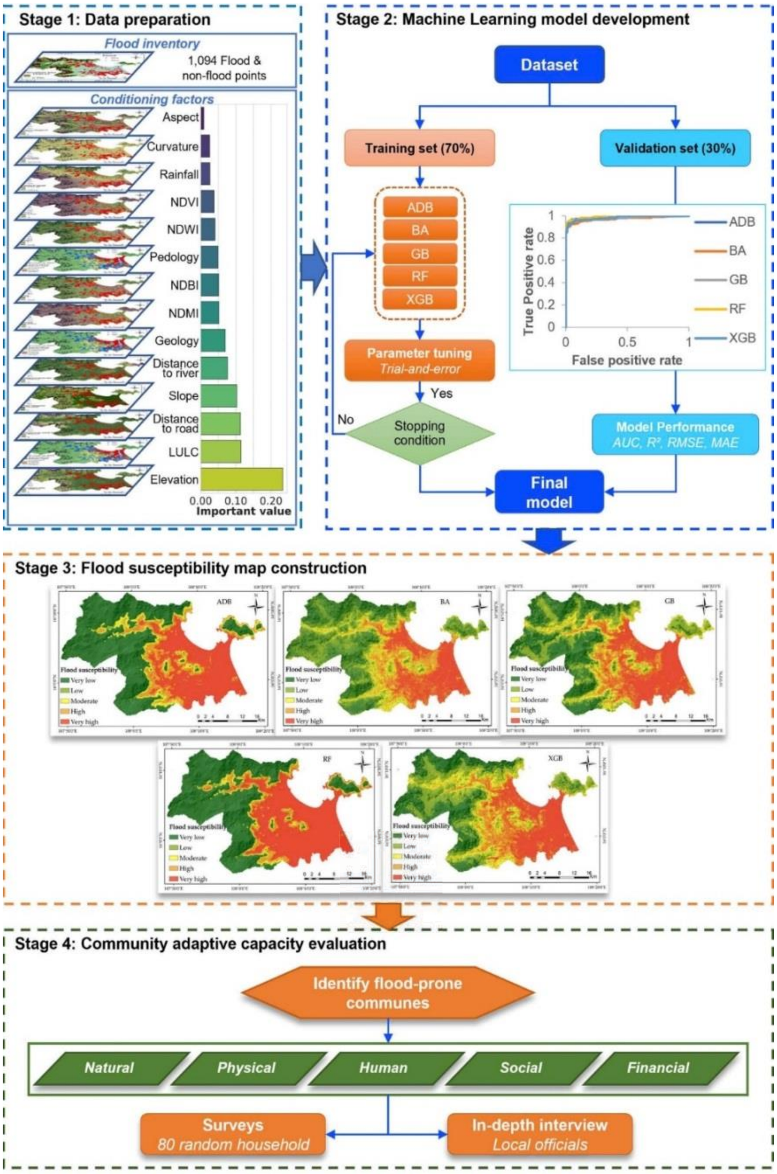


Figure 2. Flow chart used for flood susceptibility prediction

2.2.1. Random forest

Random forests are a popular machine learning algorithm proposed by Breiman in 2001. This algorithm is based on decision tree assembly and can be used to solve regression and classification problems (Rigatti, 2017). A decision tree has a top-down structure with nodes and branches. A forest is made up of several independent decision trees. Each tree presents a different perspective on a given problem. The Random Forest algorithm uses Tree Bagging, a process that randomly selects observations with replacement to construct the training subsets for each tree. Feature sampling, meanwhile, is a process of randomly selecting variables that allows a tree to be constructed from each distinct subset. Ultimately, all the decision tree ideas were assembled, and the Random Forest prediction was made by voting on all the trees (Chen, Li et al., 2020). The basic idea behind this algorithm is quite intuitive. Instead of using a complex estimator that can do everything, Random Forests uses a set of simple estimators, each offering a different perspective on a problem. These estimators are then combined to form a more powerful prediction method (Lee, Kim et al., 2017).

The accuracy of a random forest model depends on three key hyperparameters. These parameters must be established before training: the maximum number of nodes, the number of trees, and the number of sampled features (Vafakhah, Mohammad Hasani Loor et al., 2020). (i) The number of trees determines the maximum depth of each subtree in the forest, or the maximum number of leaf nodes. A high value for this parameter means that each tree is learned in detail in the training dataset. Conversely, a small value simplifies the model but can lead to underfitting if the tree is not deep enough. (ii) The number of trees is the number of subtrees in the random forest. A large number of trees makes the model more stable and the predictions more accurate. On the other hand, a low number of trees can lead to a model that lacks randomness and is less accurate. (iii) This parameter represents the number of input variables randomly selected at each split. A lower value increases model randomness, thereby enhancing tree diversity.

2.2.2. Adaboost

Freund and Schapire proposed the AdaBoost model. This algorithm can solve both classification and regression problems. Adaboost works by combining several weak models to form a model with better predictive capabilities. The idea is first to establish a model based on the training data set and then build another model to correct the errors of the first. This process is repeated until the errors are minimized (Li, Wang et al., 2008). The Adaboost model works in the following steps (Jahanbani, Vahidnia et al., 2024): (i) Initially, Adaboost randomly generates subsets; (ii) the Adaboost model repeatedly and iteratively trains the subsets based on the correct predictions from the previous training; (iii) the incorrectly predicted observations are given higher weights to minimize the level of error in subsequent training sessions. (iv) This process is repeated until there are no more errors or a specified maximum number of estimates is reached; (v) all submodels are combined by voting. The accuracy of the Adaboost model depends on parameters such as n_estimators, the learning rate, the base estimator, and the loss function. In this case, the n_estimator is the number of weak classifiers trained in the Adaboost model. If the value of this parameter is low, the model will not learn enough, leading to underfitting. However, a value that is too high can lead to overfitting. The learning rate is a coefficient that adjusts the contribution of weak models. This parameter controls each tree's influence on the final result. Its value is usually between 0 and 1, meaning that when each tree has little influence on the training model, more trees are needed to achieve the desired accuracy and make the model more stable. The base estimator is the basic model, which affects its complexity. This means that overly complex trees can complicate the model and lead to overfitting. On the other hand, simple trees will increase the model's generalization ability.

2.2.3. Bagging

Bagging is considered a more popular machine learning algorithm. This model is an ensemble model designed to improve the accuracy of individual prediction models (Chen, Li et al., 2020). The bagging model involves generating multiple subsets from the training data by random sampling with replacement. These subsets are then used to train multiple base models. In the prediction process, the results of the base models are combined by majority voting to create the final model (Chapi, Singh et al., 2017). Bagging works in the following steps: (i) Generating multiple data samples: From the original dataset, bagging generates multiple data samples by random sampling with replacement (bootstrap sampling). Each set of samples should be the same size as the original dataset, but data points may be duplicated or removed. (ii) Training the model: Each sample is used to train a separate model. The models can be decision trees, logistic regression, or any other type of model. iii) Prediction aggregation: When a prediction is needed, each model makes its own prediction. The final result is determined by vote (Chen, Li et al., 2020; Zhang, Fu et al., 2022; Pal, Saha et al., 2024). The accuracy of the bagging model depends on the parameters n_estimators, Base_estimator, max_samples, and max_features. Here, the n_estimator represents the number of submodels created and trained in parallel on bootstrap samples. Normally, a high value of this parameter helps stabilise the model and improve prediction accuracy. The Base_estimator is the base model used repeatedly during bagging. Shallow trees help increase the model's generalisability, while deep trees make it more vulnerable to overfitting. Max_sample represents the number of bootstrap samples used to create submodels from the training model. A low value of this parameter helps increase model diversity, but too low a value makes models less stable. Max_features is chosen randomly for each submodel.

2.2.4. Gradient boost

Gradient Boosting is a powerful machine learning algorithm that combines multiple weak learners into a single model to improve predictive accuracy. This algorithm is based on the principle that each submodel learns the errors of previous models (Lai, Nguyen et al., 2023). With each iteration, new models improve their error and prediction capabilities. By combining multiple decision tree regressors, gradient boosting can effectively capture complex feature relationships (Di, Zhang et al. 2019). The Gradient Boosting model works in the following three steps: (i) Base model initialization: The Gradient Boosting model starts by initializing the first weak training models. This model makes initial predictions, which serve as the basis for error correction in subsequent steps (Lee, Kim et al., 2017; Lai, Nguyen et al., 2023, Ait Naceur, Abdo et al., 2024); (ii) error calculation and construction of new training models: At each iteration, the error value is calculated on the predicted value and the actual value. This error value is then used to train a new model, focusing on correcting model errors from previous iterations. This process allows the model to improve after each iteration. (iii) Update of the ensemble model: After training the base models, they will be aggregated into an ensemble model using majority voting.

The accuracy of the gradient-boosting model depends on the adjustment of the following parameters: n_estimators, learning_rate, max_depth, min_samples_split, and subsample. N_estimators represents the number of weak models. A high number minimizes errors, but can easily lead to overfitting. A too-low number can lead to underfitting. The learning rate indicates the influence of each model on the overall model.

A low value indicates a longer training time, but the model is more general. Max_depth limits the number of splits in each decision tree. Deep trees can learn complex problems, but are prone to overfitting. Min_samples_split represents the minimum number of sample splits. The sub-sample is the fraction of the data randomly selected to train the submodel.

2.2.5. Xgboost

XGBoost is one of the most powerful and popular machine learning algorithms today, capable of solving classification and regression problems. This algorithm belongs to the family of gradient-boosting frameworks, ensemble methods that sequentially build weak models, with new models correcting the errors of previous models (Abedi, Costache et al., 2022). Specifically, XGBoost uses decision trees as the base model and builds them via gradient boosting (Yuan, Wang et al., 2024). The accuracy of the XGBoost model depends on parameter tuning, such as the learning rate, which is one of the most important hyperparameters of the gradient boosting algorithm. Its value is usually between 0 and 1 (Kavzoglu and Teke, 2022). The learning rate determines the influence of each new tree added to the model at each iteration. A low value allows the model to learn more slowly and more consistently by reducing each tree's contribution, thereby reducing the risk of overfitting and improving the model's generalization ability (Stokanović, Đukić et al., 2024). The number of trees (n_estimators) determines the total number of decision trees (submodels) built during training. Each iteration adds a new tree to learn from the errors of previous trees. Increasing n_estimators allows the model to learn more complex patterns in the data, but also increases the risk of overfitting if not accompanied by tuning mechanisms such as decreasing the learning rate (Hancock and Khoshgoftaar, 2020). The gamma parameter controls the minimum reduction in loss required to split a leaf node. Low gamma values make the model more divisible, increasing its complexity and susceptibility to overfitting. On the contrary, high gamma values require very useful splits, which simplify the model but may omit some important structures in the data (Lam, Zhang et al., 2019). The maximum tree depth (max_depth) determines the maximum number of levels that each tree in the model can reach. The deeper the tree, the more complex relationships the model can learn from the data. However, increasing max_depth increases the complexity of the model, which can lead to overfitting, especially with noisy data or small sample sizes (Sagi and Rokach, 2021).

3. Results

3.1. Feature Importance

Figure 3 presents the relative importance of each factor proposed for the construction of the flood susceptibility map in Da Nang City. In this study, the Random Forest algorithm was used to evaluate the contribution of each factor. The model is based on the relationships (often nonlinear) between the values of the sampling points and the conditioning factors to determine their importance. Random Forest assigns a weight to each factor, and the higher this value, the more determining the factor is considered in the flood susceptibility modeling. Elevation was the most important factor in flooding in Da Nang City. Floods tend to occur in lower elevation areas. In Da Nang City, the eastern region was lower, so this region was often affected by flooding. Then, LULC has a significant influence on flooding because, in recent years, urban growth in Da Nang City has increased rapidly, leading to greater infiltration areas and surface runoff. All of this increases the probability of flooding in the study area. The variable “distance to the road” ranks third in importance in Da Nang City. Indeed, areas near roads are generally flatter and concentrate a large number of infrastructures and buildings. These developments alter the natural flow of water and reduce the soil's infiltration capacity. This phenomenon tends to worsen as urbanization expands rapidly. The slope follows from the influence of flat or gently sloping land, which is more likely to accumulate rainwater.

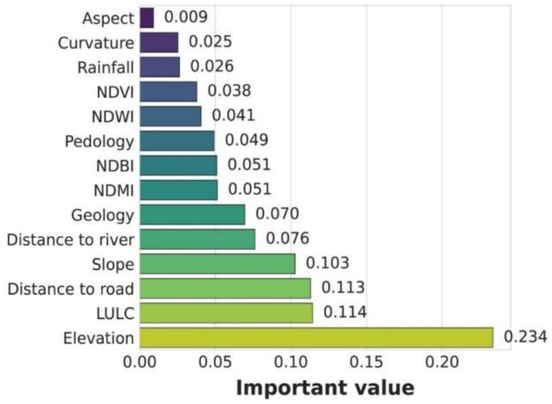


Figure 3. Importance of Flood Conditioning Factors Using the RF Method

In Da Nang City, the eastern region has a flatter slope; therefore, it is more affected by flooding. Similarly, regions near rivers, such as the lower Han River, or areas affected by the Vu Gia-Thu Bon river system, are more prone to flooding. Although geology, pedology, NDBI, and NDMI have a moderate influence on the probability of flooding in Da Nang City. It directly influences the infiltration capacity. Although previous studies have shown that NDVI and precipitation are two important factors that affect the probability of flooding in many areas, for Da Nang City, NDVI has a lower level of influence because the role of vegetation in reducing runoff in Da Nang is relatively limited compared to other factors (especially considering LULC and moisture/water indices).

Furthermore, precipitation in Da Nang City is relatively uniform during the rainy season, so this factor does not differ much from one area to another. Therefore, their importance is not as great as other factors, such as topography or LULC. Finally, exposure and curvature have little influence on the probability of flooding in Da Nang; this assessment is also consistent with previous studies.

3.2. Accuracy Assessment

Table 1 presents the AUC value of five proposed models. The findings demonstrate that the RF model achieved superior performance compared to the other models, with an AUC of 0.996. The ADB model was second with an AUC score of 0.995. The BA model was third with an AUC score of 0.993. The XGB model ranked fourth (AUC = 0.991), whereas the GB model ranked fifth (AUC = 0.990) during training. For the validation process, the RF model continues to outperform the other models in terms of performance in predicting flood susceptibility in the study area with the AUC score of 0.989, followed by the ADB model 0.987, BA with the AUC score of 0.985, XGB with the AUC score of 0.984 and GB with the AUC score of 0.983, respectively. Overall, all proposed models performed well in predicting flood susceptibility in the study area.

Table 1 provides the values of other statistical metrics used to assess the performance of the proposed models. The findings demonstrate that the RF model outperformed the other models, with lower RMSE and MAE values (RMSE = 0.16 and MAE = 0.04), followed by ADB (RMSE = 0.18 and MAE = 0.08), BA (RMSE = 0.21 and MAE = 0.12), XGB (RMSE = 0.32 and RMSE = 0.31) and GB (RMSE = 0.34 and MAE = 0.33), respectively, for the training data set. In terms of the validation dataset, the RF model showed a more accurate flood susceptibility prediction ability than other models, with an RMSE score of 0.2 and an MAE score of 0.06, followed by ADB (RMSE = 0.23 and MAE = 0.08), BA (RMSE = 0.28 and MAE = 0.1), XGB (RMSE = 0.34 and MAE = 0.33) and GB (RMSE = 0.35 and MAE = 0.38), respectively.

In addition, we also used the R^2 index to ensure the performance of the proposed models. In general, all the proposed models exhibited higher R^2 values. Among them, the value of the RF model was higher than other models in both the training data set and the validation data set (R2 = 0.95 for the training data set and 0.89 for the validation data set), followed by ADB (R^2=0.94 for the traning data set and 0.87 for the validation data set), BA (R^2=0.92 for the traning data set and 0.85 for the validation data set), XGB (R^2=0.91 for the traning data set and 0.82 for the validation data set) and GB (R^2=0.88 for the traning data set and 0.81 for the validation data set).

Table 1. The performance of the model proposed during the training data set and the validation of the data set

Model	**Training dataset**				**Validating dataset**			
	RMSE	**MAE**	**AUC**	**R^2**	**RMSE**	**MAE**	**AUC**	**R^2**
ADB	0.18	0.08	0.995	0.94	0.23	0.08	0.987	0.87
BA	0.21	0.12	0.993	0.92	0.28	0.1	0.985	0.85
GB	0.34	0.33	0.99	0.88	0.35	0.38	0.983	0.81
RF	0.16	0.04	0.996	0.95	0.2	0.06	0.989	0.89
XGB	0.32	0.31	0.991	0.91	0.34	0.33	0.984	0.82

3.3. Flood susceptibility maps

Figure 5 illustrates the flood susceptibility map of Da Nang City derived from the RF model, whereas the maps produced by the other models are shown in Fig. 2A. It can be seen that the eastern regions are affected by very high and high flooding, including the Hai Chau district, Thanh Khe, Ngu Hanh Son, and part of Son Tra. These regions have low elevation and high construction density. On the other hand, mountainous areas in the west, such as Hoa Vang district and part of Lien Chieu district, are in the very low- and low-flooding affected region. More specifically, for the ADB model, 47% of the study area is in the very low flood susceptibility zone, 2% of the study area is in the low flood susceptibility zone, 6% of the study area is in the moderate zone, 11% of the study area is in the high zone, and 32.7% of the study area is in the very high flood susceptibility zone. For the BA model, 24.29% of the study area is in the very low flood susceptibility zone, 17.09% of the study area is in the low category, 13.75% of the study area is in the moderate category, 15.33% of the study area is in the high category and 19.53% of the study area is in the very high flood susceptibility category. For the RF model, 46.77% of the study area is located in the very low flood susceptibility class, 2.47% of the study area is classified in the low flood susceptibility zone, 3.71% of the study area is classified in the moderate zone, 8.45% of the study area is classified in the high zone and 38.33% of the study area is classified in the very high flood susceptibility zone. For the XGB model, the very low flood susceptibility accounts for about 23.02%, the low flood susceptibility accounts for about 23%, the moderate flood susceptibility accounts for 15%, the high flood susceptibility accounts for 14.5%, and the very high flood susceptibility accounts for 24.45%. For the GB model, about 26.98% of the study area occupies the very low flood susceptibility zone, 22.06% of the study area in the low zone, 12.59% of the study area in the moderate zone, 13.94% of the study area in the high zone, and 24.43% of the study area in the very high zone (Fig. 4).

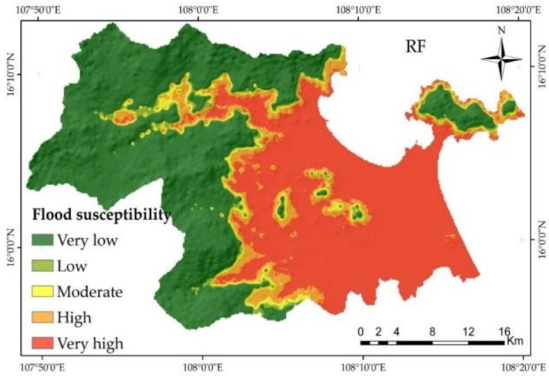


Figure 4. Flood susceptibility map produced by the RF model

3.4. Adaptation Capacity of the Community

3.4.1. Natural capital

The quality of domestic water plays an important role, directly affecting communities' resilience to flood impacts. The survey results show that 34% of households believe that water quality after floods is only slightly affected, while 35% assess that water quality is severely affected and cannot be used. The difficulty of accessing clean water increases the risk of disease and reduces communities' resilience. Furthermore, the role of rivers, lakes, and canals in flood regulation is still not widely appreciated. Only 8% of households believe that river and lake systems play a significant role in mitigating the impact of floods.

In comparison, 41% believe that rivers and lakes are no longer effective because water surface areas are shrinking due to urbanization. The survey also shows that river and lake drainage plays an important role in improving flood drainage: 45% of households said this activity is carried out regularly in their area. However, 14% of respondents still reported that their area had never been dredged, indicating the need to strengthen the management and protection of natural water storage areas.

3.4.2. Physical capital

The survey results show that housing forms and technical infrastructure play an important role in improving the resilience of residential communities to floods. Of these, 26% of the households interviewed own solid multi-story houses with good resilience, while 30% have houses with tiled or corrugated iron roofs, which are easily affected by floods. 86% of households consider raising the floor of the house or building a loft an effective solution to protect property and people during floods. However, in the study area, the flood prevention system still has many limitations: 30% of the households interviewed said that the dyke and dam system is not guaranteed, 36% said that the transportation system is paralyzed during floods, and 30% said that the road system is easily damaged, affecting the response and adaptation capacity of the community. 38% of households said that the drainage system in urban areas is overloaded due to concrete and a lack of regular maintenance. People rated improving the drainage system as the top priority (73%), followed by transportation (18%) and clean water supply (6%). After floods, although services such as electricity, water, and health care are quickly restored, drainage systems and transportation infrastructure often take more than 1 month to recover, greatly affecting people's livelihoods and access to services. Human capital

Households' ability to prepare before floods occur plays an important role in mitigating their impact. In the study area, people's ability to prepare in advance is still limited, especially in terms of evacuation plans, food storage, and medical supplies. The survey results show that only 15% of households have a full storage plan, while 43% have basic preparations but lack an evacuation plan, 16% implement only some simple measures, and, notably, 26% have no response plan at all. Although most people are aware of food preparation and health protection, epidemic prevention plans and practical skills are still limited and have not received much attention. Of the 80 households surveyed, 53 are proactive and have a clear food storage and evacuation plan. In contrast, 27 interviewed households belong to vulnerable groups such as poor households, low education, and households with disabilities, and are unable to prepare response plans before floods occur due to a lack of funds and information.

Furthermore, a large number of interviewed households lack full knowledge of climate change: 44 households believe that climate change is the cause of increasingly severe floods, and 24 households do not know the cause. Furthermore, providing timely and accurate forecast information is an important issue in improving the community's adaptive and resilient capacity. 34% of the households interviewed said they received forecast information very quickly and did not have enough time to prepare prevention plans; 16% said the information was inaccurate, making it difficult to prepare timely prevention plans. This situation affects 85% of the households interviewed regularly and continuously.

3.4.3. Social capital

Social capital is the connections and collective capacity of a community, considered an important resource for improving community resilience to floods. The survey results show that 80% of people are satisfied with the support of local authorities and charitable organizations, reflecting effective coordination between levels of government in flood prevention. 64% of households noted that mutual support during floods played an important role in first aid, food exchange, and evacuation, thereby reducing pressure on the formal relief system. However, only 41% of the households interviewed participated in community organizations related to natural disaster prevention, while 59% did not, which limited the provision of timely response information. 39% of households receiving long-term support can recover from floods more quickly and stabilize their livelihoods than households in areas without support. The main reasons are the lack of integration of support activities into local development planning and limited community capacity.

3.4.5. Financial resources

Financial resources are important factors in implementing solutions to prevent, respond to, and recover from floods, thereby improving the community's adaptive capacity. These resources can come from the state, individual businesses, and social organizations. The results show that 16% of low-income households with incomes below 6 million VND/month are the most vulnerable group, as they have limited financial accumulation capacity and live in areas near rivers. Low-income groups often have unstable jobs. Meanwhile, 64% of households with incomes above 6 million/month have better financial conditions to improve flood resilience, such as renovating homes and relocating in case of an emergency. Although not directly affected, this group may still suffer indirect losses in business operations or supply chains. This shows the need for support policies that focus on low-income groups to improve the community's adaptive capacity.

Furthermore, 56% of households do not have enough financial resources to overcome the consequences of floods when losses exceed 10%, of which 5 households have losses greater than 50%, 42 households fall into the medium damage group (10−50%). On the contrary, 33 households have low or no damage due to their proactive approach to floods. Obviously, the extent of damage is directly proportional to households' financial capacity. Therefore, there is a need for timely financial support policies to improve the community's resilience and adaptation capacity.

4. Discussions

Floods are considered among the most dangerous natural disasters, causing significant damage to people and the country's socioeconomy (Demissie, Rimal et al., 2024). Climate change and urban development have increased the frequency and intensity of flood events. Accurate assessment of flood susceptibility is vital for guiding local authorities in sustainable land-use planning. Furthermore, accurately identifying areas at risk of flooding helps people evacuate proactively, thus minimizing damage to people and property (Widya, Rezaie et al., 2024). Meanwhile, the community's adaptive capacity is considered a measure of society's resilience to natural hazards, in general, and floods in particular. This is a combination of three factors: risk awareness, organizational capacity, and proactive action. Communities with high adaptive capacity are often demonstrated through the acquisition of natural, physical, human, social, and economic capital. Therefore, this study combines the assessment of flood sensitivity and community adaptation capacity, which is not only of scientific significance, but also of high practical significance in disaster risk management in general and flood in particular. At the same time, the results of this study can support local authorities in developing community capacity-building programs to minimize the impact of floods.

Vietnam is one of the 13 countries most severely affected by climate change and one of the 6 countries most affected by floods (Nguyen et al., 2023). Therefore, the national policy on disaster mitigation, in general, and on floods in particular, in Vietnam has been established and continuously improved and updated. This shows the government's interest in responding to floods, particularly amid rising sea levels and urbanization. Specifically, the 2013 Law on Natural Disaster Prevention and Control (amended in 2020) plays an important role in legislation, regulations, and coordination mechanisms among authorities at all levels, organizations, and individuals to prevent, respond to, and mitigate the consequences of floods. The Strategy on Natural Disaster Prevention, Control, and Mitigation for the 2021-2030 period emphasizes the application of technology and big data in flood prevention and control. In addition, national programs and goals on climate change adaptation aim to invest in resilient infrastructure and to transform people's livelihoods in areas frequently affected by floods, thereby improving communities' adaptive capacity. These policies not only play an important role in disaster management but are also integrated into the urban planning system and socioeconomic development strategies.

In Vietnam, the flood risk management strategy clearly shows decentralization from the central to provincial and local levels. The provincial level focuses on directing, strategizing, and supervising work at the grassroots level, such as strengthening the dyke system and upgrading the drainage system. Meanwhile, the communal level takes on the role of preventing, responding to, and mitigating the consequences of floods. Regarding the decision-making process in Vietnam, in particular, and in developing countries in general, flood risk management still relies on a top-down model, in which the central government issues policies and directs disaster prevention activities, while the provincial level implements and guides lower levels. This model helps ensure consistency in the management system and is convenient in emergencies; however, in many cases, it still lacks flexibility and has not promoted the role of local communities in responding to natural disasters, particularly floods (Nguyen et al., 2023). Therefore, this study highlights the importance of integrating spatial analysis of flood risk with assessment of community resilience to improve the effectiveness of flood risk management. The national strategy for prevention and mitigation for the period 2021−2030 focuses on the application of advanced technologies and big data analysis in flood prevention. At the same time, climate change adaptation programs also focus on investing in resilient infrastructure systems to address floods and on transforming livelihoods in flood-prone areas to improve community resilience. These policies not only contribute significantly to flood risk management but are also integrated into urban planning and socioeconomic development strategies. In addition, Da Nang city has implemented many programs and actions to minimize flood-related damage, including the construction of 12 flood control reservoirs and the upgrading of 35 km of drainage culverts.

In addition to the national disaster mitigation policy, the proactive participation of people also plays an important role in reducing the impact of floods and generally improving the community's adaptive capacity. In fact, the community can effectively implement many adaptive solutions tailored to local conditions and available resources. One effective solution is to build flood-resistant houses, such as raising floor height and adding attics to single-story houses in low-lying areas. In addition, the community diversifies its livelihoods by taking on secondary jobs, such as handicrafts and small-scale trading, to gradually reduce dependence on agriculture, which is easily affected by floods. Not only do people resort to spontaneous measures, but they also actively cooperate with local authorities by proposing long-term solutions, such as dredging rivers or building small reservoirs. Da Nang is considered one of the fastest-growing urban areas in Vietnam and significantly affects the risk of flooding, both in intensity and frequency. Specifically, an increase in impervious surfaces increases flow, causing floods to rise faster. The filling of natural rivers, lakes, and canals to serve the construction of new urban areas has disrupted the hydrological system, exacerbating flooding in low-lying areas. This phenomenon is not only found in Vietnam but also in many cities around the world (Devi, Sridharan et al., 2019). Therefore, integrating spatial analysis of areas prone to flooding with the community's adaptive capacity is important for minimizing the negative impacts of floods and supporting policymakers and local authorities in developing reasonable solutions.

Although this study has been successful in mapping the spatial distribution of flood-prone areas and assessing the adaptive capacity of the community, floods are greatly affected by climate change and urban planning, so in the future, we will try to integrate these factors into the assessment of the susceptibility and adaptive capacity of the community, to provide a more comprehensive view and can support local authorities in coming up with effective solutions, especially in the context of climate change. Furthermore, this study collected socioeconomic data at a single point in time, which does not fully reflect the temporal variation in the community's adaptive capacity. Adding socioeconomic data across multiple time points will improve the study's precision. We will try to add this data in the future.

5. Conclusions

Flooding is one of the most severe natural hazards, causing substantial human and socioeconomic losses, especially under climate change and rapid urbanization. Most previous studies have focused on spatial flooding prediction, while very few have integrated both spatial and socioeconomic data to provide a more comprehensive understanding of the phenomenon. The objective of this study is to predict flood susceptibility and assess community adaptation capacity based on machine learning and socioeconomic data. The conclusion of this study is as follows.

(i) This study demonstrates the effectiveness of machine learning in predicting flood susceptibility in Da Nang City. The models proposed in this study can be adapted and applied to other regions with similar natural and socioeconomic characteristics to those of Da Nang City.

(ii) Among the proposed models, the RF model performed better than the other models with an AUC value of 0.989. According to this model, approximately 46.77% of the study area is located in the very low flood susceptibility class, 2.47% of the study area is classified in the low flood susceptibility zone, 3.71% of the study area is classified in the moderate zone, 8.45% of the study area is classified in the high zone and 38.33% of the study area is classified in the very high flood susceptibility zone.

(iii) This study spatially predicted flood susceptibility and integrated socioeconomic factors, including access to economic, physical, natural, social, and human resources, to assess community adaptive capacity. The results showed that easy access to resources, such as financial and social capital, can help mitigate the impact of floods by strengthening communities' adaptive capacity. Urban development and industrialization play a major role in enhancing communities' adaptive capacity. However, urban development increases population pressure and land-use changes, thereby increasing flood risks. This study developed a new theoretical framework integrating spatial analysis and socioeconomic data to inform strategies for managing future flood risks in a context of climate change and increasing urbanization.

Reference

Abass K., Buor D., Afriyie K., Dumedah G., Segbefi A.Y., Guodaar L., Garsonu E.K., Adu-Gyamfi S., Forkuor D., Ofosu A., 2020. Urban sprawl and green space depletion: Implications for flood incidence in Kumasi, Ghana. International Journal of Disaster Risk Reduction, 51, 101915. https://doi.org/10.1016/j.ijdrr.2020.101915.

Abedi R., Costache R., Shafizadeh-Moghadam H., Pham Q.B., 2022. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. Geocarto International, 37(19), 5479−5496. https://doi.org/10.1080/10106049.2021.1920636.

Ait Naceur H., Abdo H.G., Igmoullan B., Namous M., Alshehri F., Albanai J.A., 2024. Implementation of random forest, adaptive boosting, and gradient boosting decision trees algorithms for gully erosion susceptibility mapping using remote sensing and GIS. Environmental Earth Sciences, 83(3), 121. https://doi.org/10.1007/s12665-024-11424-5.

Al-Juaidi A.E., 2023. The interaction of topographic slope with various geo-environmental flood-causing factors on flood prediction and susceptibility mapping. Environmental Science and Pollution Research, 30(21), 59327−59348. https://doi.org/10.1007/s11356-023-26616-y.

Amiri A., Soltani K., Ebtehaj I., Bonakdari H., 2024. A novel machine learning tool for current and future flood susceptibility mapping by integrating remote sensing and geographic information systems. Journal of Hydrology, 632, 130936. https://doi.org/10.1016/j.jhydrol.2024.130936.

Amitrano D., Martino G. Di, Simone A. Di, Imperatore P., 2024. Flood detection with SAR: A review of techniques and datasets. Remote Sensing, 16(4), 656. https://doi.org/10.3390/rs16040656.

Babaei S., Ghazavi R., Erfanian M., 2018. Urban flood simulation and prioritization of critical urban sub-catchments using SWMM model and PROMETHEE II approach. Physics and Chemistry of the Earth, Parts A/B/C, 105, 3−11. https://doi.org/10.1016/j.pce.2018.02.002.

Breiman L., 2001. Random forests. Machine learning, 45, 5−32. https://doi.org/10.1023/A:1010933404324.

Chapi K., Singh V.P., Shirzadi A., Shahabi H., Bui D.T., Pham B.T., Khosravi K., 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environmental Modelling & Software, 95, 229−245. https://doi.org/10.1016/j.envsoft.2017.06.012.

Chen W., Li Y., Xue W., Shahabi H., Li S., Hong H., Wang X., Bian H., Zhang S., Pradhan B., 2020. Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. Science of The Total Environment, 701, 134979. https://doi.org/10.1016/j.scitotenv.2019.134979.

Demissie Z., Rimal P., Seyoum W.M., Dutta A., Rimmington G., 2024. Flood susceptibility mapping: integrating machine learning and GIS for enhanced risk assessment. Applied Computing and Geosciences, 23, 100183. https://doi.org/10.1016/j.acags.2024.100183.

Devi N.N., Sridharan B., Kuiry S.N., 2019. Impact of urban sprawl on future flooding in Chennai city, India. Journal of Hydrology, 574, 486−496. https://doi.org/10.1016/j.jhydrol.2019.04.041.

Di B., Zhang H., Liu Y., Li J., Chen N., Stamatopoulos C.A., Luo Y., Zhan Y., 2019. Assessing susceptibility of debris flow in southwest China using gradient boosting machine. Scientific Reports, 9(1), 12532. https://doi.org/10.1038/s41598-019-48986-5.

Dimitrov S., Borisova B., Ihtimanski I., Radeva K., Iliev M., Semerdzhieva L., Petrov S., 2024. Rapid Mapping: Unmanned Aerial Vehicles and Mobile-Based Remote Sensing for Flash Flood Consequence Monitoring (A Case Study of Tsarevo Municipality, South Bulgarian Black Sea Coast). Urban Science, 8(4), 255. https://doi.org/10.3390/urbansci8040255.

Elsafi S.H., 2014. Artificial neural networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan. Alexandria Engineering Journal, 53(3), 655−662. https://doi.org/10.1016/j.aej.2014.06.010.

Freund Y., Schapire R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119−139. https://doi.org/10.1006/jcss.1997.1504.

Gao W., Liao Y., Chen Y., Lai C., He S., Wang Z., 2024. Enhancing transparency in data-driven urban pluvial flood prediction using an explainable CNN model. Journal of Hydrology, 645(A), 132228. https://doi.org/10.1016/j.jhydrol.2024.132228.

Gegenleithner S., Pirker M., Dorfmann C., Kern R., Schneider J., 2025. Long short-term memory networks for enhancing real-time flood forecasts: a case study for an underperforming hydrologic model. Hydrology and Earth System Sciences, 29(7), 1939−1962. https://doi.org/10.5194/hess-29-1939-2025.

Haase D., 2013. Participatory modelling of vulnerability and adaptive capacity in flood risk management. Natural Hazards, 67(1), 77−97. https://doi.org/10.1007/s11069-010-9704-5.

Hancock J., Khoshgoftaar T.M., 2020. Performance of CatBoost and XGBoost in Medicare Fraud Detection, 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 572-579. Doi: 10.1109/ICMLA51294.2020.00095.

Hendrawan V.S.A., Rahardjo A.P., Mawandha H.G., Aldrian E., Muhari A., Komori D., 2025. Past and future climate-related hazards in Indonesia. EGUsphere, 1−34. https://doi.org/10.5194/egusphere-2025-584.

Hoang D.-V., Liou Y.-A., 2024. Assessing the influence of human activities on flash flood susceptibility in mountainous regions of Vietnam. Ecological Indicators, 158, 111417. https://doi.org/10.1016/j.ecolind.2023.111417.

Jahanbani M., Vahidnia M.H., Aghamohammadi H., Azizi Z., 2024. Flood susceptibility mapping through geoinformatics and ensemble learning methods, with an emphasis on the AdaBoost-Decision Tree algorithm, in Mazandaran, Iran. Earth Science Informatics, 17(2), 1433−1457. https://doi.org/10.1007/s12145-023-01213-2.

Kavzoglu T., Teke A., 2022. Advanced hyperparameter optimization for improved spatial prediction of shallow landslides using extreme gradient boosting (XGBoost). Bulletin of Engineering Geology and the Environment, 81(5), 201. https://doi.org/10.1007/s10064-022-02708-w.

Kaya C.M., Derin L., 2023. Parameters and methods used in flood susceptibility mapping: a review. Journal of Water and Climate Change, 14(6), 1935−1960. https://doi.org/10.2166/wcc.2023.035.

Khosravi K., Shahabi H., Pham B.T., Adamowski J., Shirzadi A., Pradhan B., Dou J., Ly H.-B., Gróf G., Ho H.L., 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.

Kundzewicz Z.W., Kanae S., Seneviratne S.I., Handmer J., Nicholls N., Peduzzi P., Mechler R., Bouwer L.M., Arnell N., Mach K., 2014. Flood risk and climate change: global and regional perspectives. Hydrological Sciences Journal, 59(1), 1−28. https://doi.org/10.1080/02626667.2013.857411.

Lai T.A., Nguyen N.T., Bui Q.T., 2023. Hyper-parameter optimization of gradient boosters for flood susceptibility analysis. Transactions in GIS, 27(1), 224−238. https://doi.org/10.1111/tgis.13023.

Lam D., Zhang X., Li H., Deshan Y., Schott B., Zhao T., Zhang W., Mutic S., Sun B., 2019. Predicting gamma passing rates for portal dosimetry-based IMRT QA using machine learning. Medical Physics, 46(10), 4666−4675. https://doi.org/10.1002/mp.13752.

Lee S., Kim J.-C., Jung H.-S., Lee M.J., Lee S., 2017. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. Geomatics, Natural Hazards and Risk, 8(2), 1185−1203. https://doi.org/10.1080/19475705.2017.1308971.

Li X., Wang L., Sung E., 2008. AdaBoost with SVM-based component classifiers. Engineering Applications of Artificial Intelligence, 21(5), 785−795. https://doi.org/10.1016/j.engappai.2007.07.001.

Liuzzo L., Sammartano V., Freni G., 2019. Comparison between different distributed methods for flood susceptibility mapping. Water Resources Management, 33, 3155−3173. https://doi.org/10.1007/s11269-019-02293-w.

Mahmoud S.H., Gan T.Y., 2018. Multi-criteria approach to develop flood susceptibility maps in arid regions of Middle East. Journal of Cleaner Production, 196, 216−229. https://doi.org/10.1016/j.jclepro.2018.06.047.

Mahmoud S.H., Gan T.Y., 2018. Urbanization and climate change implications in flood risk management: Developing an efficient decision support system for flood susceptibility mapping. Science of the Total Environment, 636, 152−167. https://doi.org/10.1016/j.scitotenv.2018.04.282.

Mirza M.M.Q., 2011. Climate change, flooding in South Asia and implications. Regional environmental change 11(1), 95−107. https://doi.org/10.1007/s10113-010-0184-7.

Nachappa T.G., Piralilou S.T., Gholamnia K., Ghorbanzadeh O., Rahmati O., Blaschke T., 2020. Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using Dempster Shafer Theory. Journal of Hydrology, 590, 125275. https://doi.org/10.1016/j.jhydrol.2020.125275.

Namara W.G., Damisse T.A., Tufa F.G., 2022. Application of HEC-RAS and HEC-GeoRAS model for flood inundation mapping, the case of Awash bello flood plain, upper Awash River Basin, oromiya regional state, Ethiopia. Modeling Earth Systems and Environment, 8(2), 1449−1460. https://doi.org/10.1007/s40808-021-01166-9.

Nguyen H.D., Nguyen Q.-H., Bui Q.-T., 2024. Solving the spatial extrapolation problem in flood susceptibility using hybrid machine learning, remote sensing, and GIS. Environmental Science and Pollution Research, 31(12), 18701−18722. https://doi.org/10.1007/s11356-024-32163-x.

Nguyen H.D., Nguyen T.H.T., Nguyen Q.-H., Nguyen T.G., Dang D.K., Nguyen Y.N., Bui T.H., Nguyen N.D., Bui Q.-T., Brecan P., 2023. Bottom-up approach for flood-risk management in developing countries: a case study in the Gianh River watershed of Vietnam. Natural Hazards, 118(3), 1933−1959. https://doi.org/10.1007/s11069-023-06098-4.

Pal S., Saha A., Gogoi P., Saha S., 2024. An Ensemble of J48 Decision Tree with AdaBoost and Bagging for Flood Susceptibility Mapping in the Sundarbans of West Bengal, India. Geomorphic Risk Reduction Using Geospatial Methods and Tools, Springer, 117−133. https://doi.org/10.1007/978-981-99-7707-9_7.

Panjavarnam B., Anish Bose S.S., Dinesh K., Nataraj P., 2024. Unmanned Aerial Vehicle-Disaster Management. 2024 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), IEEE, 1−4. https://doi.org/10.1109/ICPECTS62210.2024.10780226.

Pappenberger F., Dutra E., Wetterhall F., Cloke H.L., 2012. Deriving global flood hazard maps of fluvial floods through a physical model cascade. Hydrology and Earth System Sciences, 16(11), 4143−4156. https://doi.org/10.5194/hess-16-4143-2012.

Perivolioti T.-M., Zachopoulos K., Zioga M., Tompoulidou M., Katsavouni S., Kemitzoglou D., Terzopoulos D., Mouratidis A., Tsiaoussi V., 2024. Monitoring the impact of floods on water quality using optical remote sensing imagery: the case of lake karla (Greece). Water, 16(23), 3502. https://doi.org/10.3390/w16233502.

Rashidiyan M., Rahimzadegan M., 2024. Investigation and evaluation of land use-land cover change effects on current and future flood susceptibility. Natural Hazards Review, 25(1), 04023049. https://doi.org/10.1061/NHREFO.NHENG-1854.

Ren H., Pang B., Bai P., Zhao G., Liu S., Liu Y., Li M., 2024. Flood susceptibility assessment with random sampling strategy in ensemble learning (RF and XGBoost). Remote Sensing, 16(2), 320. https://doi.org/10.3390/rs16020320.

Rigatti S.J., 2017. Random forest. Journal of Insurance Medicine, 47(1), 31−39.

Sagi O., Rokach L., 2021. Approximating XGBoost with an interpretable decision tree. Information Sciences, 572, 522−542. https://doi.org/10.1016/j.ins.2021.05.055.

Sharma C.S., Mishra A., Panda S.N., 2014. Assessing impact of flood on river dynamics and susceptible regions: geomorphometric analysis. Water Resources Management, 28, 2615−2638. https://doi.org/10.1007/s11269-014-0630-2.

Smit B., Wandel J., 2006. Adaptation, adaptive capacity and vulnerability. Global Environmental Change, 16(3), 282−292. https://doi.org/10.1016/j.gloenvcha.2006.03.008.

Stokanović S., Đukić D., Miljković N., 2024. Robustness of XGBoost algorithm to missing features for binary classification of medical data. 2024 23rd International Symposium INFOTEH-JAHORINA (INFOTEH), IEEE, 1−6. Doi: 10.1109/INFOTEH60418.2024.10495929.

Tan X., Li Y., Wu X., Dai C., Zhang X., Cai Y., 2024. Identification of the key driving factors of flash flood based on different feature selection techniques coupled with random forest method. Journal of Hydrology: Regional Studies, 51, 101624. https://doi.org/10.1016/j.ejrh.2023.101624.

Vafakhah M., Mohammad Hasani Loor S., Pourghasemi H., Katebikord A., 2020. Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping. Arabian Journal of Geosciences, 13(11), 417. https://doi.org/10.1007/s12517-020-05363-1.

Versini P.-A., Gaume E., Andrieu H., 2010. Assessment of the susceptibility of roads to flooding based on geographical information-test in a flash flood prone area (the Gard region, France). Natural Hazards and Earth System Sciences, 10(4), 793−803. https://doi.org/10.5194/nhess-10-793-2010.

Vojtek M., Vojteková J., 2019. Flood susceptibility mapping on a national scale in Slovakia using the analytical hierarchy process. Water, 11(2), 364. https://doi.org/10.3390/w11020364.

Wahba M., Sharaan M., Elsadek W.M., Kanae S., Hassan H.S., 2024. Examination of the efficacy of machine learning approaches in the generation of flood susceptibility maps. Environmental Earth Sciences, 83(14), 429. https://doi.org/10.1007/s12665-024-11696-x.

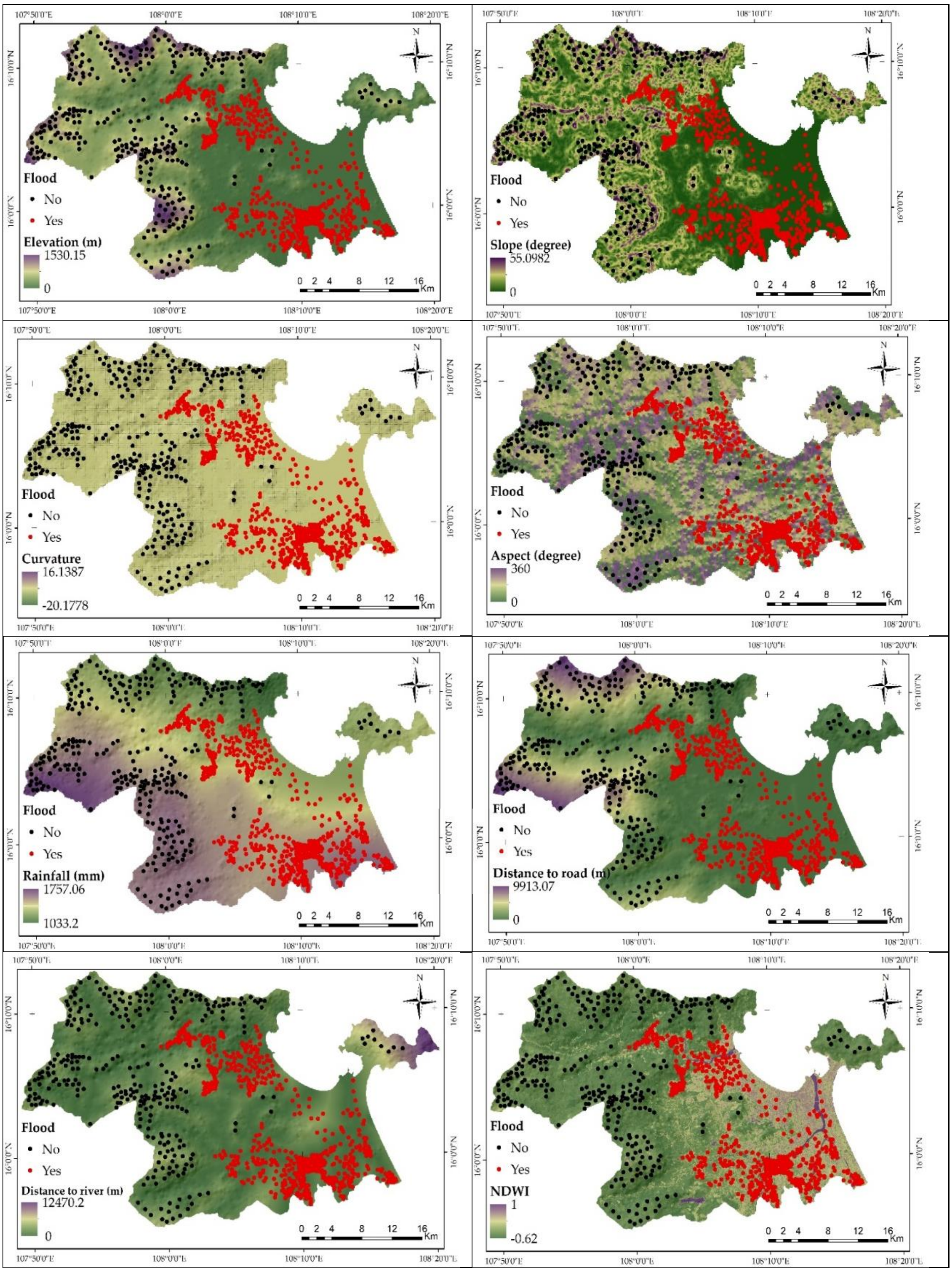
Widya L.K., Rezaie F., Lee W., Lee C.-W., Nurwatik N., Lee S., 2024. Flood susceptibility mapping of Cheongju, South Korea based on the integration of environmental factors using various machine learning approaches. Journal of Environmental Management, 364, 121291. https://doi.org/10.1016/j.jenvman.2024.121291.

Youssef A.M., Pradhan B., Dikshit A., Mahdi A.M., 2022. Comparative study of convolutional neural network (CNN) and support vector machine (SVM) for flood susceptibility mapping: a case study at Ras Gharib, Red Sea, Egypt. Geocarto International, 37(26), 11088−11115.
https://doi.org/10.1080/10106049.2022.2046866.

Yuan H., Wang M., Zhang D., Ikram R.M.A., Su J., Zhou S., Wang Y., Li J., Zhang Q., 2024. Data-driven urban configuration optimization: An XGBoost-based approach for mitigating flood susceptibility and enhancing economic contribution. Ecological Indicators, 166, 112247. https://doi.org/10.1016/j.ecolind.2024.112247.

Zhang T., Fu Q., Wang H., Liu F., Wang H., Han L., 2022. Bagging-based machine learning algorithms for landslide susceptibility modeling. Natural Hazards, 110(2), 823−846. https://doi.org/10.1007/s11069-021-04986-1.

APPENDIX


Flood
No
Yes
Elevation (m)
1530.15
0
Flood
No
Yes
Slope (degree)
55.0982
0
Flood
No
Yes
Curvature
16.1387
-20.1778
Flood
No
Yes
Aspect (degree)
360
0
Flood
No
Yes
Rainfall (mm)
1757.06
1033.2
Flood
No
Yes
Distance to road (m)
9913.07
0
Flood
No
Yes
Distance to river (m)
12470.2
0
Flood
No
Yes
NDWI
1
-0.62
0 2 4 8 12 16 Km

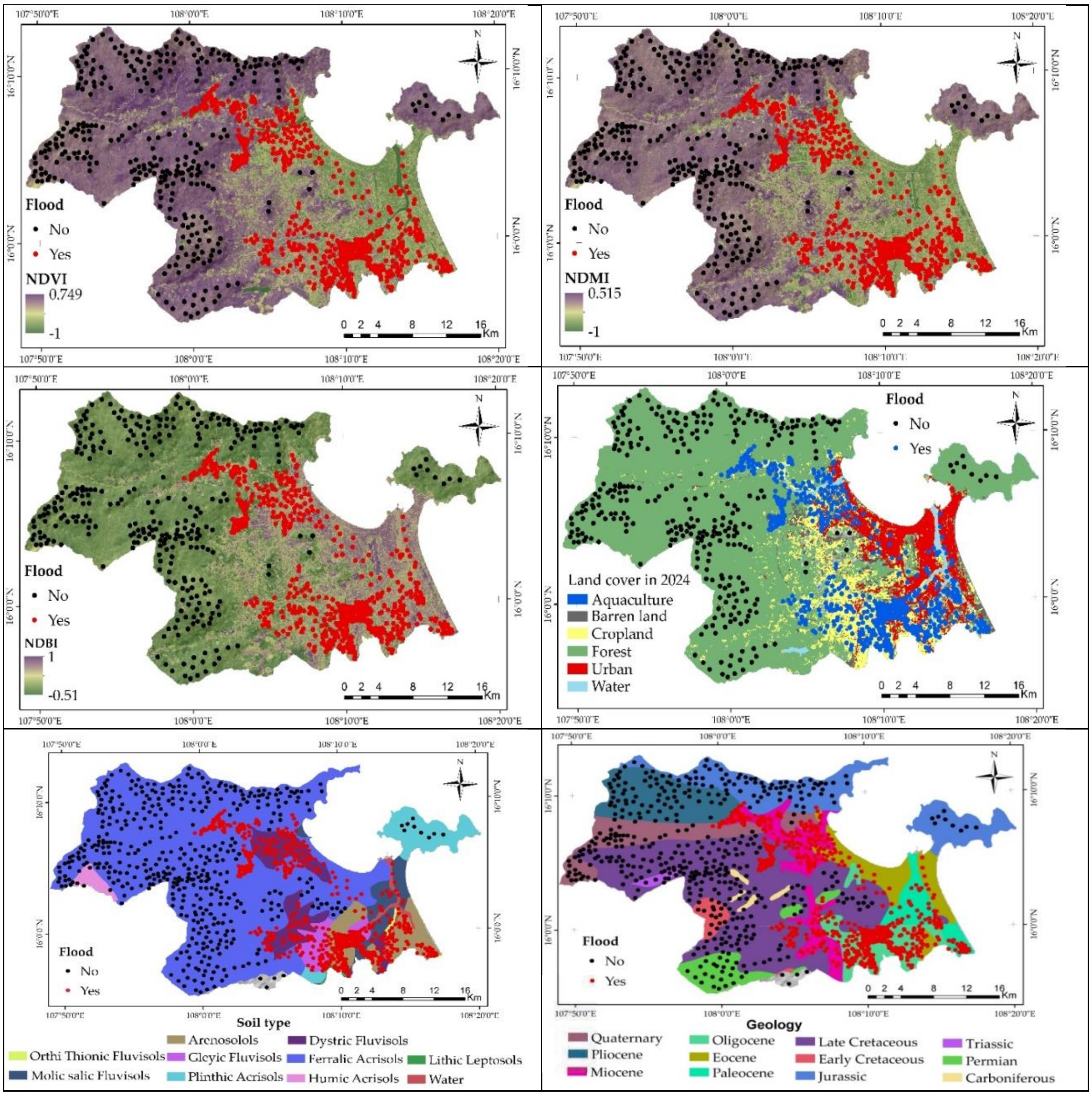


Figure 1A. Conditioning factor used to build the machine learning model in this study

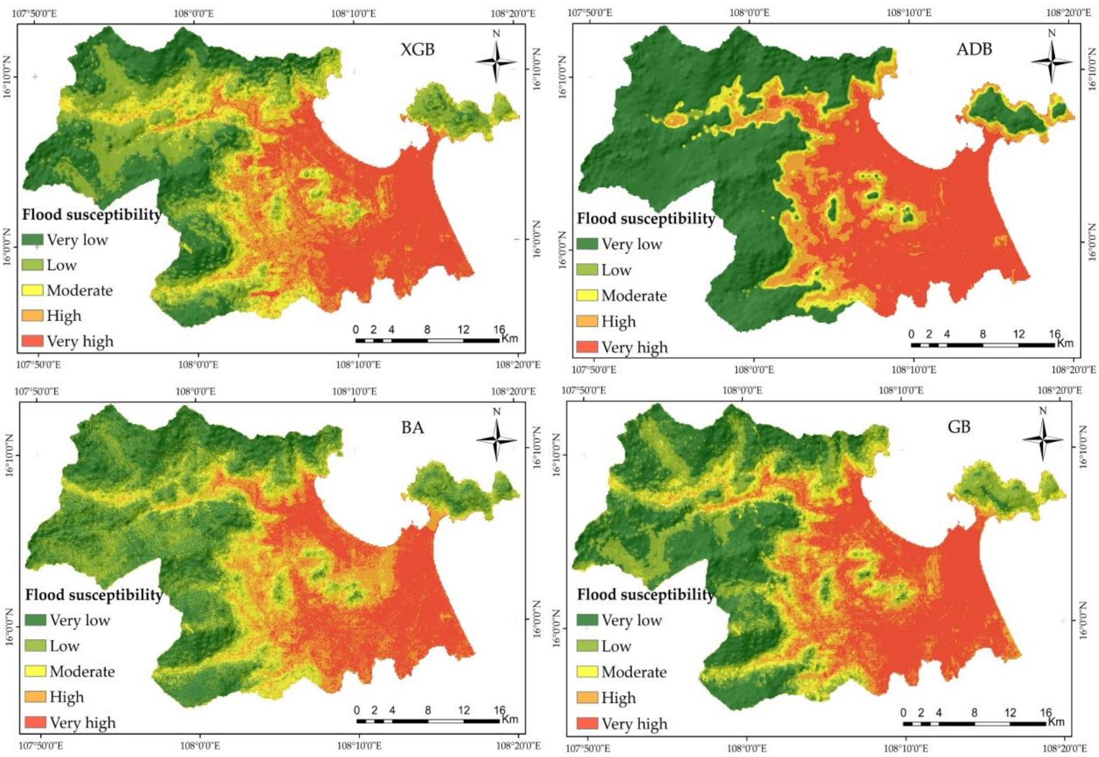


Figure 2A. Flood susceptibility map produced by the XGB, ADB, BA, GB model