ASSESSMENT OF INUNDATION SUSCEPTIBILITY IN THE CONTEXT OF CLIMATE CHANGE, BASED ON MACHINE LEARNING AND REMOTE SENSING: CASE STUDY IN VINH PHUC PROVINCE OF VIETNAM

Phan Manh HUNG [1], Huu Duy NGUYEN [2*], Chien Pham VAN [3]

DOI: 10.21163/GT_2023.182.07

ABSTRACT:

Accurate prediction of streamflow plays an important role in water resource management and the continuous assessment of inundation susceptibility in the context of climate change plays a key role in facilitating the construction of appropriate strategies for sustainable development. So far, few studies into inundation susceptibility have explicitly incorporated the effects of climate change into their methodologies. This study aimed to assess inundation susceptibility for Vinh Phuc province in Vietnam, from 2000 to 2020, using machine learning and remote sensing. The algorithms used were support vector machine, catboost, and extratrees. A geo-spatial database of 206 inundation points and 11 conditioning factors (namely elevation, slope, curvature, aspect, distance to river, distance to road, NDVI, NDBI, rainfall, soil type, and TWI) from 2000 to 2020 was developed to be used as the input data. RMSE, MAE, AUC, and R^2 were used to assess the fit of the models. The results showed that all the proposed models were a good fit, with AUC values of 0.95 and over. In general, the total area marked as very low risk or low risk has increased, with the high risk and very high-risk areas having decreased over the period studied. This change was mainly concentrated in the city of Vinh Yen where there has been strong urban growth. The models proposed in this study are a promising toolkit to assess inundation susceptibility continuously and can support decision makers involved in sustainable development. Our results highlight the benefits and consequences of planned and unplanned development. Properly planned can reduce the flood risk, while unplanned development can increase the risk. Therefore, by applying the theoretical framework in this study, decision makers or planners can build the most appropriate strategies for flood control in the context of climate change. Our approach in this study represents a theoretical framework for future research not only on inundation management but also natural hazard management, in regions around the world.

Key-words: *inundation, support vector machine, Catboost, Extratree, machine learning, Vinh Phuc.*

1. INTRODUCTION

Climate change is considered one of the three key factors for "global changes". This process is the transformation of nature to human, important influence on the biological, the natural resource such as water, land and air resource, the weather system and the natural hazard such as forest fire, flood, and inundation (Petrişor et al. 2020, Nguyen et al. 2022a). Climate change can alter the frequency and intensity of precipitation and the hydrological regime of a region and therefore can influence the occurrence and frequency of inundation. Several studies have established that inundation is the result of climate change and is influenced by climatic determinants such as increased precipitation (Roy et al. 2020, Avand and Moradi 2021, Li et al. 2021, Rajkhowa and Sarma 2021). According to the Center for Research on the Epidemiology of Disasters (CRED), inundation is one of the most damaging natural hazards to human life and property. According to the World Bank, in the period 1995-2005, approximately 150,000 floods and inundation occurred worldwide, leading to some 157,000 deaths, representing 11% of all victims of natural disasters (Janizadeh et al. 2021a, Ghosh et al. 2022c, Ghosh et al. 2022b). Multiple studies have indicated that flood and inundation affect in the region of 200 million people each year and that these numbers are likely to increase significantly by 2050 due to climate change and urban growth (Pourghasemi et al. 2020, Janizadeh et al. 2021b).

[1]*Department of Agriculture and Rural Development, Vinh Phuc, Vietnam,* hungccdd@gmail.com
[2]*Faculty of Geography,National University,Hanoi,Vietnam,*Corresponding author* nguyenhuuduy@hus.edu.vn
[3]*Thuyloi University, Dong Da, Hanoi, Vietnam,* Pchientvct_tv@tlu.edu.vn

41% of the deaths mentioned above were in Asia, and of all Asian countries, Vietnam is one of the countries most vulnerable to flood and inundation (Nguyen 2022a). For example, severe flood and inundation in the Central region of Vietnam in 1999 caused 547 deaths and damaged 630,000 homes. This is why predicting flood and inundation susceptibility has become a top priority for the scientific community and local authorities across the country.

Inundation susceptibility is the likelihood of the occurrence of high-water resources in a given region. Most studies in this area have focused on the assessment of inundation susceptibility at a specific time. However, the continuous assessment of inundation susceptibility is an effective framework to better support decision makers to understand the causes of increased inundation. Nguyen et al. (2022a) showed that constructing strategies that mitigate against inundation is more effective when susceptibility is continuously evaluated. Penning-Rowsell et al. (2013) also indicated that continuous inundation assessment plays an important role in building appropriate strategies to reduce inundation damage.

Previous studies have used different methodologies to assess inundation susceptibility. These can be divided into three main groups: hydrodynamic, remote sensing and GIS, and machine learning. The hydrodynamic approach – which features models such as Mike (Tansar et al. 2020, NGUYEN et al. 2023), HEC-RAS (Ongdas et al. 2020, Psomiadis et al. 2021), and SWAT (Jodar-Abellan et al. 2019, Liu et al. 2022)– uses mathematical equations to predict the flow of water in rivers, streams, and other waterways, and the relationship between river flows and precipitation. Although this approach has been used extensively in previous studies, the hydrodynamic model requires detailed topographical, meteorological, and other terrain data (Nguyen et al. 2021). Therefore, these models are limited to application in small areas or locations where there is plenty of data. The construction of hydrodynamic models also requires comprehensive knowledge of the relevant hydrological parameters. In recent years, with the development of remote sensing data and the data sharing policies of agencies such as NASA and NOAA, remote sensing and GIS have become a more viable option for assessing inundation susceptibility over a wide region. However, this approach is limited in its ability to identify the relationship between inundation and its causes (Wang and Xie 2018). It is clear, then, that these approaches will likely be superseded by more state-of-the-art methods.

A number of recent studies have used data-driver, integrated with remote sensing data, to continuously assess inundation susceptibility by analyzing the relationships between past inundation events and conditioning factors (Bui et al. 2020). Data-driver approach requires less input data than hydrodynamic models. This approach was separated by the statistical models and machine learning. Statistical models such as logistic regression (Al-Juaidi et al. 2018, Lim and Lee 2018), frequency ratio (Ghosh et al. 2022a), weight of evidence (Hong et al. 2018), fuzzy logic (Sahana and Patel 2019), and AHP (Swain et al. 2020) have shown promise, but in the context of climate change and urbanizations, inundation is becoming ever more complicated and non-linear; therefore, statistical models cannot achieve adequate levels of accuracy in their predictions. The machine learning approach has been receiving a great deal of attention within the scientific community. This approach includes models such as support vector machine (SVM) (Mohammady et al. 2019), random forest (Lee et al. 2017), bagging (Talukdar et al. 2020), adaboost (Tien Bui et al. 2016), artificial neural networks (Priscillia et al. 2021) and decision tree (Khosravi et al. 2018, Pham et al. 2021a). There is, however, no universal consensus as to which algorithm(s) – from the thousand available options – may best simulate inundation susceptibility in a given region.

In summary, there are two major gaps in the literature. First, most studies focus on the assessment of inundation susceptibility at a specific time, when assessment at different times has been shown to be important in the development of appropriate inundation prevention measures. Second, the selection of appropriate methods to predict the inundation susceptibility in a specific region continues to be a major challenge for the scientific community. In response to these two issues, this study has aimed to carry out (i) a continuous inundation susceptibility assessment for the period 2000–2020 and (ii) a comparison of models to select the best one for predicting inundation susceptibility. Although this study was carried out in a province in Vietnam and was focused on inundation, our findings may be applied in many other geographical regions and used to evaluate other natural hazards.

2. STUDY AREA AND DATA USE

2.1. Study area

Vinh Phuc province is located in the Red River Delta, in the Northern Midlands and Mountains region of Vietnam. It covers an area of approximately 1,236 km^2 (**Fig. 1**). The province descends from the northeast to the southwest, and its diverse topography can be divided into three categories: high mountains, middle lands, and plains. The high mountain region, with an elevation ranging from 200 to 1000 m, covers 65,500 ha of the province. The middle lands, with an elevation from 20 to 200 m, occupy 25,100 ha, and the plains, with an elevation of 0 to 20 m, occupy 33,500 ha. Vinh Phuc is located in the tropical monsoon climate region and has an average annual rainfall of 1400–1900 mm, mainly concentrated between May and October, accounting for 80% of the total rainfall. The dry season lasts from November to April and accounts for the remaining 20%. The river system in the study area is dense and the hydrological regime is mainly depending on the flow in two main river systems named the Red River and the Phan - Ca Lo River. According to statistics from the Forest Protection Department, the forested area of Vinh Phuc accounts for about 22% of the province. However, in recent years, under the pressure of a population explosion and accelerating urbanization, the forest area has been seriously reduced. This is thought to have been the main cause for the increasing severity of inundations and landslides in the province.

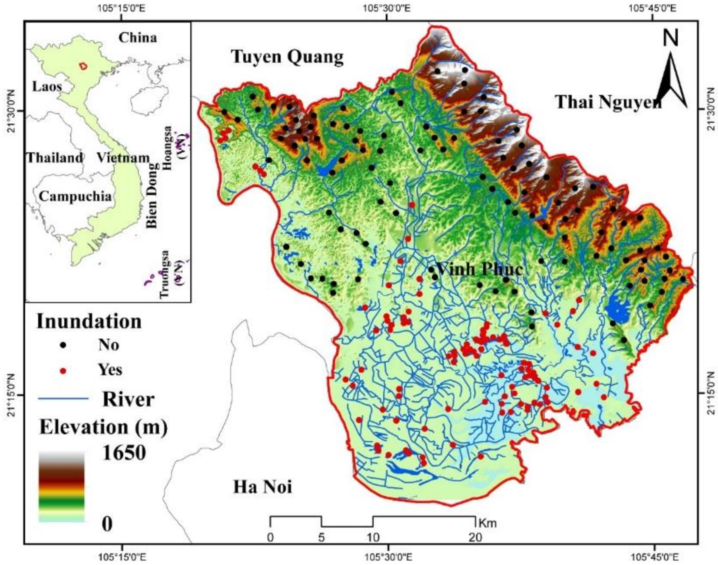


Fig. 1. The location of Vinh Phuc province.

2.2. Data use

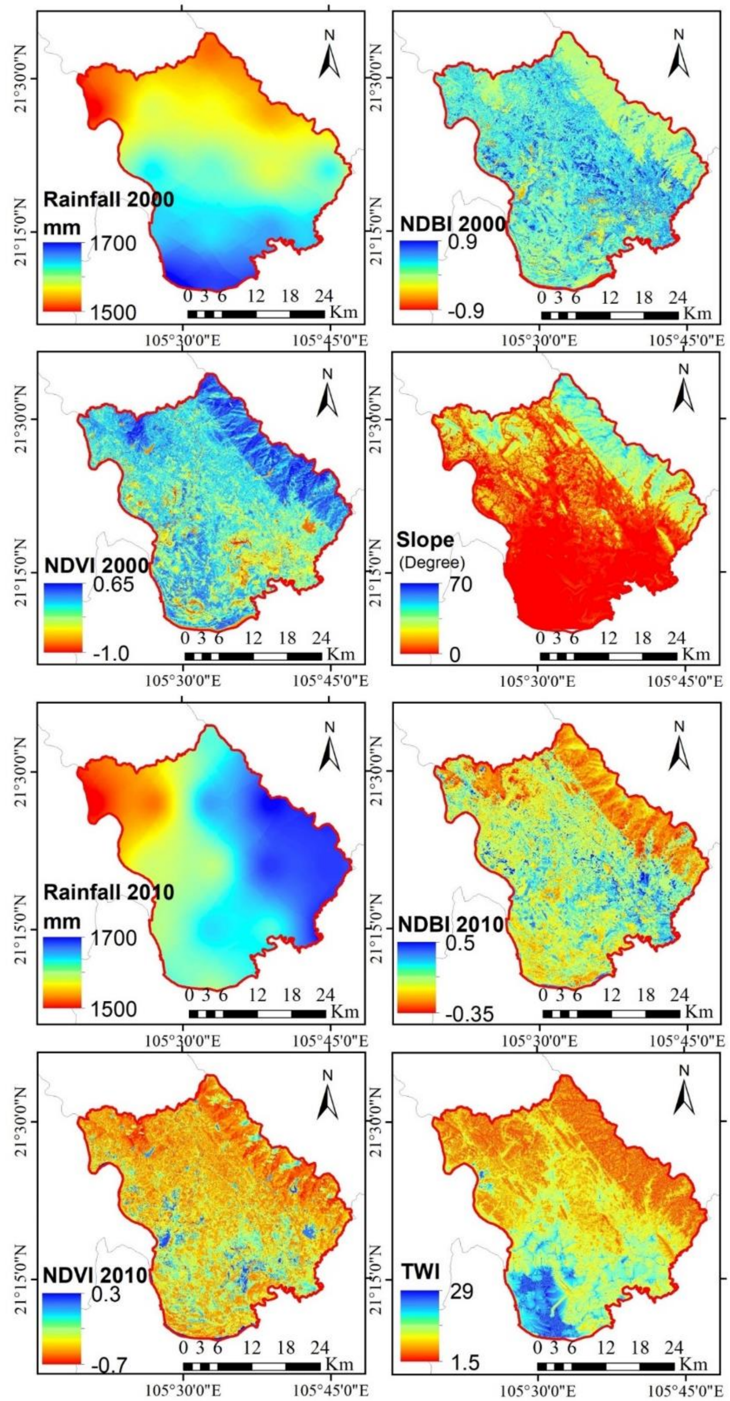
Inundation inventory

Inundation inventory preparation is essential in building a inundation susceptibility model and describes the relationships between past inundation events and conditioning factors (Nguyen et al. 2022b, Yaseen et al. 2022). In this study, 106 inundation points were collected from different sources such as the province's Department of Natural Resources and Environment, field missions in 2021 and 2022, and Sentinel 1A imagery. The construction of inundation susceptibility maps requires the binary classification of inundation inventory into two groups: inundation points and non- inundation points. Non- inundation points were taken from high-altitude locations that have never been affected by inundation. The inundation inventory was divided into two groups: 70% of the data for training and 30% for validation).

Conditioning factors

The construction of inundation susceptibility models is very complicated because it requires data regarding environmental, hydrological, climatic, and anthropic factors (Ghosh et al. 2022c, Nguyen

2022b). The selection of appropriate conditioning factors is critical and determines the accuracy of the inundation susceptibility model. After reviewing the existing literature, 11 conditioning factors were selected, namely altitude, slope, aspect, curvature, distance to river, distance to road, NDVI, NDBI, rainfall, soil type, and TWI. All conditioning factors were transformed to raster format at a resolution of 10 m (**Fig. 2**).



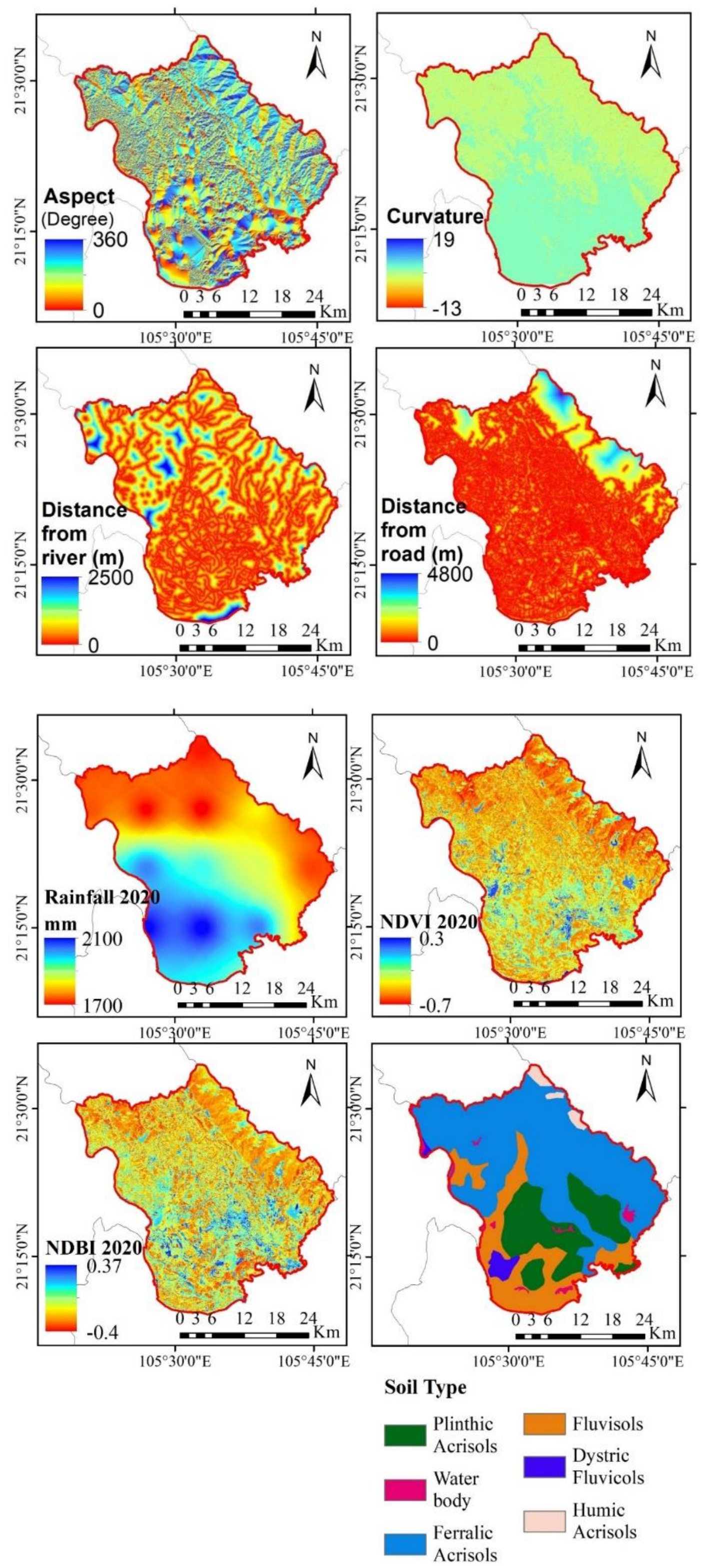


Fig. 2. Conditioning factor used for the flood susceptibility model.

Elevation, slope, aspect, curvature, and TWI were extracted from a DEM constructed by topographic map with the scale of 1:50,000 m. Note that the map was obtained from the Ministry of Natural Resources and Environment. Distance to river and distance to road were extracted from the topographic map. NDVI and NDBI in 2000, 2010, and 2020 were calculated from Landsat 7 and 8 images. Rainfall in 2000, 2010, and 2020 were downloaded from https://chrsdata.eng.uci.edu/. Soil type was collected from Ministry of Natural Resources and Environment.

Inundation likelihood is inversely proportional with altitude. The south of the area is characterized by flat and low land, so this region is often affected by inundation (Dahri et al. 2022). The altitude value in the province ranged from 0 to 1650 m. Slope directly influences the probability of inundation occurrence as it controls flow velocity. The probability of occurrence of inundation is inversely proportional with slope (Arabameri et al. 2022). In the study area, the slope value ranged from 0 to 70 degrees. Aspect is indirectly related to the occurrence of inundation because it influences the direction of flow and the maintenance of soil moisture (Nguyen 2022a). In the study area, the aspect value ranged from 0 to 360 degrees. Curvature plays an important role in assessing the likelihood of inundation occurrence. It is directly linked to the heterogeneity of each region. The curvature value is inversely proportional to the occurrence of inundation (Sachdeva and Kumar 2022). In the study area, the curvature value ranged from -13 to 19.

TWI is a measure of how likely a region is to experience inundation. The higher the value of TWI, the higher the probability of flood or inundation (Tehrany et al. 2014). In the study area, the TWI was found to range from 1.5 to 29. The type of soil in a region plays a significant role in the likelihood of inundation. This is because soil affects the rate of infiltration (water soaking into the ground) and the formation of runoff (water flowing on the surface) (Bui et al. 2019). In the study area, the soil types were classified into five groups: plinthic acrisols, ferralic acrisols, fluvisols, dystric fluvisols, and humic acrisols.

NDVI represents the density of vegetation in an area and directly influences the infiltration capacity and the speed of flow. Regions with low vegetation density are more vulnerable to inundation (Nachappa et al. 2020). In the study area, the range of NDVI values in 2000, 2010, and 2020 were -1‑0.65, 0‑0.5, and -0.7‑0.3 respectively. NDBI represents the density of construction in a specific region, which influences the probability of inundation as concrete constructions do not allow water to seep into the ground (Saha et al. 2021, Nguyen 2022b). In this study, the value ranges of NDBI in 2000, 2010, and 2020 were -0.9‑0.9, -0.35‑0.5, and -0.4‑0.37 respectively.

Rainfall was selected because precipitation of high intensity over a short time are closely linked with river flow and therefore inundation susceptibility (Band et al. 2020). In the study area, value ranges of rainfall 2000, 2010, and 2020 were 1500‑1700 mm, 1500‑1700 mm and 1700‑2100 mm respectively. Distance to road directly influences the capacity of infiltration and evacuation of water (Linh et al. 2022).

In the study area, distance to road ranged from 0 to 4800 m. Because most inundation occurs along the rivers, the further the region is from a river, the less likely it is to inundation (Chowdhuri et al. 2020). In this study, the distance to river value ranged from 0 to 2500 m.

3. METHODOLOGY

We applied machine learning and remote sensing in our study. This process was divided into four main steps: (i) collection and pre-processing of input data, (ii) construction of inundation susceptibility models, (iii) assessment of the accuracy of the proposed models, and (iv) analysis of the effects of climate change on inundation susceptibility in the period 2000‑2020 (**Fig. 3**).

(i) Collection and pre-processing of input data

The input data of the inundation susceptibility model constituted two groups: inundation inventory and conditioning factors. Inundation inventory comprises historical inundation - and non-inundation points, which were coded as 1 and 0, respectively. These points were collected from different sources: reports from the Department of Natural Resources and Environment, field missions, and Sentinel 1A imagery.

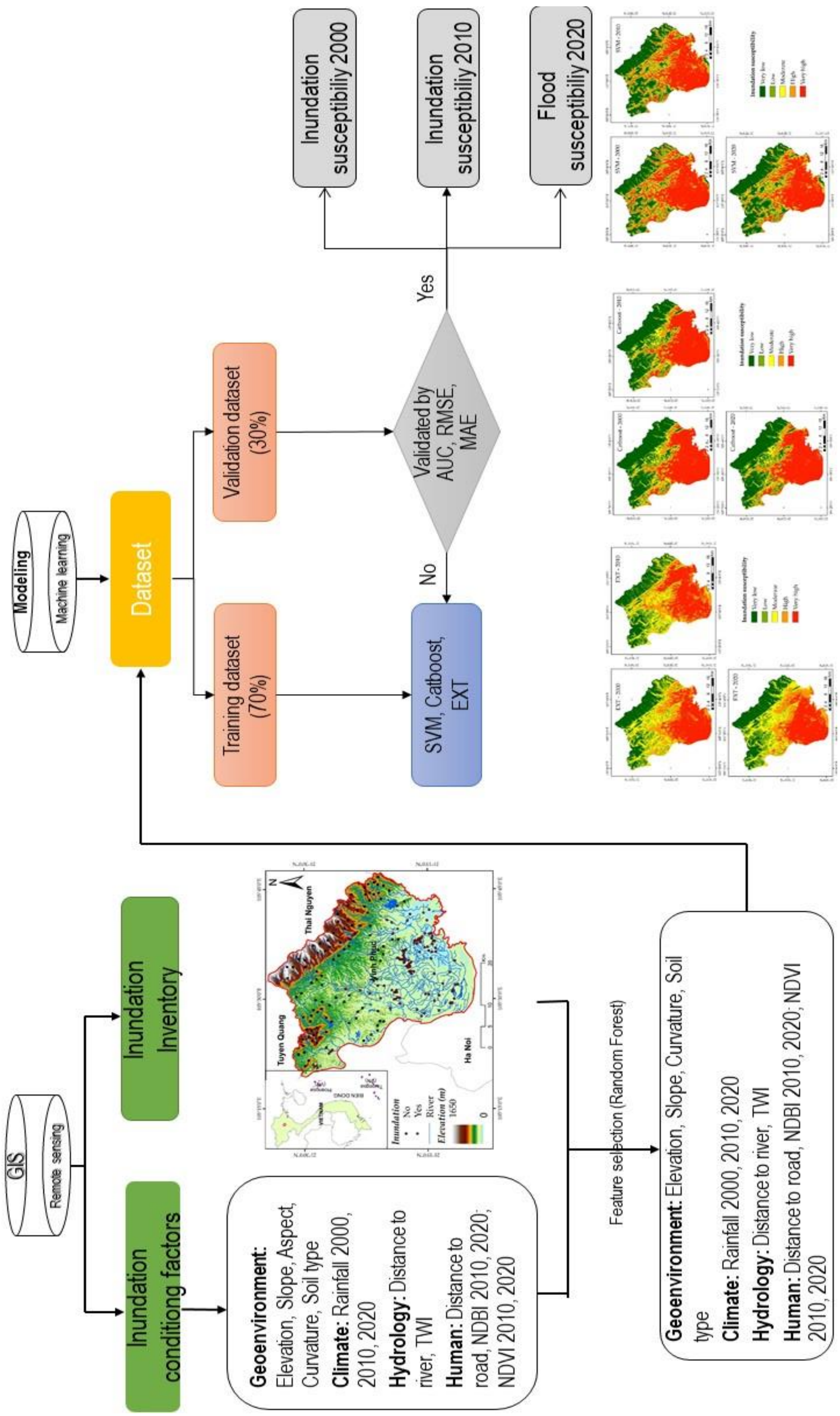


Fig. 3. The methodology used in this study.

While this study assesses inundation susceptibility continuously over the period 2000‑2020, theoretically, the conditioning factors should be collected from 2000 to 2020 to use as input data to assess changes in inundation susceptibility over this period. However, due to difficulties in data collection, we hypothesize that the topographic features did not change over the period in question, and that the relevant changes were instead in hydro-meteorological conditions and anthropic activity. These factors were calculated from the 1/50,000 m topographic map and from remote sensing data.

(ii) Building the inundation susceptibility model

206 inundation points and 11 conditioning factors were used as input data for the inundation susceptibility model. This data was divided into two groups: 70% to train the models (the adjustment of hyper-parameters) and 30% to evaluate the accuracy of the models. Other rates (50/50, 60/40, and 80/20) were tested, but 70/30 performed better. Building inundation susceptibility models can be challenging when data is limited. In order to validate the model, a 10-fold cross-validation method was applied to the dataset. The results of this validation were used to evaluate the model's performance through various performance indices. To find the best model, Bayesian optimization was used to select the optimal hyper-parameter values. Additionally, different models such as SVM, catboost (CB), EXT, adaboost, random forest, and xgboost were tested, and ultimately SVM, CB, and EXT were chosen as the best performing models. The accuracy of the models is impacted by the process of tuning their hyper-parameters. RMSE was used as the objective function to find the hyper-parameters that resulted in the highest accuracy after the training process. The hyper-parameters for the models SVM, CB, and EXT were determined through trial and error and set to C=0.1, Gamma=0.5 for SVM; depth=3, learning_rate=0.01 for CB; and max_depth=2, min_samples_split=2 for EXT.

(iii) Evaluation of the accuracy of the proposed models

The statistical indices AUC, RMSE, R^2, and MAE were applied to evaluate the inundation susceptibility models.

(iv) Analysis of the effects of climate change on the inundation susceptibility from 2000 to 2020

After validation of the inundation susceptibility models, the inundation susceptibility maps for 2000, 2010, and 2020 were generated using SVM, CB, and EXT. The maps were compared to assess changes and how they were connected with climate change.

3.1. Support vector machines (SVMs)

SVMs are supervised machine learning models that can solve mathematical discrimination and regression problems. They were conceptualized in the 1990s from a statistical learning theory proposed by (Cortes and Vapnik 1995) . The models have the ability to work with high-dimensional data and achieve good results. Requiring only a small number of parameters, they are appreciated for their ease of use. The main operating principle of the SVM model is the search for a hyperplane (feature space) in which the data is separated into several classes whose boundary is as far as possible from the data points (or "maximum margin") (Tehrany et al. 2014). To achieve this, SVMs use kernels, i.e. mathematical functions to project and separate data in vector space, with "support vectors" being the data closest to the border. The furthest boundary of all the training points is optimal and therefore presents the best capacity for generalization (Choubin et al. 2019). The accuracy of the SVM model is characterized by two main hyper-parameters: gamma and C. The adjustment of the C parameter aims to eliminate outliers in the data. Increasing the value of C makes the SVM model able to choose the hyperplane to better separate the data points, while the gamma parameter determines the number of data points in the construction of the hyperplane. If the value of gamma is low, all data points far from the median will be used to calculate the median, and vice versa (Nguyen 2022a). In this study, using trial and error, the value of C and gamma were set at 0.1 and 0.5 respectively.

3.2. Catboost (CB)

HGS is a swarm-based optimizer algorithm, first introduced by Yang et al. (2021). This algorithm is inspired by the behavior of animals in a state of starvation. In order to find food and improve their survival, animals tend to cooperate with each other. Stronger animals have a greater heart capacity to obtain food than weak animals (Yang et al., 2021). In nature, animal behavior is influenced by many different factors, a primary one being hunger. When the food source is limited, it leads to competition between animals – a "hunger game." The HGS algorithm is divided into two stages: the first stage simulates the process of cooperation between animals to find a food source; the second step describes the animals' activities in a state of starvation (AbuShanab et al., 2021; Yang et al., 2021). The HGS algorithm has proven effective in the technical assessment and analysis of natural hazards (Nguyen, 2022b).

3.3. Extratrees (EXT)

EXT is a powerful algorithm developed by an unknown creator. It is used to address both classification and regression problems by using decision trees (Sachdeva and Kumar 2022). The EXT algorithm builds multiple decision trees from the original training sample, and each node in the tree selects the best feature from a random set of K features to split the data based on mathematical indices. Like the random forest algorithm, the EXT algorithm generates several decision trees by randomly sampling the data, leading to unique patterns in each tree.

Predictions are made through majority voting from the multiple decision trees (Heddam et al. 2020, Heddam 2023). The precision of the EXT model is influenced by the hyper-parameters max_depth and min_samples_split. In this study, the hyper-parameters were determined through trial-and-error to be max_depth=2 and min_samples_split=2.

3.4. Performance assessment

Assessing the fit of a model is a crucial step in model building. In this study, the ROC, AUC, RMSE, MAE, and R^2 indices were used to evaluate the fit of the model, which have been widely used in previous studies. The ROC curve illustrates the performance of a model by plotting the true positive rate on the Y-axis and false positive rate on the X-axis.

The AUC (area under the curve) ROC summarizes the overall performance of a classification model, with a value ranging from 0 to 1. A value of 1 indicates a perfect model, while 0.5 represents a non-informative model (Nachappa et al. 2020, Rehman et al. 2022).

RMSE and MAE give insight into the dispersion or variability of prediction accuracy by measuring the errors between the predicted and observed values (Chapi et al. 2017, Wang et al. 2019):

$$RMSE = \sqrt{\frac{1}{m}\sum_{i-1}^{m}(Y_{ipredicted} - Y_{iobserved})^2}$$

$$MAE = \frac{1}{m}\sum_{i-1}^{m}|Y_{ipredicted} - Y_{iobserved}|$$

where $Y_{iobserved}$ is the value of the i^{th} observation from the validation dataset and $Y_{ipredicted}$ is the predicted value for the i^{th} observation.

4. RESULTS AND DISCUSSIONS

4.1. Rainfall change between 2000 and 2020

Fig. 4 shows the annual precipitation in Vinh Phuc province between 2000 and 2020. The annual precipitation in 2000 ranged from 1542.38 to 1636.86mm, in 2010 from 1439.8 to 1647.18 mm, and in 2020 from 1748.64 to 2025.46 mm. During the period 2000–2010, precipitation increased in the northern mountainous region and decreased in the southern region. From 2010 to 2020, precipitation became stronger and was concentrated in the southern region.

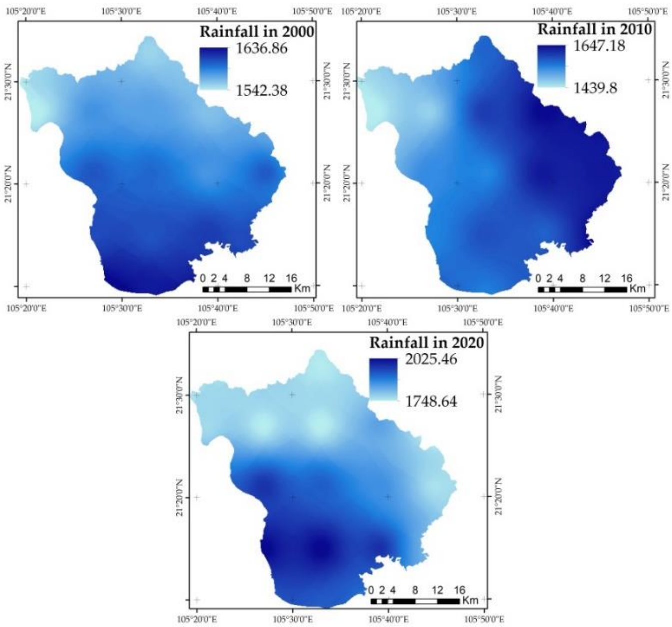


Fig. 4. Rainfall in Vinh Phuc province between 2000 and 2020.

4.2. Feature selection analysis

Assessing the importance of each conditioning factor is essential in the process of building a inundation susceptibility model. In this study the random forest technique was used to assess the importance of each factor. The results (**Fig. 5**) showed slope (0.3), altitude (0.265), TWI (0.15), and rainfall (0.11) to be the most important predictors of inundation occurrence, followed by NDVI (0.06), distance to river (0.05), curvature (0.025), soil type (0.02), distance to road (0.01), and NDBI (0.005). Aspect had no influence on the occurrence of inundation and so was eliminated from the inundation susceptibility model.

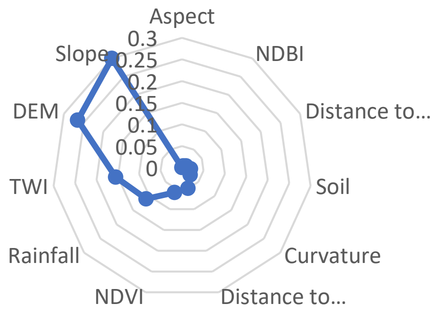
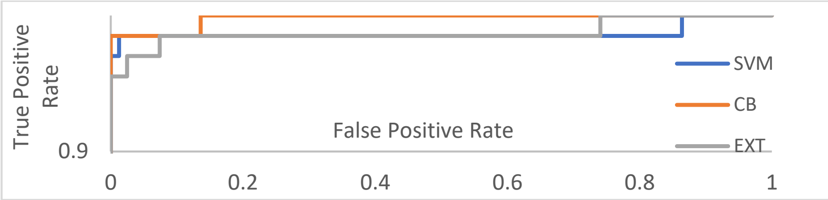


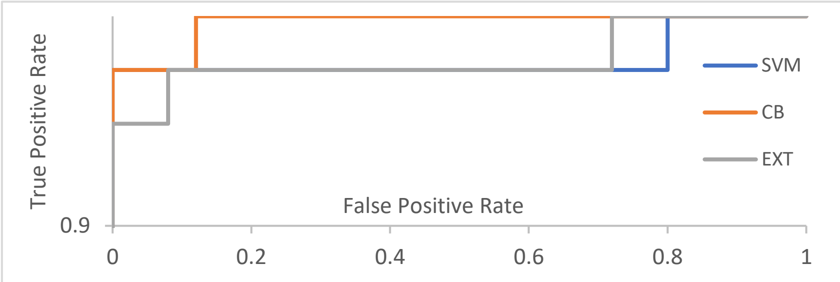
Fig. 5. Feature selection and ranking using random forest.

4.3. Model comparison

Ideally, a machine learning model should not be evaluated on the same dataset it uses for training, because the model will overfit the dataset and not perform better than the previous model. Therefore, for a model to be evaluated objectively, it must be evaluated on a test dataset that has not been used before. **Fig. 6** shows the AUC-ROC results for the training dataset and validating dataset. According to the AUC results for the training dataset, the CB model (AUC=0.99) outperformed SVM (AUC=0.98) and EXT (AUC=0.98). Similarly, for the validating dataset the AUC value for the CB model was 0.99, beating SVM (AUC=0.97) and EXT (AUC=0.97). Although all the proposed models returned results with high accuracy, the CB model was the best performing in the study area.



a)



b)

Fig. 6. The validation of three inundation susceptibility models using ROC based on training dataset (a) and validating dataset (b)

Table 1 presents the value of RMSE, MAE, and R^2 for the training dataset and validating dataset. For the CB model, the training dataset results for RMSE, MAE, and R^2 were 0.12, 0.05, and 0.95 respectively, outperforming SVM (RMSE=0.14, MAE=0.06, R^2=0.91) and EXT (RMSE=0.24 , MAE=0.21, R^2=0.75). For the validating dataset, for the CB model, the value of RMSE (0.17), MAE (0.07), and R^2 (0.96) was better than the other models, followed by SVM (RMSE=0.18, MAE=0.08 and R^2=0.96) and EXT (RMSE =0.26, MAE=0.22 and R^2=0.72), respectively.

Table 1.

The accuracy assessment of three inundation susceptibility models for the training dataset and the validating dataset.

	Training dataset				Validating dataset			
	RMSE	MAE	AUC	R^2	RMSE	MAE	AUC	R^2
SVM	0.14	0.06	0.98	0.91	0.18	0.08	0.97	0.96
CB	0.12	0.05	0.99	0.95	0.17	0.07	0.99	0.96
EXT	0.24	0.21	0.98	0.75	0.26	0.22	0.97	0.72

4.4. Inundation susceptibility prediction

After validation, the models were used to evaluate inundation susceptibility in 2000, 2010, and 2020. The models were fed pixels from the study area and the 11 conditioning factors for each of 2000, 2010, and 2020. Although there were small differences between each of the models' predictions, inundation was forecast to occur mainly in the southern region where there is low elevation and slope. The area of very low inundation susceptibility increased from 2000 to 2020, while the high and very high risk areas decreased in size.

For the SVM model, the very low area increased from 264.736 km² in 2000 to 345.537 km² in 2010 and 369.0107 km² in 2020. The low area decreased from 124.6562 km² in 2000 to 121.4723 km² in 2010 and increases to 125.3921 km² in 2020. Similarly, the moderate risk area decreased from 127.5294 km² in 2000 to 121.6532 km² in 2010 and then grew to 123.8686 km² in 2020. The high risk area decreased from 188.7407 km² in 2000 to 183.5609 km² in 2010 and again to 170.7813 km² in 2020. The very high risk area shrank from 514,3594 km² in 2000 to 447,712 km² in 2010 and then further to 430,8828 km² in 2020.

For the CB model, the very low inundation susceptibility area decreased from 414.1629 km² in 2000 to 411.6574 km² in 2010; the 2020 figure stayed almost the same at 411.689 km². The low risk area grew from 101.0637 km² in 2000 to 115.8995 km² in 2010 and decreased slightly to 114.0801 km^2 in 2020. The area of moderate inundation susceptibility decreased from 115.532 km^2 in 2000 to 105.312 km^2 in 2010 and again to 104.2839 km^2 in 2020. The high risk area increased from 125.194 km^2 in 2000 to 128.2546 km^2 in 2010 and again to 131.2367 km^2 in 2020. The very high susceptibility area shrank slightly from 464.0691 km^2 in 2000 to 458.812 km^2 in 2010 and then to 458.6458 km^2 in 2020.

For the EXT model, the very low area decreased from 235.4863 km^2 in 2000 to 232.4067 km^2 in 2010 and then increased significantly to 271.7133 km^2 in 2020. The low area grew from 171.2381 km^2 in 2000 to 177.6635 km^2 in 2010 and shrank slightly to 174.3416 km^2 in 2020. The moderate areas diminished from 225.0975 km^2 in 2000 to 191.1555 km^2 in 2010 and grew again to 213.1152 km^2 in 2020. The high inundation susceptibility area increased from 252.6806 km^2 in 2000 to 265.4405 km^2 in 2010 and decreased to 236.7902 km^2 in 2020. The very high-risk area decreased from 335.5191 km^2 in 2000 to 265.4405 km^2 in 2010 and then increased to 323.9735 km^2 in 2020 (**Table 2, Fig. 7 and 8**).

Table 2.

The distribution of five classes of inundation susceptibility zones as generated by SVM, CB, and EXT.

		Very low (Km^2)	Low (Km^2)	Moderate (Km^2)	High (Km^2)	Very high (Km^2)
SVM	2000	264.736	124.6562	127.5294	188.7407	514.3594
	2010	345.537	121.4723	121.6532	183.5609	447.712
	2020	369.0107	125.3921	123.8686	170.7813	430.8828
CB	2000	414.1629	101.0637	115.532	125.194	464.0691
	2010	411.6574	115.8995	105.312	128.2546	458.812
	2020	411.689	114.0801	104.2839	131.2367	458.6458
EXT	2000	235.4863	171.2381	225.0975	252.6806	335.5191
	2010	232.4067	177.6635	191.1555	265.4405	265.4405
	2020	271.7133	174.3416	213.1152	236.7902	323.9735

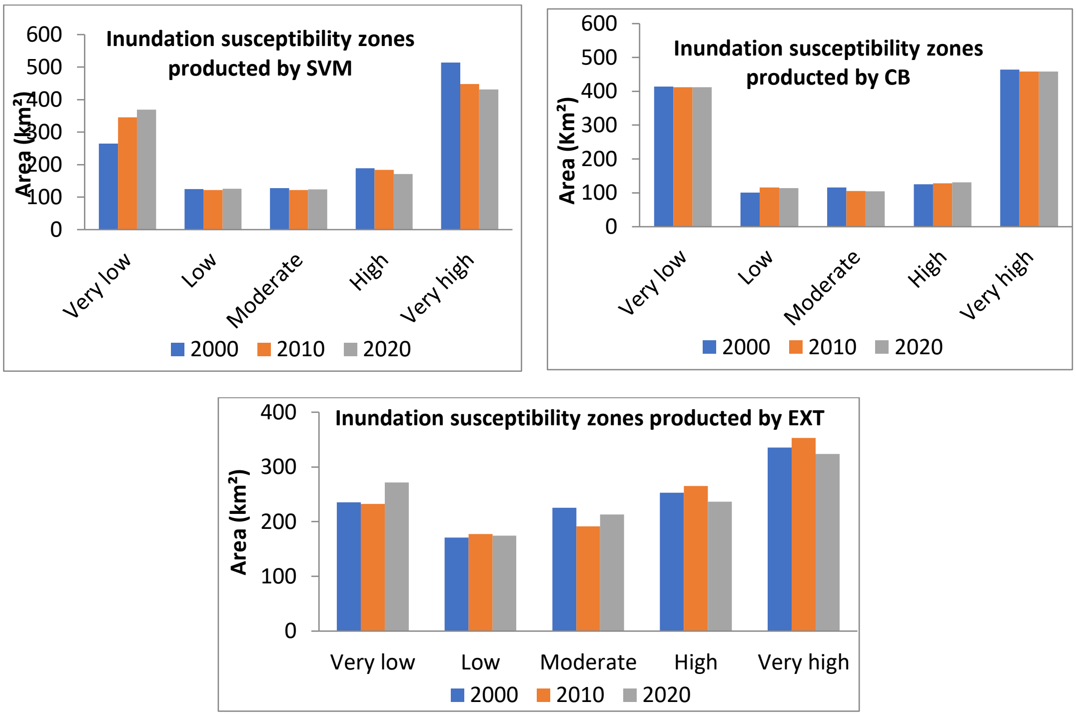
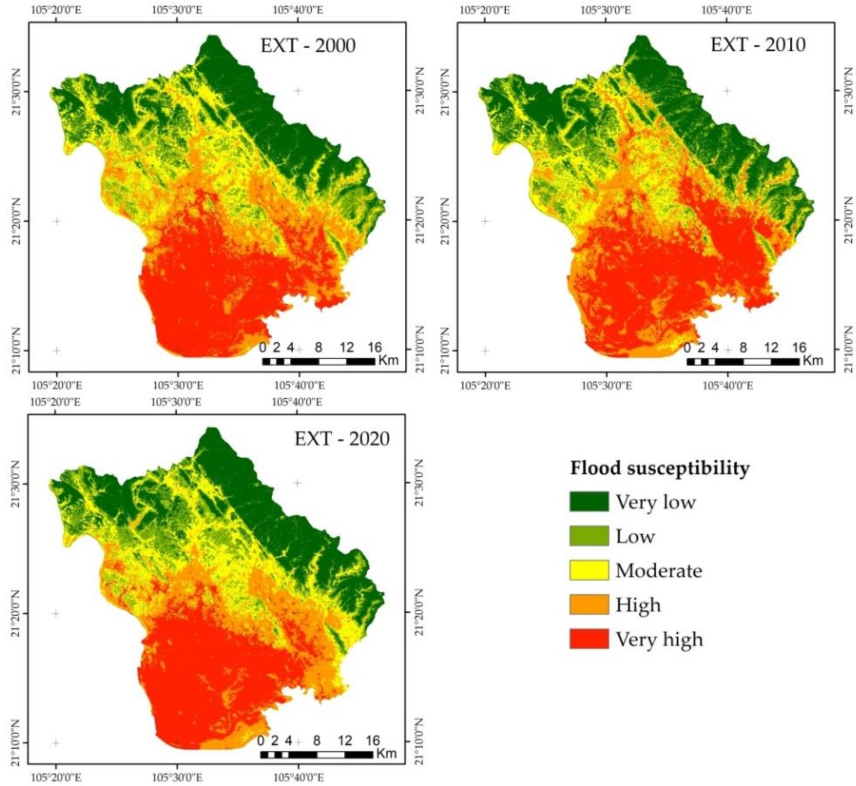
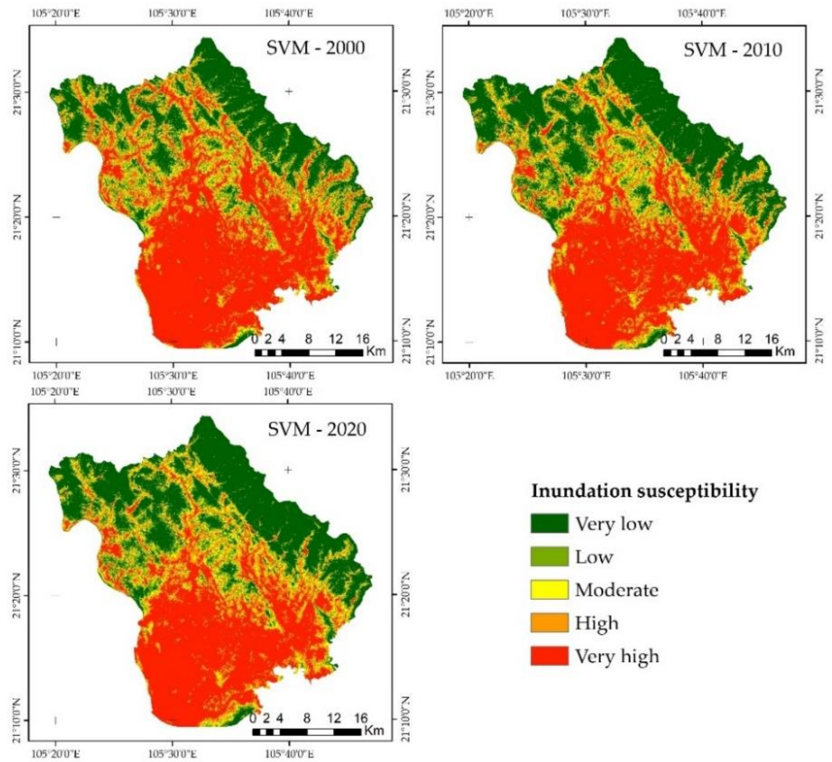


Fig.7. The inundation susceptibility zones under five classes produced by SVM, CB and EXT.


EXT - 2000
EXT - 2010
EXT - 2020
Flood susceptibility
Very low
Low
Moderate
High
Very high


SVM - 2000
SVM - 2010
SVM - 2020
Inundation susceptibility
Very low
Low
Moderate
High
Very high

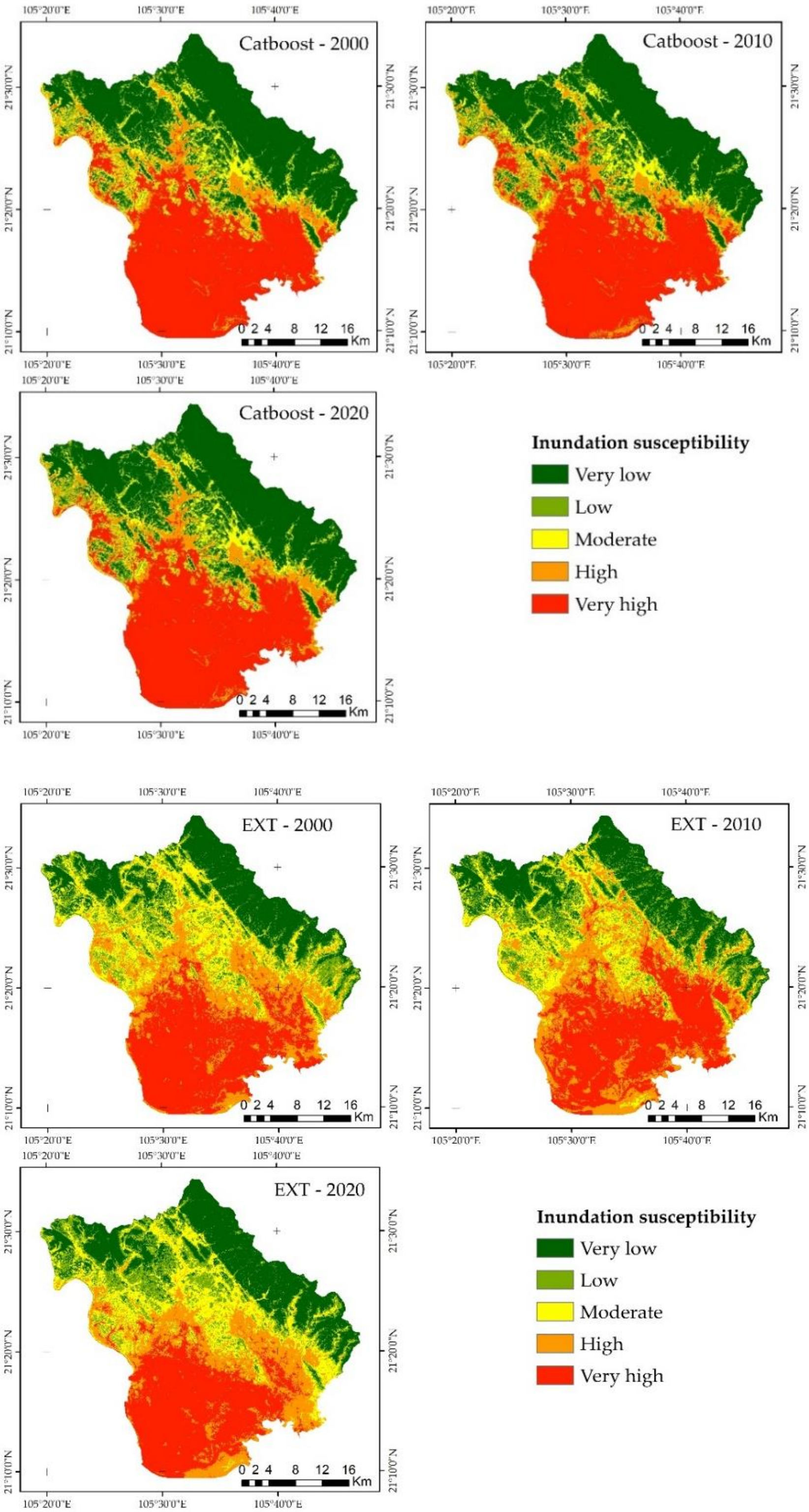


Fig. 8. Inundation susceptibility in 2000, 2010, and 2020 as generated by SVM, CB, and EXT.

5. DISCUSSION

5.1. The significance of results

The inundation is considered the most destructive natural hazard in the world, causing significant damage to human life and the economy of the country, especially in the coastal country. These damages are more and more serious in the last years due to climate change. So the understanding of effects of climate change on inundation susceptibility have received great attention from the global scientific community, particularly with a view to better support those tasked with designing appropriate strategies for inundation control in the context of climate change (Li et al. 2020, Nguyen et al. 2022a). This study develops a theoretical framework with which to assess the effects of climate change in general, and precipitation in particular, on inundation susceptibility in Vinh Phuc province, using machine learning and remote sensing. Although several previous studies have analyzed inundation susceptibility, most of these focused on the assessment of susceptibility for a specific time. Continuous inundation susceptibility assessment is necessary to consider more fully the effects of climate change on future flood or inundation. The results in our study indicate that the relationships between climate change and inundation play a key role in the strategy of flood risk management in the future, can be used by decision makers and local authorities in the construction of flood risk reduction strategies.

5.2. Inner validation of the results

The results support the first hypothesis in the introduction section that the inundation susceptibility exhibits strong relationships with natural and human factors as the slope, elevation, TWI and rainfall. For inundation susceptibility model development, studies have indicated that low quality of data has a negative effect on the quality of inundation susceptibility models and that, therefore, pre-processing of the data is critical (Du et al. 2023). In this study, random forest was used to assess the importance of each of 11 conditioning factors. It should be noted that the importance of conditioning factors depends on the natural and socio-economic condition of each region and the method used. Previous study, Islam et al. (2021) used information gain ratio to assess the importance of 10 conditioning factors.

The results highlighted that LULC, distance to road, and elevation were three most important factors in the construction of a model of inundation susceptibility for the Teesta sub-catchment of northern Bangladesh. Pham et al. (2021b) applied frequency ratio to rank 16 conditioning factors and the results showed that elevation, rainfall, and slope were most important for their model for the Vu Gia‑Thu Bon watershed in Vietnam. Huu Duy Nguyen (2022) used the random forest technique to analyze the importance of 13 conditioning factors and reported that LULC, elevation, and slope were most important for his map in Vietnam’s Ha Tinh province.

In this study, slope, elevation, TWI, and rainfall were the most important factors on the probability of inundation occurrence in Vinh Phuc province. Slope and elevation scored highest as they have critical effects on flow regime and flow velocity. TWI was ranked third as it describes the wetness of a given area and therefore the soil saturation capacity.

Rainfall was ranked fourth in importance. In Vietnam, there are three main types of inundation: fluvial, coastal, and pluvial. Coastal inundation does not exist in the study area because Vinh Phuc province is not influenced by the sea and the tide. Fluvial inundation is insignificant because the study area is not affected by major rivers. The province is often hit by heavy rainfall over a short period, which causes inundation. In recent years, inundations have also increased in intensity and number due to the reduction of surface vegetation, which influences the flow regime, and soil saturation capacity. This is why NDVI was ranked fifth of the factors featured.

Distance to river, curvature, soil type, distance to road, and NDBI were ranked from six to ten respectively, because the inundation in the study area mainly occurs in the low altitude area during the period of heavy rainfall. Heavy rainfall and low altitude are the main factors for the occurrence of inundation. Aspect did not influence the probability of inundation occurrence in the study area. It is worth noting that the models used in this study are driver-models, therefore, the importance of each of the conditioning factors depends on the statistical relationship between past events and conditioning factors. In this way, inundation in the study area can be seen mainly to depend on four factors: elevation, slope, TWI, and rainfall.

This study also confirms that all machine learning models were successful in assessing the effects of climate change in inundation susceptibility in Vinh Phuc province. Previous studies have demonstrated that model accuracy depends on model structure and the data characteristics of each region (Bui et al. 2020). Therefore, the selection of models is one of the most important steps. In this study, the models SVM, CB, and EXT were selected. Of the three, CB performed best. CB is considered to be one of the most powerful and precise machine learning tools as, unlike other machine learning tools, it does not need to run multiple trials to get good results. It provides optimum models from the first round (Lu et al. 2022). The SVM model was the second best performing. It not only performs well with large datasets, but also with small datasets. Moreover, SVM has the advantage of reducing dataset noise and has high generalizability (Deka 2014). EXT was third best performing. Although it has advantages such as generalizability and bias reduction, it struggles with nonlinear problem solving (Geurts et al. 2006).

5.3. The external validation of results

The results of this study were unexpected, because the initial hypothesis was that inundation susceptibility in the study area had increased from 2010 to 2020 due to the change in precipitation in the context of climate change. This trend is typical of several regions in the world: Nguyen et al. (2022a) pointed out that inundation susceptibility in Nhat Le‑Kien Giang had increased due to the change in precipitation in the context of climate change. Previous study, Chen et al. (2022) described an increase in inundation risk in Taiwan over the period 1979‑2099. Javari (2022) found inundation risk in Iran to have grown due to the increase in precipitation between 1975 and 2017. However, in this study, although the increase in precipitation from 2010 to 2020, however, in this period, good planning can reduce the effects of inundation.

5.4. The importance of results

The results of this study contribute important aspects of flood risk management strategies. The construction of the models in this study represents a high quality tool to assess the effects of climate change on floods or inundations. The models proposed in this study showed a more reliable performance than the previously cited studies. We can conclude that this is the first time that these models are used to build a inundation susceptibility map and assess the effects of climate change on inundation susceptibility. In addition, our results are important in answering the first question which has not been addressed in previous studies. Good land planning plays an essential role in reducing the flood risk. However, the success of putting in place good planning depends on determining the areas prone to flood and the impact of climate change. This can be achieved by understanding inundation susceptibility in a complex way in the context of climate change, as well as limiting new construction in areas with high or very high inundation susceptibility.

5.5. Limitation and Future Research

This study presents the general limitation related to the use of the data. This study uses Landsat 08 with the resolution of 30m to extract the NDVI and NDBI in 2020. However, several studies have pointed out that the Sentinel 2A can present the information in more detail. In addition, non-flood or non-inundation points were selected in the never-flooded area. However, given the context of climate change and the non-linear nature of the flood or inundation phenomenon, this leads to uncertainties in the selection of these points. Ultimately, flood intensity is strongly influenced by climate change scenarios and land use. Therefore, it is necessary to evaluate the scenarios of this change on floods in future studies. These results may be important for policy makers and local authorities to promote sustainable land use planning in the future.

6. CONCLUSIONS

Every year, the flood causes major damage to human life and the country's economy in the world. With the geographical location, Vietnam is considered a country most affected by this natural hazard. Among the provinces, Vinh Phuc province is one of the provinces most damaged by inundation, especially in the context of climate change. This study assesses the impacts of climate change in general and rainfall change in particular on inundation susceptibility in Vinh Phuc province over the period 2000‑2020, using machine learning (SVM, CB, and EXT) and remote sensing. The results of this study can help managers and planners understand the causes of changes in inundation susceptibility and thus develop appropriate strategies and policies to reduce future inundation damage.

All the models proposed in this study used good fit to assess the effects of climate change and rainfall on inundation susceptibility from 2000 to 2020. All three achieved AUC of over 0.95. Of the three, CB was most accurate (AUC=0.99). SVM and EXT both scored AUC=0.98. The models proposed in this study can be used to assess the effects of climate change on inundation susceptibility in other regions of the world, particularly in regions with limited data.

In general, the size of the areas of very low and low inundation susceptibility increased in from 2000 to 2020, while the high and very high risk areas decreased, due to changes in rainfall, vegetation, and infrastructure.

With the development of remote sensing data covering the entire planet, the approach used in this study can be easily applied to assess the effects of climate change on other natural hazard in different regions of the world. The models proposed in this study provide good results at large scales, which can help decision-makers and local authorities to determine the regions likely to be affected by climate change in large-scale.

The findings of this study may support managers and planners to implement appropriate measures in areas with high and very high inundation susceptibility, to reduce damage to human life and property. Moreover, the theoretical framework in this study highlights the effects of the change in rainfall on inundation susceptibility.

REFERENCES

Al-Juaidi, A.E., Nassar, A.M. & Al-Juaidi, O.E., 2018. Evaluation of flood susceptibility mapping using logistic regression and gis conditioning factors. Arabian Journal of Geosciences, 11, 1-10.

Arabameri, A., Seyed Danesh, A., Santosh, M., Cerda, A., Chandra Pal, S., Ghorbanzadeh, O., Roy, P. & Chowdhuri, I., 2022. Flood susceptibility mapping using meta-heuristic algorithms. Geomatics, Natural Hazards and Risk, 13 (1), 949-974.

Avand, M. & Moradi, H., 2021. Using machine learning models, remote sensing, and gis to investigate the effects of changing climates and land uses on flood probability. Journal of Hydrology, 595, 125663.

Band, S.S., Janizadeh, S., Chandra Pal, S., Saha, A., Chakrabortty, R., Melesse, A.M. & Mosavi, A., 2020. Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. Remote Sensing, 12 (21), 3568.

Bui, D.T., Ngo, P.-T.T., Pham, T.D., Jaafari, A., Minh, N.Q., Hoa, P.V. & Samui, P., 2019. A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. Catena, 179, 184-196.

Bui, Q.-T., Nguyen, Q.-H., Nguyen, X.L., Pham, V.D., Nguyen, H.D. & Pham, V.-M., 2020. Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. Journal of Hydrology, 581, 124379.

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.

Chen, Y.-J., Lin, H.-J., Liou, J.-J., Cheng, C.-T. & Chen, Y.-M., 2022. Assessment of flood risk map under climate change rcp8. 5 scenarios in taiwan. Water, 14 (2), 207.

Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F. & Mosavi, A., 2019. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Science of the Total Environment, 651, 2087-2096.

Chowdhuri, I., Pal, S.C. & Chakrabortty, R., 2020. Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern india. Advances in Space Research, 65 (5), 1466-1489.

Cortes, C. & Vapnik, V., 1995. Support vector machine. Machine learning, 20 (3), 273-297.

Dahri, N., Yousfi, R., Bouamrane, A., Abida, H., Pham, Q.B. & Derdous, O., 2022. Comparison of analytic network process and artificial neural network models for flash flood susceptibility assessment. Journal of African Earth Sciences, 193, 104576.

Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. Applied soft computing, 19, 372-386.

Dorogush, A.V., Ershov, V. & Gulin, A., 2018. Catboost: Gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

Du, Q.V.V., Nguyen, H.D., Pham, V.T., Nguyen, C.H., Nguyen, Q.-H., Bui, Q.-T., Doan, T.T., Tran, A.T. & Petrisor, A.-I., 2023. Deep learning to assess the effects of land use/land cover and climate change on landslide susceptibility in the tra khuc river bassin of vietnam. Geocarto International, (just-accepted), 1-40.

Geurts, P., Ernst, D. & Wehenkel, L., 2006. Extremely randomized trees. Machine learning, 63, 3-42.

Ghosh, A., Dey, P. & Ghosh, T., 2022a. Integration of rs-gis with frequency ratio, fuzzy logic, logistic regression and decision tree models for flood susceptibility prediction in lower gangetic plain: A study on malda district of west bengal, india. Journal of the Indian Society of Remote Sensing, 50 (9), 1725-1745.

Ghosh, S., Saha, S. & Bera, B., 2022b. Flood susceptibility zonation using advanced ensemble machine learning models within himalayan foreland basin. Natural Hazards Research, 2 (4), 363-374 Available from: https://www.sciencedirect.com/science/article/pii/S2666592122000245.

Ghosh, S., Saha, S. & Bera, B., 2022c. Flood susceptibility zonation using advanced ensemble machine learning models within himalayan foreland basin. Natural Hazards Research.

Hai Ly, N., Nguyen, H.D., Loubiere, P., Van Tran, T., Șerban, G., Zelenakova, M., Brețcan, P. & Laffly, D., 2022. The composition of time-series images and using the technique smote enn for balancing datasets in land use/cover mapping. Acta Montanistica Slovaca, 27, 2.

Heddam, S., 2023. Extremely randomized trees versus random forest, group method of data handling, and artificial neural network. Handbook of hydroinformatics. Elsevier, 291-304.

Heddam, S., Ptak, M. & Zhu, S., 2020. Modelling of daily lake surface water temperature from air temperature: Extremely randomized trees (ert) versus air2water, mars, m5tree, rf and mlpnn. Journal of Hydrology, 588, 125130.

Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.-X. & Chen, W., 2018. Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of poyang county, china. Science of the total environment, 625, 575-588.

Islam, A.R.M.T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K.U., Pham, Q.B., Kuriqi, A. & Linh, N.T.T., 2021. Flood susceptibility modelling using advanced ensemble machine learning models. Geoscience Frontiers, 12 (3), 101075.

Janizadeh, S., Chandra Pal, S., Saha, A., Chowdhuri, I., Ahmadi, K., Mirzaei, S., Mosavi, A.H. & Tiefenbacher, J.P., 2021a. Mapping the spatial and temporal variability of flood hazard affected by climate and land-use changes in the future. Journal of Environmental Management, 298, 113551 Available from: https://www.sciencedirect.com/science/article/pii/S0301479721016133.

Janizadeh, S., Pal, S.C., Saha, A., Chowdhuri, I., Ahmadi, K., Mirzaei, S., Mosavi, A.H. & Tiefenbacher, J.P., 2021b. Mapping the spatial and temporal variability of flood hazard affected by climate and land-use changes in the future. Journal of Environmental Management, 298, 113551.

Javari, M., 2022. Rainfall random variability and its effects on flood risk management in iran. Modeling Earth Systems and Environment, 8 (1), 1109-1133.

Jodar-Abellan, A., Valdes-Abellan, J., Pla, C. & Gomariz-Castillo, F., 2019. Impact of land use changes on flash flood prediction using a sub-daily swat model in five mediterranean ungauged watersheds (se spain). Science of the Total Environment, 657, 1578-1591.

Khosravi, K., Pham, B.T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., Prakash, I. & Bui, D.T., 2018. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at haraz watershed, northern iran. Science of the Total Environment, 627, 744-755.

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.

Li, S., Wang, Z., Lai, C. & Lin, G., 2020. Quantitative assessment of the relative impacts of climate change and human activity on flood susceptibility based on a cloud model. Journal of hydrology, 588, 125051.

Li, Z., Song, K. & Peng, L., 2021. Flood risk assessment under land use and climate change in wuhan city of the yangtze river basin, china. Land, 10 (8), 878.

Lim, J. & Lee, K.-S., 2018. Flood mapping using multi-source remotely sensed data and logistic regression in the heterogeneous mountainous regions in north korea. Remote Sensing, 10 (7), 1036.

Linh, N.T.T., Pandey, M., Janizadeh, S., Bhunia, G.S., Norouzi, A., Ali, S., Pham, Q.B., Anh, D.T. & Ahmadi, K., 2022. Flood susceptibility modeling based on new hybrid intelligence model: Optimization of xgboost model using ga metaheuristic algorithm. Advances in Space Research, 69 (9), 3301-3318.

Liu, Y., Xu, Y., Zhao, Y. & Long, Y., 2022. Using swat model to assess the impacts of land use and climate changes on flood in the upper weihe river, china. Water, 14 (13), 2098.

Lu, C., Zhang, S., Xue, D., Xiao, F. & Liu, C., 2022. Improved estimation of coalbed methane content using the revised estimate of depth and catboost algorithm: A case study from southern sichuan basin, china. Computers & Geosciences, 158, 104973.

Mohammady, M., Pourghasemi, H.R. & Amiri, M., 2019. Assessment of land subsidence susceptibility in semnan plain (iran): A comparison of support vector machine and weights of evidence data mining algorithms. Natural Hazards, 99, 951-971.

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.

Nguyen, H.D., 2022a. Flood susceptibility assessment using hybrid machine learning and remote sensing in quang tri province, vietnam. Transactions in GIS, 26 (7), 2776-2801.

Nguyen, H.D., 2022b. Gis-based hybrid machine learning for flood susceptibility prediction in the nhat le-kien giang watershed, vietnam. Earth Science Informatics, 1-18.

Nguyen, H.D., Dang, D.K., Nguyen, Q.-H., Bui, Q.-T. & Petrisor, A.-I., 2022a. Evaluating the effects of climate and land use change on the future flood susceptibility in the central region of vietnam by integrating land change modeler, machine learning methods. Geocarto International, 1-36.

Nguyen, H.D., Nguyen, Q.-H., Du, Q.V.V., Nguyen, T.H.T., Nguyen, T.G. & Bui, Q.-T., 2021. A novel combination of deep neural network and manta ray foraging optimization for flood susceptibility mapping in quang ngai province, vietnam. Geocarto International, 1-25.

Nguyen, H.D., Quang-Thanh, B., Nguyen, Q.-H., Nguyen, T.G., Pham, L.T., Nguyen, X.L., Vu, P.L., Thanh Nguyen, T.H., Nguyen, A.T. & Petrisor, A.-I., 2022b. A novel hybrid approach to flood susceptibility assessment based on machine learning and land use change. Case study: A river watershed in vietnam. Hydrological Sciences Journal, 67 (7), 1065-1083.

Nguyen, H.D., Trinh, T.T.H. & Dang, D.K., 2023. Flood hazard and resilience in the watershed nhat le-kien giang in vietnam. Urbanism. Architecture. Constructions/Urbanism. Arhitectura. Constructii, 14 (1).

Ongdas, N., Akiyanova, F., Karakulov, Y., Muratbayeva, A. & Zinabdin, N., 2020. Application of hec-ras (2d) for flood hazard maps generation for yesil (ishim) river in kazakhstan. Water, 12 (10), 2672.

Penning-Rowsell, E., Yanyan, W., Watkinson, A., Jiang, J. & Thorne, C., 2013. Socioeconomic scenarios and flood damage assessment methodologies for the t aihu b asin, c hina. Journal of Flood Risk Management, 6 (1), 23-32.

Petrișor, A.-I., Sirodoev, I. & Ianoș, I., 2020. Trends in the national and regional transitional dynamics of land cover and use changes in romania. Remote Sensing, 12 (2), 230.

Pham, B.T., Jaafari, A., Van Phong, T., Yen, H.P.H., Tuyen, T.T., Van Luong, V., Nguyen, H.D., Van Le, H. & Foong, L.K., 2021a. Improved flood susceptibility mapping using a best first decision tree integrated with ensemble learning techniques. Geoscience Frontiers, 12 (3), 101105.

Pham, B.T., Luu, C., Van Phong, T., Trinh, P.T., Shirzadi, A., Renoud, S., Asadi, S., Van Le, H., Von Meding, J. & Clague, J.J., 2021b. Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? Journal of hydrology, 592, 125615.

Pourghasemi, H.R., Kariminejad, N., Amiri, M., Edalat, M., Zarafshar, M., Blaschke, T. & Cerda, A., 2020. Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. Scientific reports, 10 (1), 3203.

Priscillia, S., Schillaci, C. & Lipani, A., 2021. Flood susceptibility assessment using artificial neural networks in indonesia. Artificial Intelligence in Geosciences, 2, 215-222.

Psomiadis, E., Tomanis, L., Kavvadias, A., Soulis, K.X., Charizopoulos, N. & Michas, S., 2021. Potential dam breach analysis and flood wave risk assessment using hec-ras and remote sensing data: A multicriteria approach. Water, 13 (3), 364.

Rajkhowa, S. & Sarma, J., 2021. Climate change and flood risk, global climate change. Global climate change. Elsevier, 321-339.

Rehman, S., Hasan, M.S.U., Rai, A.K., Rahaman, M.H., Avtar, R. & Sajjad, H., 2022. Integrated approach for spatial flood susceptibility assessment in bhagirathi sub-basin, india using entropy information theory and geospatial technology. Risk Analysis.

Roy, P., Pal, S.C., Chakrabortty, R., Chowdhuri, I., Malik, S. & Das, B., 2020. Threats of climate and land use change on future flood susceptibility. Journal of Cleaner Production, 272, 122757.

Saber, M., Boulmaiz, T., Guermoui, M., Abdrabo, K.I., Kantoush, S.A., Sumi, T., Boutaghane, H., Nohara, D. & Mabrouk, E., 2021. Examining lightgbm and catboost models for wadi flash flood susceptibility prediction. Geocarto International, 1-26.

Sachdeva, S. & Kumar, B., 2022. Flood susceptibility mapping using extremely randomized trees for assam 2020 floods. Ecological Informatics, 67, 101498.

Saha, T.K., Pal, S., Talukdar, S., Debanshi, S., Khatun, R., Singha, P. & Mandal, I., 2021. How far spatial resolution affects the ensemble machine learning based flood susceptibility prediction in data sparse region. Journal of Environmental Management, 297, 113344.

Sahana, M. & Patel, P.P., 2019. A comparison of frequency ratio and fuzzy logic models for flood susceptibility assessment of the lower kosi river basin in india. Environmental Earth Sciences, 78, 1-27.

Swain, K.C., Singha, C. & Nayak, L., 2020. Flood susceptibility mapping through the gis-ahp technique using the cloud. ISPRS International Journal of Geo-Information, 9 (12), 720.

Talukdar, S., Ghose, B., Salam, R., Mahato, S., Pham, Q.B., Linh, N.T.T., Costache, R. & Avand, M., 2020. Flood susceptibility modeling in teesta river basin, bangladesh using novel ensembles of bagging algorithms. Stochastic Environmental Research and Risk Assessment, 34, 2277-2300.

Tansar, H., Babur, M. & Karnchanapaiboon, S.L., 2020. Flood inundation modeling and hazard assessment in lower ping river basin using mike flood. Arabian Journal of Geosciences, 13, 1-16.

Tehrany, M.S., Pradhan, B. & Jebur, M.N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in gis. Journal of hydrology, 512, 332-343.

Tien Bui, D., Ho, T.-C., Pradhan, B., Pham, B.-T., Nhu, V.-H. & Revhaug, I., 2016. Gis-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with adaboost, bagging, and multiboost ensemble frameworks. Environmental Earth Sciences, 75, 1-22.

Wang, X. & Xie, H., 2018. A review on applications of remote sensing and geographic information systems (gis) in water resources and flood risk management. Water, 10 (5), 608.

Wang, Y., Hong, H., Chen, W., Li, S., Panahi, M., Khosravi, K., Shirzadi, A., Shahabi, H., Panahi, S. & Costache, R., 2019. Flood susceptibility mapping in dingnan county (china) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. Journal of environmental management, 247, 712-729.

Yaseen, A., Lu, J. & Chen, X., 2022. Flood susceptibility mapping in an arid region of pakistan through ensemble machine learning model. Stochastic Environmental Research and Risk Assessment, 36 (10), 3041-3061.