Vietnam Journal of Earth Sciences, 45(4), 456-478, https://doi.org/10.15625/2615-9783/18644

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

A framework for flood depth using hydrodynamic modeling and machine learning in the coastal province of Vietnam

Huu Duy Nguyen[1*], Dinh Kha Dang[1], Y Nhu Nguyen[1], Chien Pham Van[2], Quang-Hai Truong[3], Quang-Thanh Bui[1], Alexandru-Ionut Petrisor[4,5,6,7]

[1]*VNU University of Science, Vietnam National University, Hanoi, Vietnam*
[2]*Thuy Loi University, 175 Tay Son, Dong Da, Hanoi, Vietnam*
[3]*Institute of Vietnamese Studies & Development Sciences, Vietnam National University (VNU), Hanoi 10000, Vietnam*
[4]*Ion Mincu University of Architecture and Urbanism, 10014 Bucharest, Romania*
[5]*Technical University of Moldova, Faculty of Urbanism and Architecture, Department of Architecture, 2004 Chisinau, Moldova*
[6]*National Institute for Research and Development in Tourism, 50741 Bucharest, Romania*
[7]*National Institute for Research and Development in Constructions, Urbanism and Sustainable Spatial Development URBAN-INCERC, Bucharest, Romania, 021652 Bucharest, Romania*

Received 10 April 2023; Received in revised form 27 July 2023; Accepted 10 August 2023

ABSTRACT

Flood models based on traditional hydrodynamic modeling encounter significant difficulties with real-time predictions, require enormous computational resources, and perform poorly in data-limited regions. The difficulties are compounded as flooding worldwide worsens due to the increasing frequency of short-term torrential rain events, making it more challenging to predict floods over the long term. This study aims to address these challenges by developing a rapid flood forecasting model combining machine learning algorithms (support vector regression, XGBoost regression, CatBoost regression, and decision tree regression) with hydrodynamic modeling in Quang Tri province in Vietnam. 560 flood depth locations were obtained by hydrodynamic modeling, and several locations measured in the field were used as input data for the machine learning models to build a flood depth map for the study area. The statistical indices used to evaluate the performance of the four proposed models were the receiver operating characteristic (ROC) curve, area under the ROC curve, root mean square error, mean absolute error, and coefficient of determination (R^2). The results showed that all four models successfully constructed a flood depth map for the study area. Among the four proposed models, CatBoost regression performed best, with an R^2 value of 0.86. This was followed by XGBoost regression (R^2=0.84), decision tree regression (R^2=0.72), and then support vector regression (R^2=0.7). This integration of hydrodynamic modeling and machine learning complements the framework in much of the existing literature. It can provide decision-makers and local authorities with an advanced flood warning tool and contribute to improving sustainable development strategies in this and similar regions.

Keywords: Flood depth, machine learning, hydrodynamics, Quang Tri, Vietnam.

1. Introduction

Flood is one of the most common natural disasters and, every year, causes significant damage to economies, injury, and loss of life (Hens et al., 2018; Shafizadeh-Moghadam et al., 2018; Nguyen, 2022). According to EM-DAT data, approximately 175,000 people have died, and 2.2 billion have been affected

*Corresponding author, Email: nguyenhuuduy@hus.edu.vn

by floods over the past 27 years (Nevo et al., 2021). Due to both urban growth and climate change, extreme precipitation events occurring over a short period have become more frequent over recent years, increasing the flood risk (Chakrabortty et al., 2021; Luu et al., 2022; Nguyen et al., 2022). This particularly impacts developing countries, where mitigation measures are often inadequate or low-quality. Knowledge of flood depth is essential in reducing the damage caused by floods.

Although most parts of the world are affected by floods, Asian countries are particularly vulnerable due to the high population density in flood-prone areas (Doocy et al., 2013; Luu et al., 2021). Vietnam is most affected due to its location in a humid, tropical climate zone. Vietnam's central plains are particularly high-risk due to heavy rainfall and high population density (Luu et al., 2021; Nguyen, 2022). Engineering measures to reduce the problem is a significant challenge, with many mitigation measures prohibitively expensive (Hall et al., 2003; Johnson and Priest, 2008; Nguyen et al., 2022). Therefore, efforts are being made to determine flood depth to help decision-makers in the sustainable management of land use.

The literature shows that researchers determining flood depth encounter several challenges. For example, although hydrodynamic models such as MIKE FLOOD (Kadam and Sen, 2012; Mani et al., 2014; Tran et al., 2019), SWAT (Vilaysane et al., 2015; Farooq et al., 2019), and HEC-RAS (Namara et al., 2022) have been proven to be effective, their application is limited to large areas, due to the demand for detailed meteorological, hydrological, and topographical data. Hydrodynamic models are also not appropriate for application in urgent situations or for repetitive tasks because of the time and cost of calculation. Van den Honert and McAneney (2011) claimed that the failure of the hydrodynamic model to determine flood levels has been one of the causes of significant floods in Queensland, Australia. Most researchers have recently used remote sensing to assess the flood. Although this rapidly developing method can effectively identify floodplains, it cannot determine the causes and flood level. More robust methods are needed to determine flood levels over a wide area.

As progress in computer science accelerates, researchers have increasingly employed data-driven models based on machine learning to assess flood susceptibility (Islam et al., 202; Nguyen et al., 2022). These have been developed based on an analysis of the relationships between past flood events and conditioning factors and have several advantages: (i) these models can better solve nonlinear structures with high accuracy, (ii) these models perform better in data-limited regions, and iii) these models can easily integrate with traditional models such as remote sensing and hydrodynamic modeling (Nguyen 2022; Gharakhanlou and Perez, 2023; Youssef et al., 2023). The selection of the appropriate algorithm for a given task from hundreds of possible candidates represents an enduring challenge for the scientific community. Some machine learning models that have been applied to assess flood, landslides, forest fire, and groundwater include support vector machines (Tehrany et al., 2014; Tehrany et al., 2015), random forest (Nachappa et al., 2020; Abu El-Magd 2022), bagging (Talukdar et al., 2020), adaboost (Nguyen 2022), and artificial neural networks (Falah et al., 2019; Costache et al., 2020). So far, there has been no established use of algorithms to assess flood depth. The provision of accurate flood depth data aids decision-makers in building appropriate strategies in the case both of emergencies and long-term sustainable land-use planning. (Nguyen et al., 2022) Pointed out that determining flood depth is considered an important task in developing new infrastructure. (Hosseiny et al., 2020)

Reported that the identification of flood area and flood depth with high accuracy is an essential step in flood characteristic analysis to support decision-makers in sustainable land-use planning. (Ghanbari et al., 2021) Showed that knowledge of flood depth is crucial for planning adaptation measures to flood risks in urban areas, especially in coastal regions vulnerable to storms and floods.

This study aims to develop novel hybrid models by integrating machine learning with hydrodynamic modeling to estimate flood extent and flood depth in Quang Tri province of Vietnam. Quang Tri province has the general characteristics of the littoral province in Vietnam. The terrain is lower from west to east, with the Truong Son range to the west, followed by hills, coastal plains, and sand dunes. The rainy season in the region extends mainly from September to December.

Existing hydrodynamic modeling is well-developed and can be used as input data for the machine learning model. The integration between machine learning and hydrodynamic modeling can improve the quality of predictions. The methodology was explicitly developed: (i) to use hydrodynamic models to determine the flood depth in the Thach Han River watershed and (ii) to use machine learning to determine the flood depth for the entire province of Quang Tri. The framework proposed in this study is a new approach to describing flood characteristics that allow large-scale, efficient, and inexpensive assessment. This is the first these models are developed and applied in a Vietnam province where floods are often affected. The results of this study can be integrated into hydrological forecast models for real-time flood analysis with different scenarios.

2. Materials and methods

2.1. Study area

Quang Tri province is located in the Central region of Vietnam, at 16°45'-17°30′N, 106°10′-107°12′E. It covers an area of 4740 km² with a population of approximately 630,000 (Fig. 1).

The topography of Quang Tri is very complex and diverse but can be divided into three main geographical areas: mountains, midlands, and plains. The mountainous zone forms part of the Truong Son range and is located in the west of the province, and accounts for about 50% of the province's area. Its altitude ranges from 250 to 1700 m. The midlands have an altitude of between 50 and 250 m, representing about 30% of the province's area. The plains are mainly distributed along the coast and have an elevation between 0 and 30 m; they account for about 20% of the study area.

Quang Tri Province has a dense river system, with an average density of 0.8-1.0 km/km^2. Ben Hai, Thach Han, and O Lau are three central river systems.

Located in the tropical monsoon zone, the study area experiences a climate of two distinct seasons. The dry season begins in March with hot and dry wind from the southwest, causing an increase in temperature and decreased humidity until August. The rainy season runs from September to February and is influenced by the northeast monsoon, which causes a drop in temperature and is accompanied by storms and floods. The average annual rainfall in the province is 1900-2500 mm. The heaviest rainfall is concentrated between September and December, accounting for 65-75% of annual precipitation. Heavy rains concentrated within a short period have caused significant flooding in Quang Tri province, especially in 1999, 2010, and 2020.

Quang Tri Province has reduced damage to human life and property through flood prevention measures such as a levee system over 180 km long, 157 pumping stations, and 131 reservoirs.

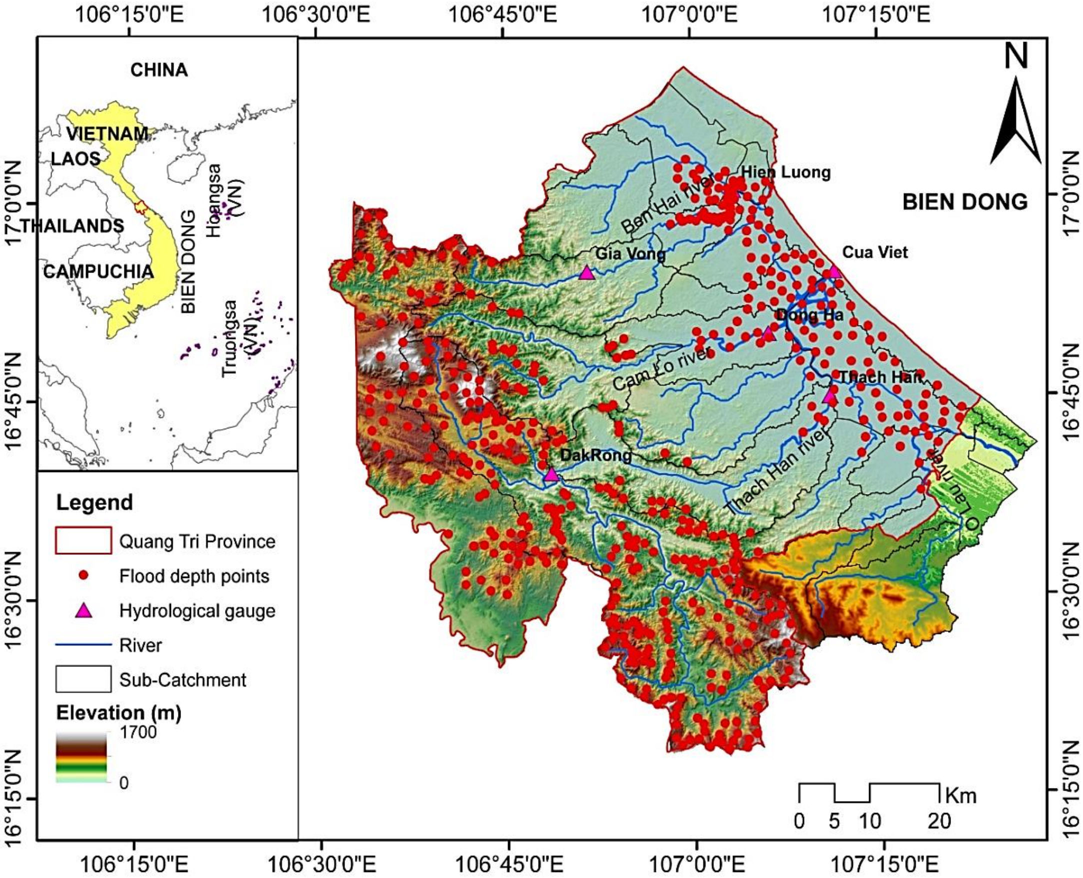

Figure 1. Location of Quang Tri province

2.2. Flood inventory

Preparing a flood inventory is an essential step in evaluating the characteristics of floods using machine learning. In this study, 150 flood marks in the flood event in 1999 and 2005 were collected on a field mission in 2022. These points were measured using a geodimeter, including coordinates and flood depth. We used hydrodynamic modeling to add further flood points to consolidate the data. Several authors have pointed out that developing a hydrodynamic model with high accuracy can be an excellent solution to complete the input data for the machine learning model. Ultimately, 380 flood depth points were added from the hydrodynamic model under Mike Flood.

2.2.1. Hydrodynamic model

A coupling between one-dimensional section averaged (1D), and two-dimensional depth-averaged (2D) models were implemented to simulate flood depth in the study area. The 1D model solves the Saint-Venant equations consisting of continuous and momentum equations using an implicit finite difference scheme developed by (Abbott and Ionescu, 1967). In the 2D model, shallow-water equations (well-known governing equations) are solved using the cell-centered finite-volume method for spatial deviations and a second-order Runge Kutta or an explicit Euler method for temporal integrations. The computational domain in the 2D model is discretized by the subdivision of the continuum into non-overlapping triangle elements. An approximate Riemann solver was also used to compute the convective fluxes, making it possible to handle discontinuous solutions between the edges of elements in the computational grid (Tansar et al., 2020). The coupling between 1D and 2D models allows representation of the domain of interest in the model as much as it enables the simulation of continuous interaction and lateral transfer flow processes between rivers and tributaries and adjacent floodplains.

In terms of numerical setup and implementation of the model for the domain of interest, the 1D model was used to represent the three river systems: Ben Hai, Thach Han, and O Lau. In the Ben Hai system, the area studied included the main channel of the river (from Gia Vong gauging station to the river mouth, a length of 23.4 km) and the Sa Lung branch (from Sa Lung dam to the confluence with Ben Hai River near Hien Luong Bridge, a length of 15.7 km.

In the Thach Han River system, the model included the Thach Han River (from the confluence of the Dakrong River and Rao Quan stream to Tram Dam and up to the East Sea at Cua Viet, a length of about 77 km) and the Hieu or Cam Lo River (from the Cam Tuyen bridge to the confluence with the Thach Han River at Gia Do, near Dong Ha city, a length of about 20.3 km). In the O Lau system, the main river, from Pho Trach Bridge to Tam Giang Lagoon in Cua Lac (a length of about 31.8 km), was modeled.

The adjacent floodplains along rivers and channels of the three river systems above were represented in the 2D model using an unstructured triangle grid (with a total of 78,234 cells or triangles and 39,772 nodes). Note that the triangles and cells ranged between 150 and 200 m in length. The elevation at each grid node was interpolated based on a digitized 1/10,000 map. Figure 2 is a simple diagram of the multiple rivers and channels represented in the 1D model, Fig. 3 presents the calculation domains for the 2D model and coupling between the 1D and 2D models.

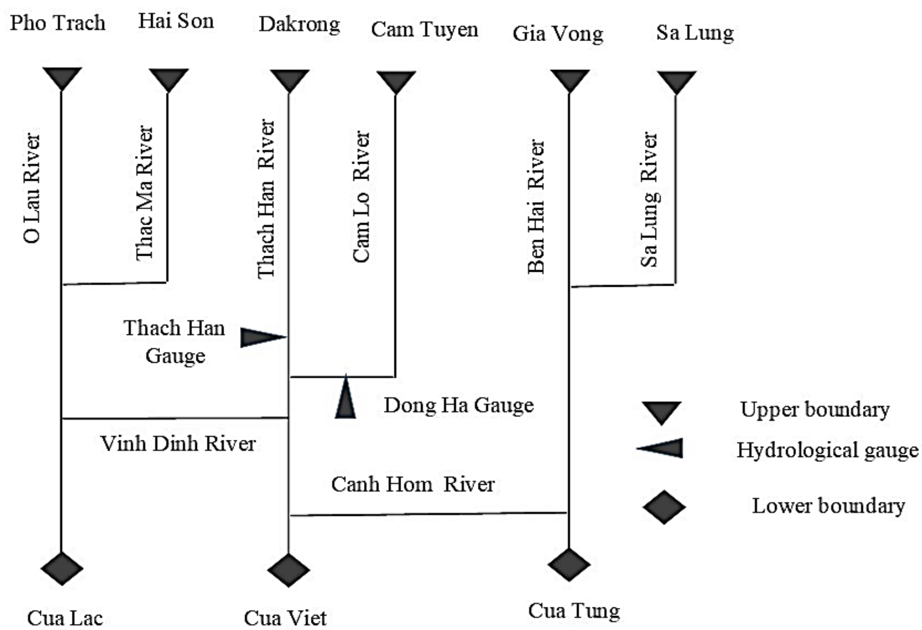

Figure 2. Simplified diagram of the rivers and channels considered in the 1D model

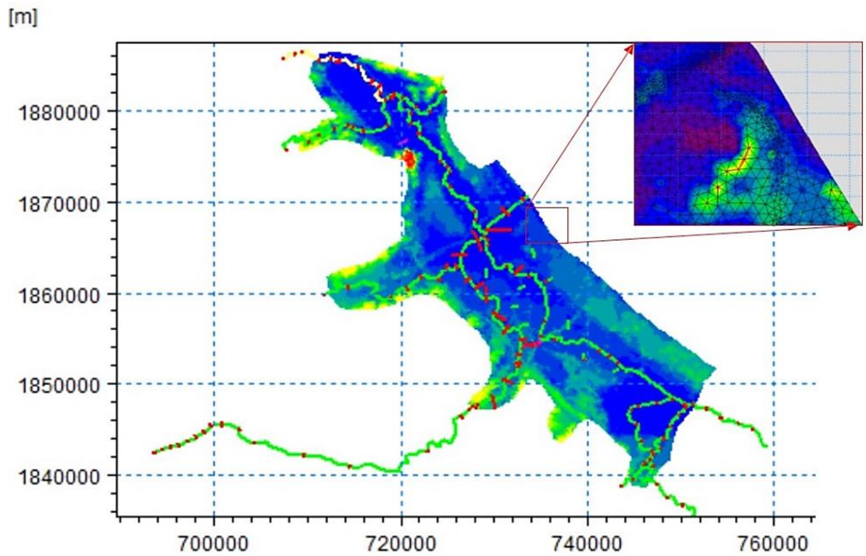

Figure 3. Coupling between 1D and 2D models in the computational domain

The hourly time series of the water discharge was prescribed at upstream boundaries (i.e., Pho Trach, Hai Son, Dakrong, Cam Tuyen, Gia Vong, and Sa Lung). In contrast, the hourly datasets of water elevation were prescribed at downstream boundaries named Cua Lac, Cua Viet, and Cua Tung. Note that water discharge calculated from the rainfall-runoff model was used at several specific upstream boundaries where hydrological data were unavailable. Different simulations were performed for calibration, validation, and related purposes. Detailed simulated results are presented in the next section.

2.2.2. NAM model

In the whole of Quang Tri province and neighboring river basins, flow data was only available for the Gia Vong station upstream of Ben Hai River, so we first calibrated and verified the NAM model with the Gia Vong data and then used that as a basis from which to generate runoff at other sub-basins in the study area. The Quang Tri River basin was divided into 21 sub-basins based on the DEM and the available meteo-hydrological data (Fig. 1). The area of the sub-basins varied from 20 to 1046 km^2.

To accurately describe the flood-forming conditions in Quang Tri province, the hydrodynamic network included:

In the Ben Hai River system: the main channel of the Ben Hai River from the Gia Vong gauge station to the river mouth (23.4 km long), the Sa Lung branch (15.7 km in length) from Sa Lung dam to the confluence with the Ben Hai River near Hien Luong bridge.

In the Thach Han River system: the main channel of the Thach Han River from the confluence of the Dakrong River with the Rao Quan stream, flowing out the Tram dam, and up to the Eastern Sea at Cua Viet (77 km in length); and the Hieu River (Cam Lo River; about 23.4 km in length) starting from the Cam Tuyen bridge and meeting the Thach Han River at the confluence of Gia Do near Dong Ha city.

In the O Lau River system: O Lau main river from Pho Trach bridge to Tam Giang lagoon in Cua Lac, 31.8 km in length.

The connection between Ben Hai and Thach Han River and Canh Hom River, being about 16.1 km long, must also be included. Vinh Dinh River from the Viet Yen culvert on Thach Han River (in Trieu A commune), flowing through Trieu Phong, Hai Lang districts up to O Lau River with a length of 37.6 km.

According to the network that was constructed above, there are 6 upstream discharge boundaries - Sa Lung Bridge, Gia Vong, Cam Tuyen, Dakrong, Hai Son, and Pho Trach and 3 downstream water level boundaries at Cua Tung, Cua Viet, and Cua Lac. Based on observed data conditions, discharge data is only observed at the upper boundary at Gia Vong, and the upper boundaries must utilize discharge generated by the hydrological model (NAM) using calibrated and verified.

2.3. Conditioning factors

Selecting the appropriate conditioning factors is essential when using machine learning to assess the flood. However, there is no standard guide or consensus on selecting these factors (Zhao et al., 2019; Prasad et al., 2022). Several studies have found that the selection of conditioning factors depends on data availability, data which are often grouped under topography, climate, hydrology, and human activity. In this study, 13 conditioning factors were selected to use as the input data for the machine learning model, namely elevation, slope, aspect, curvature, distance to river, distance to road, flow direction, rainfall, land use, normalized difference vegetation index (NDVI), normalized difference built-up index (NDBI), soil type, and sediment transport index (STI) Fig. 4).

Elevation is a crucial conditioning factor when assessing flood depth in any region. Low-elevation areas are more susceptible to deeper floods. In the province of Quang Tri, the low-altitude region in the east is often affected by flood depth (Eslaminezhad et al., 2022). The elevation value in the study area (as a whole) ranges from 0 to 1700 m.

The slope is another topographic factor and is considered indispensable to flood analyses. It affects water flow, velocity, and accumulation capacity (Yariyan et al., 2020). All these elements influence the depth of the flood. In the study area, the slope value ranged between 0 and 76 degrees.

Curvature is essential in a flood depth model because it significantly influences drainage. The higher the curvature value, the greater the degree of flood (Mirzaei et al., 2021). In this study, curvature ranged from -16 to 21.

Aspect strongly affects evaporation and vegetation distribution and is inversely proportional to flood depth (Costache et al., 2022). Aspects in Quang Tri province ranged from 0 to 360 degrees.

Distance to the river affects flow velocity, flood magnitude, and flood depth. Regions near the river are flooded more deeply than those further away (Khosravi et al., 2019). In the study area, a flood occurs along the Ben Hai, Thach Han, and O Lau, and flood depth is highest in the areas close to these rivers.

The road distance is an essential factor for a flood depth model because roads influence permeability and roughness (Pradhan and Youssef, 2011; Tehrany and Kumar, 2018; Mind'je et al., 2019). In the province of Quang Tri, in recent years, the construction of roads has impacted flood depth. Areas near roads tend to flood more.

Rainfall is indispensable in any flood pattern because heavy precipitation over a short time causes floods with high intensity (Pradhan and Youssef, 2011; Pham et al., 2021; Hoang et al., 2022). The province of Quang Tri has a tropical monsoon climate and is often affected by significant typhoons accompanied by heavy rains, a central cause of floods. In the study area, the rainfall value 2021 ranged from 1680 to 3566 mm.

Land use significantly affects river flow, roughness, and velocity (Nguyen et al., 2022). On recent field missions, we have witnessed a rapid change in land use across Quang Tri province, leading to a higher floodwater level and faster flow speed than before.

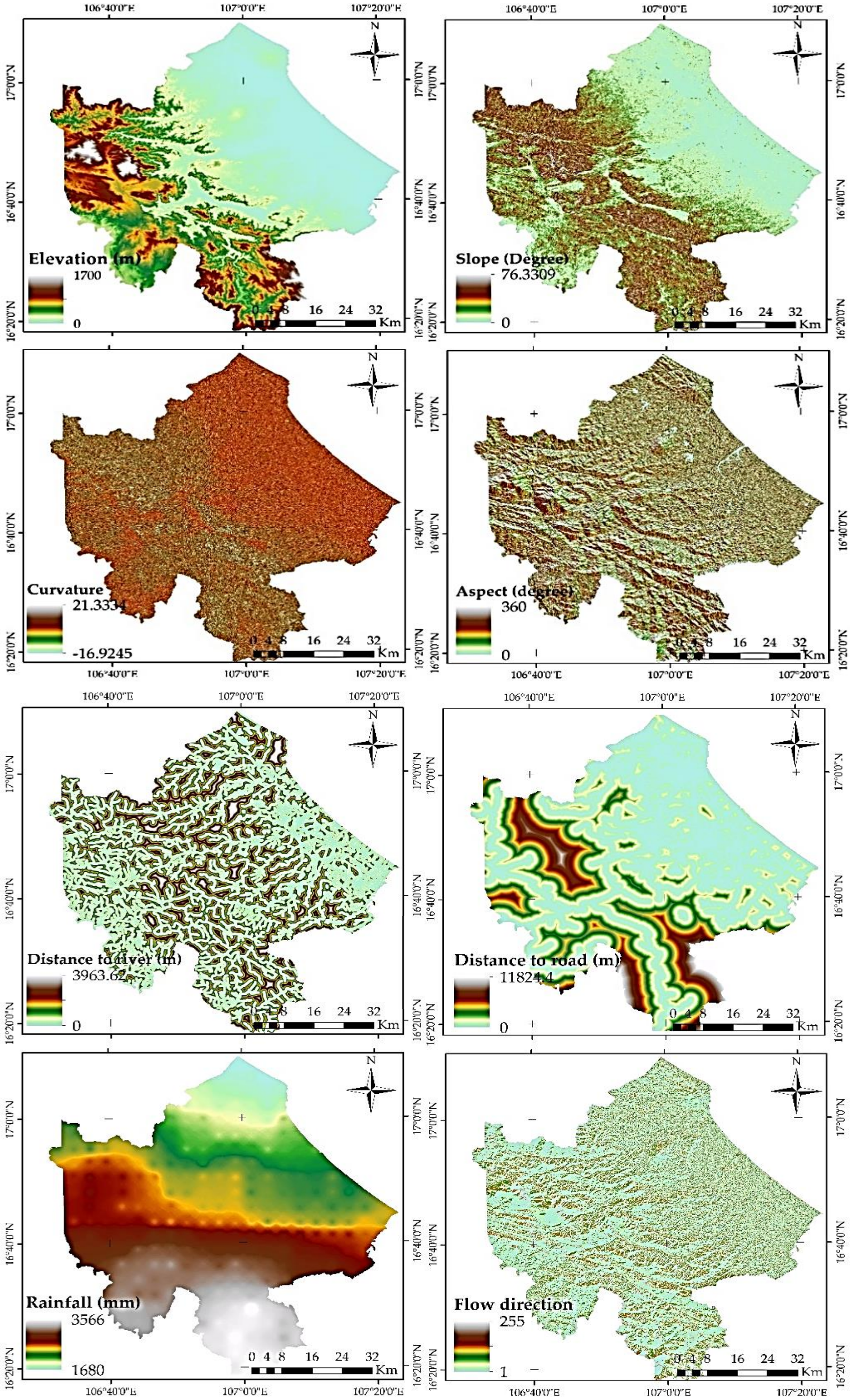

Figure 4. Conditioning factors used for the flood depth model

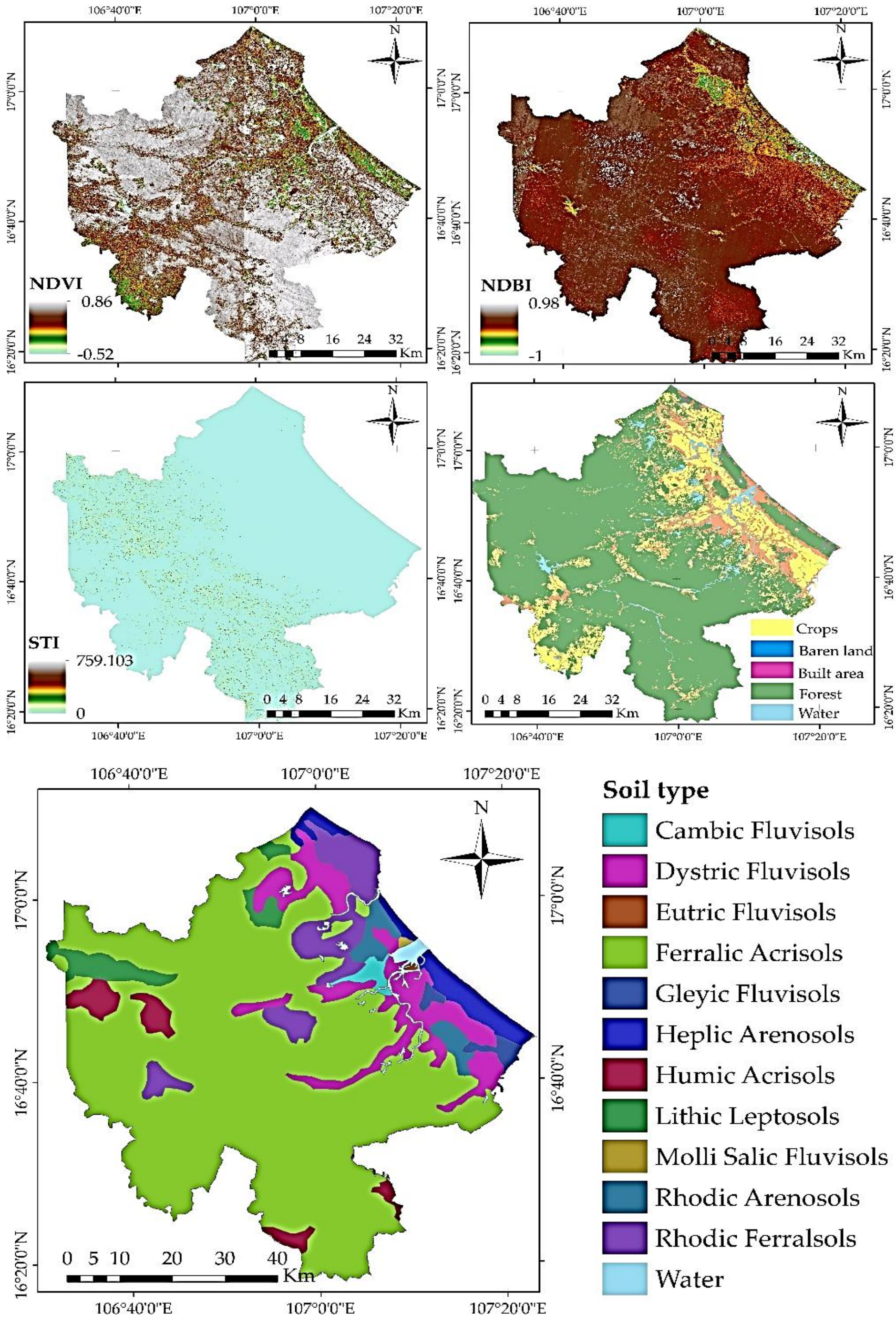

Figure 4. Cont.

NDVI and *NDBI* show vegetation and building density. The higher the vegetation density, the lower the flood probability and the shallower the flood depth. Meanwhile, the higher the density of buildings, the higher the flood probability and the greater the flow rate and flood depth (Nguyen, 2022). The value of NDVI and NDBI in the study area ranges from -1 to 1.

Flow direction is indispensable in constructing a map, as it affects the water storage capacity and influences the flood depth in an area (Liuzzo et al., 2019; Luu et al., 2022). In the province of Quang Tri, the value of this factor ranges from 1 to 255.

STI influences the capacity of water accumulation (Chen et al., 2019). The value of STI in Quang Tri province ranges from 0 to 759.

Soil type is crucial in flood assessment because it influences water flow and the rainfall-runoff mechanism (Hammami et al., 2019). In the study area, soil was divided into 11 types: cambic fluvisols, dystric fluvisols, eutric fluvisols, ferralic acrisols, gleyic fluvisols, heplic arenosols, humic acrisols, lithic leptosols, molli salic fluvisols, rhodic arenosols, and rhodic ferralsols.

2.4. Machine learning methods

The methodology used to estimate the flood depth in this study comprised five main steps: (i) collection of flood depth locations, (ii) selection of conditioning factors, (iii) construction of machine learning models, (iv) evaluation of the proposed models, and (v) construction and analysis of the flood depth map (Fig. 5).

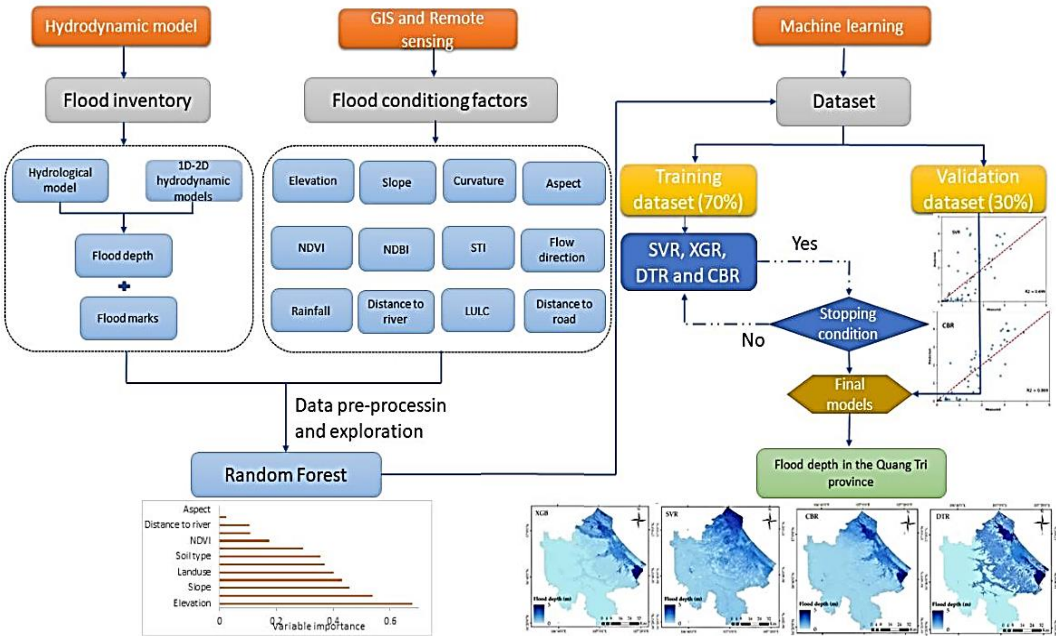

Figure 5. Methodology used for the flood depth model

(i) In this study, the flood depth locations were collected from two sources: flood marks measured on the field mission and hydrodynamic modeling. 540 flood depth locations were used to build the machine learning model.

(ii) Conditioning factors were selected from four groups: topographic, hydrological, climatic, and anthropogenic. We selected all available factors and then used random forests to eliminate unnecessary ones.

Ultimately, 540 flood depth locations and 12 conditioning factors were used as the model input data. These data were divided into two groups: 60% to train the models and 40% to validate the models.

(iii) Four machine learning models, namely support vector regression (SVR), XGBoost regression (XGR), CatBoost regression (CBR), and decision tree regression (DTR), were used to build flood depth maps. The precision of SVR depended on the parameters C, gamma, and epsilon; DTR used the criterion and splitter parameters; C.B. used depth; and XGB worked on n_estimators, eta min_child_weight, and max_depth.

In this study, model parameters were C=10, gamma='scale', epsilon=0.1, and max_iter=500 for SVR; max_depth=5, criterion='absolute_error', and splitter='best' for DTR; iterations=500, depth=3, train_dir=”logloss”, loss_function='RMSE' for CBR, and n_estimators=500, eta=0.3, min_child_weight=1, and max_depth=6 for XGR.

(iv) The statistical indices R^2, RMSE, and MAE were used to evaluate the proposed models.

(v) After validation of the proposed models, they were used to build the flood depth map by assigning 30 million pixels for all the study areas with each of the 12 conditioning factors.

2.4.1. Support vector regression (SVR)

Support vector machines (SVMs) represent a family of machine learning algorithms capable of solving discrimination and mathematical regression problems. The first SVMs were developed in the 1990s and were used to solve regression problems in 1996 by Vapnik et al. (1997).

The main ideas of SVR are similar to those of SVMs. They consist of discriminating using a hyperplane in which the data is separated into several classes, the boundary of which is as far as possible from the data points (or "maximum margin"). To achieve this goal, SVR uses kernels, i.e., mathematical functions for projecting and separating data in vector space, the "support vectors" being the data closest to the border. The optimization aims to find the furthest boundary of all training points, which is, the most optimal and has the best generalization capacity. Using the kernel trick allows the use of linear classifiers and solves nonlinear problems (Tehrany et al., 2015; Choubin et al., 2019).

More precisely, SVR operates by finding the function f(x), which has at most one deviation ε concerning the training examples (xi, yi), for i = 1,…, N, and which is as flat as possible. This amounts to not considering errors more minor than ε and prohibiting those larger than ε. Maximizing the function's flatness minimizes the model's complexity, which affects its generalization performance. The accuracy of the SVR model depends on the adjustment of the C and gamma parameters. These parameters were optimized using trial and error (Naghibi et al., 2017; Nachappa et al., 2020).

2.4.2. XGBoost regression (XGR)

XGBoost (eXtreme Gradient Boosting) is a prevalent machine learning model and enhances the Gradient Boosting algorithm. It is used to solve both classification and regression problems and reduce errors in predictive data analysis (Abedi et al., 2022; Nguyen et al., 2022). It is a weak decision tree set that predicts residuals and correct errors from previous decision trees. This algorithm's particularity resides in using the decision tree (Costache et al., 2022). Weak decision trees that do not perform well enough are 'pruned' until they entirely play their role (Ma, Zhao, et al. 2021). XGR performance depends on ground, max_depth, eta, gamma, colsample_bytree, min_child_weight, subsample (Abedi et al., 2022). In this study, these parameters were determined using the trial-and-error technique.

2.4.3. CatBoost regression (CBR)

CatBoost was developed by Yandex and is based on gradient boosting. This technique promotes learning by turning weak learners into strong learners by improving old patterns and reducing errors. Each decision tree is an evolution of the first set of data (Abujayyab et al., 2022; Nguyen et al., 2022). CBR is powerful for two reasons: it provides accurate results even with (i) little training data and (ii) a raw dataset (Sahin 2022).

Although CBR is based on gradient descent, it has some additional features that make it more robust: (i) it implements a symmetric tree which reduces prediction time and also has a shallower tree depth by default; (ii) CBR uses random permutations, similar to how XGR has a random parameter (Nguyen et al., 2022).

2.4.4. Decision tree regression (DTR)

D.T.s are popular machine learning tools that can solve regression and classification problems. They have a tree-like hierarchical structure consisting of a root node, branches, internal nodes, and leaf nodes (Ghosh and Maiti, 2021). A decision tree starts with a root node that has no incoming branches. Branches out of the root node lead to internal nodes, also known as decision nodes. Depending on the available characteristics, both nodes perform evaluations on homogeneous subsets, which are referred to as leaf or terminal nodes. Leaf nodes represent all possible outcomes in the dataset (Sahani and Ghosh, 2021). The decision tree model is constructed in two main steps: (i) building the tree and (ii) pruning the tree. In many cases, tree pruning may be required to remove inappropriate knots. A good algorithm builds a large tree and then reduces it to an appropriate size (Nefeslioglu et al., 2010; Bui et al., 2012).

In this study, the method used to construct the decision tree includes: (i) selection of conditional factors, (ii) selecting the root node as the first internal node, (iii) dividing the input dataset into subsets to construct the subnodes, (iv) estimating the value of nodes using the random forest method, and (v) selection of the best nodes.

2.5. Prediction performance evaluation

Evaluation of the performance of the prediction model is indispensable and the most important of all the steps in the construction of the model. Model performance was evaluated using statistical indices ROC, AUC, RMSE, MAE, and R2.

RMSE and MAE are popular quantitative methods for determining the predictive ability of models by measuring the absolute error between the observation value and the estimate value, that is, the prediction errors (Bui et al., 2019). The following equation computes them:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i-1}^{n}(X_{Sim} - Y_{obs})^2}$$

$$MAE = \frac{1}{n}\sum_{i-1}^{n}|X_{Sim} - Y_{obs}|$$

X_{obs} represents the observation value in the training or validation dataset; X_{sim} represents the output value of the prediction model; n is the sample number. Machine learning and optimization algorithms use RMSE as the objective function to optimize model parameters. The smaller the value of RMSE, the better the model.

R^2 is a popular index for evaluating the quantitativeness of a regression model by measuring the proportion of the total variance of the dependent variable computed by a model. The closer the R^2 value is to 1, the better the model explains the variance of the dependent variable (the closer the relationship is between the independent variable and the dependent variable). Meanwhile, the value of R^2 is zero, which means that the model does not explain the variance of the dependent variable (the relationship between the independent variable and the dependent variable is discrete) (Nguyen et al., 2021).

NASH is used to evaluate system errors in a long-term simulation by measuring the accuracy rate between simulation and observation value (Damadi et al., 2021). The following equation computes it:

$$NASH = 1 - [\frac{\sum_{i=1}^{n}\left(X_i^{obs} - X_i^{sim}\right)^2}{\sum_{i=1}^{n}\left(X_i^{obs} - \bar{X}\right)^2}]$$

3. Results

3.1. Hydrodynamic model calibration

Using flood data from an event in

November 1999 (it is considered a historic flood of the province), different simulations were performed to calibrate the modeling parameters in the coupling between 1D and 2D models. Different values were tested to determine those appropriate for the modeling parameter. Figure 6 shows the comparison between simulated and observed water level at Thach Han and Dong Ha stations.

The model reproduces the observed data very well at both locations. The values for the NASH coefficient are equal to 0.95 and 0.92 at Thach Han and Dong Ha stations, respectively. Furthermore, the discrepancy in water elevation between the simulated results and observed data ranges from 15 to 19 cm, revealing that the model accurately represented the maximum value of water elevation at both stations.

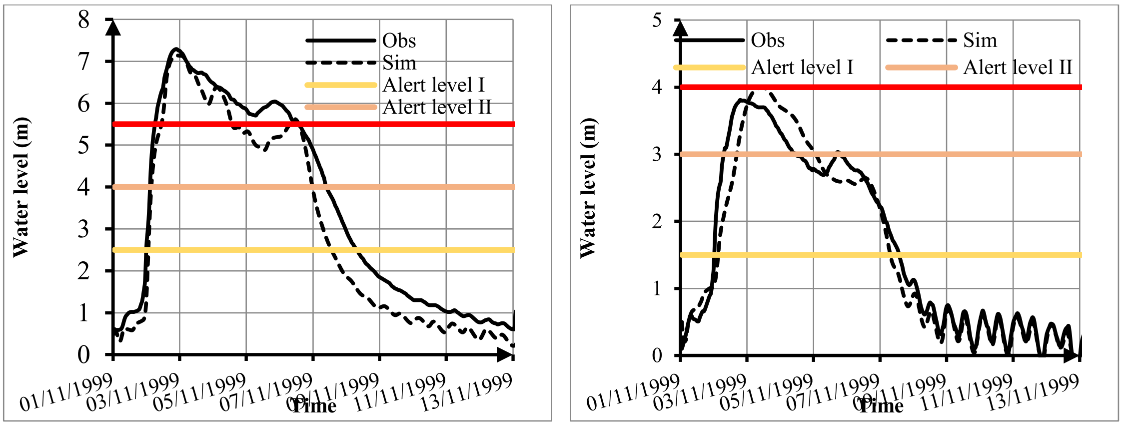

Figure 6. Comparison of simulated and observed water level at Thach Han and Dong Ha station in Nov. 1999 flood event

Figure 7 shows the validation results of the model for the flood event in October 2005. Similar to calibration results, the coupling between 1D and 2D models also represents very well the water elevation of the flood at both stations. The NASH coefficient of water elevation ranged from 0.85 to 0.90, while the discrepancy of the highest value of water elevation changes from 2 to 16 cm at both locations. These results suggest that the values of the modeling parameters used in the simulation are accepted. The model can be applied in further investigations related to flood depth.

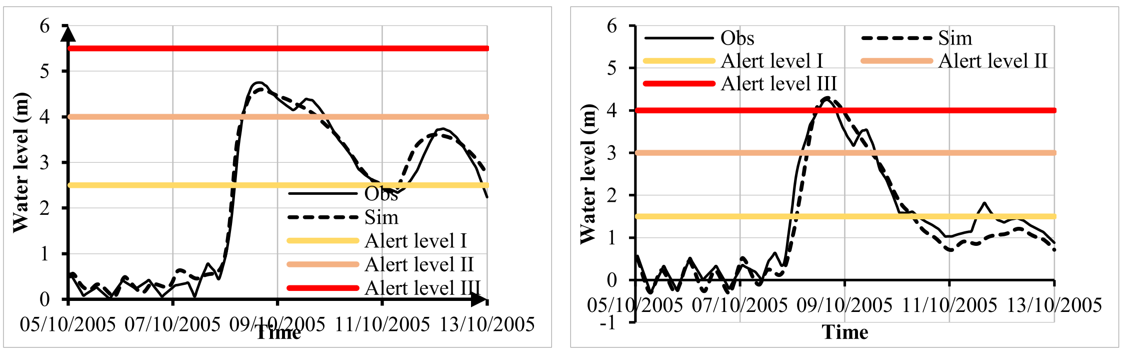

Figure 7. Comparison of simulated and observed water level at Thach Han and Dong Ha station in Oct. 2005 flood event

3.2. Conditioning factor assessment

Assessing the quality of input data is an essential step in improving model performance. One way to improve the quality of input data is to select appropriate conditioning factors because data redundancy can reduce model accuracy. This study used random forest to assess the importance of 12 conditioning factors used for the flood depth model. The results showed that elevation (0.67), rainfall (0.54), slope (0.46), distance to the road (0.43), and land use (0.4) were the most critical factors for flood depth in Quang Tri province, followed by STI (0.37), soil type (0.35), NDBI (0.29), NDVI (0.18), curvature (0.1), distance to the river (0.1) and flow direction (0.025). However, the aspect did not influence flood depth, with an R.F. value of 0 (Fig. 8).

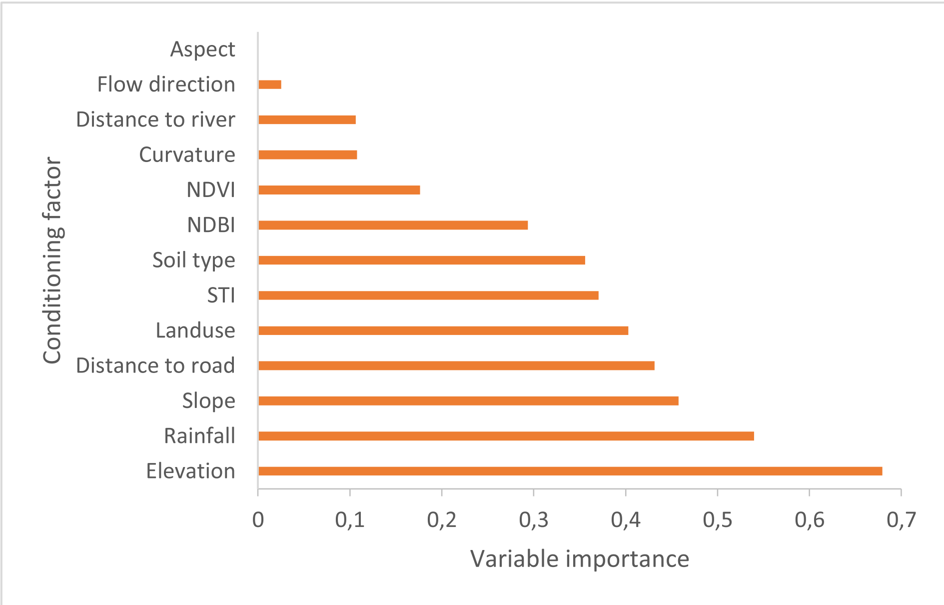

Figure 8. The importance of conditioning factors using random forest

3.3. Model performance

In this study, the R^2 coefficient of determination was used to evaluate the performance of flood depth models during the training and validation processes. The results showed that for the learning process, the XGR model had the most accurate flood depth estimation capability among the four models with an R^2 value of 0.99, followed by C.B. with an R^2 value of 0.94, D.T. with an R^2 value of 0.85 and SVR with an R^2 value of 0.7. For the validation process, the CBR model performed best, with an R^2 value of 0.86, followed by XGR (0.84), DTR (0.75), and SVR (0.7) (Fig. 9).

RMSE and MAE were also used to evaluate the performance of the models. For the learning process, the XGR model had the best fit, with an RMSE value of 0.16 and an MAE value of 0.09, followed by CBR (RMSE=0.25, MAE=0.14), DTR (RMSE=0.42, MAE=0.17), and SVR (RMSE=0.6, MAE=0.31). For the validation process, the CBR model was the most accurate, with an RMSE value of 0.39 and an MAE value of 0.21, followed by XGR (RMSE=0.43, MAE=0.2), DTR (RMSE=0.58, MAE=0.23), and SVR (RMSE=0.6, MAE=0.32) (Table 1).

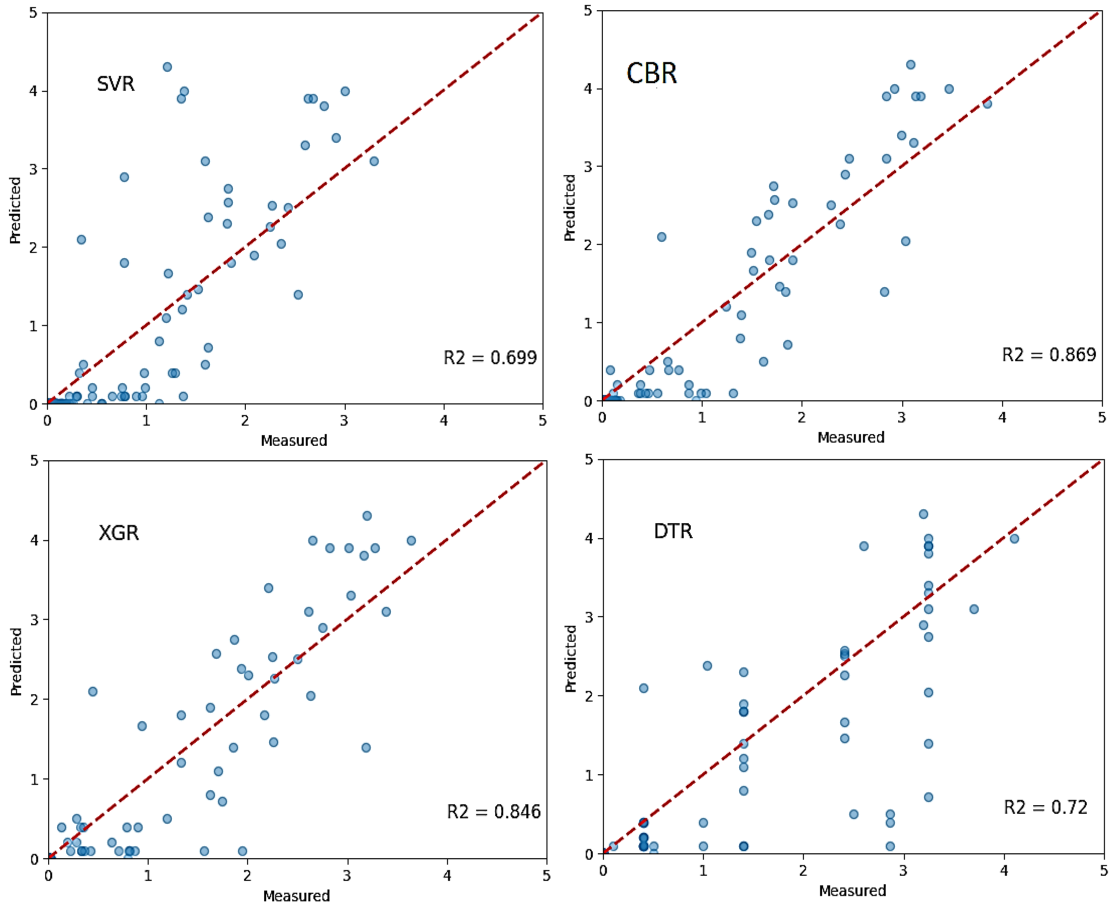

Figure 9. the R^2 values for SVR, XGR, CBR, and DTR

Table 1. Performance of the four models

	Training dataset			Validating dataset		
	RMSE	MAE	R²	RMSE	MAE	R²
SVR	0.604018	0.316811	0.700956	0.60094	0.322496	0.699445
CBR	0.258044	0.146606	0.945421	0.396621	0.213205	0.869078
XGR	0.000836	0.000529	0.999999	0.43078	0.201412	0.845556
DTR	0.424725	0.169214	0.85214	0.580535	0.236999	0.719509

3.4. Flood depth mapping evaluation

Figure 10 shows the flood depth level in Quang Tri province according to CBR, DTR, SVR, and XGR models. In this study, the maximum flood depth was 5 m. It should be noted that there is a difference between measuring the water level and the flood depth. This is because the water level was measured against the national elevation mark at Hon Dau (in the northern part), which is generally considered the mean sea level for many years. Flood depth is measured by taking the water level (7 meters) minus the topographic height of the location. So the depth is 5 meters.

SVR identified about 2590 km² of Quang Tri province to be flooded by less than 1 m, 129 km² flooded by 1-1.5 m, 19 km² flooded by 1.5-2 m, 66 km² 2-2.5 m, and 210 km² flooded by more than 2.5 m.

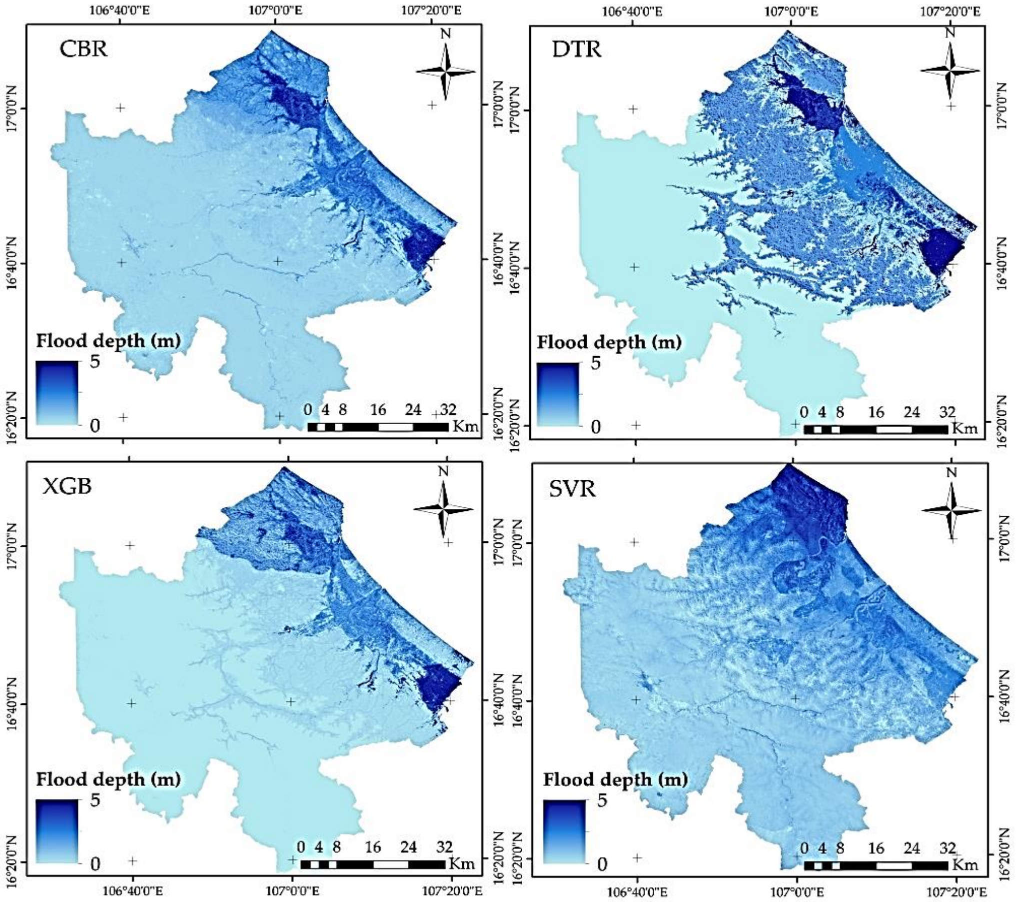

Figure 10. Flood depth produced by CBR, DTR, XGR, and SVR models

According to the XGR model, about 2559 km² was flooded by 0-1 m of water, 143 km² by 1-1.5 m, 111 km² by 1.5-2 m, 87 km² by 2-2.5 m, and 114 km² by more than 2.5 m.

For the CBR model, 2570 km² of the study area was flooded by between 0 and 1 m, 200 km² by 1-1.5 m, 124 km² by 1.5-2 m, 62 km² by 2-2.5 m, and 58 km² by over 2.5 m.

Finally, according to the DTR model, 2365 km^2 saw a flood of 0-1 m, 280 km² 1-1.5 m, 173 km² 1.5-2 m, 89 km² 2-2.5 m, and 111 km² were flooded by more than 2.5 m of water.

4. Discussions

The growing flood risk in the context of climate change has received significant attention from the scientific community (Bronstert et al., 2002; Saghafian et al., 2008). Various studies have been developed to reduce flood damage. However, there are not yet universal guides for model selection that can solve all problems in all regions because each region has different environmental conditions, climates, hydrology, and human activities. In particular, the integration between machine learning and the hydrodynamic model to estimate flood depth is still missing in the literature and is not yet used in the study area. So, in this study, we indicated that integrating hydrodynamic modeling and machine learning shows promise in obtaining results accurately and reliably (Nguyen et al., 2022; Nguyen et al., 2022). The results of this study can support decision-makers or planners in the sustainable development of the territory, for example, the avoidance of new constructions in the region at high flood risk.

In general, the performance of all four proposed models was acceptable. We found the CBR model to be more potent than the others, with an R^2 value of 0.89. CBR makes models with optimum precision and does not require high computational resources. Adjusting CBR parameters is easy, and CBR can be optimized with default parameters (Li et al., 2022; Lu et al., 2022). With an R^2 value of 0.83, the model of XGR was second. In addition to rapid convergence, XGR has the advantage of not requiring any data preprocessing, since it has built-in routines that allow processing missing data, so implementation is straightforward. In addition, this algorithm helps to process large datasets efficiently while reducing their attributes (Devan and Khare, 2020; Zhong et al., 2020; Hajek et al., 2022). The DTR model came third, with an R^2 value of 0.72. In addition to being easy to install, DTR makes decisions using a targeted approach, which can make significant improvements to performance (Bansal et al., 2022; Wen et al., 2022). With an R^2 of 0.7, SVR had the lowest accuracy. This algorithm is easy to interpret and relatively simple, hence not very computationally expensive, and it effectively solves problems in dimensional space. However, the main drawback of SVR is that it is inefficient when working on a large dataset. Another drawback is that it works less efficiently when a dataset has much noise, as in this study (Lin et al., 2006; Marín et al., 2022).

Compared with the inundation depth map obtained using another method in previous studies, the results of this study agree well with this map (Nguyen, 2022; Doan et al., 2022). The floodplains concentrate mainly in the eastern regions with 0 to 5 m depth.

In general, the approach proposed in this study can provide an accurate flood prediction model on a large scale, especially in regions with limited data. The results of this study may represent an essential tool in the early warning of flood. They can support decision-makers and local authorities in the construction of appropriate strategies to reduce damage to human life and property.

This study applied different machine learning algorithms, namely SVR, XGR, CBR, and DTR, to construct flood depth maps in Quang Tri province in Vietnam. As expected, machine learning succeeded in building a highly accurate map. This map plays a crucial role in informing processes such as limiting construction in highly flood-prone areas and flood risk reduction planning. Previous studies have shown that urban growth significantly influences flood risk. Although it is not a primary cause, it has crucial effects on flood intensity. In this sense, planners should be meticulous about changes to land use, especially the transformation of agricultural land into urban areas, a growing trend in coastal areas in Vietnam.

Planning cannot prevent hazards but it can minimize their damage. In more detail, there are two planning implications Planners can impose restrictions to minimize the effects of floods and are also able to minimize the exposure to floods by restricting new constructions and high densities of population in areas with high flood levels. At the same time, planners are responsible for avoiding land use changes, particularly those attempting deforestation to make room for urban developments, which are frequent in the coastal areas of Vietnam (Petrisor et al., 2020). In general, our study provides additional evidence indicating that planners need to integrate the results of scientific research into their work (Petrisor et al., 2021, 2022).

Although this study was successfully used to construct precise flood depth maps in the Quang Tri province of Vietnam, it encountered general limitations related to the use of the data. The collection of flood depth data is challenging due to the data-sharing policy of Vietnam (similar issues are faced in other countries) and the limited financial resources of the project. This study used the hydrodynamic model to add training data for the machine learning model. Also, in this study, the topographic data was extracted from the 1:50,000 scale topographic map. However, this data only reflects the terrain. Other techniques, such as drones (UAV) or LIDAR, can provide more information, such as slope, flow direction, or anthropogenic factors such as buildings, dyke networks, etc. Ultimately, Quang Tri province is subject to two types of floods: river floods and coastal floods. However, coastal flooding is insignificant because the study area is influenced by a micro-tidal regime and storm surges, contributing only marginally to flooding. Previous studies have confirmed this. Thus, in this study, we do not consider the risk related to the sea. However, several studies have pointed out that using more data makes the models more reliable.

While the hydrodynamic model was constructed with high accuracy, there are still some remaining uncertainties. This study used RMSE as the objective function, so it was necessary to address the overfitting problem. To this end, we used several techniques, such as a k-fold search limit or search benchmark. In future research, additional training data would improve the accuracy of the models. The more training data, the more accurate the prediction model.

5. Conclusions

Flood is one of the most dangerous natural disasters, and the flood impact has worsened in the context of climate change. Accurate prediction of flood depth and extent plays an important role in flood early warning systems that reduce damage to human life and property. Therefore, this study aims to develop a rapid flood forecasting model by combining machine learning algorithms (SVR, XGB, C.B., and D.T.) and hydrodynamic modeling in Quang Tri province in Vietnam. The results of this study can support decision-makers or local authorities in evacuating people in the event of an emergency; equally, it can inform the sustainable development of the territory.

Integration of the hydrodynamic with machine learning models can build precise flood depth maps. This approach can be used to build a map in any region on any scale, especially in data-limited regions.

Of the 13 conditioning factors used as model input data, elevation, and rainfall were the most influential, while the aspect was the least. Among the proposed models, CBR outperformed the other models with an R^2 value of 0.86 for simulating flood depth, followed by XGR (R^2=0.84), DTR (R^2=0.72), and SVR (R^2=0.7), respectively. All models can provide results accurately and quickly enough to meet the demands of a flood early warning system.

The flood depth maps indicated areas of high inundation located in the east of the study area, with low elevation and a dense river network.

Although this study has explained in detail the scientific significance and all the steps of flood depth model construction, completing all these processes locally is considered a significant challenge. Future studies can use the trained models to assess flood risk in Quang Tri province and other provinces in Vietnam in general.

Acknowledgments

This work was funded by the GCRF Networking Grant (REF: GCRFNGR4\1165) from the Academy of Medical Sciences.

References

Abbott M.B., Ionescu W., 1967. On the numerical computation of nearly horizontal flows. Journal of Hydraulic Research, 5(2), 97-117.

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.

Abu El-Magd S.A., 2022. Random forest and naïve Bayes approaches as tools for flash flood hazard susceptibility prediction, South Ras El-Zait, Gulf of Suez Coast, Egypt. Arabian Journal of Geosciences, 15(3), 217.

Abujayyab S.K., Kassem M.M., Khan A.A., Wazirali R., Coşkun M., Taşoğlu E., Öztürk A., Toprak F., 2022. Wildfire Susceptibility Mapping Using Five Boosting Machine Learning Algorithms: The Case Study of the Mediterranean Region of Turkey. Advances in Civil Engineering, 3959150.

Bansal M., Goyal A., Choudhary A., 2022. A comparative analysis of K-Nearest Neighbour, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. Decision Analytics Journal, 100071.

Bronstert A., Niehoff D., Bürger G., 2002. Effects of climate and land-use change on storm runoff generation: present knowledge and modelling capabilities. Hydrological Processes, 16(2), 509-529.

Bui D.T., Ngo P.-T.T., Pham T.D., Jaafari A., Nguyen Q.M., Pham V.H., 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.

Chakrabortty R., Pal S.C., Janizadeh S., Santosh M., Roy P., Chowdhuri I., Saha A., 2021. Impact of climate change on future flood susceptibility: an evaluation based on deep learning algorithms and GCM model. Water Resources Management, 35, 4251-4274.

Chen W., Hong H., Li S., Shahabi H., Wang Y., Wang X., Ahmad B.B., 2019. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. Journal of Hydrology, 575, 864-873.

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.

Costache R., Arabameri A., Costache I., Crăciun A., Pham T.B., 2022. "New machine learning ensemble for flood susceptibility estimation. Water Resources Management, 36(12), 4765-4783.

Costache R., Arabameri A., Moayedi H., Pham Q. B., Santosh M., Nguyen H., Pandey M., Pham T.B., 2022. Flash-flood potential index estimation using fuzzy logic combined with deep learning neural network, naïve Bayes, XGBoost and classification and regression tree. Geocarto International, 37(23), 6780-6807.

Costache R., Pham Q.B., Avand M., Nguyen T.T.L., Vojtek M., Vojteková J., Lee S., Dao N.K., Pham T.T.N., Tran D.D., 2020. Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment. Journal of Environmental Management, 265, 110485.

Damadi S., Dehvari A., Dahmardeh Ghaleno M.R., Ebrahimiyan M., 2021. Flood hazard zonation using HEC-RAS hydraulic model in Sarbaz River, Sistan and Baluchestan Province. Watershed Engineering and Management, 13(3), 590-601.

Devan P., Khare N., 2020. An efficient XGBoost–DNN-based classification model for network intrusion detection system. Neural Computing and Applications, 32, 12499-12514.

Doang Q.T., Pham T.N., 2022. Research and develop integrated flood forecasting and warning tools for 03 river basins: Thach Han, Vu Gia-Thu Bon and Tra Khuc-Song Ve. Hydrometeorology Journal, 736, 93-110.

Doocy S., Daniels A., Murray S., Kirsch T.D., 2013. The human impact of floods: a historical review of events 1980-2009 and systematic literature review. PLoS currents 5(5).

Eslaminezhad S.A., Eftekhari M., Azma A., Kiyanfar R., Akbari M., 2022. Assessment of flood susceptibility prediction based on optimized tree-based machine learning models. Journal of Water and Climate Change, 13(6), 2353-2385.

Falah F., Rahmati O., Rostami M., Ahmadisharaf E., Daliakopoulos I.N., Pourghasemi H.R., 2019. Artificial neural networks for flood susceptibility mapping in data-scarce urban areas. In: Spatial modeling in GIS and R for Earth and Environmental Sciences, Eds, Pourghasemi H.Z, Elsevier, 323-336.

Farooq M., Shafique M., Khattak M.S., 2019. Flood hazard assessment and mapping of River Swat using HEC-RAS 2D model and high-resolution 12-m TanDEM-X DEM (WorldDEM). Natural Hazards, 97, 477-492.

Ghanbari M., Arabi M., Kao S.C., Obeysekera J., Sweet W., 2021. Climate change and changes in compound coastal-riverine flooding hazard along the U.S. coasts. Earth's Future, 9(5), e2021EF002055.

Gharakhanlou N.M., Perez L., 2023. Flood susceptible prediction through the use of geospatial variables and machine learning methods. Journal of Hydrology, 671, 129121.

Ghosh A., Maiti R., 2021. Soil erosion susceptibility assessment using logistic regression, decision tree and random forest: study on the Mayurakshi river basin of Eastern India. Environmental Earth Sciences, 80, 1-16.

Hajek P., Abedin M.Z., Sivarajah, U., 2022. Fraud detection in mobile payment systems using an xgboost-based framework. Information Systems Frontiers, 1-19.

Hall J.W., Meadowcroft I.C., Sayers P.B., Bramley M.E., 2003. Integrated flood risk management in england and wales. Natural Hazards Review, 4(3), 126-135.

Hammami S., Zouhri L., Souissi D., Souei A., Zghibi A., Marzougui A., Dlala M., 2019. Application of the GIS based multi-criteria decision analysis and analytical hierarchy process (AHP) in the flood susceptibility mapping (TUNISIA). Arabian Journal of Geosciences, 12, 1-16.

Hens L., Nguyen A.T., Hanh T.H., Cuong N.S., Lan T.D., Van Thanh N., Le D.T., 2018. Sea-level rise and resilience in Vietnam and the asia-pacific: A synthesis. Vietnam J. Earth Sci., 40(2), 126-152. https://doi.org/10.15625/0866-7187/40/2/11107.

Hoang-Cong H., Ngo-Duc T., Nguyen-Thi T., Trinh-Tuan L., Jing Xiang C., Tangang F., Jerasorn S., Phan-Van T., 2022. A high-resolution climate experiment over part of Vietnam and the Lower Mekong Basin: performance evaluation and projection for rainfall. Vietnam J. Earth Sci., 44(1), 92-108. https://doi.org/10.15625/2615-9783/16942.

Hosseiny H., Nazari F., Smith V., Nataraj C., 2020. A framework for modeling flood depth using a hybrid of hydraulics and machine learning. Scientific Reports, 10(1), 1-14.

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

Johnson C.L., Priest S.J., 2008. Flood risk management in england: A changing landscape of risk responsibility? International Journal of Water Resources Development, 24(4), 513-525.

Kadam P., Sen D., 2012. Flood inundation simulation in ajoy river using mike-flood. ISH Journal of Hydraulic Engineering, 18(2), 129-141.

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.

Li L., Zhang Z., Xiong Y., Hu Z., Liu S., Tu B., Yao Y., 2022. Prediction of hospital mortality in mechanically ventilated patients with congestive heart failure using machine learning approaches. International Journal of Cardiology, 358, 59-64.

Lin J.-Y., Cheng C.-T., Chau K.-W., 2006. Using support vector machines for long-term discharge prediction. Hydrological Sciences Journal, 51(4), 599-612.

Liuzzo L., Sammartano V., Freni G., 2019. Comparison between different distributed methods for flood susceptibility mapping. Water Resources Management, 33, 3155-3173.

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.

Luu C., Nguyen D.D., Amiri M., Van P.T., Bui Q.D., Prakash I., Pham B.T., 2022b. Flood susceptibility modeling using radial basis function classifier and fisher's linear discriminant function. Vietnam J. Earth Sci., 44(1), 55-72. https://doi.org/10.15625/2615-9783/16626.

Luu C., Nguyen D.-D., Van Phong T., Prakash I., Pham B.T., 2021. Using decision tree j48 based machine learning algorithm for flood susceptibility mapping: A case study in Quang Binh province, Vietnam. CIGOS 2021, Emerging Technologies and Applications for Green Infrastructure: Proceedings of the 6th International Conference on Geotechnics, Civil Engineering and Structures, Springer, 1927-1935.

Luu C., Pham B.T., Van Phong T., Costache R., Nguyen H.D., Amiri M., Bui Q.D., Nguyen L.T., Van Le H., Prakash I., 2021. Gis-based ensemble computational models for flood susceptibility prediction in the Quang Binh province, Vietnam. Journal of Hydrology, 599, 126500.

Ma M., Zhao G., He B., Li Q., Dong H., Wang S., Wang Z., 2021. Xgboost-based method for flash flood risk assessment. Journal of Hydrology, 598, 126382.

Mani P., Chatterjee C., Kumar R., 2014. Flood hazard assessment with multiparameter approach derived from coupled 1D and 2D hydrodynamic flow model. Natural Hazards, 70, 1553-1574.

Marín A., Martínez-Merino L.I., Puerto J., Rodríguez-Chía A.M., 2022. The soft-margin support vector machine with ordered weighted average. Knowledge-Based Systems, 237, 107705.

Mind'je R., Li L., Amanambu A.C., Nahayo L., Nsengiyumva J.B., Gasirabo A., Mindje M., 2019. Flood susceptibility modeling and hazard perception in rwanda. International Journal of Disaster Risk Reduction, 38, 101211.

Mirzaei S., Vafakhah M., Pradhan B., Alavi S.J., 2021. Flood susceptibility assessment using extreme gradient boosting (egb), Iran. Earth Science Informatics, 14, 51-67.

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.

Naghibi S.A., Ahmadi K., Daneshi A., 2017. Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. Water Resources Management, 31, 2761-2775.

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.

Nefeslioglu H., Sezer E., Gokceoglu C., Bozkir A., Duman T., 2010. Assessment of landslide susceptibility by decision trees in the metropolitan area of istanbul, turkey. Mathematical Problems in Engineering, 901095.

Nevo S., Morin E., Rosenthal A.G., Metzger A., Barshai C., Weitzner D., Voloshin D., Kratzert F., Elidan G., Dror G., 2021. Flood forecasting with machine learning models in an operational framework. arXiv preprint arXiv: 2111.02780.

Ngo H.L., Nguyen H.D., Loubiere P., Van Tran T., Şerban G., Zelenakova M., Breţcan 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), 342-359.

Ngo T.T.L., 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.

Nguyen H.D., 2022. Flood susceptibility assessment using hybrid machine learning and remote sensing in Quang Tri province, Vietnam. Transactions in GIS, 26, 2776-2801.

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., 2022c. Spatial modeling of flood hazard using machine learning and gis in ha tinh province, Vietnam. Journal of Water and Climate Change, 14(1), 200-222.

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., Pham V.D., Lan Vu P., Thanh Nguyen T.H., Nguyen Q.-H., Nguyen T.G., Dang D.K., Tran V.T., Bui Q.-T., Lai T.A., Petrişor A.-I., 2022b. Cropland abandonment and flood risks: Spatial analysis of a case in North central Vietnam. Anthropocene, 38, 100341.

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., 2022c. 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 T.G., Tran N.A., Vu P.L., Nguyen Q.-H., Nguyen H.D., Bui Q.-T., 2021. Salinity intrusion prediction using remote sensing and machine learning in data-limited regions: A case study in Vietnam's Mekong Delta. Geoderma Regional, 27, e00424.

Pham-Thi T.-H., Matsumoto J., Nodzu M.I., 2021. Evaluation of the Global Satellite Mapping of Precipitation (GSMaP) data on sub-daily rainfall patterns in Vietnam. Vietnam J. Earth Sci., 44(1), 33-54. https://doi.org/10.15625/2615-9783/16594.

Pradhan B., Youssef A., 2011. A 100-year maximum flood susceptibility mapping using integrated hydrological and hydrodynamic models: Kelantan river corridor, malaysia. Journal of Flood Risk Management, 4(3), 189-202.

Prasad P., Loveson V.J., Das B., Kotha M., 2022. Novel ensemble machine learning models in flood susceptibility mapping. Geocarto International, 37(16), 4571-4593.

Saghafian B., Farazjoo H., Bozorgy B., Yazdandoost F., 2008. Flood intensification due to changes in land use. Water Resources Management, 22, 1051-1067.

Sahani N., Ghosh T., 2021. Gis-based spatial prediction of recreational trail susceptibility in protected area of Sikkim Himalaya using logistic regression, decision tree and random forest model. Ecological Informatics, 64, 101352.

Sahin E.K., 2022. Comparative analysis of gradient boosting algorithms for landslide susceptibility mapping. Geocarto International, 37(9), 2441-2465.

Shafizadeh-Moghadam H., Valavi R., Shahabi H., Chapi K., Shirzadi A., 2018. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. Journal of Environmental Management, 217, 1-11.

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., Kumar L., 2018. The application of a dempster-shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods. Environmental Earth Sciences, 77, 1-24.

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.

Tehrany M.S., Pradhan B., Jebur M.N., 2015. Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. Stochastic Environmental Research and Risk Assessment, 29, 1149-1165.

Thai T.H., Tri D.Q., 2019. Combination of hydrologic and hydraulic modeling on flood and inundation warning: Case study at Tra Khuc-Ve river basin in Vietnam. Vietnam J. Earth Sci., 41(3), 240-251. https://doi.org/10.15625/0866-7187/41/3/13866.

Tien Bui D., Pradhan B., Lofman O., Revhaug I., 2012. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and naive bayes models. Mathematical problems in Engineering, 974638.

Van Den Honert R.C., Mcaneney J., 2011. The 2011 brisbane floods: Causes, impacts and implications. Water, 3(4), 1149-1173.

Vapnik V., Golowich S.E., Smola A., 1997. Support vector method for function approximation, regression estimation, and signal processing. Advances in Neural Information Processing Systems, 281-287.

Vilaysane B., Takara K., Luo P., Akkharath I., Duan W., 2015. Hydrological stream flow modelling for calibration and uncertainty analysis using swat model in the Xedone river basin, Lao pdr. Procedia Environmental Sciences, 28, 380-390.

Wen T.J., Chuan C.-H., Yang J., Tsai W.S., 2022. Predicting advertising persuasiveness: A decision tree method for understanding emotional (in) congruence of ad placement on youtube. Journal of Current Issues & Research in Advertising, 43(2), 200-218.

Yariyan P., Avand M., Abbaspour R.A., Torabi Haghighi A., Costache R., Ghorbanzadeh, O., Janizadeh, S., Blaschke, T., 2020. Flood susceptibility mapping using an improved analytic network process with statistical models. Geomatics, Natural Hazards and Risk, 11(1), 2282-2314.

Youssef A.M., Mahdi A.M., Al-Katheri M.M., Pouyan S., Pourghasemi H.R., 2023. Multi-hazards (landslides, floods, and gully erosion) modeling and mapping using machine learning algorithms. Journal of African Earth Sciences, 197, 104788.

Zhao G., Pang B., Xu Z., Peng D., Xu L., 2019. Assessment of urban flood susceptibility using semi-supervised machine learning model. Science of the Total Environment, 659, 940-949.

Zhong R., Johnson Jr R., Chen Z., 2020. Generating pseudo density log from drilling and logging-while-drilling data using extreme gradient boosting (xgboost). International Journal of Coal Geology, 220, 103416.