Spatial modeling of flood hazard using machine learning and GIS in Ha Tinh province, Vietnam

Huu Duy Nguyen

Faculty of Geography, University of Science, Vietnam National University, Hanoi, Vietnam
E-mail: nguyenhuuduy@hus.edu.vn

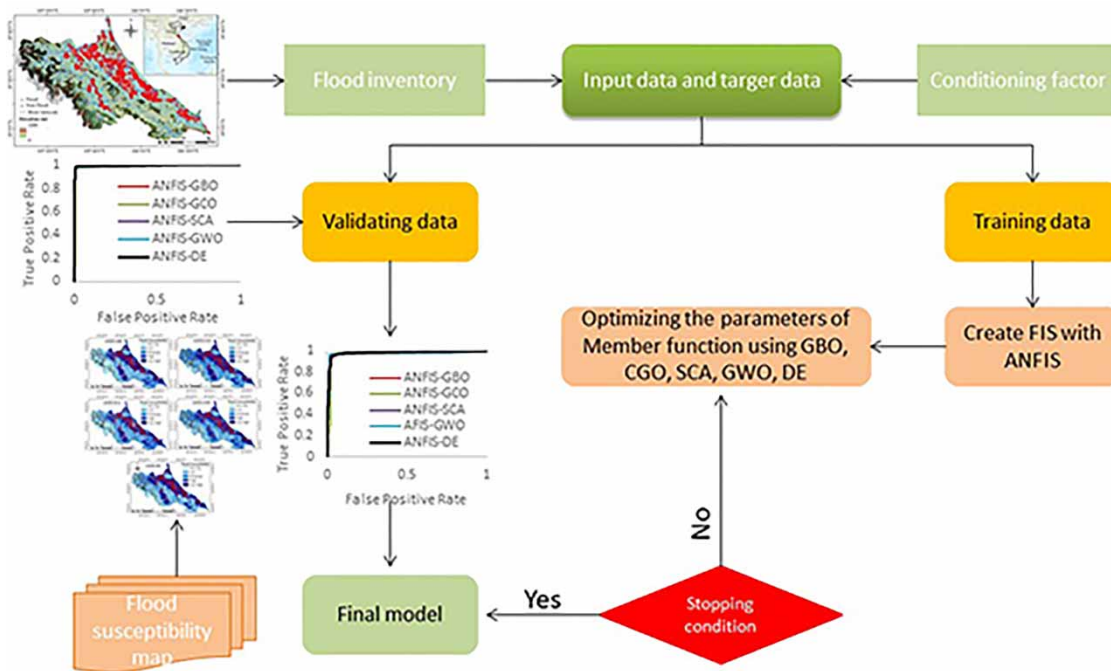HDN, 0000-0001-9306-4641

ABSTRACT

The objective of this study was the development of an approach based on machine learning and GIS, namely Adaptive Neuro-Fuzzy Inference System (ANFIS), Gradient-Based Optimizer (GBO), Chaos Game Optimization (CGO), Sine Cosine Algorithm (SCA), Grey Wolf Optimization (GWO), and Differential Evolution (DE) to construct flood susceptibility maps in the Ha Tinh province of Vietnam. The database includes 13 conditioning factors and 1,843 flood locations, which were split by a ratio of 70/30 between those used to build and those used to validate the model, respectively. Various statistical indices, namely root mean square error (RMSE), area under curve (AUC), mean absolute error (MAE), accuracy, and R1 score, were applied to validate the models. The results show that all the proposed models performed well, with an AUC value of more than 0.95. Of the proposed models, ANFIS-GBO was the most accurate, with an AUC value of 0.96. Analysis of the flood susceptibility maps shows that approximately 32–38% of the study area is located in the high and very high flood susceptibility zone. The successful performance of the proposed models over a large-scale area can help local authorities and decision-makers develop policies and strategies to reduce the threats related to flooding in the future.

Key words: adaptive neuro-fuzzy inference system, flood, Ha Tinh, Vietnam

HIGHLIGHTS

- Flood susceptibility modeling was done using hybrid machine learning approaches.
- The proposed models have achieved great precision and have surpassed the reference models.
- ANFIS-GBO and ANFIS-SCA were the best models.

GRAPHICAL ABSTRACT

1. INTRODUCTION

The consequences of global warming and urban sprawl contribute to both the growing severity and number of floods, landslides, and droughts (Mukherjee *et al.* 2018; Feng *et al.* 2020; Ghorbanzadeh *et al.* 2020). Particularly, climate change leads to an increase in the intensity and frequency of torrential rains, which are the main aggravators of flooding (Prein *et al.* 2017; Nachappa *et al.* 2020; Tabari 2020). According to statistics from the World Bank, the occurrence of flooding has increased by more than 40% in the last two decades, leading to serious economic consequences (Alfieri *et al.* 2017; Andaryani *et al.* 2021). Globally, flooding affected an estimated 109 million people between 1995 and 2015 and caused damage worth approximately USD 75 billion each year (Andaryani *et al.* 2021). This figure may increase to USD 300 billion by 2050, by which time an estimated 1.3 billion people will be living on floodplains (Giang *et al.* 2020).

Vietnam has a coastline of more than 3,200 km; it has 2,360 rivers, with a total length of about 41,900 km. This is why it is one of the countries most affected by floods. Between 1989 and 2015, floods in Vietnam led to at least 14,867 deaths or missing persons (Luu *et al.* 2018). The Central region of Vietnam is particularly affected by flooding as it has a relatively high rainfall of 1,800–2,500 mm/year and is densely populated (Pham *et al.* 2021f). An example of this vulnerability came in October 2020, when flooding resulted in an estimated 129 deaths or missing persons and 214 injured. 111,200 houses were damaged, constituting costs of approximately USD 1 billion.

Despite the efforts of experts and policymakers to reduce the effects of flooding in recent years, its impact is in fact increasing, all over the world (Pham *et al.* 2021d, 2021e). We urgently need a better understanding of what makes a given area vulnerable to flooding. The term 'flood susceptibility' describes the spatial and temporal probability of a flood event, and flood susceptibility maps are considered crucial to the management of flood risk in the future (Pham *et al.* 2020). The literature shows that there are three central methods to determine areas at risk of flooding: remote sensing, physics-based models, and data-driven models.

In the case of remote sensing, researchers have used both radar and optical data to detect flood zones (Sun *et al.* 2000; Schumann *et al.* 2007; Anusha & Bharathi 2020). Although this approach can generate highly accurate flood maps at a low cost, especially when the model is integrated with GIS, it has significant limitations related to spatial and temporal resolutions and is affected by clouds. In addition, remote sensing requires the manual adjustment of the threshold parameters to

obtain a good result; this process is time-consuming and laborious. Critically, remote sensing is also unable to present the original causes of the flood (Nguyen *et al.* 2020).

Physics-based models – such as MIKE FLOOD (Patro *et al.* 2009; Kadam & Sen 2012) and HEC-RAS (Shustikova *et al.* 2019; Ongdas *et al.* 2020) – have received considerable attention from researchers looking to map flood risk. Although this approach has proven capable of modeling future flood scenarios, unfortunately these models are affected by changes in topography, initial or boundary conditions, or other parameters such as coefficients of friction, diffusion, or degradation (often inaccessible to direct measurement). Moreover, these models require detailed data at location – such as hydro-geomorphological data – requiring intensive calculations and thus making short-term forecasting difficult (Eslaminezhad *et al.* 2022). Previous research has demonstrated that physics-based models only have short-term predictive ability (Eslaminezhad *et al.* 2022; Prasad *et al.* 2022). One final challenge is that the establishment of these models requires a thorough understanding of hydrological parameters (Pham *et al.* 2021f).

The drawbacks of physics-based models have led many to the use of data-based models such as statistical modeling and machine learning. Statistical modeling has been used extensively to predict flood risk, with methods including frequency ratio (FR; Samanta *et al.* 2018; Tehrany & Kumar 2018), logistic regression (LR; Nandi *et al.* 2016), fuzzy logic (Pulvirenti *et al.* 2011; Perera & Lahat 2015), and weight of evidence (Rahmati *et al.* 2016). These models have proven effective in assessing flood susceptibility in several regions around the world. However, the inundation process is often nonlinear and time-varying, which presents a significant challenge for statistical models, particularly at the regional scale (Ha *et al.* 2022).

Some scientists have used machine learning, integrated with data from satellite images, to determine areas that are susceptible to floods. Machine learning is quicker and requires a smaller amount of input data than physics-based models. Moreover, it can resolve the nonlinear characteristics of past flood events just from the data; it does not need to understand the physical processes involved (Islam *et al.* 2021). The use of machine learning is predicated on the relationship within input dataset and flooding remaining unchanged in the future (Bui *et al.* 2019). Examples of this methodology include Support Vector Machine (Tehrany *et al.* 2014; Pham *et al.* 2019), Random Forest (RF; Lee *et al.* 2017; Chen *et al.* 2020), Adaboost (Hong *et al.* 2018a; Pham *et al.* 2021g), Artificial Neural Networks (Falah *et al.* 2019; Bui *et al.* 2020), and Adaptive Neuro-Fuzzy Logic (Hong *et al.* 2018b; Tabbussum & Dar 2021).

Several previous studies have claimed that machine learning is more appropriate than traditional approaches to flood prediction (Mosavi *et al.* 2017), but traditional machine learning methods are becoming less effective. A particular problem is that of overfitting: machine learning is good at prediction in the training process, because systems learn tasks based on data from the past, but if data are missing or insufficiently diverse, the systems cannot predict accuracy in the testing process (Mosavi *et al.* 2018; Bui *et al.* 2020; Chou *et al.* 2021).

In recent years, hybrid models have become a more popular way to resolve the problems of mapping flood susceptibility (Nguyen *et al.* 2021a; Yen *et al.* 2021). Hybrid models combine individual models with metaheuristic algorithms. The advantage is that hybrid models can eliminate the weak points of individual models, to obtain more accurate results (Saha *et al.* 2021). Moreover, metaheuristic algorithms explore the entire search space, thus limiting local optimization problems (Zhao *et al.* 2020). Optimization algorithms have been applied effectively in different domains, such as economics, earth science, and engineering. These can be divided into three groups: evolution-based – such as Genetic Algorithm (GA) (Shahabi *et al.* 2021) and DE (Razavi-Termeh *et al.* 2021) – physics-based – such as Henry Gas Solubility Optimization (HGSO; Nguyen *et al.* 2021b), Atom Search Optimization (ASO; Mundra *et al.* 2022) – and swarm-based – such as Whale Optimization Algorithm (WOA; Liu *et al.* 2020), GWO (Nguyen *et al.* 2021a), and Bat Algorithm (BA; Ahmadlou *et al.* 2019).

One example is that of Termeh *et al.* (2018), who integrated ANFIS with three optimization algorithms (Ant Colony Optimization, GA, and Particle Swarm Optimization (PSO)) to produce flood susceptibility maps for the Fars province of southern Iran. The hybrid models were successful, outperforming the reference models. Bui *et al.* (2019) combined the Extreme Learning Machine model with PSO to evaluate flood susceptibility in a mountain region of northern Vietnam. This combination displayed promising precision, with an AUC of over 0.95. Bui *et al.* (2020) utilized the swarm intelligence algorithms, namely GWO, Grasshopper Optimization Algorithm, and Social Spider Optimization algorithm to improve the performance of the DNN model to determine flood susceptibility in an area of Lai Chau province. The hybrid models outperformed the Support Vector Machine and RF reference models.

Such excellent results have inspired the implementation and development of hybrid models in the modeling of many natural hazards. However, key limitations of machine learning models remain, namely the generalization problem (i.e. models can

perform well above the range of the training dataset but cannot predict beyond the range) and the no-free-lunch problem (i.e. there are no models that can solve all problems in all regions).

The researchers recommend developing new ways to build flood susceptibility maps and related natural hazard simulations, and so the work of this study was to conceive of new hybrid models based on ANFIS and five optimization algorithms to predict flood susceptibility in Ha Tinh province, Vietnam. ANFIS is a popular neuro-fuzzy model for solving the highly complex problem of nonlinearity. In addition, ANFIS has the ability to adapt automatically to problems. ANFIS is based on the fuzzy model and is particularly strong at interpreting problems, which in turn improves its ability to generalize. ANFIS has gained recognition due to its strength in transforming fuzzy sets into clean inputs and providing clean output from fuzzy rules (Chopra *et al.* 2021). The model has been proven to be effective in several different areas, such as flood susceptibility construction and streamflow prediction (Adnan *et al.* 2022). However, ANFIS also has disadvantages: the selection of the type and the number of membership functions, as well as the position of the membership functions (Chopra *et al.* 2021). Therefore, it is necessary to integrate ANFIS with optimization algorithms like Gradient-Based Optimizer (GBO), Chaos Game Optimization (CGO), and Sine Cosine Algorithm (SCA). These optimization algorithms have proven effective in several different areas although, as yet, they have rarely been applied in the field of environmental science and natural hazard management.

We have created new methods, based on ANFIS and three novel optimization algorithms (GBO, CGO, and SCA), to generate spatial flood susceptibility maps with minimal field data for the study area, which is a province that is regularly affected by flooding. The performance of these models was compared with the reference models ANFIS-DE and ANFIS-GWO. Our initial hypothesis was that the three proposed models would successfully build a flood susceptibility map and that their performance would surpass the reference models ANFIS-DE and ANFIS-GWO.

The novel elements of this study are twofold. This is the first time flood susceptibility in Ha Tinh has been mapped in such a rigorous way; it is also the first time ANFIS has been integrated with GBO, CGO, and SCA. To the knowledge of the author, these hybrid models have not previously been used to construct a map, and so the finding of this study represents a novel approach to the use of deep-learning algorithms in generating flood susceptibility maps.

2. STUDY AREA AND DATA USE

2.1. Study area

The province of Ha Tinh lies in the Central region of Vietnam, whose geographical coordinates extend from 17°54′ to 18°37′ north and from 106°30′ to 105°07′ east (Figure 1). The average slope is 1.2%, and the area is divided into four main land types: high mountains, midlands, plains, and coastal areas. The mountainous terrain has an average height of 700–900 m and accounts for nearly 80% of the province's surface. The midlands have an elevation of 12.4–47.5 m, the plains 1.25–8.5 m, and the littoral zone −0.9–5.5 m.

Due to the combined effects of tropical depression, cold waves, and high slope, the orographic precipitation over the Central region of Vietnam is very high (Le *et al.* 2021). The study area lies in the tropical monsoon region and experiences two different seasons: the rainy season (November–March) and the dry season (April–October). The average annual precipitation for the period 1958–2018 was relatively high – between 1,142 and 4,391 mm, with 75% of the precipitation concentrated in the rainy season. Peak rainfall is in September and October, when the monthly average is 500–800 m, which is very high compared with other regions of Vietnam (Le *et al.* 2021). The study area has a dense river network with a total length of more than 400 km and an average annual flow of 195 m^3/s. There are three main river systems: Ngan Sau, Ngan Pho, and the coastal river system.

Forests currently cover about 66% of the surface of the province, although illegal deforestation, infrastructure construction, and forest fires are causing this figure to rapidly decrease. According to the Department of Agriculture and Rural Development, the forested area decreased by 1,728.17 ha between 2019 and 2020. This increases the flood risk in the study area. According to the Institute of Science and Technology, between 1961 and 2015, 44 storms affected Ha Tinh province. Storms can combine with heavy precipitation over a large area to result in serious floods. In October 2020, the province suffered two particularly destructive floods, which killed 6 people, destroyed 3,765 houses, and caused damage of VND 723 billion.

In recent years, the provincial government's flood strategy has focused on two main objectives: the adoption of structural measures – such as containment or flood mitigation through actions like the construction of dams – and the integration of
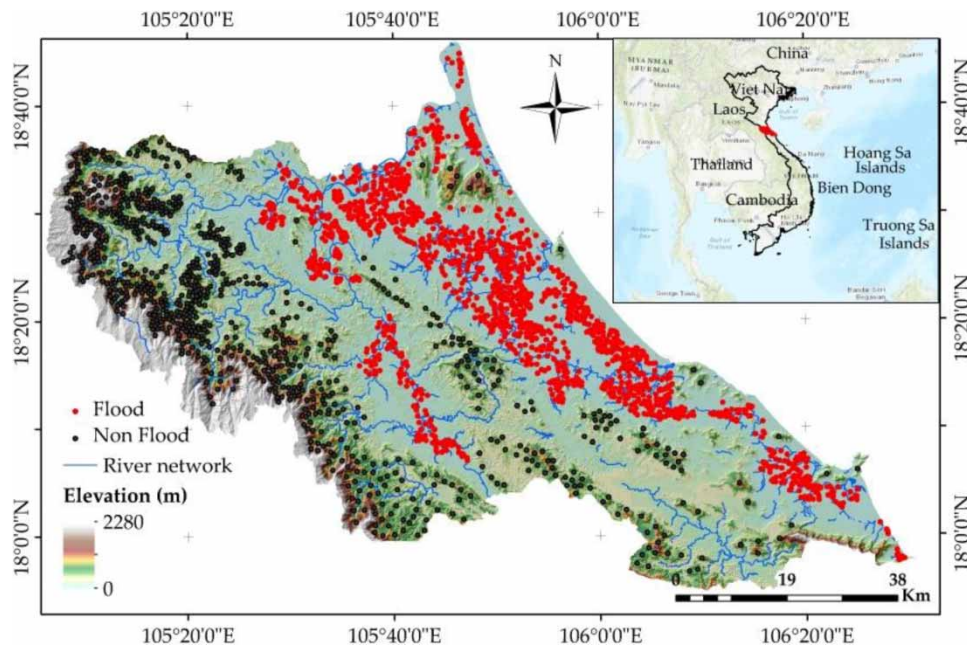
Figure 1 | Location of the study area.

flood depth maps, produced using hydrodynamic modeling, into land-use planning strategies to minimize future damage to human life and the economy. However, as mentioned above, hydrodynamic modeling has critical limitations that influence the accuracy of the flood depth map. Van den Honert & McAneney (2011) pointed out that the inability of hydrodynamic modeling to accurately predict floods had led to major damage in Queensland, Australia. Therefore, in this study, machine learning was selected for the construction of the map.

2.2. Geospatial data

2.2.1. Flood inventory

The first step when using a machine learning approach to create a flood susceptibility model is building the flood inventory map, which details past flood locations and relations with conditioning factors at those locations (Mojaddadi *et al.* 2017; Nachappa *et al.* 2020; Pham *et al.* 2021a). In this study, the locations of floods were collected from sources obtained from the Department of Natural Resources and Environment of Ha Tinh (DONRE of Ha Tinh). In addition, field missions were carried out in 2020 and 2021 to measure flood marks, and Landsat 8 OLI and Sentinel 1A satellite images were used to detect floodplains during flooding in October 2010 and October 2020. Details of the methodology were presented by Kumar (2019) and Tiampo *et al.* (2022). A total of 1,843 samples were acquired to establish the model for Ha Tinh, including 943 samples for flood and 900 samples for non-flood. Flood and non-flood inventory were coded as 1 and 0, respectively. These data were split into two sets: 70% of the dataset was utilized to create the model, 30% to verify the model.

2.2.2. Conditioning factors

The origin of flooding is a combination of geographical factors; however, the literature offers no universal guide to choosing the most appropriate. We selected 13 factors: elevation, aspect, slope, curvature, Normalized Difference Built-Up Index (NDBI), Normalized Difference Vegetation Index (NDVI), density to river, density to road, Compound Topographical Index (CTI), Topographic Wetness Index (TWI), land use, rainfall, and soil (Figure 2). Aspect, slope, curvature, CTI, and TWI were calculated using a DEM with a resolution of 10 m. The DEM was built using a topographic map with a scale of 1:10,000 m, which was obtained from the national Ministry of Natural Resource and Environment. NDBI and NDVI with 10 m resolution were computed from Sentinel 2A images from September 2021. Land use in 2020 and soil type, both with a 10 m resolution, were acquired from DONRE of Ha Tinh. Density to river and density to road, with a resolution of
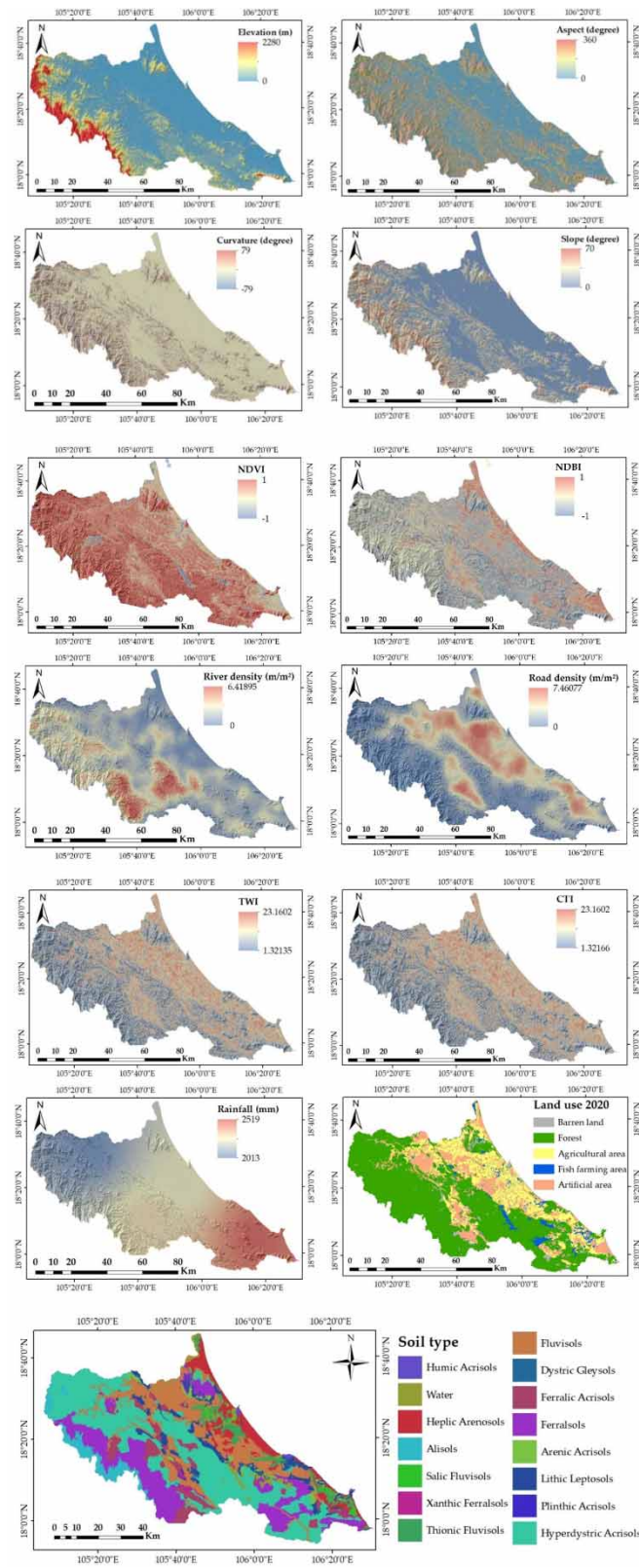
Figure 2 | Flood conditioning factors.

10 m, were extracted from the topographic map with a scale of 1:10,000 m. Rainfall data was downloaded from https://chrsdata.eng.uci.edu/ and resampled to 10 m.

Elevation is an important factor because it is linked to the capacity of water accumulation. Flooding is most likely to happen in lower-lying areas. Elevation affects flood regulation because with decreasing altitude, rainfall and river flow increase (Choubin *et al.* 2019; Nachappa *et al.* 2020). In the study area, the altitude value is 0–2,280 m.

Slope is the most vital conditioning factor, because it has effects on flow velocity and water accumulation capacity. Gentle slopes are usually more prone to flooding, as precipitation runs off steeper slopes into lowland water bodies (Yariyan *et al.* 2020). The slope value varied from 0° to 70° in this study.

Aspect is an important factor in hydrological response. It influences the local climatic conditions, soil moisture, and infiltration. Although aspect affects flooding indirectly, several researchers have considered it essential to susceptibility maps (Nachappa *et al.* 2020). The aspect value varied from 0° to 360° in this study.

Several researchers have underlined the critical role of *curvature* in flood susceptibility modeling since it directly impacts flow accumulation (Mirzaei *et al.* 2021). In Ha Tinh province, curvature varied from −79 to 79.

NDVI is one of the indices to measure vegetation density in the study area. The higher the vegetation density is, the lower the probability of flooding, and conversely (Dodangeh *et al.* 2020; Nachappa *et al.* 2020). The NDVI value ranged from −1 to 1 in this study.

NDBI calculates building density. This is important because buildings directly affect surface permeability (Nguyen *et al.* 2021a). The NDBI value ranged from −1 to 1 in this study.

Density to river was critical, as we were looking at fluvial flooding, where the river plays an essential role in flood expansion. Most areas near rivers are prone to flooding, as water overflows the riverbank (Vojtek & Vojteková 2019; Chowdhuri *et al.* 2020). The density to river value varied from 0 to 6.4 m in this study.

Density to road is also a primary factor as it directly affects the infiltration capacity of the surface (Ahmadlou *et al.* 2021; Linh *et al.* 2022). In Ha Tinh province, the national Highway 1A acts as a dam that blocks the flow of water toward the sea. The density to road value varied from 0 to 7.4 m in this study.

TWI describes soil moisture and the ability of the soil to erode spatially (Nachappa *et al.* 2020). The TWI value varied from 1.3 to 23.1 in this study.

CTI is the quotient between the slope and the flow accumulation. It describes the capacity of water resources per unit area. A low CTI value represents a basin with a steep slope and a small water surface (Azedou *et al.* 2021). The CTI value varied from 1.3 to 23.1 in this study.

Land use is key when predicting flood occurrence because each type of land use has a different water impermeability capacity (Andaryani *et al.* 2021; El-Haddad *et al.* 2021). Beckers *et al.* (2013) pointed out that the evolution of land cover is the main cause of increased flood risk. We categorized land use as either forest, water, agricultural area, artificial area, barren land, or fish-farming areas.

Various previous studies have highlighted the relationship between *precipitation* and flood occurrence. Rainfall is a trigger factor for flooding and an increase in the intensity of rainfall can lead to an increase in the intensity of a flood (Chapi *et al.* 2017; Pham *et al.* 2021f). The rainfall value varied from 2,013 to 2,519 mm in this study.

Although *soil type* influences the process of flooding indirectly, studies have shown its importance in modeling as it controls the process of water infiltration from the surface (Bui *et al.* 2019; Costache *et al.* 2020). Therefore, in this study, soil type was divided into humic acrisols, heplic arenosols, alisols, salic fluvisols, xanthic ferralsols, thionic fluvisols, fluvisols, dystric gleysols, ferralic acrisols, ferralsols, arenic acrisols, lithic leptosols, plinthic acrisols, and hyperdystric acrisols.

3. METHODS

Our methodology consisted of four steps: (i) preparation of input data, (ii) building the models, (iii) model validation, and (iv) construction of the flood susceptibility map (Figure 3).

(i) We collected flood inventory data from several field missions and from Landsat 8OLI and Sentinel 1A imagery. The 13 conditioning factors selected were elevation, aspect, slope, curvature, NDBI, NDVI, density to river, density to road, CTI,
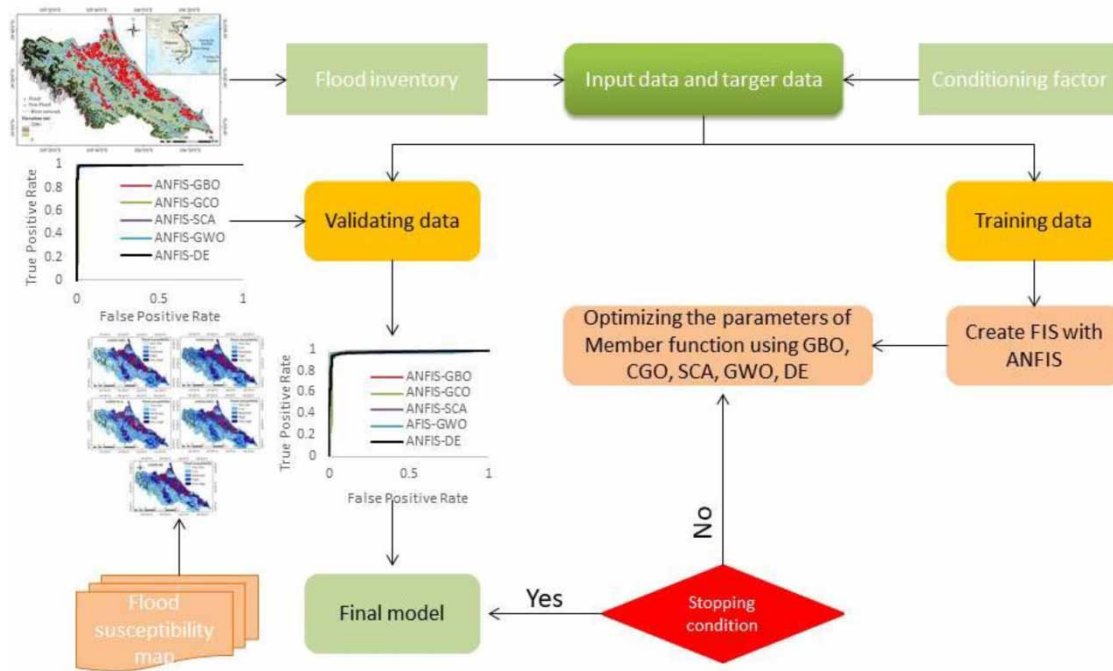
Figure 3 | Flowchart of methodology.

TWI, land use, rainfall, and soil type. Data were normalized using the min–max normalization technique:

$$X_{\text{Normalized}} = \frac{(x_i - \min)}{(\max - \min)}$$

In addition, we used RF to select important factors and filter out non-predictive factors, because data redundancy could increase the complexity of models and could reduce the accuracy of prediction models.

(ii) Two sets of data must be considered in any statistical model. The first is used to train the models and a separate second set is used to evaluate the models. In this study, a total of 1,843 flood and non-flood and 13 conditioning factor points were divided into sets for training (70%) and validation (30%). We used five different models to develop the flood susceptibility map using Python. For neural network usage, the network topology was determined by the dataset. The model parameter initialization process comprised two steps: the initialization of the ANFIS model hyper-parameters using the trial-and-error method and the initialization of the five optimization algorithms.

The structure of the ANFIS model comprises five layers, of which the first is the input layer, which consisted of 13 conditioning factors (see Figure 2). The final output layer contains two classes of results: flood and non-flood. The ANFIS model is influenced by hyper-parameters (number of fuzzy memberships, member of function, batch size, optimizer, and loss).

Optimization algorithms allow us to enhance the accuracy of the models by modifying the hyper-parameters. In this study, ANFIS models were developed using five optimization algorithms (GBO, CGO, SCA, GWO, and DE) to create the fuzzy interface system (FIS). After the generation of FIS, the parameters of the membership function were generated and stored in a master. In the ANFIS algorithm, both the fuzzy and defuzzy layers had parameters, while the rules and normalization layers did not. It should be noted that at each iteration, ANFIS was trained with the problem size created by the optimization algorithms. Each generated different solutions starting with the set of input weights and other specific parameters. RMSE was utilized as an objective function to help to determine the best optimization solution using five optimization algorithms. In the end, if the last condition of the model was met, producing the best result with a small RMSE, the optimization would stop. Otherwise, the optimization process was repeated until the model achieved the best results.

See Table 1 for the parameters of the proposed models.

The five proposed algorithms were used to calculate the weights of the model. The total value of the weights in the calculated models was 114,740.

(iii) The ROC curve is one of the techniques most often used in assessing the accuracy of a flood susceptibility model, by validating the dataset. The curve was generated by overlaying the validation dataset onto the prediction results of the five proposed models. AUC represents the accuracy of the models.

(iv) After constructing the proposed models, the map was generated by inputting into the models all the pixels of the study area, along with the value of conditioning factors. The output value is the flood susceptibility index and lies within a range of 0 to 1. These values were split for five classes (applying the natural break method): very low, low, moderate, high, and very high.

3.1. Adaptive Neuro-Fuzzy Inference System

ANFIS is an optimized FIS in an adaptive network. FIS is widely used in the solution of a variety of real-world problems, thanks to its ability to manage complex processes by applying IF-THEN rules. However, FIS is limited when providing accurate results for unforeseen situations. Therefore, FIS was optimized with an artificial neural network system to become ANFIS. This algorithm has advantages in solving linear load problems: FIS develops a model in a multilayer transition network, and the parameters of FIS can be established by learning from the input data by combining the method of least squares and reducing the backpropagation slope. The ANFIS network is composed of five main layers plus one input layer. If two fuzzy sets are associated with each input variable, then the system generates 16 inference rules. In Layer 1 (of fuzzification), each node behaves like a membership function (such as Gaussian or sigmoid). In Layer 2, each node realizes a fuzzy T-norm. The product operator is usually used. In Layer 3, the outputs of the T-norms are normalized. After the linear combination of input variables (in Layer 4), the predicted output is obtained in Layer 5 via a weighted average of the outputs of the different rules (Figure 4).

3.2. Gradient-Based Optimizer

GBO is a novel metaheuristic optimization based on the integration of Newton's method, gradient search rules, and local operators. This algorithm uses a set of vectors and two main operators (gradient search rule (GSR) and local escaping operators) to explore all the search space. GBO works in three main stages: initialization, GSR, and local escaping operator (LEO) (Ahmadianfar $et\ al.$ 2020). GSR helps GBO consider stochastic behavior in the optimization process to ease the exploration process and avoid local optimization. Also, in the algorithm, the direction movement has been added in GSR to improve the algorithm's convergence speed (Ahmadianfar $et\ al.$ 2020). Based on GSR and direction movement, the position of vector (x_n^m) is calculated by the following equation:

$$X1_n^m = x_n^m - randn \times p_1 \times \frac{2\Delta x \times x_n^m}{x_{\text{worst}} - x_{\text{best}} + \varepsilon} + rand \times p_2 \times (x_{\text{best}} - x_n^m)$$

Table 1 | The parameters of ANFIS optimized by GBO, CGO, SCA, GWO, and DE

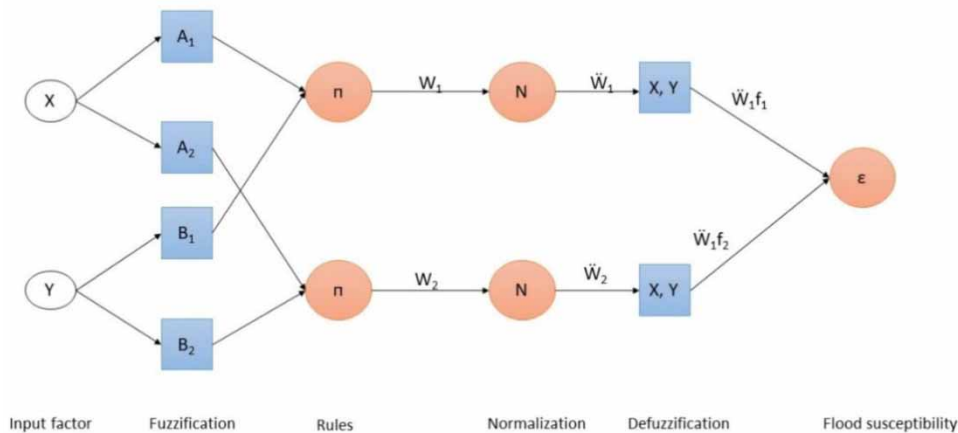| Parameters	ANFIS-GBO	ANFIS-CGO	ANFIS-SCA	ANFIS-GWO	ANFIS-DE
Inputs	13	13	13	13	13
Number of fuzzy memberships	2	2	2	2	2
Batch size	32	32	32	32	32
Member of function	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Optimizer	SGD	SGD	SGD	SGD	SGD
Loss	RMSE	RMSE	RMSE	RMSE	RMSE
Epochs	200	200	200	200	200
Problem size	114,740	114,740	114,740	114,740	114,740

Figure 4 | The structure of ANFIS for flood susceptibility models.

where n is the number of iterations, m is the total number of iterations, *randn* is normally a distributed random number, and ε is a small number with the value range (0, 0.1).

LEO enables GBO to solve complex, large-scale problems. LEO creates a solution with superior performance (X_{LEO}^m) using several solutions, including the best position (x_{best}). The solution is calculated by the following equation:

$$X_{\text{LEO}}^m = \begin{cases} X_n^{m+1} + f_1(u_1 x_{\text{best}} - u_2 x_k^m) + f_2 p_1 \left(u_3(X2_n^m - X1_n^m)\right) + \dfrac{u_2(x_{r1}^m - x_{r2}^m)}{2}, & \text{if } rand < 0.5, \\[2ex] X_n^{m+1} + f_1(u_1 x_{\text{best}} - u_2 x_k^m) + f_2 p_1 \left(u_3(X2_n^m - X1_n^m)\right) + \dfrac{u_2(x_{r1}^m - x_{r2}^m)}{2}, & \text{otherwise} \end{cases}$$

where f_1 is a uniform random number within the range $(-1, 1)$, and f_2 is a random number from a normal distribution with a mean of 0 and a standard division of 1. Pr is a probability value; u_1, u_2, u_3 are random numbers. More details of this algorithm were presented by Ahmadianfar *et al.* (2020). GBO has performed well in a variety of contexts, including economics (Deb *et al.* 2021) and energy (Ismaeel *et al.* 2021; Zhou *et al.* 2021). Prior to this study, this algorithm had not been applied in the field of geography.

3.3. Chaos Game Optimization

CGO is a metaheuristic optimization proposed by Talatahari & Azizi (2021) and is based on chaos theory. Chaos theory explains how a small change in initial conditions can cause an extreme deviation later. Based on this theory, the current state of a system can be used to determine the future state of that system. In mathematics, chaos play is the methodology of constructing fractals using an initial polygon and a randomly selected initial point. This process aims to create a sequence of points by applying the method iteratively (Talatahari & Azizi 2021). The main shape of the fractal is the vertex of a polygon, so the vertex of the polygon must be appropriately defined. Next, an initial point is randomly chosen to start the fractal construction. Subsequent points are determined based on the distance of the original point, and the remaining vertex of the polygon is randomly selected after each iteration. This process is repeated continuously and, with random selection of initial points and vertices of the polygon after each iteration, a fractal is constructed (Talatahari & Azizi 2021). The triangle was constructed:

$X = X1, X2, \ldots, \ Xi, \ Xn = x11, x12, \ldots, x1j, \ldots, x1d, \ x21, x22, \ldots, x2j, \ldots, x2d, \ \ldots, \ xi1, \ xi2, \ldots, \ xij, \ldots, xid, \ldots, \ xn1, xn2, \ldots, xnj, \ldots, xnd,$
$i = 1, 2, \ldots, n, \ j = 1, 2 \ldots, d.$

$$x_i^j(0) = x_{i,\min}^j + rand \cdot (x_{i,\max}^j - x_{i,\min}^j), \begin{cases} i = 1, 2, \ldots, n \\ j = 1, 2 \ldots, d \end{cases}$$

where d is the problem dimension, n is the total number of initializations, and $x_i^j(0)$ is the jth design variable at point i in the search space. $x_{i,\min}^j$ is the lower bounds of the decision variables, while $x_{i,\max}^j$ is the upper bound of the decision variables. *rand* is the random number, which ranked from 0 to 1.

More details of this algorithm can be found in Talatahari & Azizi (2021). CGO has already been successfully applied in fields such as engineering and energy.

3.4. Sine Cosine Algorithm

SCA was first proposed by Mirjalili (2016) and is a population-based metaheuristic. It uses sine and cosine to search for the global optimum from a set of random solutions and update the positions of these solutions in the direction of the optimal solution. Regions in the search space are explored when the values of the sine and cosine functions are greater than 1 or less than −1, whereas potential regions are exploited in the search space when the sine and cosine values are between −1 and 1. SCA has the advantage of discovering different regions in the search space in order to minimize the local optimization problem. In addition, this algorithm can also efficiently exploit potential regions in the search space (Talatahari & Azizi 2021). SCA was defined by the following equation:

$$X_{ij}^{t+1} = X_{ij}^t + r_1 * \sin(r_2) * |r_3 P_{ij}^t - X_{ij}^t|$$

$$X_{ij}^{t+1} = X_{ij}^t + r_1 * \cos(r_2) * |r_3 P_{ij}^t - X_{ij}^t|$$

X_{ij}^t represents the individual I at iteration t in the dth dimension; P_{ij}^t is the best position of the individual at iteration t in the dth dimension; and r_1, r_2, and r_3 are random parameters.

More details of this algorithm can be found in Gabis et al. (2021). SCA has performed well in contexts such as energy and earth sciences.

3.5. Grey Wolf Optimization

Grey Wolf Optimization (GWO) is a swarm intelligence proposed by Mirjalili et al. (2014). It is based on the predatory nature and social hierarchy of grey wolves. Grey wolves live in packs of 5–12 and are divided into four groups: alpha (α, the most dominant), beta (β), delta (δ), and omega (ω, the least dominant; Darabi et al. 2021). GWO works by dividing a set of solutions into four groups. The first three results are considered the best, belonging to α, β, and δ, and the fourth result belongs to ω. To implement this mechanism, in each iteration, GWO updates based on the three best solutions. The algorithm is divided into three clear processes: siege, hunt, and attack. GWO has a small number of parameters and is easy to implement. In addition, the convergence speed of GWO is faster than that of other layout algorithms (Zou et al. 2019). The digital model can be represented by the following equation:

$$D = |C \cdot X_{\text{victim}}(t) - X_{\text{wolf}}(t)|$$

$$X_{\text{wolf}}(t+1) = X_{\text{victim}}(t) - A * D$$

A, C, and D are defined as the coefficient vectors; $X_{\text{victim}}(t)$ and $X_{\text{wolf}}(t)$ are the current positions of the victim and the wolf, respectively, and t is the current iteration.

More details of this algorithm can be found in Liu et al. (2021). GWO has been successfully applied in many different applications including the prediction of forest fires, landslides, and floods.

3.6. Differential Evolution

Differential Evolution (DE) is a biologically inspired algorithm first introduced by Storn & Price (1997). The DE mechanism is divided into four main stages: (i) initialization, (ii) mutation, (iii) recombination, and (iv) selection (Storn & Price 1997). During the initialization phase, the model creates a primitive population by considering the maximum and minimum amplitudes. During the mutation stage, the first generation of infants is formed by crossover and mutation. The mutation phase also performs the construction of the partner vectors at random (Arora et al. 2021). This process is represented by the following equation:

$$V_{i,\,G+1} = X_{r1,\,G} + F \times (X_{r2,G} - X_{r3,G})$$

r_1, r_2, $r_3 \in \{0, 1, 2,\ldots, \text{NP}\}$ are the three random unequal integers. F is the coefficient and varies from 0 to 2.

The recombination is formed randomly after the partners are calculated and tested. This process is represented by the following equation:

$$u_{ji,\,G+1} = \begin{cases} V_{ji,\,G+1}, & \text{if } (rand) < C_r \text{ or } j = j_{rand} \\ X_{ji,\,G}, & \text{otherwise} \end{cases}$$

Then, the newborn set generation is compared with the mother set generation using objective functions, and the best members are selected and act as the next generation (Razavi-Termeh *et al.* 2021). The process of selection is represented by the following equation:

$$X_{ji,\,G+1} = \begin{cases} U_{ji,\,G+1}, & \text{if } f(u_{i,\,G+1}) \leq f(X_{i,G}) \\ X_{ji,\,G}, & \text{otherwise} \end{cases}$$

In the DE algorithm, all members have an equal chance of becoming parents, and the best members get to the next step compared with the parent generation. The DE algorithm has the advantage of not requiring any optimization function to achieve the best solution; however, to tune the assigned problems, the DE algorithm maintains a large number of existing solutions and initiates new solutions in parallel with existing solutions (Al-Sudani *et al.* 2019). More details of this algorithm can be found in Razavi-Termeh *et al.* (2021).

3.7. Accuracy assessment

Models must be evaluated and validated before they are used to make predictions. We used the indices receiver operating curve (ROC), RMSE, MAE, R^2, accuracy, and K to verify the accuracy of the proposed models. Ours is not the first study to use these indices (Arora *et al.* 2021; Razavi-Termeh *et al.* 2021). ROC is denoted by 1 – specificity on the X-axis and sensibility on the Y-axis. The advantage of this index is that there is no structural assumption that has been made about the shape of the graph, and there is no need to determine the underlying distribution of results for these two groups (Nguyen *et al.* 2022b). In addition, the performance of the model is measured by the area under the ROC curve.

$$\text{AUC} = \sum \text{TP} + \sum \frac{\text{TN}}{P} + N$$

TP constitutes the number of flood pixel classified correctly as flood, while TN is the number of non-flood pixel correctly classified as non-flood. P is the number of flood pixel. N is the number of non-flood pixel.

In machine learning, the evaluation of model performance is paramount, whether during training, validation, or testing. RMSE and MAE are the indices most commonly used for this. These indices are intuitive to understand, as well as compatible with several statistical assumptions (Karunasingha 2022). They evaluate the differences between observed and predicted values (Willmott & Matsuura 2005). These values range from 0 to 1. When the value is 0, it means that the model has no error.

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i-1}^{m} (Y_{i\text{predicted}} - Y_{i\text{observed}})^2}$$

$$\text{MAE} = \frac{1}{m} \sum_{i-1}^{m} |Y_{i\text{predicted}} - Y_{i\text{observed}}|$$

where m is the number of samples, $Y_{i\text{predicted}}$ is the prediction value at sample i, and $Y_{i\text{observed}}$ is the observation value at sample i.

Accuracy is the percentage of samples (flood and non-flood) that are classified correctly out of the total sample. Accuracy is considered a more reliable measure for the process of building a model. K is used to evaluate the percentage of consensus among categorical components. The main advantage of the kappa index is its simplicity and its ability to apply to multi-

class problems (Sadeghbeygi *et al.* 2021).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$K = \frac{P_p - P_{\exp}}{1 - P_{\exp}}$$

FP and FN are the number of flood and non-flood points that are incorrectly classified as flood and non-flood, respectively. P_p is the number of samples classified correctly by flooding or non-flooding. $P_{\exp}$ is the number of expected agreements.

R^2 is a popular statistical index and measures the accuracy of the linear regression model. R^2 is calculated as the square of the correlation coefficient between a dependent variable and one or more independent variables. It expresses the part of the variance of the dependent variable that comes from those of the independent variables. The value of R^2 varies from -1 to 1. If the value of R^2 is negative, the model is malfunctioning. If the value of R^2 equals 0, it means that the model cannot explain the variability of the response data around its mean. If the value of R^2 equals 1, the model is perfect. These are the main advantages of the R^2 index compared with other indices such as AUC, RMSE, or MAE (Chicco *et al.* 2021).

4. RESULTS

4.1. Feature selection analysis

Sensitivity analysis is one of the most important steps to take before constructing the flood susceptibility map, to balance the model, and to highlight important factors in the model building process (Andaryani *et al.* 2021). In the ANFIS approach to identify regions susceptible to flooding, 13 factors influencing the occurrence of flooding were selected, based on a review of the existing literature and an analysis of the sensitivity of each factor. Figure 5 shows the results of the sensitivity analysis using the RF method. The results showed that land use (0.25), elevation (0.2), slope (0.14), and distance from road (0.13) were most related to flood occurrence. In other words, when eliminating land use from the model, the accuracy of the model decreases by 25%, with corresponding decreases for elevation, slope, and distance from road. CTI (0.12), NDBI (0.04), soil (0.037), and TWI (0.03) demonstrate moderate levels of importance. Aspect (0.024), distance from river (0.009), rainfall (0.002), and curvature (0.0018) had the least influence. Remarkably, the NDVI factor value was equal to 0, meaning that it did not impact the prediction of floods in the study area at all. Therefore, this factor has been removed, as data redundancy can lead to increased model complexity and may reduce the accuracy of the flood susceptibility model.
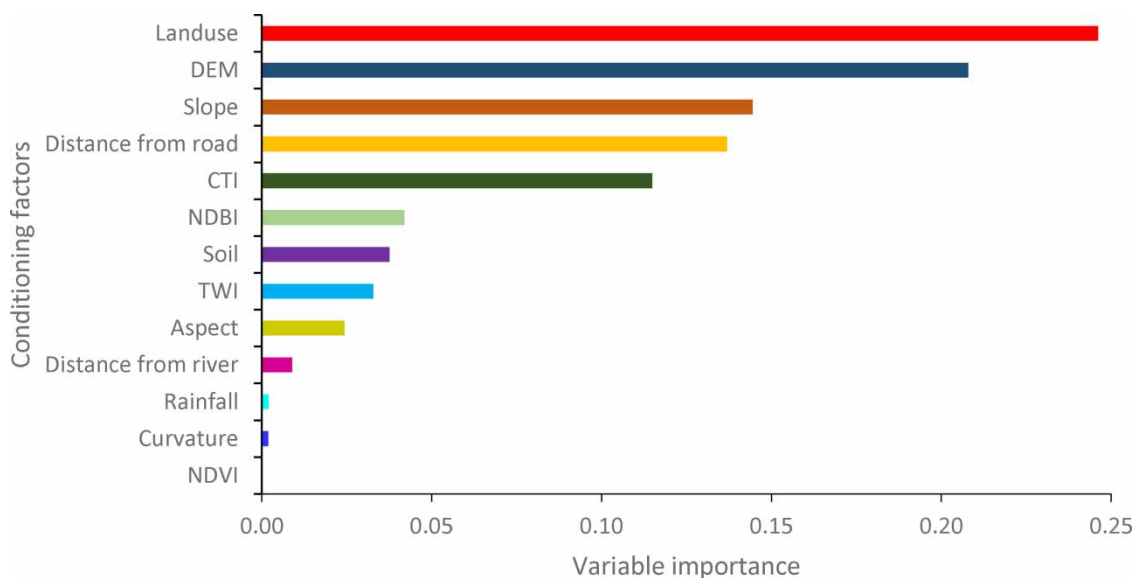

Figure 5 | The importance of conditioning factors using the RF model.

4.2. Model performance

Five models in this study were validated using ROC, based on training and validation points (Figure 6). Validation was done using the datasets with the formation points and the validation points. This is why the training and validation points have been identified on the flood susceptibility map. The ANFIS-SCA model resulted in the highest AUC value (0.996), followed by ANFIS-GBO (AUC = 0.994), ANFIS-GWO (0.994), ANFIS-DE (AUC = 0.993), and ANFIS-CGO (AUC = 0.992) in the training process. In the validation, the AUC value of the ANFIS-SCA model was 0.994, which exceeded ANFIS-GBO (AUC = 0.99), ANFIS-GWO (0.987), ANFIS-DE (0.984), and ANFIS-CGO (AUC = 0.98).
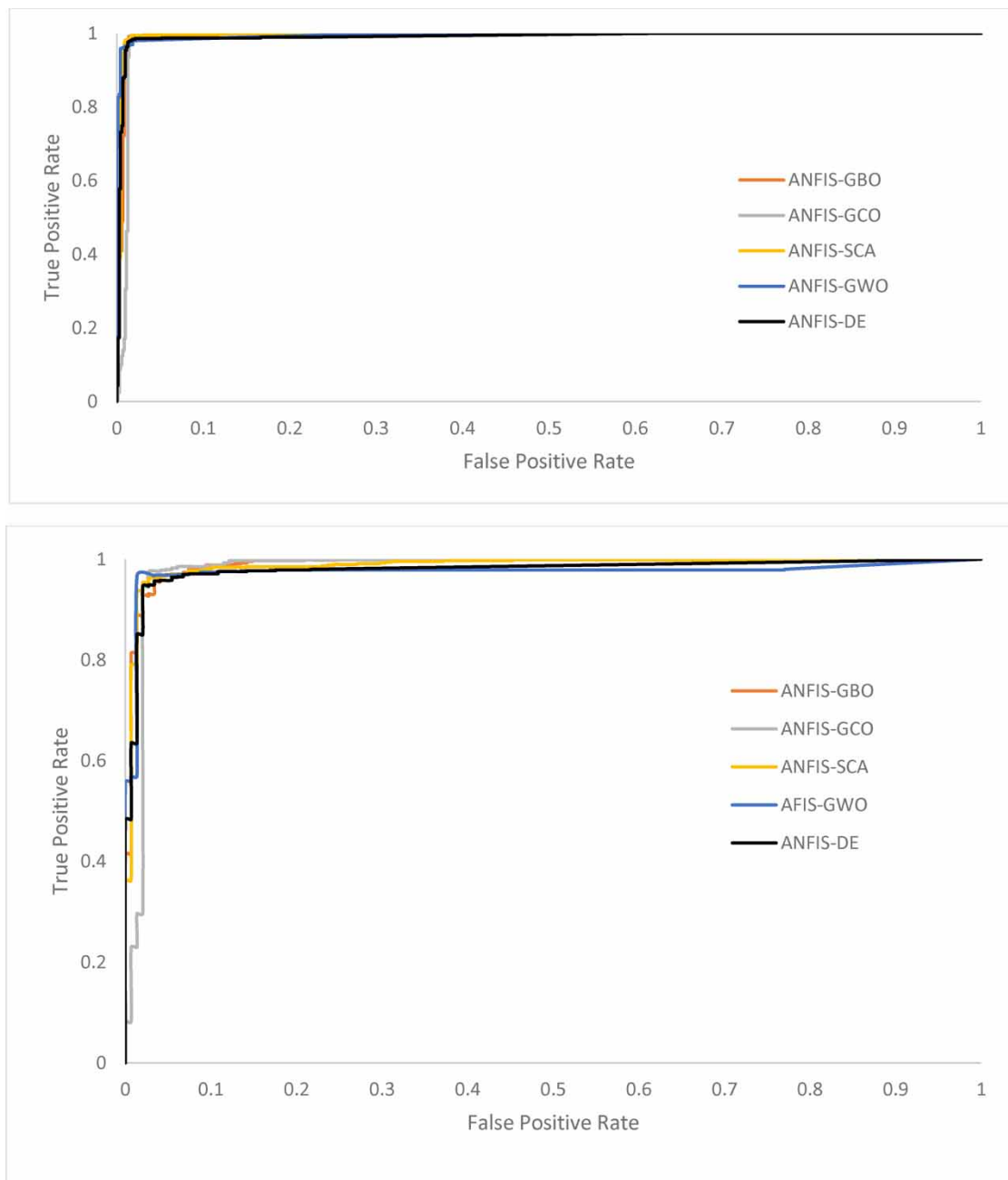

Figure 6 | The validation of five models using ROC based on training point and validation point.

The accuracy of the five models based on training and validation data was analyzed by the RMSE, MAE, ACC, K, and R^2 (Table 2). The results show that for the validation process, the proposed ANFIS-GBO model outperformed the other models, followed by ANFIS-SCA, ANFIS-GWO, ANFIS-DE, and ANFIS-CGO, respectively. In validation, the ANFIS-GBO model was also more precise than the other models, followed by ANFIS-SCA, ANFIS-GWO, ANFIS-DE, and then ANFIS-CGO. The ANFIS-GBO model had the best overall performance and ANFIS-CGO the poorest.

4.3. Flood susceptibility maps

The five machine learning models ANFIS-GBO, ANFIS-CGO, ANFIS-SCA, ANFIS-GWO, and ANFIS-DE were used to calculate flood susceptibility for each pixel in the study region. Figure 7 shows the flood susceptibility map produced by each of the five proposed models. Table 3 shows the distribution of each level of flood susceptibility of ANFIS-GBO, ANFIS-CGO, ANFIS-SCA, ANFIS-GWO, and ANFIS-DE. It can be seen that although there were small differences between the models, the areas affected by flooding at high and very high risk levels were located along the river and on the eastern coastal plain.

5. DISCUSSION

Accurate estimation of the level of flood susceptibility in a given area is vital to protect inhabitants and develop effective mitigation strategies. Flood susceptibility mapping has become central to global flood risk management efforts (Pham et al. 2021b; Shahabi et al. 2021; Towfiqul Islam et al. 2021). Powerful new techniques continue to be developed, improving the accuracy of results and better supporting those responsible for flood risk management. In this study, several models – namely ANFIS, ANFIS-GBO, ANFIS-CGO, ANFIS-SCA, ANFIS-GWO, and ANFIS-DE – were developed to prepare flood susceptibility maps for a region that is often affected by flooding but does not have in place adequately powerful or effective measures to mitigate against flood damage.

In this study, 1,843 historic locations of flooding were gathered both from satellite images and field missions. Several different techniques highlighted in previous studies have been used to detect flood locations. For example, we can derive indices like Enhanced Vegetation Index (EVI), NDWI, and Normalized Difference Surface Water Index (NDSWI) to detect flood locations using optical remote sensing data such as Landsat imagery (Tong et al. 2018; Du et al. 2021). However, these data may be influenced by cloud cover; they also suffer from limited spatial and temporal resolutions. Several researchers have applied radar images like Sentinel 1 to detect flood locations (Martinis et al. 2018). These images use long wavelengths so they are not influenced by atmospheres and can obtain information day and night. Moreover, these images have the advantage of being able to detect water and non-water areas using threshold determination. For these reasons, we used radar data to collect flood locations in this study.

This study utilized RF to assess the importance of each of the 13 factors used to create the flood susceptibility model. The importance depends on the study area's geo-environmental, climatic, hydrological, and anthropic activities, and on the methodology utilized. In addition, the flood susceptibility model used in this study was data-driven, so the importance of each factor depends on data distribution (Andaryani et al. 2021; Luu et al. 2021). Importance was calculated by assigning weights to each factor. The higher the weight, the more important the factor, and vice versa. If the weight equals 0, the factor does not influence the probability of the occurrence of the flood. Shahabi et al. (2021) pointed out that slope, distance to river, drainage

Table 2 | Model performance of ANFIS-GBO, ANFIS-CGO, ANFIS-SCA, ANFIS-GWO, and ANFIS-DE

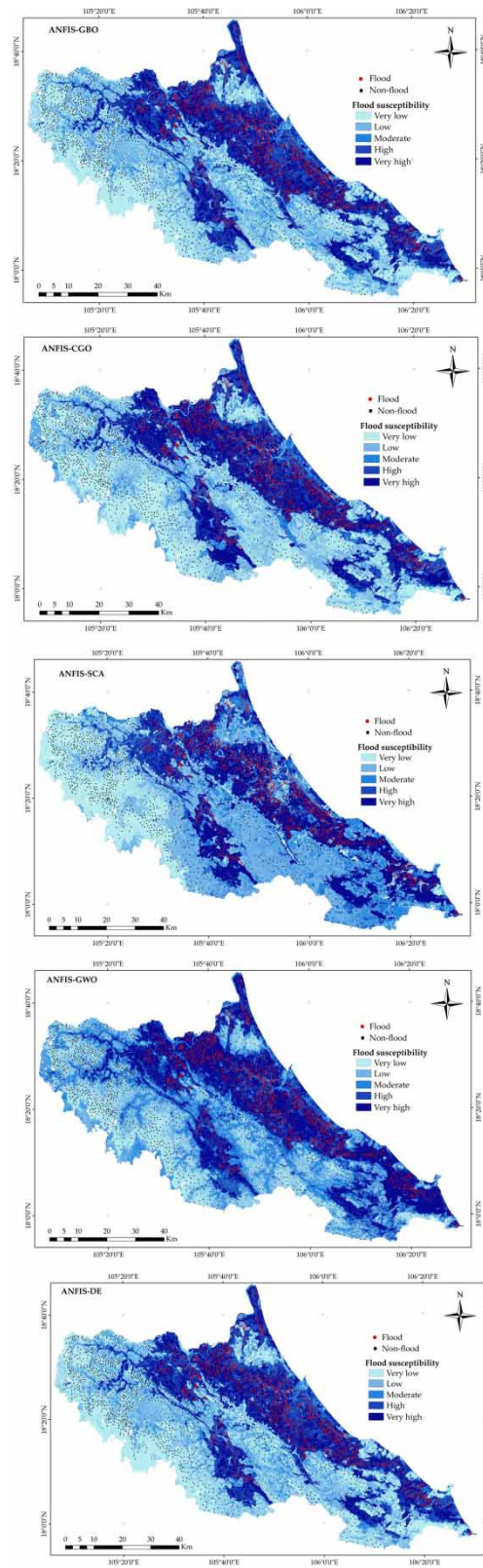|	Training dataset						Validating dataset					
	RMSE	MAE	AUC	ACC	K	R^2	RMSE	MAE	AUC	ACC	K	R^2
Proposed model												
ANFIS-GBO	0.14	0.09	0.994	0.98	0.97	0.91	0.23	0.16	0.99	0.96	0.86	0.63
ANFIS-CGO	0.2	0.15	0.992	0.98	0.96	0.82	0.26	0.2	0.98	0.95	0.83	0.54
ANFIS-SCA	0.17	0.12	0.996	0.98	0.97	0.88	0.25	0.17	0.99	0.95	0.85	0.58
Reference model												
ANFIS-GWO	0.17	0.14	0.994	0.98	0.97	0.87	0.25	0.17	0.987	0.95	0.85	0.57
ANFIS-DE	0.18	0.14	0.993	0.98	0.97	0.86	0.26	0.19	0.984	0.95	0.84	0.56

Figure 7 | Flood susceptibility maps in Ha Tinh province, Vietnam.

Table 3 | The distribution of each class of ANFIS-GBO, ANFIS-CGO, ANFIS-SCA, ANFIS-GWO, and ANFIS-DE

	Very low (%)	Low (%)	Moderate (%)	High (%)	Very high (%)
ANFIS-GBO	28.49	28.55	7.9	19.71	15.35
ANFIS-CGO	13.52	27.43	20.83	19.43	18.78
ANFIS-SCA	24.64	28.58	14.55	17.54	14.68
ANFIS-GWO	19.56	25.95	18.07	19.5	16.92
ANFIS-DE	18.79	24.14	18.59	18.65	19.82

density, and TWI were the most important of the 10 factors used to construct a flood map of Iran's Haraz watershed, using the ORAE technique. Pham et al. (2021a) reported that soil type, distance from river, river density, and geology were most influential on flood occurrence in Vietnam's Nghe An province, using the RF technique. Luu et al. (2021) used OneR to assess the importance of conditioning factors when building flood susceptibility models for the Quang Binh province in Vietnam; they pointed out that land use, geology, slope, and rainfall were the most important factors.

Our RF analysis showed that the most important factors that cause flooding in Ha Tinh province were, in descending order, land use (0.25), elevation (0.2), slope (0.14), and distance from road (0.13). These results are consistent with previous studies. Land use ranks highly because flooding can only occur in the region after the soil has been saturated. The clearing of forests in the mountains and the construction of infrastructure allowing for the development of floodplains are the main causes of aggravation of floods. The reduction in the regulation and retention capacity of lakes, which is directly related to land use, is a crucial factor in the intensification of floods. The results show that all residential and agricultural areas are located in places with high and very high flood susceptibility. Residential areas are more vulnerable to floods because they have impermeable surfaces. In addition, agricultural areas are located on the low-lying coastal plains. Elevation and slope came second and third, and are still important: these physical characteristics strongly influence the occurrence of flooding related to the determination capacity of surface runoff. The higher the elevation and the slope, the faster the flow velocity and the faster the water accumulates in the delta area (Abedi et al. 2021). Towfiqul Islam et al. (2021) used machine learning to assess the flood susceptibility in the Teesta River Basin of Bangladesh. The authors emphasized that land use, distance to road, elevation, and slope are the most important factors contributing to the probability of flood occurrence. El-Haddad et al. (2021) reported that distance to river, land use, lithology, and slope were the most influential factors on the likelihood of flood occurrence in Wadi Qena Basin of Egypt. Arora et al. (2021) showed that distance to river, curvature, slope, river density and land use were the most important factors for the flood susceptibility model in the Ganga River Basin of India. Pham et al. (2021c) used 16 conditioning factors to establish flood susceptibility maps for the Quang Nam province of Vietnam. Of the factors used, elevation, rainfall, slope, and land cover/land use had the biggest effect on the likelihood of flood.

Finally, there is a question linked to the importance of precipitation in the flood susceptibility model. Why is it less influential than other factors such as land use, elevation, slope, and distance from road, while still being a flood trigger factor? Based on different types of floods, an appropriate conditioning factor is chosen for the flood susceptibility model. Moreover, for each type of flood, the importance of each conditioning factor is different. Towfiqul Islam et al. (2021) and Saha et al. (2021) eliminated precipitation factors when assessing flood susceptibility in the Teesta River Basin in northern Bangladesh. Luu et al. (2021) reported that rainfall ranked fourth out of the ten factors used to establish flood susceptibility maps in Quang Binh province of Vietnam. In the study of Nguyen et al. (2021a), rainfall ranked sixth out of the 13 factors used to map flood susceptibility in Quang Ngai province.

In the Ha Tinh province, there are fluvial and coastal floodplains. Coastal flooding has less of an impact because Ha Tinh province is less influenced by a micro-tidal regime and storm surges, which are the two important factors that cause flooding. The river flooding in the area is caused by torrential rains in the mountainous area and the intensity of the rains tends to reduce when they reach the plain. Moreover, several previous studies have pointed out that flooding can only occur after soil saturation (Luu et al. 2021). That is why, in most cases, floods usually occur after heavy rains or rainy seasons. The national Highway 1A acts as a dam which stores the waters evacuated toward the seas. This is one of the important causes of flooding in Ha Tinh province. Therefore, the importance of rainfall is less than that of land use, elevation, slope, or distance from road. This has been substantiated by several previous studies. NDVI had less influence on the model. The importance of land use, elevation, slope, and distance from road can be related to the diversity of characteristics of

these factors. As mentioned above, the model applied in this study is data-driven; the distribution of the data strongly influences the results. So, the lower impact of NDVI may be due to the large area of vegetation (RF = 0).

The final result confirmed our initial hypothesis that the models proposed in this study would successfully build flood susceptibility maps for Ha Tinh and that they would outperform the reference models.

One of the advantages of metaheuristic algorithms is the usefulness of the stochastic system, which helps avoid local optimization problems in order to converge on a near-optimal solution (Lu *et al.* 2021). The objective here is not to obtain the best solution, but to have a near-optimal solution within a reasonable computation time. The system works first by exploration (to discover the promising spaces within which to search) and exploitation (the search for high-quality solutions in these promising spaces). Improving the performance of a metaheuristic algorithm requires a balance between the process of exploration and exploitation (Bui *et al.* 2020). The GBO and SCA algorithms had the advantage of a balance between these two processes (Abualigah & Diabat 2021; Deb *et al.* 2021). That is why two models, ANFIS-GBO and ANFIS-SCA, performed better than others.

Besides the simple structure, fast search speed, and high search precision, a big advantage of the GWO algorithm is that there are fewer setting parameters and it is easier to combine with other algorithms (Arora & Banyal 2021). The ANFIS-GWO model came third in terms of performance.

ANFIS-DE was fourth. Besides its simple structure and ease of use, DE has the advantage of avoiding the local optimization problem (Rout *et al.* 2013).

The ANFIS-CGO algorithm was less accurate than all the other models (RMSE = 0.26, MAE = 0.2, ACC = 0.95, K = 0.84, R^2 = 0.54). CGO faces key limitations in the process of exploring the search space. This leads to local optimization problems (Talatahari & Azizi 2021).

The ANFIS model and associated hybrids have been applied in various studies to predict the likelihood of flood occurrence around the world. Costache *et al.* (2020) used ANFIS and hybrids thereof to analyze flood susceptibility in the Trotuş River Basin in Romania. The precision of the results varied from 0.85 to 0.94 for the value of AUC. Termeh *et al.* (2018) evaluated flood susceptibility in the Jahrom Township in Fars province in Iran using ANFIS and three optimization algorithms (Ant Colony Optimization, GA, and PSO). In this case, the value of AUC ranged from 0.91 to 0.94. Another study (Hong *et al.* 2018b) used ANFIS and two optimization algorithms (GA and DE) to predict flood susceptibility in Hengfeng County in China. In this study, the maximum AUC value was 0.87. Vafakhah *et al.* (2020) analyzed flood susceptibility in the Gilan Province of Iran using ANFIS. The accuracy score was 63%. We can conclude that the accuracy of the models proposed in this study was consistent with the accuracy of previous significant studies.

The flood susceptibility levels in Ha Tinh province are also consistent with the flood susceptibility maps obtained using the AHP method (Nguyen *et al.* 2022a). They are also corroborated by previous studies showing the extent to which the coastal plains in the Central region of Vietnam are often subject to large floods (Luu *et al.* 2021; Pham *et al.* 2021e; Nguyen *et al.* 2022b). Therefore, the flood susceptibility maps constructed in this study can be seen to be a suitable alternative solution for local organizations responsible for assessing flood susceptibility.

Flood management is now a critical task. As the climate changes, sea levels rise, and floods become more common and more destructive. In Ha Tinh and elsewhere, informed land-use planners must limit new construction or the concentration of populations in areas with high and very high probability of flooding. The accuracy of the models in this study surpassed the reference models in previous studies. Therefore, the results can support planners to develop necessary strategies to diminish the damage. Although this study applies to Vietnam, the results may be applied to other regions around the world that are similarly affected by flooding.

6. CONCLUSION

In the context of climate change, the severity and violence of floods are increasing in Asia, and especially in Vietnam. Therefore, building a flood susceptibility model with high accuracy is one of the most important tools available to policy makers who are responsible for formulating strategies to reduce impact. This study proposed five hybrid models – ANFIS-GBO, ANFIS-CGO, ANFIS-SCA, ANFIS-GWO, and ANFIS-DE – to determine areas with probability of occurrence of flooding for Ha Tinh province in Vietnam.

ANFIS and its hybrids are the most powerful models to determine areas with probability of occurrence of flooding; however, the accuracy depends on the structure of the input data, so data preprocessing must be done properly.

This was the first time these five models have been used to build such maps, which represents the main novelty of this study. The models are characterized by high precision (AUC > 0.95; RMSE < 0.26) and complement the current literature on flood susceptibility. Additionally, the complete proposed models can be used to assess flood susceptibility levels in any region of the world, particularly where data are limited.

Ha Tinh province was divided according to five levels of flood susceptibility. Between 32 and 38% of the study area was located in the high and very high flood susceptibility zone. These regions are mainly concentrated along the river and on the eastern coastal plain, where there is a high density of population and infrastructure.

Once potential flood-prone areas have been identified, local government agencies can take appropriate action to reduce the extreme negative impacts of flooding. The results of this study can also be applied to the prediction and assessment of other types of disasters.

This study still has general limitations related to model input data. However, reducing uncertainties will further improve the accuracy of the predictions. More conditioning factors must be selected; then one of the selection factor methods must be used to remove the non-predicting factors. Furthermore, a limited level of detail is available, particularly regarding flood depth and velocity. Plus, the flood inventory used in this study used binary values (0, 1) to present flood and non-flood points, and did not include flood frequency. Therefore, the flood locations were given equal weight when used to predict flood susceptibility. In future research, we will record the flood frequency at each flood point and put that data to use.

Flooding is changing due to climate change and land-use changes; therefore, the application of machine learning to assess the effects of these phenomena on flooding is extremely useful for decision-makers and others tasked with building effective strategies to reduce future damage to property and loss of life.

FUNDING STATEMENT

No funding was received for this study.

AUTHOR CONTRIBUTIONS STATEMENT

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by H.D.N. The first draft of the manuscript was written by H.D.N. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abedi, R., Costache, R., Shafizadeh-Moghadam, H. & Pham, Q. B. 2021 Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto International* **37**, 1–18.

Abualigah, L. & Diabat, A. 2021 Advances in sine cosine algorithm: a comprehensive survey. *Artificial Intelligence Review* **54** (4), 2567–2608.

Adnan, R. M., Mostafa, R. R., Elbeltagi, A., Yaseen, Z. M., Shahid, S. & Kisi, O. 2022 Development of new machine learning model for streamflow prediction: case studies in Pakistan. *Stochastic Environmental Research and Risk Assessment* **36** (4), 999–1033.

Ahmadianfar, I., Bozorg-Haddad, O. & Chu, X. 2020 Gradient-based optimizer: a new metaheuristic optimization algorithm. *Information Sciences* **540**, 131–159.

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

Ahmadlou, M., Al-Fugara, A. k., Al-Shabeeb, A. R., Arora, A., Al-Adamat, R., Pham, Q. B., Al-Ansari, N., Linh, N. T. T. & Sajedi, H. 2021 Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks. *Journal of Flood Risk Management* **14** (1), e12683.

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., Wyser, K. & Feyen, L. 2017 Global projections of river flood risk in a warmer world. *Earth's Future* **5** (2), 171–182.

Al-Sudani, Z. A., Salih, S. Q. & Yaseen, Z. M. 2019 Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. *Journal of Hydrology* **573**, 1–12.

Andaryani, S., Nourani, V., Haghighi, A. T. & Keesstra, S. 2021 Integration of hard and soft supervised machine learning for flood susceptibility mapping. *Journal of Environmental Management* **291**, 112731.

Anusha, N. & Bharathi, B. 2020 Flood detection and flood mapping using multi-temporal synthetic aperture radar and optical data. *The Egyptian Journal of Remote Sensing and Space Science* **23** (2), 207–219.

Arora, N. & Banyal, R. K. 2021 Workflow scheduling using particle swarm optimization and gray wolf optimization algorithm in cloud computing. *Concurrency and Computation: Practice and Experience* **33** (16), e6281.

Arora, A., Arabameri, A., Pandey, M., Siddiqui, M. A., Shukla, U., Bui, D. T., Mishra, V. N. & Bhardwaj, A. 2021 Optimization of state-of-the-art fuzzy-metaheuristic ANFIS-based machine learning models for flood susceptibility prediction mapping in the Middle Ganga Plain, India. *Science of the Total Environment* **750**, 141565.

Azedou, A., Lahssini, S., Khattabi, A., Meliho, M. & Rifai, N. 2021 A methodological comparison of three models for gully erosion susceptibility mapping in the rural municipality of El Faid (Morocco). *Sustainability* **13** (2), 682.

Beckers, A., Dewals, B., Erpicum, S., Dujardin, S., Detrembleur, S., Teller, J., Pirotton, M. & Archambeau, P. 2013 Contribution of land use changes to future flood damage along the river Meuse in the Walloon region. *Natural Hazards and Earth System Sciences* **13** (9), 2301–2318.

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, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., Wang, X., Bian, H., Zhang, S. & Pradhan, B. 2020 Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. *Science of the Total Environment* **701**, 134979.

Chicco, D., Warrens, M. J. & Jurman, G. 2021 The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science* **7**, e623.

Chopra, S., Dhiman, G., Sharma, A., Shabaz, M., Shukla, P. & Arora, M. 2021 Taxonomy of adaptive neuro-fuzzy inference system in modern engineering sciences. *Computational Intelligence and Neuroscience* **2021** (1), 6455592.

Chou, T. Y., Hoang, T. V., Fang, Y. M., Nguyen, Q. H., Lai, T. A., Pham, V. M., Vu, V. M. & Bui, Q. T. 2021 Swarm-based optimizer for convolutional neural network: an application for flood susceptibility mapping. *Transactions in GIS* **25** (2), 1009–1026.

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.

Costache, R., Țîncu, R., Elkhrachy, I., Pham, Q. B., Popa, M. C., Diaconu, D. C., Avand, M., Costache, I., Arabameri, A. & Bui, D. T. 2020 New neural fuzzy-based machine learning ensemble for enhancing the prediction accuracy of flood susceptibility mapping. *Hydrological Sciences Journal* **65** (16), 2816–2837.

Darabi, H., Haghighi, A. T., Rahmati, O., Shahrood, A. J., Rouzbeh, S., Pradhan, B. & Bui, D. T. 2021 A hybridized model based on neural network and swarm intelligence-grey wolf algorithm for spatial prediction of urban flood-inundation. *Journal of Hydrology* **603**, 126854.

Deb, S., Abdelminaam, D. S., Said, M. & Houssein, E. H. 2021 Recent methodology-based gradient-based optimizer for economic load dispatch problem. *IEEE Access* **9**, 44322–44338.

Dodangeh, E., Choubin, B., Eigdir, A. N., Nabipour, N., Panahi, M., Shamshirband, S. & Mosavi, A. 2020 Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Science of the Total Environment* **705**, 135983.

Du, J., Kimball, J. S., Sheffield, J., Pan, M., Fisher, C. K., Beck, H. E. & Wood, E. F. 2021 Satellite flood inundation assessment and forecast using SMAP and Landsat. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **14**, 6707–6715.

El-Haddad, B. A., Youssef, A. M., Pourghasemi, H. R., Pradhan, B., El-Shater, A.-H. & El-Khashab, M. H. 2021 Flood susceptibility prediction using four machine learning techniques and comparison of their performance at Wadi Qena Basin, Egypt. *Natural Hazards* **105** (1), 83–114.

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* (Pourghasemi, H.R. ed.). Elsevier, Amsterdam.

Feng, S., Lu, H., Tian, P., Xue, Y., Lu, J., Tang, M. & Feng, W. 2020 Analysis of microplastics in a remote region of the Tibetan Plateau: implications for natural environmental response to human activities. *Science of the Total Environment* **739**, 140087.

Gabis, A. B., Meraihi, Y., Mirjalili, S. & Ramdane-Cherif, A. 2021 A comprehensive survey of sine cosine algorithm: variants and applications. *Artificial Intelligence Review* **54** (7), 5469–5540.

Ghorbanzadeh, O., Blaschke, T., Aryal, J. & Gholaminia, K. 2020 A new GIS-based technique using an adaptive neuro-fuzzy inference system for land subsidence susceptibility mapping. *Journal of Spatial Science* **65** (3), 401–418.

Giang, P. Q., Trang, N. T. M., Anh, T. T. H. & Binh, N. T. 2020 Prediction of economic loss of rice production due to flood inundation under climate change impacts using a modeling approach: a case study in Ha Tinh Province. *Vietnam Clim Change* **6**, 52–63.

Ha, M. C., Vu, P. L., Nguyen, H. D., Hoang, T. P., Dang, D. D., Dinh, T. B. H., Şerban, G., Rus, I. & Brețcan, P. 2022 Machine learning and remote sensing application for extreme climate evaluation: example of flood susceptibility in the Hue Province, Central Vietnam Region. *Water* **14** (10), 1617.

Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., Zhu, A.-X., Chen, W. & Ahmad, B. B. 2018a Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena* **163**, 399–413.

Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A.-X., Chen, W., Kougias, I. & Kazakis, N. 2018b Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Science of the Total Environment* **621**, 1124–1141.

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.

Ismaeel, A. A., Houssein, E. H., Oliva, D. & Said, M. 2021 Gradient-based optimizer for parameter extraction in photovoltaic models. *IEEE Access* **9**, 13403–13416.

Kadam, P. & Sen, D. 2012 Flood inundation simulation in Ajoy River using MIKE-FLOOD. *ISH Journal of Hydraulic Engineering* **18** (2), 129–141.

Karunasingha, D. S. K. 2022 Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences* **585**, 609–629.

Kumar, R. 2019 Flood inundation and hazard mapping of 2017 floods in the Rapti River Basin using Sentinel-1A Synthetic Aperture Radar images. In: *Applications and Challenges of Geospatial Technology* (Kumar, P., Rani, M., Chandra Pandey , P., Sajjad, H. & Chaudhary, B., eds.). Springer, Cham.

Le, T., Lan-Anh, N., Daskali, V., Verbist, B., Vu, K., Anh, T., Nguyen, Q., Nguyen, V. & Willems, P. 2021 Urban flood hazard analysis in present and future climate after statistical downscaling: a case study in Ha Tinh city, Vietnam. *Urban Water Journal* **18** (4), 257–274.

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.

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, D., Fan, Z., Fu, Q., Li, M., Faiz, M. A., Ali, S., Li, T., Zhang, L. & Khan, M. I. 2020 Random forest regression evaluation model of regional flood disaster resilience based on the whale optimization algorithm. *Journal of Cleaner Production* **250**, 119468.

Liu, R., Peng, J., Leng, Y., Lee, S., Panahi, M., Chen, W. & Zhao, X. 2021 Hybrids of support vector regression with grey wolf optimizer and firefly algorithm for spatial prediction of landslide susceptibility. *Remote Sensing* **13** (24), 4966.

Lu, P., Ye, L., Zhao, Y., Dai, B., Pei, M. & Tang, Y. 2021 Review of meta-heuristic algorithms for wind power prediction: methodologies, applications and challenges. *Applied Energy* **301**, 117446.

Luu, C., Von Meding, J. & Kanjanabootra, S. 2018 Assessing flood hazard using flood marks and analytic hierarchy process approach: a case study for the 2013 flood event in Quang Nam, Vietnam. *Natural Hazards* **90** (3), 1031–1050.

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.

Martinis, S., Plank, S. & Ćwik, K. 2018 The use of Sentinel-1 time-series data to improve flood monitoring in arid areas. *Remote Sensing* **10** (4), 583.

Mirjalili, S. 2016 SCA: a sine cosine algorithm for solving optimization problems. *Knowledge-based Systems* **96**, 120–133.

Mirjalili, S., Mirjalili, S. M. & Lewis, A. 2014 Grey wolf optimizer. *Advances in Engineering Software* **69**, 46–61.

Mirzaei, S., Vafakhah, M., Pradhan, B. & Alavi, S. J. 2021 Flood susceptibility assessment using extreme gradient boosting (EGB), Iran. *Earth Science Informatics* **14** (1), 51–67.

Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N. & Ghazali, A. H. b. 2017 Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Geomatics, Natural Hazards and Risk* **8** (2), 1080–1102.

Mosavi, A., Rabczuk, T. & Varkonyi-Koczy, A. R. 2017 Reviewing the novel machine learning tools for materials design. In *International Conference on Global Research and Education*. Springer, pp. 50–58.

Mosavi, A., Ozturk, P. & Chau, K.-w. 2018 Flood prediction using machine learning models: literature review. *Water* **10** (11), 1536.

Mukherjee, S., Mishra, A. & Trenberth, K. E. 2018 Climate change and drought: a perspective on drought indices. *Current Climate Change Reports* **4** (2), 145–163.

Mundra, P., Arya, A. & Gawre, S. K. 2022 An efficient model for forecasting renewable energy using ensemble LSTM based hybrid chaotic atom search optimization. *Neural Processing Letters* 1–23.

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.

Nandi, A., Mandal, A., Wilson, M. & Smith, D. 2016 Flood hazard mapping in Jamaica using principal component analysis and logistic regression. *Environmental Earth Sciences* **75** (6), 1–16.

Nguyen, H. X., Nguyen, A. T., Ngo, A. T., Phan, V. T., Nguyen, T. D., Do, V. T., Dao, D. C., Dang, D. T., Nguyen, A. T. & Nguyen, T. K. 2020 A hybrid approach using GIS-based fuzzy AHP–TOPSIS assessing flood hazards along the south-central coast of Vietnam. *Applied Sciences* **10** (20), 7142.

Nguyen, H. D., Nguyen, Q.-H., Du, Q. V. V., Nguyen, T. H. T., Nguyen, T. G. & Bui, Q.-T. 2021a 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, Q.-H., Chou, T.-Y., Yeh, M.-L., Hoang, T.-V., Nguyen, H.-D. & Bui, Q.-T. 2021b Henry's gas solubility optimization algorithm in formulating deep neural network for landslide susceptibility assessment in mountainous areas. *Environmental Earth Sciences* **80** (11), 1–10.

Nguyen, D.-L., Chou, T.-Y., Chen, M.-H., Hoang, T.-V., Tran, T.-P., Ngo, T.-S., Do Thi, N.-A., Bui, T.-H. & Nguyen, Q.-H. 2022a Flood susceptibility assessment in central region of Vietnam based on ensemble fuzzy logic and multivariate analysis. *Urbanism. Arhitectura. Constructii* **13** (1), 29–44.

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.

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.

Patro, S., Chatterjee, C., Mohanty, S., Singh, R. & Raghuwanshi, N. 2009 Flood inundation modeling using MIKE FLOOD and remote sensing data. *Journal of the Indian Society of Remote Sensing* **37** (1), 107–118.

Perera, E. D. P. & Lahat, L. 2015 Fuzzy logic based flood forecasting model for the Kelantan River basin, Malaysia. *Journal of Hydro-Environment Research* **9** (4), 542–553.

Pham, B. T., Prakash, I., Khosravi, K., Chapi, K., Trinh, P. T., Ngo, T. Q., Hosseini, S. V. & Bui, D. T. 2019 A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. *Geocarto International* **34** (13), 1385–1407.

Pham, B. T., Phong, T. V., Nguyen, H. D., Qi, C., Al-Ansari, N., Amini, A., Ho, L. S., Tuyen, T. T., Yen, H. P. H. & Ly, H.-B. 2020 A comparative study of kernel logistic regression, radial basis function classifier, multinomial naïve Bayes, and logistic model tree for flash flood susceptibility mapping. *Water* **12** (1), 239.

Pham, B. T., Jaafari, A., Phong, T. V., Yen, H. P. H., Tuyen, T. T., Luong, V. V., Nguyen, H. D., Le, L, H. V. & Foong, 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., Jaafari, A., Van Phong, T., Yen, H. P. H., Tuyen, T. T., Van Luong, V., Nguyen, H. D., Van Le, H. & Foong, L. K. 2021b 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., Phong, T. V., Trinh, P. T., Shirzadi, A., Renoud, S., Asadi, S., Le, H. V., von Meding, J. & Clague, J. J. 2021c Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? *Journal of Hydrology* **592**, 125615.

Pham, B. T., Luu, C., Van Dao, D., Van Phong, T., Nguyen, H. D., Van Le, H., von Meding, J. & Prakash, I. 2021d Flood risk assessment using deep learning integrated with multi-criteria decision analysis. *Knowledge-based Systems* **219**, 106899.

Pham, B. T., Luu, C., Van Phong, T., Nguyen, H. D., Van Le, H., Tran, T. Q., Ta, H. T. & Prakash, I. 2021e Flood risk assessment using hybrid artificial intelligence models integrated with multi-criteria decision analysis in Quang Nam Province, Vietnam. *Journal of Hydrology* **592**, 125815.

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. 2021f Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? *Journal of Hydrology* **592**, 125615.

Pham, Q. B., Pal, S. C., Chakrabortty, R., Norouzi, A., Golshan, M., Ogunrinde, A. T., Janizadeh, S., Khedher, K. M. & Anh, D. T. 2021g Evaluation of various boosting ensemble algorithms for predicting flood hazard susceptibility areas. *Geomatics, Natural Hazards and Risk* **12** (1), 2607–2628.

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.

Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P. & Holland, G. J. 2017 The future intensification of hourly precipitation extremes. *Nature Climate Change* **7** (1), 48–52.

Pulvirenti, L., Pierdicca, N., Chini, M. & Guerriero, L. 2011 An algorithm for operational flood mapping from Synthetic Aperture Radar (SAR) data using fuzzy logic. *Natural Hazards and Earth System Sciences* **11** (2), 529–540.

Rahmati, O., Pourghasemi, H. R. & Zeinivand, H. 2016 Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto International* **31** (1), 42–70.

Razavi-Termeh, S. V., Shirani, K. & Pasandi, M. 2021 Mapping of landslide susceptibility using the combination of neuro-fuzzy inference system (ANFIS), ant colony (ANFIS-ACOR), and differential evolution (ANFIS-DE) models. *Bulletin of Engineering Geology and the Environment* **80** (3), 2045–2067.

Rout, U. K., Sahu, R. K. & Panda, S. 2013 Design and analysis of differential evolution algorithm based automatic generation control for interconnected power system. *Ain Shams Engineering Journal* **4** (3), 409–421.

Sadeghbeygi, A., Moravej, K. & Delavar, M. A. 2021 Replacing kappa index with quantitative and spatial agreement and disagreement components for the accuracy assessment of different thematic maps. *Scientific-Research Quarterly of Geographical Data (SEPEHR)* **29** (116), 77–87.

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.

Samanta, S., Pal, D. K. & Palsamanta, B. 2018 Flood susceptibility analysis through remote sensing, GIS and frequency ratio model. *Applied Water Science* **8** (2), 1–14.

Schumann, G., Matgen, P., Hoffmann, L., Hostache, R., Pappenberger, F. & Pfister, L. 2007 Deriving distributed roughness values from satellite radar data for flood inundation modelling. *Journal of Hydrology* **344** (1–2), 96–111.

Shahabi, H., Shirzadi, A., Ronoud, S., Asadi, S., Pham, B. T., Mansouripour, F., Geertsema, M., Clague, J. J. & Bui, D. T. 2021 Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. *Geoscience Frontiers* **12** (3), 101100.

Shustikova, I., Domeneghetti, A., Neal, J. C., Bates, P. & Castellarin, A. 2019 Comparing 2D capabilities of HEC-RAS and LISFLOOD-FP on complex topography. *Hydrological Sciences Journal* **64** (14), 1769–1782.

Storn, R. & Price, K. 1997 Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* **11** (4), 341–359.

Sun, X., Mein, R. G., Keenan, T. D. & Elliott, J. 2000 Flood estimation using radar and raingauge data. *Journal of Hydrology* **239** (1–4), 4–18.

Tabari, H. 2020 Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports* **10** (1), 1–10.

Tabbussum, R. & Dar, A. Q. 2021 Performance evaluation of artificial intelligence paradigms – artificial neural networks, fuzzy logic, and adaptive neuro-fuzzy inference system for flood prediction. *Environmental Science and Pollution Research* **28** (20), 25265–25282.

Talatahari, S. & Azizi, M. 2021 Chaos Game Optimization: a novel metaheuristic algorithm. *Artificial Intelligence Review* **54** (2), 917–1004.

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** (13), 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.

Termeh, S. V. R., Kornejady, A., Pourghasemi, H. R. & Keesstra, S. 2018 Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Science of the Total Environment* **615**, 438–451.

Tiampo, K. F., Huang, L., Simmons, C., Woods, C. & Glasscoe, M. T. 2022 Detection of flood extent using Sentinel-1A/B synthetic aperture radar: an application for Hurricane Harvey, Houston, TX. *Remote Sensing* **14** (9), 2261.

Tong, X., Luo, X., Liu, S., Xie, H., Chao, W., Liu, S., Liu, S., Makhinov, A., Makhinova, A. & Jiang, Y. 2018 An approach for flood monitoring by the combined use of Landsat 8 optical imagery and COSMO-SkyMed radar imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* **136**, 144–153.

Towfiqul Islam, A. R. M., 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.

Vafakhah, M., Mohammad Hasani Loor, S., Pourghasemi, H. & Katebikord, A. 2020 Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping. *Arabian Journal of Geosciences* **13** (11), 1–16.

Van den Honert, R. C. & McAneney, J. 2011 The 2011 Brisbane floods: causes, impacts and implications. *Water* **3** (4), 1149–1173.

Vojtek, M. & Vojteková, J. 2019 Flood susceptibility mapping on a national scale in Slovakia using the analytical hierarchy process. *Water* **11** (2), 364.

Willmott, C. J. & Matsuura, K. 2005 Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* **30** (1), 79–82.

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.

Yen, H. P. H., Pham, B. T., Van Phong, T., Ha, D. H., Costache, R., Van Le, H., Nguyen, H. D., Amiri, M., Van Tao, N. & Prakash, I. 2021 Locally weighted learning based hybrid intelligence models for groundwater potential mapping and modeling: a case study at Gia Lai province, Vietnam. *Geoscience Frontiers* **12** (5), 101154.

Zhao, W., Zhang, Z. & Wang, L. 2020 Manta ray foraging optimization: an effective bio-inspired optimizer for engineering applications. *Engineering Applications of Artificial Intelligence* **87**, 103300.

Zhou, W., Wang, P., Heidari, A. A., Zhao, X., Turabieh, H. & Chen, H. 2021 Random learning gradient based optimization for efficient design of photovoltaic models. *Energy Conversion and Management* **230**, 113751.

Zou, Q., Liao, L., Ding, Y. & Qin, H. 2019 Flood classification based on a fuzzy clustering iteration model with combined weight and an immune grey wolf optimizer algorithm. *Water* **11** (1), 80.