Assessment of flood susceptibility areas using Optimized BP neural network models over the Tumen River Basin, China

He Chun Quan (hcquan@ybu.edu.cn)
Yanbian university https://orcid.org/0000-0002-4812-4198
Ri Jin
Yanbian University
Aifen Jin
Yanbian University
Guangri Jin
Yanbian University
Weihong Zhu
Yanbian University
Zhehao Lin
Yanbian University

Research Article

Keywords: Flood susceptibility, BP neural network, GIS, Natural hazards, Machine learning

Posted Date: April 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1510126/v1

Assessment of flood susceptibility areas using Optimized BP neural network models over the Tumen River Basin, China

Hechun Quan[a]*, Ri Jin[a]*, Aifen Jin[a], Guangri Jin[b], Weihong Zhu[a], Zhehao Lin[a]

[a] College of Geography and Ocean Sciences，Yanbian University，Yanji 133002，Jilin，China

[b] College of Engineering，Yanbian University，Yanji 133002，Jilin，China

ABSTRACT

Floods caused by rainstorms occur almost every year in Tumen River Basin. The floods caused severe damage to local residents, property and the ecological environment. For reducing the damage of floods, it is necessary to produce the flood susceptibility map. For mapping the flood prone areas, four algorithms such as genetic algorithm (GA), particle swarm optimization (PSO), artificial fish swarm algorithm (AFSA) and artificial bee colony algorithm (ABC) were used to optimize back propagation (BP) neural network respectively. The accuracy of the model was evaluated by Receiver Operating characteristic (ROC), root mean square error(RMSE) and mean absolute error (MAE). A total number of 222 flooded and un-flooded locations were identified. 155 locations of which were randomly selected to training the model, and the remaining 67 locations were used to validate the model. A total of thirteen flood conditioning factors，including altitude, slope, aspect, curvature, Topographic Wetness Index (TWI), Stream Power Index (SPI), Sediment Transport Index (STI), distance to river, landuse, rainfall, lithology, runoff coefficient and soil type were used in the proposed models. The area under the curve (AUC) obtained from ROC indicated that AFSA-BP showed a high accuracy (99.91%), and ABC-BP had the lowest prediction accuracy (97.28%). According to the generated flood susceptibility map, about 13% of Tumen River Basin is under very high flood susceptibility. The distance to river, rainfall, altitude and land use have the greatest impact on flood susceptibility. The results showed that the BP model optimized by the four algorithms is an effective and reliable tool for mapping the flood susceptibility. The flood susceptibility map can be used in flood prevention strategies.

Keywords: Flood susceptibility; BP neural network; GIS; Natural hazards; Machine learning

1. Introduction

Flood hazards result from a combination of physical exposure and human vulnerability to flooding. Floods are the most common and widespread of all natural disasters. It accounts for 55% of the deaths and 31% of the economic losses caused by natural disasters(Hong et al.，2018a; Chen et al.,2019).

Continued precipitation and changes in land use caused by human activities such as urbanization, deforestation and loss of wetlands may increase the risk of flooding in areas vulnerable to climate change. In order to effectively manage and reduce flood risk, it is very important to identify flood prone areas(Tehrany et al.，2015a). Flood susceptibility mapping (FSM) is considered an important step in the prevention of damage caused by flooding(Kourgialas and Karatzas, 2011). Because flood susceptibility can mitigate human and socio-economic losses by identifying flood vulnerable sites and providing people with sufficient lead time to respond to floods(Zhao et al.，2018).

In the current literature, the following main types of flood susceptibility assessment can be found:

i) Hydrologic and hydraulic models. These models provide information about the flood inundation extent, the water depth and the velocity based on 1-D or 2-D models (Gharbi et al., 2016; Hong et al., 2018b) such as MIKE system hydrological european(MIKE SHE)、HEC-RAS（Alaghmand et al.，2012）、TOPMODEL(Vincendon B et al.，2010)、soil and water assessment tool (SWAT) (Toosi et al.，2019).

ii) Empirical, statistical and probabilistic methods. It is mainly divided into qualitative and quantitative

* Corresponding author at: College of Geography and Ocean Sciences，Yanbian University，Yanji 133002，Jilin，China.
E-mail address: hcquan@ybu.edu.cn(HC Quan); rijin@ybu.edu.cn(R Jin)

methods. Analytic hierarchy process and fuzzy logic are the main qualitative techniques commonly used in flood susceptibility assessment. Frequency ration (FR) (Tehrany et al., 2014a; Rahmati et al., 2016a, Arabameri et al., 2019), weights-of evidence(WOE) (Rahmati et al., 2016a; Tehrany et al.,2014b; Hong et al., 2018a), Generalized Linear Model (GLM) (Chapi et al., 2017),logistic regression (LR) (Nandi et al.,2016; Hong et al., 2018a; Bui et al., 2019a) were widely applied as quantitative methods.

iii) Nonlinear machine learning algorithms. Artificial neural network (ANN) (Kia et al., 2012, Jahangir et al., 2019), decision tree (DT) (Tehrany et al., 2019), support vector machine (SVM) (Tehrany et al., 2015b; Hong et al., 2018a; Tehrany et al., 2019),naïve bayes (Liu et al., 2015), boosted regression trees(BRT)(Rahmati and Pourghasemi, 2017),random forest(RF) (Chapi et al., 2017; Hong et al., 2018a; Bui et al., 2019a), k-nearest neighbor (KNN) (Liu et al., 2016) and neural-fuzzy approaches (Bui et al., 2016).

iv）Hybrid and ensemble methods. Recently, some hybrid and integrated methods have been introduced in flood assessment. These methods are based on models combining various base classifiers. Hybrid and ensemble methods can improve predictive performance and deal with complex high-dimensional problems in a more robust way(Bui et al., 2019b). Typical examples of hybrid methods are the use of adaptive network-based fuzzy inference system (ANFIS) model and evolutionary and meta-heuristic optimization algorithms such as genetic algorithm (GA) (Termeh et al., 2018; Hong et al., 2018a), Differential Evolution particle(DE)(Hong et al., 2018a),particle swarm optimization (PSO) (Termeh et al., 2018), biogeographic based optimization (BBO)(Wang et al., 2019), imperialistic competitive algorithm (ICA) (Wang et al., 2019), bat algorithm (BA) (Ahmadlou et al., 2018),Ant colony algorithm (ACO) (Lai et al., 2016; Termeh et al., 2018) and firefly algorithm (FA) (Bui et al., 2018).

v) A method combining remote sensing and GIS with data-driven methods (Chen et al., 2017). Nowadays, due to the progress of remote sensing, geographic information system (GIS), machine learning (ML) and statistical model, it is very feasible to create more accurate flood susceptibility map. Models such as statistics, probability and machine algorithms combined with remote sensing (RS) and GIS have been used regularly by many researchers (Chapi et al., 2017; Zhao et al.,2018, Moghadam et al.,2018).

In addition to the selection of evaluation model, the selection of flood impact factors is also very important in the flood susceptibility mapping. There are many conditioning factors commonly used in the literature. In order to effectively improve the accuracy of susceptibility assessment, 13 factors that have a greater impact on flood were selected through the investigation of the paper and the field situation, including altitude, slope, aspect, curvature, TWI, SPI, STI, distance to river, landuse, rainfall, lithology, runoff coefficient and soil type. In this study， the flood susceptibility mapping was carried out by combining hybrid machine algorithm, remote sensing and GIS. Four algorithms such as GA, PSO, AFSA and ABC algorithm were used to optimize BP neural network respectively.

2. Study area

Tumen River Basin is located at the border of China, North Korea and Russia. It is bounded by Tumen River in the south, facing North Korea across the river and bordering Russia in the East. The geographical coordinates are 41°59′N～44°30′N、127°27′E～131°18′E, as shown in Fig. 1. The terrain of tumen River Basin is generally higher in the west and lower in the east. The central region is lower than the north and south. The total basin area is 33,168 km^2. The total river drop is 1297m with an average gradient of 1.2‰. The annual average temperature in the Tumen River Basin is between 2°C and 6°C. The annual average rainfall is in the range of 551-700mm. It belongs to the middle temperate humid monsoon climate with abundant and concentrated rain which is easy to produce flood disasters. The highest annual rainfall is more than 600mm in the north of tumen River Basin. The rugged topography, various geomorphology and the dense distribution of rivers lead to a high frequency of flood disasters under heavy rainfall conditions.

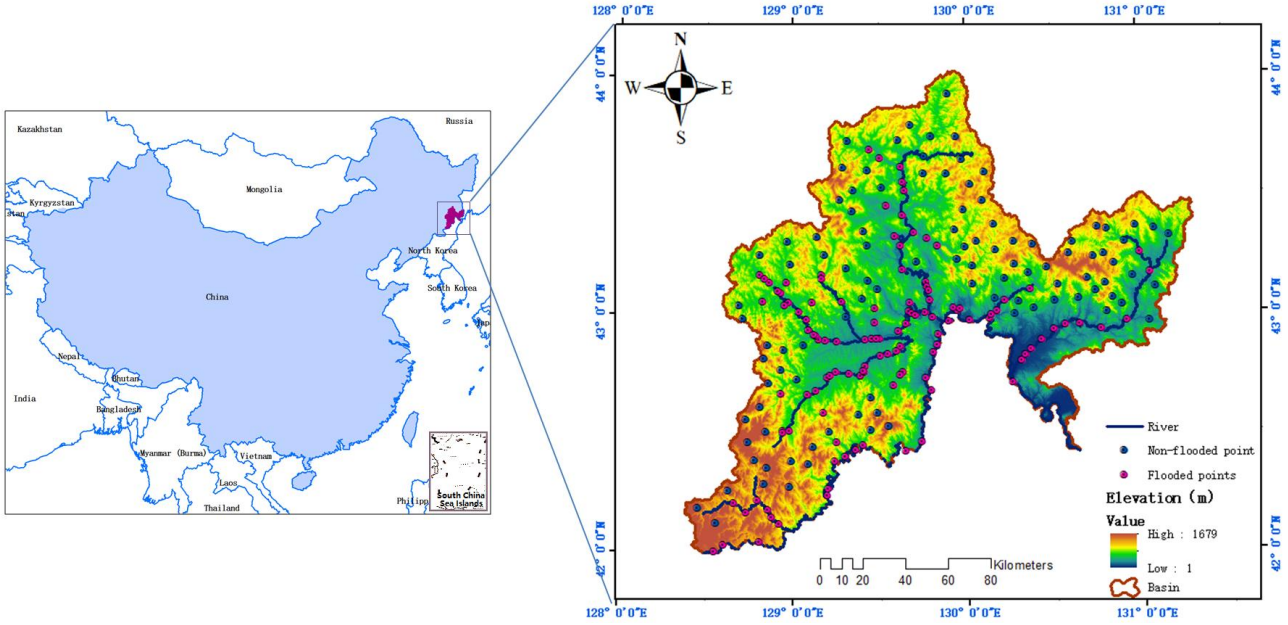


Fig.1. Location of the study area.

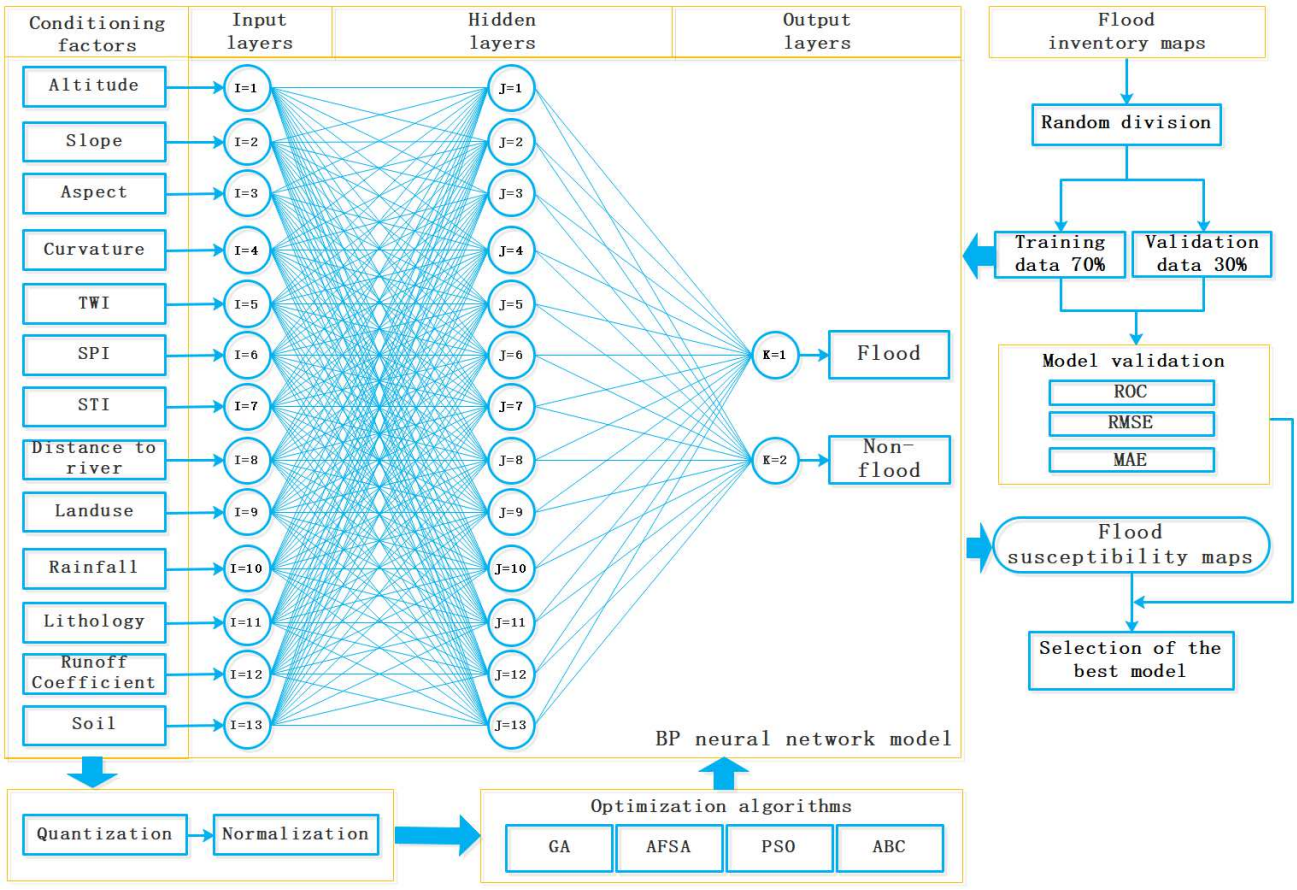


Fig.2. Methodological flow chart adopted in this study.

3. Materials and methodology

In this study, GA, PSO, ABC and AFSA were used to optimize BP neural network respectively for the flood susceptibility mapping in the study area. The main process consists of the following steps. i) The selection of impact factors； ii) Quantification and normalization; iii)the optimization of BP neural network by using training

data and four optimization methods; iv) The generation of flood susceptibility map; v) Analyze and verify the susceptibility map generated by the model using the test data; vi) Finally analyze the most effective BP neural network optimization method(Fig. 2).

3.1 Flood inventory

In order to effectively analyze the flood susceptibility of an area, it is necessary to analyze historical flood events. The inventory Map includes historical flood location, date, type and scale, etc., which is an important data for predicting flood occurrence and analyzing the relationship between conditioning factors. The flood inventory map of tumen River Basin was drawn based on literature and field investigation. A total of 222 flood locations are included, of which 119 points are flooded areas and 103 points are non-flooded areas. The data of the flood inventory is divided into training and test data, accounting for 70% and 30% respectively(Hong et al., 2018b).

3.2 Flood conditioning factors

The flood susceptibility is closely related to disaster formative environment. Which flood conditioning factors are used to analyze the formative environment is very important for the susceptibility mapping. According to the references（Termeh et al.，2018; Hong et al.，2018b; Yariyan et al., 2020; Andaryani et al., 2021）, data availability, geographical characteristics and expert opinions, thirteen factors such as altitude, slope, aspect, curvature, TWI, SPI, STI, distance to river, landuse, rainfall, lithology, runoff coefficient and soil type were selected as conditioning factors(Fig.3).

Topography is closely related to flood disaster and is the main factor in the hydrological process of the basin. The steep terrain enhances the rapid concentration of water flow(Destro et al.， 2018). Aster data were used to generate a digital elevation model (DEM) with a spatial resolution of 30m in the study area. Seven geomorphic factors were obtained from DEM which were altitude, slope, aspect, curvature, TWI, SPI and STI.

At higher elevations, flooding is relatively infrequent. However, the lower plains are at greater risk of flooding (Bui et al., 2019a). The altitude map generated using the DEM was divided into five categories: 0-275, 275-483, 483-688, 688-937 and >937m (Fig.3a).

Slope is an important factor in flood susceptibility mapping. Slope controls the rate of surface runoff and steep slopes increase runoff(Andaryani et al., 2021). The slope map was divided into five categories: 0-5, 5-12, 12-20, 20-30 and >30° (Fig.3b).

The aspect is closely related to the convergence and direction of water flow (Andaryani et al., 2021). The aspect map was generated by DEM and divided into nine categories: Flat, North, Northeast, East, Southeast, South, Southwest, West and Northwest (Fig.3c).

Curvature reflects the shape of the surface which can be divided into concave, flat and convex types. It has a significant impact on the ponding and infiltration of water in a given area(Cao et al., 2016). The curvature was generated by DEM and divided into three categories: convex, flat and convave(Fig.3 d).

TWI is a geomorphic parameter used to quantify local topographic relief and quantitatively evaluate runoff. TWI is the ratio of the area of a particular watershed to the slope. TWI reflects the spatial variation of wetness in a watershed. It describes the impact of topography on soil water distribution in a certain region(Rahmati et al.，2016b；Yariyan et al., 2020). TWI was divided into five categories: 0-4, 4-7, 7-10, 10-14 and >14 (Fig.3e).

SPI is an important hydrological factor. It can measure erosive power and the intensity of surface runoff. When its value is high, it indicates the power of the flood（Chapi et al., 2017). On the contrary, it indicates potential areas with flood storage within the basin (Termeh et al., 2018). SPI was divided into five categories: 0-50, 50-100, 100-150, 150-500 and >500 (Fig.3 f).

STI is a useful comprehensive topographic variable to characterize water and sediment transport in a particular

landscape. Areas with low STI value are mostly located in flat terrain and the risk of flood is relatively high(Hong et al.，2018a). The frequency of floods also receives impact of sediment dynamic processes. STI was divided into five categories: 0-5, 5-10, 10-30, 30-50 and >50 (Fig.3g).

TWI, SPI and STI are calculated by formula (1), (2) and (3) respectively:

$$\mathrm{TWI} = \ln(\mathrm{A_S}/\tan\beta) \quad (1)$$

$$\mathrm{SPI} = \mathrm{A_S}\tan\beta \quad (2)$$

$$\mathrm{STI} = (\mathrm{A_S}/22.13)^{0.6}(\sin\beta/0.0896)^{1.3} \quad (3)$$

Where As is the specific catchment area (m2m -1), and β (radian) is the slope gradient (in degrees).

Distance to river is one of the important factors in flood susceptibility mapping. After rainfall, large amounts of runoff quickly flow into rivers and cause flooding. The closer the exposure is to the river, the greater the risk of flooding (Mahmoud and Gan, 2018). According to the river and disaster situation of Tumen River basin, buffer zone was carried out for the main river. Distance to river categorized into five classes: (1) 0-400; (2) 400-800; (3) 800-1200; (4) 1200-1600; (5) > 1600m (Fig.3h).

Landuse map is essential to identify areas vulnerable to floods(Karlsson et al.，2017;Komolafe et al.，2018; Peng et al., 2019). Different types of land use have different effects on the formation of surface runoff. In particular, it is closely related to the velocity, interception and infiltration of runoff(Das and Gupta，2021). Under the same rainfall conditions, different land use types have different flood susceptibility. In this study, the land use map categorized into seven classes: water body, residential area, grassland, wetland, farmland, forest land and bare land(Fig.3i).

Rainfall is the main cause of flooding. Streams, ditches, rivers and other flood drainage pipes have limited drainage capacity. When the rainfall is too heavy, the inflow of the drainage pipe exceeds the displacement, resulting in flood disaster. The magnitude and intensity of rainfall determine the severity of flood disaster (Pham et al., 2019；Chițu et al., 2020). The data of precipitation are selected from the national average precipitation data provided by the Chinese Academy of Sciences. The natural breakpoint method was used to divide the layers of precipitation into five levels: 525-557, 557-579, 579-599, 599-619 and 619-653mm(Fig.3j).

The impermeability of lithology is one of the internal factors that determine the occurrence of flood disaster. Weak rocks are more susceptible to weathering and erosion and more permeable than hard rocks. The hard rock with weak permeability is easy to form surface runoff which is conducive to the occurrence of floods (Santos and Reis, 2018). According to the degree of weakness, lithology map was divided into six types: Metamorphic, Basalt, Sedimentary, Andesite, Diorite and Granite (Fig.3k).

The Runoff coefficient is mainly affected by the topography, watershed, slope, vegetation and soil characteristics of the catchment area. The larger the runoff coefficient is, the water is not easy to be absorbed by soil and the surface runoff is easy to form(Toosi et al.，2019). The runoff coefficient was divided into five levels: 0-0.24, 0.24-0.37, 0.37-0.5, 0.5-0.56 and >0.56 (Fig.3l).

Different types of soil have different effects on flood disasters. If there is more clay and less sand in the soil, the water permeability will be poor. This will easily produce surface runoff which will help the occurrence of flood disasters(Towfiqul Islam et al., 2021). The soil was divided into five levels: GrayBrown, Aeolian, Dark-brown earths, Albic, Meadow, Podzolic, Limestone, Litho, Bog, Paddy, Peat, Newly, deposited and Alluvial soil(Fig.3m).

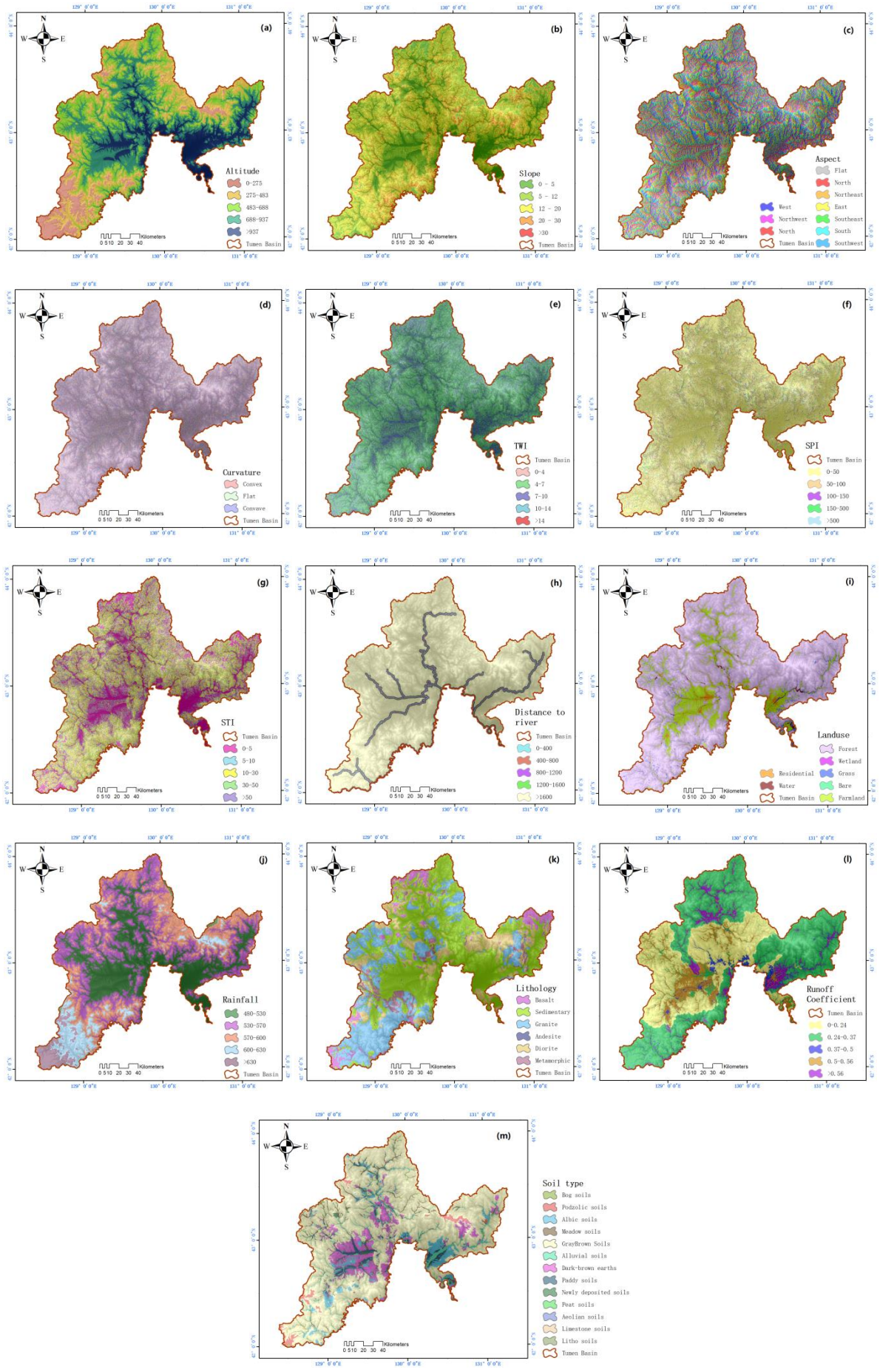


Fig.3. Flood conditioning factor (a)Altitude, (b)Slope, (c)Aspect, (d) Curvature, (e) TWI, (f) SPI, (g) STI, (h) Distance to river, (i) Landuse, (j) Rainfall, (k) Lithology, (l) Runoff Coefficient, (m) Soil type.

3.3 Methodology

3.3.1 BP neural network

BP neural network has strong non-linear data processing, fitting, learning and memory ability. It has strong modeling ability for complex nonlinear systems which is widely used in artificial intelligence, disaster prediction and many other fields. BP neural network is composed of input layer, hidden layer and output layer(Quan and Lee, 2012). The neurons between adjacent layers are connected, but the neurons between the same layer are not connected. In this paper, 13 conditioning factors were taken as input variables. The number of hidden layers was also set as 13. The number of nodes in the output layer is 2. In case of flood disaster, the output value is 1. Otherwise, the output value is 0. The transfer function between neurons is sinusoidal S-type.

In the operation of BP neural network, the unreasonable setting of initial weight often results in large error. GA, PSO, AFSA and ABC algorithms can find the optimal solution in the whole. So these algorithms are often used to optimize several models that are easy to fall into the local extremum. Combining BP neural network with the above algorithms is to learn from each other and make use of the advantages of their algorithms to mapping the flood susceptibility of the basin. It provides an effective and feasible method for accurately predicting the flood susceptibility.

3.3.2 Genetic algorithms（GA）

Genetic algorithm was first proposed by Professor J. Holland of Michigan University in 1975. It is an algorithm to search for the global optimal solution by simulating the natural evolution process. The global optimal solution was obtained by searching multiple individuals as possible solutions in the global range. In this paper, GA was used to optimize the parameters of BP neural network. The global optimal solution was searched to generate the optimal structure of BP neural network model. Genetic algorithm includes three basic operations: selection and reproduction, crossover and mutation (Chen et al., 2017). Genetic algorithms have been widely used to solve optimization problems and learning processes. Details on GA can be found in other related references (Termeh et al., 2018; Hong et al., 2018b).

3.3.3 Particle swarm optimization (PSO)

Kennedy and Eberhart proposed the classical PSO algorithm in 1995 and Shi et al modified it into the current standard PSO algorithm. Particle swarm optimization has fast convergence speed, strong robustness and local search ability and is easy to implement (Termeh et al., 2018； Bui et al., 2019a). So in this paper, PSO is used to improve the performance of BP neural network.

Firstly, the position and velocity of the initial particles constitute the initial population. The position and velocity of the ith particle in the n-dimensional solution space can be expressed as $Xi=（xi1,xi2,\dots,xid）$ and $Vi=（vi1,vi2,\dots,vid）$, respectively; Then, the optimal solution is found through successive iteration. In each iteration, the particle updates its speed and position by tracking two extreme values. An extreme value is the optimal solution of the particle itself so far. The individual extreme value can be expressed as $Pbi=（Pbi1,Pbi2,\dots,Pbid）$. Another extreme value is the optimal solution found so far in the neighborhood of the particle. The optimal particle can be expressed as $Nbesti=(Nbesti1,Nbesti2,\dots,Nbestid)$； Finally, the particle updates its velocity and position according to equations (4) and (5) :

$$\mathrm{V_i} = \mathrm{V_i} + \mathrm{c_i rand}(0,1)(\mathrm{Pbest_i} - \mathrm{X_i}) + \mathrm{c_i rand}(0,1)(\mathrm{Nbest_i} - \mathrm{X_i}) \quad (4)$$

$$\mathrm{X_i} = \mathrm{X_i} + \mathrm{V_i} \quad (5)$$

Where, c_1 and c_2 are acceleration constants. Rand (0, 1) is a random number between 0 and 1.

The flight speed of particles in each dimension needs to meet the following requirements:

$$Vi \leqslant V\max \quad (6)$$

Where: Vmax is the maximum speed of particles flying in each dimension. The smaller the Vmax, the stronger the local search ability of particle algorithm.

3.3.4 Artificial bee colony (ABC)

In 2005, karaboga group proposed the artificial bee colony algorithm to solve the multivariable function optimization problem. Artificial bee colony algorithm is an optimization method which imitates bee behavior and is an application of swarm intelligence(Karaboga et al., 2007; Zhang and Liu， 2015). The process of artificial bee colony algorithm is divided into four stages: initialization, recruitment, observation and detection.

Initialization: randomly generate SN food sources. The specific formula is shown in (7):

$$X_{i,j} = X_{min,j} + rand(0,1)(X_{max,j} - X_{min,j}) \quad (7)$$

Where the individual $i \in \{1,2... ,SN\}$, dimension $j \in \{1,2... ,D\}$, $x_{min,j}$ and $x_{max,j}$ represent the upper and lower bounds of each dimension randomly generated， Rand (0,1) is a random number between 0 and 1.

In this stage, in addition to initialization of food sources, adaptation values are also calculated, as shown in Formula (8) :

$$fit_i = \begin{cases} 1/(1+f_i) & f_i \geq 0 \\ 1+|f_i| & otherwise \end{cases} \quad (8)$$

Where, fit_i represents the adaptive value of food source, and f_i is the objective function value.

Recruitment：each hire bee will fly to a food source, and then search around the food source through Equation (9) to generate a new candidate solution:

$$V_{i,j} = X_{i,j} + \emptyset_{i,j}(X_{i,j} - X_{k,j}) \quad (9)$$

Where, X_i represents the selected food source to be modified, X_k represents the food source different from X_i, j represents the randomly selected dimension, $\phi_{i,j}$ represents the random number generated from [-1,1].

Greedy selection is made for newly generated candidate solutions V_i and X_i.

Observation: Observation bees obtained information about food sources by employing bees and calculated the probability of food selection. The specific formula is shown in (10) :

$$p_i = \frac{fit_i}{\sum_{j=1}^{SN} fit_i} \quad (10)$$

Detection: Select the food source i that has not been updated for the longest time. If it has not been updated more times than the preset fixed limit, the corresponding hire bees will transform into scout bees and randomly generate a new food source to replace food source i by formula (7).

3.3.5 Artificial fish swarm algorithm (AFSA)

AFSA is an algorithm that simulates the behavior of real fish swarm such as foraging, clustering and tail chasing and searches for the global optimal solution step by step from local to global(Luan et al., 2016). The main calculation steps are as follows:

i) Initialize the parameters. Including the total number of artificial fish M, crowding factor σ, swimming step, visual field range R_v, maximum foraging times N_{try}, and maximum iteration times N_{iter}.

ii) The initial population of M artificial fish was randomly generated. Calculate the initial state of all individuals in the initial fish swarm, and assign the optimal value to the bulletin board.

iii) Each artificial fish performed foraging behavior, clustering behavior and tail-chasing behavior in turn. Then update its status and compare it with the bulletin board. If it is better than the record on the bulletin board, the optimal value on the bulletin board is updated.

Suppose that the current state of the artificial fish is X_i, and the artificial fish X_i randomly selects a state X_j in the field of vision to calculate and compare the fitness values of the artificial fish. If Y_j is found to be better than Y_i (Y_j and Y_i are the fitness values of X_j and X_i respectively), X_i moves one step towards X_j by formula (11).

$$X_j = X_i + rand(0,1) \times V_s \quad (11)$$

Otherwise, the artificial fish X_i continues to select the next X_{next} within its field of view by formula (12).

$$X_{next} = X_i + rand(0,1) \times S_t \times (X_j - X_i)/\|X_j - X_i\| \quad (12)$$

Assume that the current position of a virtual artificial fish is X, V_S is the field of vision of the artificial fish, and position X_v is the viewpoint position of the artificial fish X_i at a certain time as shown in Fig. 4. If the food concentration at this position is higher than the current position, the artificial fish moves forward one step and reaches X_{next}; otherwise, the artificial fish X continues to seek other positions in the field of vision to judge whether the forward condition is met. After repeated attempts T_r times, if the forward condition is still not met, the artificial fish will randomly continue to move forward one step by formula (13).

$$X_{next} = X_i + rand(0,1) \times S_t \quad (13)$$

iv) The number of iterations is added by one and compared with the maximum number of iterations N_{iter}. If the maximum number of iterations is not reached, step (iii) is returned; otherwise, the optimization process is ended and the results are output.

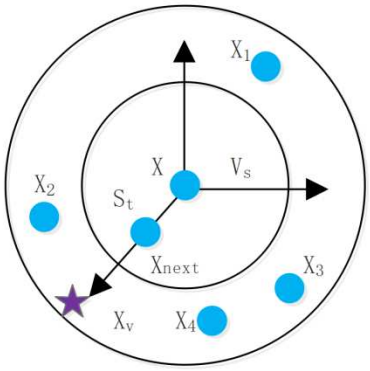


Fig. 4. Schematic diagram of visual field and moving step of artificial fish.

4. Results and discussion

4.1 Selection of flood-conditioning factors

The possible multicollinearity among conditioning factors was assessed using VIF, Tolerances and IGR. If the value of VIF exceeds 10 or tolerance is less than 0.1, possible multicollinearity is considered (Khosravi et al., 2018). VIF and tolerance values of each factor were less than 10 and greater than 0.1, respectively (Table 1), indicating that there was no multicollinearity problem among the factors. IGR is used also to evaluate the importance of various factors in flood simulation (Table 1). The results show that the distance from the river (IGR = 0.561) has the greatest impact on flood occurrence, followed by land use (IGR = 0.454), rainfall (IGR = 0.433), elevation (IGR = 0.419), curvature (IGR = 0.411), runoff coefficient (IGR = 0.374), soil type (IGR = 0.322), slope (IGR = 0.313), STI (IGR = 0.275), TWI(IGR = 0.231), SPI (IGR = 0.174), Lithology (IGR = 0.076), aspect (IGR = 0.048).

Table 1

Multicollinearity analysis.

Index	VIF	Tolerance	IGR
Altitude	4.968	0.201	0.419
Slope	5.028	0.199	0.313
Aspect	1.038	0.963	0.048
Curvature	1.319	0.758	0.411
TWI	2.238	0.447	0.231
SPI	4.754	0.210	0.174
STI	8.629	0.116	0.275
Distance to river	1.848	0.541	0.561
Landuse	3.520	0.284	0.454

Rainfall	3.837	0.261	0.433
Lithology	1.079	0.927	0.076
Runoff coefficient	2.462	0.406	0.374
Soil type	1.667	0.600	0.322

The factors that have a great impact on flood susceptibility include altitude, distance to river, land use and rainfall. Areas with low altitude, close to the river and small topographic relief have high risk and susceptibility of flood disaster. The area with small rainfall, high altitude， large topographic relief and woodland are not conducive to the occurrence of flood. Therefore, the susceptibility level in these areas is low and relatively safe.

4.2 Model validation and comparison

In this study, RMSE, MAE, Kappa, Precision and ROC curves were used to verify the accuracy and predictive validity of the four models. The comparison results showed that AFSA-BP was the optimal model with the highest kappa (96.9%), precision (98.51%), lowest RMSE (0.159) and MAE (0.103) in the validation stage (Table 2). The second optimal model was PSO-BP with the value of kappa (96.9%), precision (98.51%), RMSE (0.228) and MAE (0.161). Similarly, kappa and precision showed that the optimal models were AFSA-BP and PSO-BP in the training stage. In the training and validation stage, the lowest Kappa index is 0.906. It indicated that the prediction accuracy of the four optimization models was relatively high on the whole.

Table 2

Comparison of model performances(using training and validation dataset).

Index	Training dataset				Validation dataset			
	GA-BP	PSO-BP	AFSA-BP	ABC-BP	GA-BP	PSO-BP	AFSA-BP	ABC-BP
TN	60	61	61	58	39	40	40	39
TP	90	91	92	90	25	26	26	25
FN	3	2	1	3	1	0	0	1
FP	2	1	1	4	2	1	1	2
Precision	96.77	98.06	98.71	95.48	95.52	98.51	98.51	95.52
Sensitivity	96.77	97.85	98.92	96.77	96.15	100.00	100.00	96.15
Specificity	96.77	98.39	98.39	93.55	95.12	97.56	97.56	95.12
AUROC	98.94	99.13	99.91	97.28	99.34	99.91	99.91	99.25
RMSE	0.200	0.172	0.181	0.192	0.203	0.228	0.159	0.216
MAE	0.137	0.110	0.125	0.082	0.142	0.161	0.103	0.104
Kappa	0.933	0.960	0.973	0.906	0.906	0.969	0.969	0.906

The ROC curve obtained from the training and validation data (Chen et al.,2019; Towfiqul Islam et al., 2021) is shown in Fig.5. The detailed parameters of the ROC curve are shown in Table 2. The results show that the AUC of AFSA-BP model is the highest with the value of 0.9991. The second was PSO-BP model with AUC value of 0.9913, followed by GA-BP model with AUC value of 0.9894, and the lowest was ABC-BP model with AUC value of 0.9728. From the analysis results, the four models all showed reasonable ROC curves. Among them, AFSA-BP model has the best performance in predicting flood susceptibility followed by PSO-BP, GA-BP and ABC-BP models. The ROC diagram obtained from the validation data showed that the AUC values of AFSA-BP and PSO-BP models were the highest with the value of 0.9991. GA-BP model followed with an AUC value of 0.9934 and ABC-BP model had the lowest AUC value of 0.9925. The results showed that the AFSA-BP and PSO-BP models have the best

performance in predicting flood susceptibility followed by GA-BP and ABC-BP models. It is consistent with the results of Kappa and other evaluation indicators.

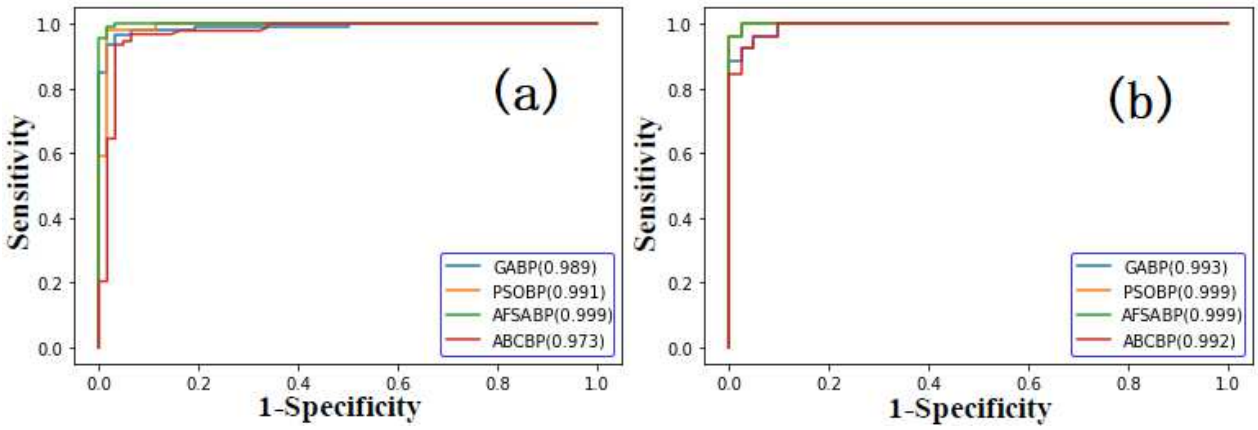


Fig. 5. ROC curves and AUC values of models. (a) Training dataset; (b) Validation dataset.

4.3 Generation of flood susceptibility maps

The four flood susceptibility maps generated by AFSA-BP, PSO-BP, GA-BP and ABC-BP models are shown in Fig. 6. Natural breakpoint method was used to reclassify the susceptibility maps into five flood susceptibility levels (Chen et al., 2019; Das and Gupta，2021): extremely low, low, medium, high and extremely high. The area percentages of each level in the four flood susceptibility maps are shown in Fig. 7. In AFSA-BP model, the area percentage of low category was the very largest (43.85%), followed by low category (38.64%), very high category (10.46%), moderate category (4%), and high category (3.06%). For the PSO-BP model, the area percentages of very low, low, moderate, high and very high grades were 46.55%, 34.21%, 9.27%, 3.02% and 6.95% respectively. In GA-BP model, the area percentages of very low, low, moderate, high and very high grades were 41.67%, 19.64%, 16.57%, 6.13% and 15.99% respectively. In the ABC-BP model, the percentage of very low, low, moderate, high and very high grades were 49.42%, 28.08%, 3.71%, 2.98% and 15.81% respectively. Four flood susceptibility maps were compared using extremely high and high FR values. The results showed that FR value was the highest in PSO model (9.97), followed by AFSA model (8.16) and ABC model (6.39), and the lowest was GA model (5.92) as shown in Table 3.

This result mainly depends on the advantages and disadvantages of each model. PSO algorithm has no crossover and mutation operation and relies on particle velocity to complete the search. In iterative evolution, only the optimal particles transmit information to other particles. PSO algorithm has the advantages of fast search speed, memorization, few parameters to be adjusted, simple structure and easy implementation.

AFSA does not require the objective function to have special analytical properties. It has the characteristics of parallel processing ability and good randomness. However, AFSA is not perfect and immature compared with GA, ANN and PSO. The study of AFSA is in the preliminary stage and many problems such as initial parameter selection and convergence rate need to be further solved.

As a promising bionic algorithm, ABC has strong global optimization ability and fast convergence speed. It has the advantages of few control parameters, strong robustness and easy implementation. However, it still has the disadvantage of poor local development ability. When approaching the global optimal solution, it is easy to fall into local optimal solution.

GA has the ability of fast random search. The search starts from the group and has the potential parallelism which can compare multiple individuals simultaneously. It is extensible and easy to combine with other algorithms. But the programming implementation of GA is complicated. The algorithm has certain dependence on the selection of initial population. The potential capability of parallel mechanism of algorithm has not been

fully utilized which is also a hot research direction of GA.

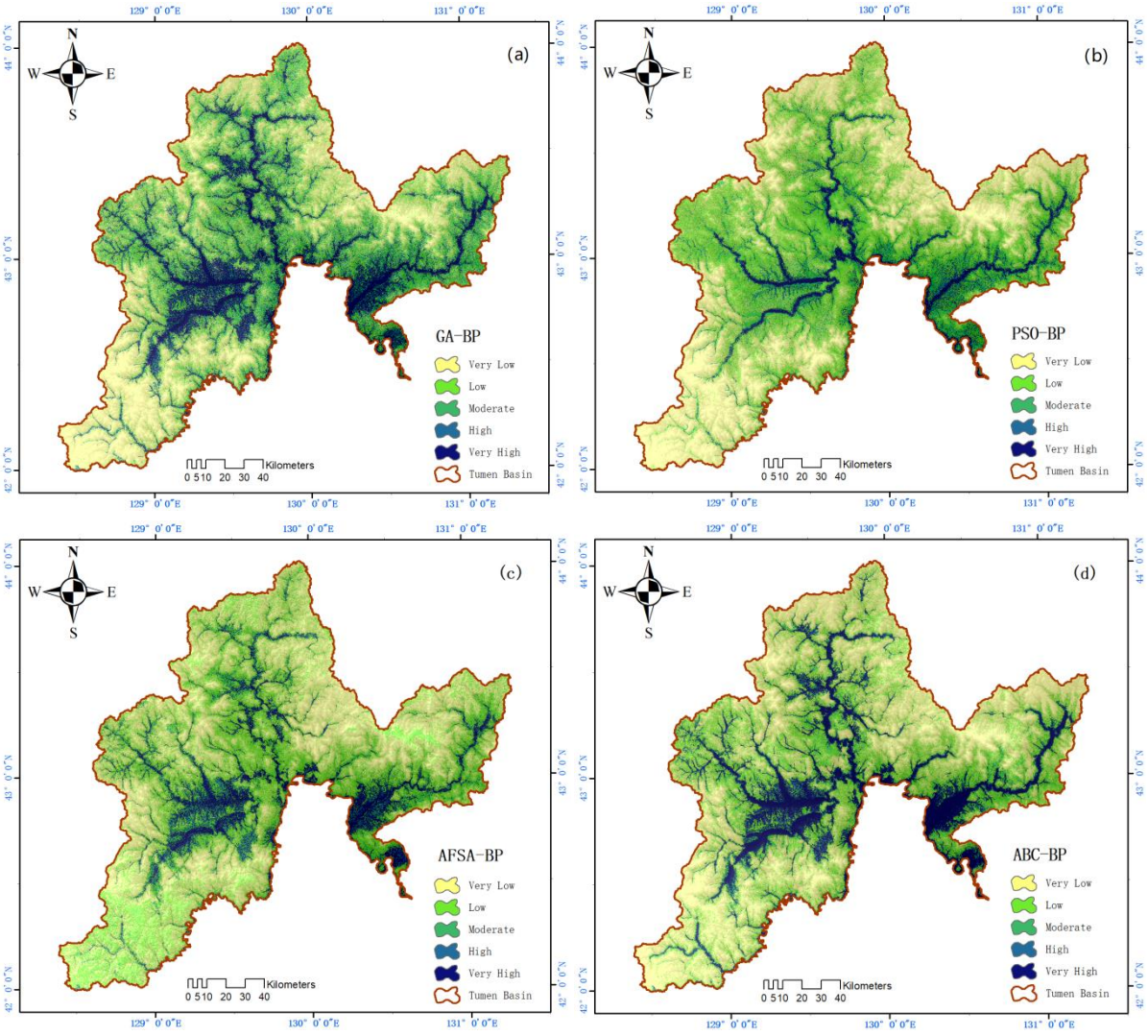
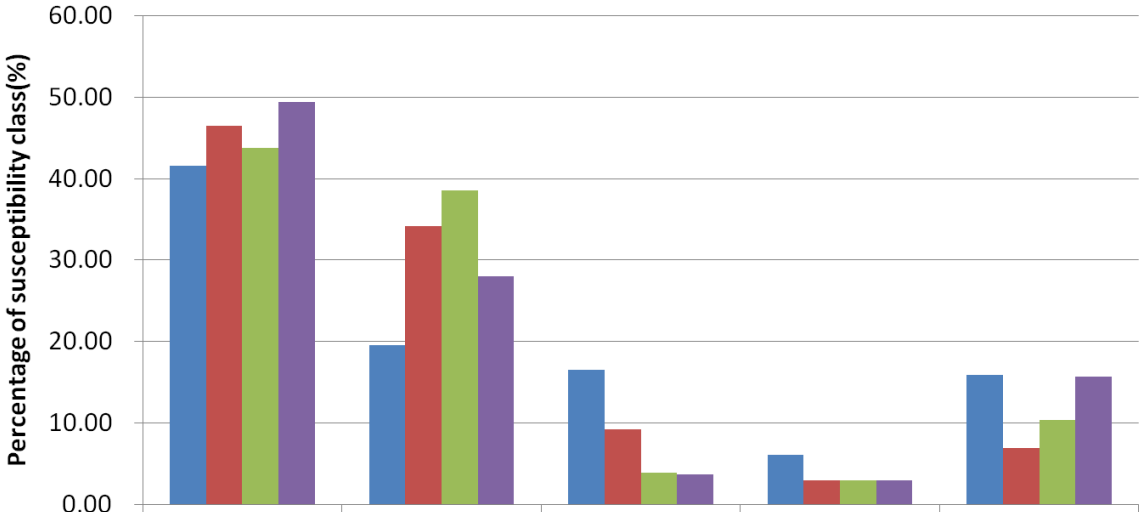


Fig. 6. Flood susceptibility maps produced by a. GA-BP model; b. PSO-BP model; c. AFSA-BP model; d. ABC-BP model.



	Very low	Low	Moderate	High	Very high
GA-BP	41.67	19.64	16.57	6.13	15.99
PSO-BP	46.55	34.21	9.27	3.02	6.95
AFSA-BP	43.85	38.64	4.00	3.06	10.46
ABC-BP	49.42	28.08	3.71	2.98	15.81

Fig. 7. Percentages of different flood susceptibility classes.

Table 3

Areas of susceptibility classes along with FR and SCAI indicators in models.

Model	Class	No. of pixels in domain	Percentage of domain(%)	No. of floods	Percentage of floods(%)	FR	SCAI
GA-BP	Very low	10506723	41.67	10	8.40	0.20	4.96
	Low	4953006	19.64	4	3.36	0.17	5.84
	Moderate	4178220	16.57	10	8.40	0.51	1.97
	High	1545323	6.13	11	9.24	1.51	0.66
	Very high	4032640	15.99	84	70.59	4.41	0.23
PSO-BP	Very low	11738737	46.55	10	8.40	0.18	5.54
	Low	8626452	34.21	17	14.29	0.42	2.39
	Moderate	2336500	9.27	16	13.45	1.45	0.69
	High	760829	3.02	5	4.20	1.39	0.72
	Very high	1753394	6.95	71	59.66	8.58	0.12
AFSA-BP	Very low	11056086	43.85	5	4.20	0.10	10.44
	Low	9742313	38.64	27	22.69	0.59	1.70
	Moderate	1008607	4.00	12	10.08	2.52	0.40
	High	772196	3.06	11	9.24	3.02	0.33
	Very high	2636710	10.46	64	53.78	5.14	0.19
ABC-BP	Very low	12462728	49.42	3	2.52	0.05	19.60
	Low	7080353	28.08	6	5.04	0.18	5.57
	Moderate	934682	3.71	7	5.88	1.59	0.63
	High	752151	2.98	4	3.36	1.13	0.89
	Very high	3985998	15.81	99	83.19	5.26	0.19

5. Conclusion

In this study, GA, PSO, AFSA and ABC algorithms were used to optimize the BP neural network model for the mapping of flood susceptibility in Tumen River Basin. The model included thirteen conditioning factors of flood susceptibility as well as 119 flooded and 103 non-flood data. The results showed that there was no multicollinearity among the selected impact factors. IGR test showed that distance to river, land use, rainfall and altitude were more sensitive to flood disaster. The accuracy of the model was evaluated by the following statistical indicators: Precision, RMSE, MAE, AUC, and Kappa. The validation results show that all the models (GA, PSO, AFSA and ABC) perform well, but the BP neural network model optimized by AFSA performs best. The flood susceptibility map generated by AFSA-BP model shows that about 13% of the study areas have a high risk of flood disaster and about 87% have a low risk. The flood susceptibility map generated in this study can identify potential flood risk areas. It provides useful information for land use planning and flood risk management in flood-affected areas.

Acknowledgement

We express our gratitude to editor of Natural Hazards, and the anonymous reviewers for their valuable

comments and suggestions that improved the quality of our paper. This work was supported by the National Natural Science Foundation of China (42067065).

References

Ahmadlou, M., Karimi, M., Alizadeh, S., Shirzadi, A., Parvinnejhad, D., Shahabi, H., Panahi,M., 2018. 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, 1252-1272. https://doi.org/10.1080/10106049.2018.1474276.

Alaghmand, S., Bin Abdullah, R., Abustan, I., Eslamian, S., 2012. Comparison between capabilities of HEC-RAS and MIKE11 hydraulic models in river flood risk modeling (a case study of Sungai Kayu Ara River basin, Malaysia). Int. J. Hydrol. Sci. Technol. 2(3), 270–291.

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.

Arabameri, A., Rezaei, K.，Cerdà, A., Conoscenti, C., Kalantari, Z.， 2019. A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. Science of the Total Environment. 660，443–458.

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

Bui, D.T., Panahi, M., Shahabi, H., Singh, V.P., Shirzadi, A., Chapi, K., Khosravi, K., Chen, W., Panahi, S., Li, S., Ahmad, B.B., 2018. Novel hybrid evolutionary algorithms for spatial prediction of floods. Scientific reports. 8(1), 15364.

Bui, D.T., Pradhan, B., Nampak, H., Bui, Q.T., Tran, Q.A., Nguyen, Q.P., 2016. Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. J. Hydrol. 540, 317–330.

Bui, D.T., Tsangaratos, P., Ngo, P.T.T., Pham, T.D., Pham, B.T., 2019b,Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods, Science of the Total Environment, 668,1038-1054.

Cao, C., Xu, P., Wang, Y., Chen, J., Zheng, L., Niu, C., 2016. Flash flood hazard sus-ceptibility mapping using frequency ratio and statistical index methods in coalmine subsidence areas. Sustainability. 8 (9), 948.

Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Bui, D.T., Pham, B.T., Pham, B.T., Khosravi, K., 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environ.Model. Softw. 95, 229–245.

Chen, W., Hong, H., Li, S., Shahab, H., Wang, Y., Wang, X., Ahmad, 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.

Chen, W., Pourghasemi, H.R., Panahi, M., Kornejady, A., Wang, J., Xie, X., Cao, S., 2017. Spatial prediction of landslide susceptibility using an adaptive neuro-fuzzy inference system combined with frequency ratio, generalized additive model, and support vector machine techniques. Geomorphology. 297, 69–85.

Chițu, Z., Tomei, F., Villani, G., Di Felice, A., Zampelli, G., Păltineanu, I.C., Vișinescu, I., Dumitrescu, A., Bularda, M., Neagu, D., Costache, R., Luca, E., 2020. Improving irrigation scheduling using MOSES short-term irrigation forecasts and in situ water resources measurements on alluvial soils of lower Danube floodplain, Romania. Water. 12(2), 520. https://doi.org/10.3390/w12020520.

Das, S., Gupta, A., 2021, Multi-criteria decision based geospatial mapping of flood susceptibility and temporal hydro-geomorphic changes in the Subarnarekha basin, India. Geoscience Frontiers. 12,101206.

Destro, E.，Amponsah, W.，Nikolopoulos, E.，Marchi, L.，Marra, F.，Zoccatelli, D.，Borga, M. 2018. Coupled prediction of flash flood response and debris flow occurrence: application on an alpine extreme flood event. J. Hydrol. 558, 225–237.

458 Gharbi, M., Soualmia, A., Dartus, D., Masbernat, L., 2016. Comparison of 1D and 2D hy-draulic models for floods
459 simulation on the Medjerda Riverin Tunisia. J. Mater. Envi-ron. Sci. 7, 3017–3026.
460 Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A.X., Chen, W., Kougias, I.,Kazakis, N., 2018a. Flood
461 susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic
462 algorithm and differential evolution. Sci. Total Environ. 621, 1124 – 1141.
463 Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A., Chen, W., 2018b, Application of fuzzy weight of evidence and data
464 mining techniques in construction of flood susceptibility map of Poyang County, China. Science of the Total
465 Environment. 625,575–588.
466 Jahangir, M.H., Reineh, S.M.M., Abolghasemi, M., 2019,Spatial predication of flood zonation mapping in Kan River
467 Basin, Iran, using artificial neural network algorithm. Weather and Climate Extremes. 25, 1-11.
468 https://doi.org/10.1016/ j.wace.2019.100215.
469 Karaboga, D., Akay, B., Ozturk, C., 2007. Artificial bee colony (ABC) optimization algorithm for training
470 feed-forward neural networks. Modeling Decisions for Artificial Intelligence. 4617, 318 – 329.
471 Karlsson, C.S., Kalantari, Z., Mörtberg, U., Olofsson, B., Lyon, S.W., 2017. Natural hazard susceptibility assessment
472 for road planning using spatial multi-criteria analysis. Environ. Manag. 60 (5), 823–851.
473 Khosravi, K., Pham, B.T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., Prakash, I., Bui, D.T., 2018. A comparative
474 assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed. northern
475 Iran. Sci. Total Environ. 627, 744 – 755.
476 Kia, M.B., Pirasteh, S., Pradhan, B., Mahmud, A.R., Sulaiman, W.N.A., Moradi, A., 2012. An artificial neural network
477 model for flood simulation using GIS: Johor River Basin, Malaysia. Environ. Earth Sci. 67, 251–264.
478 Komolafe, A.A., Herath, S., Avtar, R., 2018. Methodology to assess potential flood damages in urban areas under
479 the influence of climate change. Nat. Hazards Rev. 19 (2), 05018001.
480 Kourgialas, N.N., Karatzas, G.P., 2011. Flood management and a GIS modelling method to assess flood-hazard
481 areas-a case study. Hydrol. Sci. J. 56, 212–225.
482 Lai, C., Shao, Q., Chen, X., Wang, Z., Zhoua, X., Yang, B., Zhang, L., 2016. Flood risk zoning using a rule mining
483 based on ant colony algorithm. J. Hydrol. 542, 268–280.
484 Liu, K., Li, Z., Yao, C., Chen, J., Zhang, K., Saifullah, M., 2016. Coupling the k-nearest neighbor procedure with the
485 Kalman filter for real-time updating of the hydraulic model in flood forecasting. Int. J. Sediment Res. 31,
486 149–158.
487 Liu, R., Chen, Y.,Wu, J., Gao, L., Barrett, D., Xu, T., Li, L., Huang, C., Yu, J., 2015. Assessing spatial likelihood of
488 flooding hazard using naïve Bayes and GIS: a case study in Bowen Basin, Australia. Stoch. Environ. Res. Risk
489 Assess. 30, 1575–1590.
490 Luan, X.Y., Li, Z.P., Liu, T.Z., 2016. A novel attribute reduction algorithm based on rough set and improved artificial
491 fish swarm algorithm. Neurocomputing. 174, 522–529.
492 Mahmoud, S.H., Gan, T.Y., 2018. Multi-criteria approach to develop flood susceptibility maps in arid regions of
493 Middle East. J. Clean. Prod. 196, 216 – 229.
494 Moghadam, H.S., Valavi, R., Shahabi, H., Chapi, K., Shirzadi, A., 2018, Novel forecasting approaches using
495 combination of machine learning and statistical models for flood susceptibility mapping. Journal of
496 Environmental Management, 217,1-11.
497 Nandi, A., Mandal, A.,Wilson, M., Smith, D., 2016. Flood hazard mapping in Jamaica using principal component
498 analysis and logistic regression. Environ. Earth Sci. 75, 465.
499 Peng, Y., Wang, Q., Wang, H., Lin, Y., Song, J., Cui, T., Fan, M., 2019. Does landscape pattern influence the intensity
500 of drought and flood. Ecol. Ind. 103, 173–181.
501 Pham, Q.B., Abba, S.I., Usman, A.G., Linh, N.T.T., Gupta, V., Malik, A., Costache, R., Vo,N.D., Tri, D.Q., 2019.
502 Potential of hybrid data-intelligence algorithms for multi-station modelling of rainfall. Water Resour. Manage.
503 33 (15), 5067–5087. https://doi.org/10.1007/s11269-019-02408-3.
504 Quan, H.C., Lee, B.G., 2012. GIS-based landslide susceptibility mapping using analytic hierarchy process and

artificial neural network in Jeju(Korea) . KSCE Journal of Civil Engineering. 16(7)， 1258-1266,DOI: 10.1007/s12205-012-1242-0.
Rahmati, O., Pourghasemi, H.R., 2017. Identification of critical flood prone areas in data-scarce and ungauged regions: a comparison of three data mining models. Water Resour. Manag. 31, 1473-1487.
Rahmati, O., Pourghasemi, H.R., Zeinivand, H., 2016a. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. Geocarto Int. 31 (1), 42–70.
Rahmati, O., Zeinivand, H., Besharat, M., 2016b. Flood hazard zoning in Yasooj region,Iran, using GIS and multi-criteria decision analysis. Geomatics, Nat. Hazards Risk. 7 (3), 1000-1017.
Santos, P.P., Reis, E., 2018. Assessment of stream flood susceptibility: a cross-analysis between model results and flood losses. J. Flood Risk Manag. 11, 1038–1050.
Tehrany, M., Pradhan, B., Mansor, S., Ahmad, N., 2015a. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. Catena. 125,91–101.
Tehrany, M.S., Jones, S., Shabani, F., 2019. Identifying the essential flood conditioning factors for flood prone area. Catena. 175,174–192.
Tehrany, M.S., Lee, M.J., Pradhan, B., Jebur, M.N., Lee, S., 2014a. Flood susceptibility mapping using integrated bivariate and multivariate statistical models. Environ. Earth Sci. 72(10), 4001–4015.
Tehrany, M.S., Pradhan, B., Jebur, M.N., 2014b. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J. Hydrol. 512, 332–343.
Tehrany, M.S., Pradhan, B., Jebur, M.N., 2015b. Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. Stoch. Env. Res. Risk A. 29, 1149–1165.
Termeh R.S.V., Kornejady, A., Pourghasemi, H.R., Keesstra, S., 2018. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. Sci. Total Environ. 615, 438–451.
Toosi, A. S., Calbimonte, G. H., Nouri, H., Alaghmand, S.， 2019. River basin-scale flood hazard assessment using a modified multi-criteria decision analysis approach: A case study. Journal of Hydrology, 574,660–671.
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, 101075. https://doi.org/10.1016/j. gsf.2020.09.006.
Vincendon, B., Ducrocq, V., Saulnier, G. M., Bouilloud, L., Chancibault, K., Habets, F., Noilhan，J., 2010. Benefit of coupling the ISBA land surface model with a TOPMODEL hydrological model version dedicated to Mediterranean flash-floods. Journal of Hydrology. 394, 256-266.
Wang, Y., Hong, H.Y, Chen, W., Li, S.J., Panahi, M., Khosravi, K., Shirzadi, A., Shahabi, H., Panahi, S., Costache, R., 2019, Flood susceptibility mapping in Dingnan County (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm, Journal of Environmental Management. 247,712–729.
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. Geom. Nat. Hazards Risk. 11(1), 2282-2314.
Zhang, S., liu, S., 2015. Artificial Bee Colony Algorithm with Improved Search Equations. Journal of Information and Computational Science. 20, 4069 – 4076.
Zhao, G., Pang, B., Xu, Z., Yue, J., Tu, T., 2018. Mapping flood susceptibility in mountainous areas on a national scale in China. Science of the Total Environment. 615,1133–1142.

Declaration of Interest Statement

We declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The paper published in this journal is original research results obtained by the author independently, and none of them has submitted more than one manuscript; The content of the paper does not involve state secrets; It has never been published in any form or in any language at home or abroad; The content of the paper shall not infringe upon the copyright and other rights of others. In case of multiple contributions, infringement, leakage and other issues, the author of the paper will take full responsibility.

Hechun Quan