Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Flood risk assessment based on machine learning and multi-criteria decision analysis

Guoyi LI

1969687116@qq.com

Shandong Agriculture and Engineering University

SHAO Weiwei

China Institute of Water Resources and Hydropower Research

ZHAO Yang

Shanghai Nuclear Engineering Research & Design Institute Co.,LTD

Yong LI

Shandong Agriculture and Engineering University

Yi ZHANG

Shandong Agriculture and Engineering University

Article

Keywords: Flood risk assessment, Machine learning, Flood susceptibility, Qianshan River Basin

Posted Date: March 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9004660/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Flood risk assessment based on machine learning and multi-criteria decision analysis

LI Guoyi[1*], SHAO Weiwei[2], ZHAO Yang[3], LI Yong[1], ZHANG Yi[1]

(1. Shandong Agriculture and Engineering University, Jinan 250100, China

2. China Institute of Water Resources and Hydropower Research, Beijing 100044, China

3. Shanghai Nuclear Engineering Research & Design Institute Co.,LTD, Shanghai, 200233, China)

Corresponding author at: No.866 Nongganyuan Road, LiCheng District, 250100 Jinan, China

E-mail addresses: 1969687116@qq.com

Abstract:

In this study, three machine learning algorithms, including Multilayer perceptron (MLP), Radial basis function (RBF) and Back propagation (BP), to analyse the flood susceptibility of the study area. Considering the flood susceptibility, hazard and exposure of the study area, the Multi-Criteria Decision Analysis (MCDA) was used to conduct a flood risk assessment of the study area. Collection of information on the physical geography and socio-economic distribution of the study area, including Rainfall, DEM, Slope, Land use/Land cover (LULC), Distance from the rivers, Curve number (CN), Manning coefficient, topographic wetness index (TWI), population density and Gross Domestic Product (GDP) distribution. A machine learning approach to assess flood susceptibility in the study area, using six standard statistical indices (accuracy, precision, recall, F-score, Root Mean Square Error and Kappa index) and Receiver Operating Characteristic (ROC) curves called AUG to evaluate the performance of the machine learning model. The results show that all the proposed models perform well, but performance of the MLP model is the best (AUC=0.951). Using the Analytic Hierarchy Process (AHP) method to determine the weights, the flood susceptibility map obtained from the MLP is overlaid with hazard and vulnerability for flood risk assessment. The results show that integrated machine learning algorithms and MCDA for flood analysis are a promising approach to flood risk assessment.

Keywords: Flood risk assessment; Machine learning; Flood susceptibility; Qianshan River Basin;

1.Introduction

Floods are recognized as the most prevalent natural disaster[1]. In the past few decades, with the acceleration of climate change and urbanization, the frequency and intensity of floods have increased, posing a huge threat to global development[2]. Governments have put in place a large number of measures and policies to deal with floods, but floods still cause casualties and huge property losses every year[3]. Urban development has led to dramatic changes to the underlying surface, including reduced vegetation coverage, hardened of the ground, increased impermeability, and imperfect urban drainage facilities. These changes have resulted in increased surface runoff in urban areas and further increased risk of urban flooding[4, 5]. Experts predict that with climate change, the situation of urban floods will become more severe[6]. Flood risk assessment is an

important non-engineering measure to deal with flood disasters, and an important measure to deal with flood risks. The main purpose of this method is to identify areas prone to flood disasters by drawing flood risk maps, and provide information for flood prevention planning and mitigation plans[7].

At present, there are many methods for conducting flood risk assessment. The scenario simulation method refers to construction of flood model (hydrological and hydrodynamic model), designing future rainfall scenarios, considering climate change, population and economic changes, simulating the entire flood process, and analyzing possible future flood disasters. Commonly used models include SWMM model[8-10], Mike model[11], TELEMAC-2D model[12, 13] and Info Work ICM model[14], etc. These models require high-precision geospatial data and powerful computer computing power in the study area when performing calculations[15]. The flood risk assessment method based on the index system constructs a flood loss index system from the aspects of hazard, exposure and vulnerability, uses RS technology and GIS technology to process remote sensing data, and determines the weight values of different index factors by using the AHP method to assess the flood risk[16, 17]. The multi-criteria decision analysis (MCDA) method is a simple and effective flood risk assessment method[18-20], but it also has some drawbacks, such as relying too much on the empirical knowledge of experts, which will bring subjectivity to the model evaluation results[21].

In recent years, with the continuous improvement of computer computing power and algorithms, many machine learning methods have been used to predict natural disasters, which are widely used in flood risk analysis, including artificial neural networks [22, 23], support vector machines[24, 25], logic Regression[26, 27], Random Forest[28, 29], etc. The relationship between flood influencing factors and floods can be found through historically recorded flood points. Machine learning can predict the flood zoning of the entire study area, and can avoid the influence of subjectively determined impact factor weights. In conclusion, machine learning has excellent predictive ability and broad application prospects in flood risk assessment.

China is located in the east of Asia and on the west coast of the Pacific Ocean. Affected by the monsoon climate, rainfall is unevenly distributed throughout the year, and rainfall is concentrated in summer, which is prone to flood disasters. With China's rapid economic development and accelerated urbanization, it is urgent to take effective measures for flood prevention and disaster reduction. In this paper, three machine learning methods are used to assess the risk of urban flooding to compensate for the shortcomings in previous studies. The three machine learning algorithms are MLP, RBF and BP models, which have a sound theoretical foundation and are applied in many fields, and have excellent model learning ability. The main objectives of the paper: (1) Select the best machine learning method to evaluate the flood sensitivity of the Qianshan River Basin; (2) Expose the main influencing factors of urban flood sensitivity; (3) Determine different flood risk levels in the study area; (4) Provide an effective management basis for urban flood prevention and mitigation.

Section 2 provides a detailed introduction to the research area. Section 3 introduces the research method and acquisition and pre-treatment of data. Section 4 analyzes the flood risk assessment results. Section 5, summarizes the main conclusions and discusses the contributions, drawbacks and issues to be addressed in the next step.

2.Study area

The Qianshan River Basin (113˚20' ~114˚18' E, 21˚48' ~22˚27' N) straddles Zhuhai and Zhongshan City in Guangdong Province, and its geographical location is shown in Figure 1. The administrative area of the Qianshan River Basin covers four towns and streets, including Nanping Town, Qianshan Street, Gongbei Street and Wanzai Street in Xiangzhou District of Zhuhai, and Sanxiang Town and Tanzhou Town in Zhongshan City, which are important economic areas in the Pearl River Delta region.

The Qianshan River Basin is a subtropical maritime monsoon climate with mild climatic conditions and abundant annual rainfall, with an average annual precipitation of about 2065.0mm. The rainfall is concentrated during the flood season, mostly heavy rainfall with high intensity and short duration, especially from May to June, which accounts for more than 30% of the total annual rainfall. The total catchment area of the river basin is about 328.18km^2. The cross-border area of the Qianshan River Basin is a densely built-up area of the two cities. With the development of urbanization, the impermeable surface area has been increasing, which has changed the slope surface production and confluence conditions and the river network confluence system. The flooding is also exacerbated by the top of the outer river, but heavy rainfall in the basin is the root cause of extensive flooding. Therefore, it was of certain practical significance to choose the Qianshan river basin as the study area.

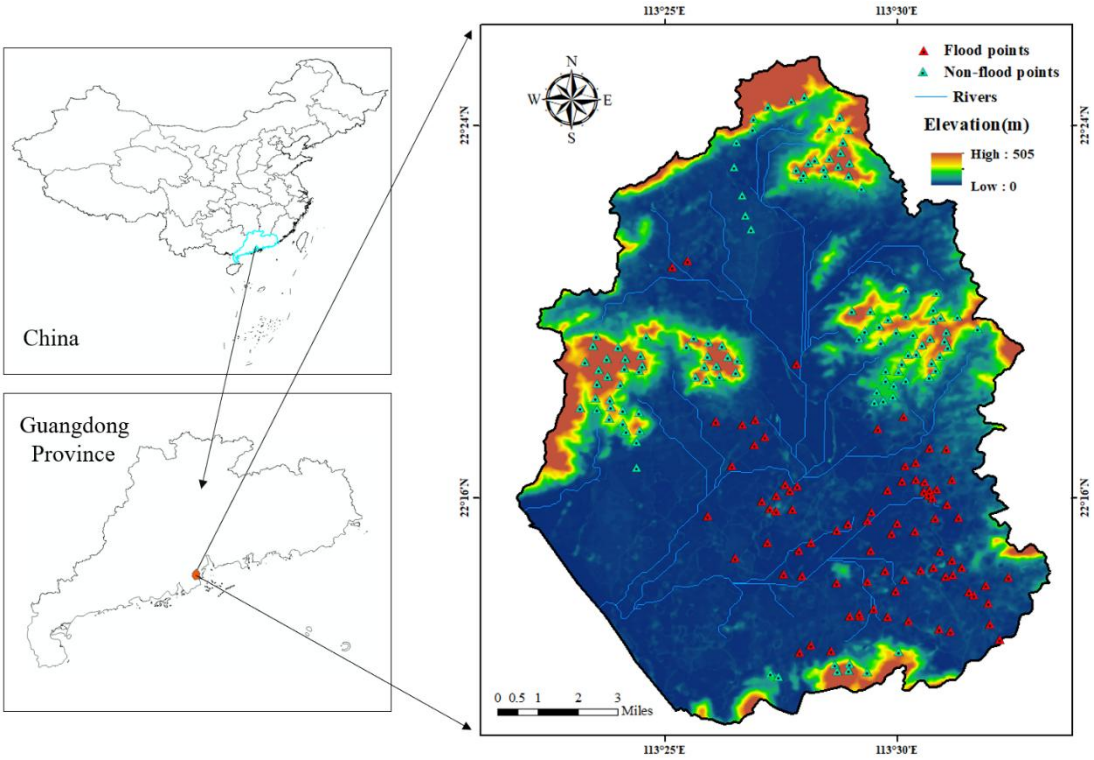


Figure 1. Location and physiographic information of the Qianshan River Basin.

3.Methods and material

Flood risk assessment based on machine learning method and MCDA includes the

following three steps (Fig.2). Firstly, collect flood point data and geospatial information in the study area to build a sample set for machine learning model training and testing. The flood point is the target variable of the machine learning model, and the index factor is the independent variable used to predict the target. Secondly, three machine learning models were selected to analyze the flood susceptibility of the study area, and to evaluate the machine learning model prediction effect; Finally, the MCDA is used to superimpose the obtained flood susceptibility map with the hazard and exposure, and the flood risk maps in GIS environment.

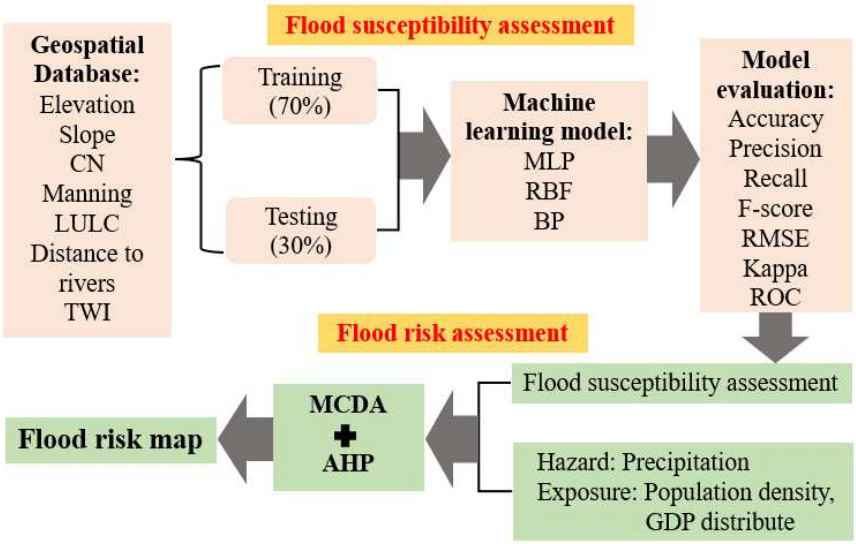


Figure 2. Flood risk assessment framework for the Qianshan River Basin.

3.1 Data Sources

The reasonable selection of flood impact indicators is a crucial step in risk assessment and analysis. Referring to previous research results[24, 28, 30] and the special conditions of the study area, this paper selects ten influencing factors for flood risk assessment, including rainfall, elevation, slope, land use type, manning coefficient, CN, distance from river, TWI, population density and GDP distribution, the sources of data are shown in the Table 1.

Rainfall is the most direct factor that causes floods. Due to the small number of meteorological stations in the selected study area, this paper uses the average rainfall data in Guangdong Province for many years to obtain the rainfall distribution in Guangdong Province, and then cuts out the rainfall distribution map in the study area. Elevation and slope are the most important topographical factors, the smaller the value, the more vulnerable to the threat of flooding. The slope is calculated from the elevation data using the hydrological analysis tool in the GIS software. The land use type data is obtained through the interpretation of remote sensing images. The land use types in the Qianshan River Basin are divided into five categories: buildings, forest, water, farmland and others. The Manning coefficient is mainly used to characterize the surface roughness, which is generally assigned by the type of land use (water:0.015, forest:0.5, farmland:0.05, buildings:0.2, others:0.05). The CN represents the characteristics of the surface and soil

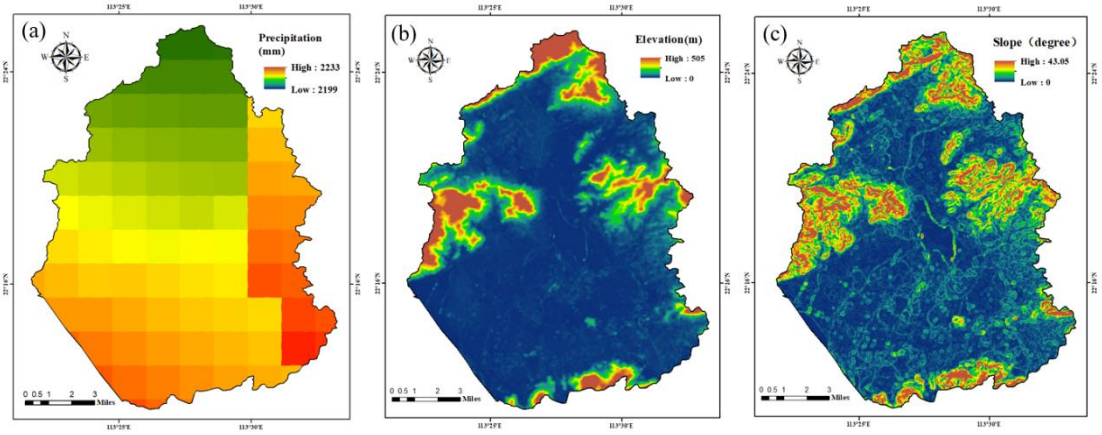
before rainfall, and its value can be determined in combination with the land use type and the results of hydrological soil classification[31]. The distance from the rivers can be calculated by GIS software, the smaller the value, the closer to the river, the higher the risk of flooding[32]. The TWI is widely used in the field of hydrology [33, 34], and it mainly reflects the distribution characteristics of soil hydrology in the basin, as shown in the Eq.(1). Population density and GDP distribution are important socio-economic attributes. The greater the density, the greater the losses caused by floods. (Fig.3)

$$TWI = \ln(\frac{a}{\tan b}) \quad (1)$$

where a is the specific catchment area ($m^2\ m^{-1}$), and b (radian) is the slope gradient (in degrees).

Table 1. The database for the ten indices.

Name	Source	Spatial resolution	Data types
Rainfall	https://www.resdc.cn/	1km ×1km	Raster
Elevation	https://www.gscloud.cn/	30 m × 30m	Raster
Slope	https://www.gscloud.cn/	30 m × 30m	Raster
Land use	https://www.resdc.cn/	16 m ×16m	Raster
Manning coefficients	https://www.resdc.cn/	16 m ×16m	Raster
CN	Harmonized World Soil Database	1km ×1km	Raster
Distance from river	https://www.gscloud.cn/	30 m × 30m	Raster
TWI	https://www.gscloud.cn/	30 m × 30m	Raster
Population density	https://www.resdc.cn/	1km ×1km	Raster
GDP distribute	https://www.resdc.cn/	1km ×1km	Raster



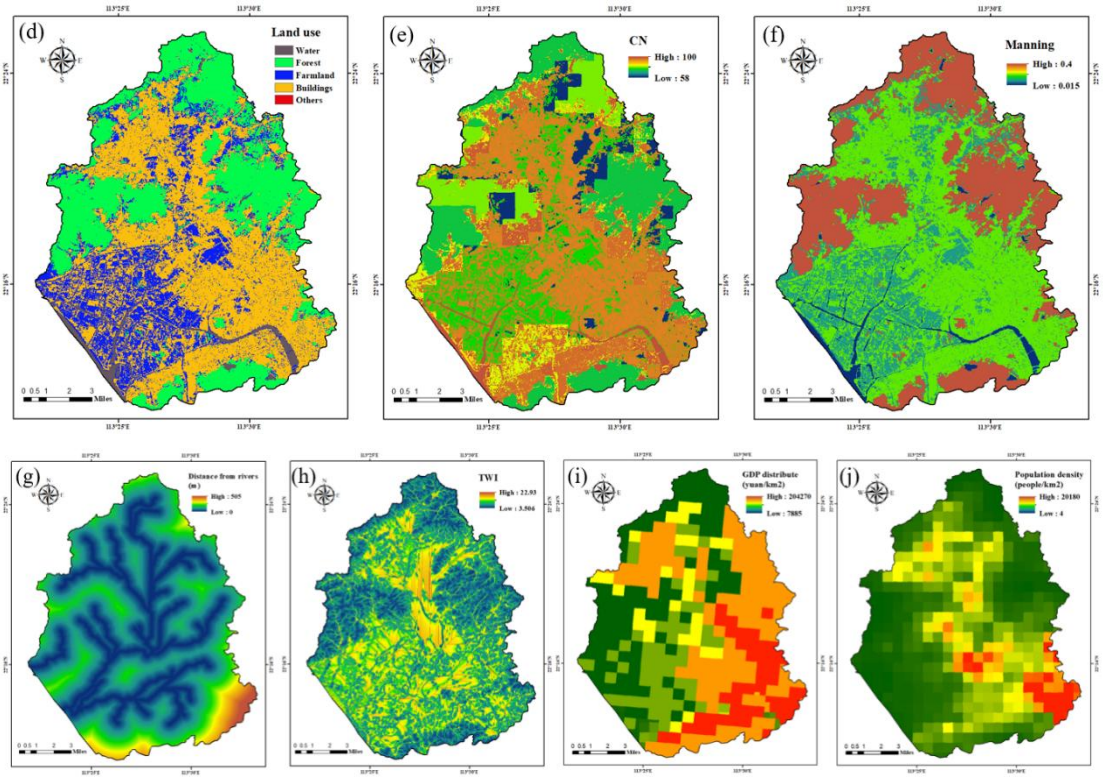


Figure 3. Maps of flood influencing factors: (a)Precipitation; (b)Elevation; (c)Slope; (d)Land use; (e)CN; (f)Manning; (g)Distance from rivers; (h)TWI; (i)GDP distribute; (j)Population density

3.2 Data pre-processing

Based on GIS software the collected index data are assigned to the same coordinate system to get the index values at different spatial locations. The correlation between different indicators and flood risk varies due to the different magnitudes of the individual indicator values, making the range of values between the indicator data very different, and this situation can have a great impact on the results of data analysis. In order to eliminate differences between different data indicators, it is necessary to pre-process the collected data to solve the problem of comparability between different indicator data. The data normalization processing method is used to process the indicator data and transform the value of each indicator into a number between 0 ~ 1. Positive indicators are calculated using equation (2), Negative indicators are calculated using equation (3). The raw data are normalized that the data values of each indicator are in the same order of magnitude, which facilitates the subsequent flood risk assessment.

$$y_{ij} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

$$y_{ij} = \frac{x_{\max} - x_i}{x_{\max} - x_{\min}} \tag{3}$$

where: y_{ij} is the i value of the j indicator after standardisation, x_i is the original value,

x_{max} and x_{min} are the maximum and minimum values of the j indicator, respectively.

3.3 Urban flood inventories

The representativeness and size of the sample set are important factors affecting machine learning and are an important step in predicting flood susceptibility[35, 36]. The flood inventory map in the Qianshan River Basin contains data on 84 flood points, which are from the Guangdong Water Conservancy and Hydropower Research Institute (red triangle in Fig.1). The data of 121 non-flood points are randomly selected to form a flood risk set (green triangle in Fig.1). The submerged locations of the flood sample set were geolocated, and the flood points were assigned a value of 1, and the non-flood points were assigned a value of 0. The sample set was randomly divided into two groups according to the ratio of 7:3, 70% of the point set was used for model training, and the remaining 30% of the point set was used for model testing.

3.4 Multi-Criteria decision analysis (MCDA)

Several popular MCDA methods used in flood risk assessment are Analytical Hierarchy Process (AHP), Ordered Weighted Averaging (OWA), and Analytic Network Process (ANP). The AHP to determine the indicator weights includes the following four calculation steps: (1) establish the hierarchical structure model; (2) construct the judgment matrix; (3) calculate the indicator weight values: (4) conduct consistency test, when CR<0.1, it can be considered that the judgment matrix passes the consistency test and the weight values are reasonable. Otherwise, the consistency test needs to be conducted again. In this study, the flood risk assessment system was constructed according to the object layer, index layer and indicator layer. When the CR values were less than 0.1, and the results obtained were reliable and reasonable through the consistency test. The obtained weight values of each indicator are shown in the Table 2.

Table 2. Weights of the urban flood risk assessment indicators

Object layer	Index layer	Relative weight	Indicator layer	Relative weight
Urban flood risk assessment	Hazard	0.38	Precipitation	1
	Exposure	0.41	Population density	0.52
			GDP distribute	0.48
	Susceptibility	0.21	Elevation	\
			Slope	\
			CN	\
			Manning	\
			LULC	\

Distance to rivers	\
TWI	\

3.5 Model descriptions

Three machine learning models were selected for flood susceptibility analysis, the basic principles are as follows.

3.5.1 Multilayer perceptron（MLP）

MLP is a feedforward artificial neural network model, which can be divided into input layer, hidden layer and output layer, and different layers are fully connected by neurons with nonlinear activation function. Each connection is assigned different weights, the input of each layer of neurons is the weighted sum of the output values of the neurons in the previous layer (Eq. (4)). MLP is a classic model in deep learning. Many scholars use the MLP model to analyze practical problems in the environmental field[37-39]. It also has practical applications in water resources[40, 41], but relatively few applications in the field of urban flood analysis. This paper innovatively uses the MLP model for Urban flood analysis to test its practicability in the field of flood risk analysis.

$$\mathrm{y}_i^l = f\left(\sum_{j=1}^{n} y_j^{l-1} w_{ji}^{l-1} + b_i^l\right) \tag{4}$$

Where y_i^l is the output of the i^{th} neuron at layer l; f is the activation function of the l^{th} layer; n is the number of neurons in the $(l-1)^{\text{th}}$ layer; y_j^{l-1} is the output of the j^{th} neuron at layer $(l-1)$; w_{ji}^{l-1} is the connection weight of the j^{th} neuron and the i^{th} neuron at layer $l-1$; b_i^l is the deviation of the i^{th} neuron at layer l.

3.5.2 Radial Basis Function (RBF)

RBF is a kind of three-layer feedforward network, which is a nonlinear transformation from the input space to the hidden layer space, and a linear transformation from the hidden layer space to the output layer space. Compared with other types of artificial neural networks, the RBF network has the advantages of simple structure, fast learning speed, and great real-time requirements. RBF model has been applied in the field of water conservancy[42, 43]. Compared with the BP neural network, the RBF model adds radial basis functions in the hidden layer, without back-propagation and without bias terms. The activation function R of the Gaussian radial basis (Eq. (5)):

$$R\left(x_i - c_j\right) = \exp\left(-\frac{1}{2\sigma_j^2}\left\|x_i - c_j\right\|^2\right) \tag{5}$$

$$x_i = \left[x_i^1, x_i^2, \cdots, x_i^h, \cdots, x_i^M \right]^T \tag{6}$$

$$y_i = \sum_{j=1}^{Q} w_{jk} \exp\left(-\frac{1}{2\sigma_j^2} \left\| x_i - c_j \right\|^2 \right) \tag{7}$$

Where x_i is the input signal of the RBF model; h is the node number of the input layer, $h = 1, 2, \cdots, M$; i is the number of the embedded sample; c_j is the center of the Gaussian radial basis function of the j th hidden layer node, $j = 1, 2, \cdots, Q$; σ is the variance of the Gaussian radial basis function; w_{jk} is the connection weight from the j th hidden layer to the k th output layer.

3.5.3 Back Propagation (BP)

BP neural network is a concept proposed by scientists led by Rumelhart and McClelland in 1986. It is a multi-layer feedforward neural network trained according to the error back propagation algorithm, and is one of the most widely used neural network models[44]. In this paper, the three-layer BP neural network algorithm is used, which is widely used in geological disaster prediction[45, 46], environmental assessment[47, 48] and urban flood analysis[49, 50]. The forward transfer process of the algorithm is using Eq. (8):

$$S_{\mathrm{j}} = \sum_{i=0}^{n} w_{ij} x_i + b_j \tag{8}$$

$$x_i = f\left(S_j \right) \tag{9}$$

Where S_{j} is the input signal of the j node, which is passed through the transfer function; n is the total number of nodes on the layer above the j node; w_{ij} is the weight between the i node and the j node; b_j is the threshold of the j node; Linear function; f is the activation function, generally a S function or a linear function is selected.

3.6 Model comparison method

To validation the performance of different machine learning models (MLP, RBF and BP) in flood risk assessment, a variety of standard quantitative indices are used to evaluate the training set and testing set respectively. The training set evaluation results represent the fitting effect of the data, and the testing set evaluation results represent the model simulation prediction effect[51, 52]. Standard quantitative indices include: accuracy (ACC), precision (P), recall (R), F-score (F), Kappa (K) and Root Mean Square Error (RSME), as shown in Eq. (10)-(15). ACC represents the proportion of correct predictions (flood points

and non-flood points) in the sample; P, R and F are used to characterize the ability to identify flood points; Kappa coefficient (K) is an effective statistical measurement method, which has used to measure stochastic agreement between categorical components; RMSE indicates the degree of dispersion of predicted and actual values, and model is considered perfect when the RMSE value is small[53]. The ROC curve is extensively used indicators by researchers to evaluate the predictive ability and classification quality of the model[54]. The area under the ROC curve (AUC) is often used to quantitatively evaluate the predictive ability of the model. The larger AUC values indicates that the model performs well[25, 55].

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$F = \frac{2 \times P \times R}{P + R} \tag{13}$$

$$K = \frac{P_o - P_e}{1 - P_e} \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ei} - x_{oi}\right)^2} \tag{15}$$

Where TP (true positive) represents the number of correct predictions in the set of flood points; FP (false positive) represents the number of incorrect predictions in the set of flood points; TN (true negative) represents the number of correct predictions in the set of non-flood points; FN (false negative) indicates the number of prediction errors in the set of non-flood points; P_o is number of pixels to be classified correctly as flood or non-flood pixels; P_e are expected agreements; x_{oi} and x_{ei} are the ith observation and model prediction, respectively, n is the number of points in the sample set.

4.Results

4.1 Model evaluation

According to the evaluation method proposed in Section 3.4, the evaluation results of the three models of MLP, RBF and BP are shown in Table 3. The MLP model has the highest ACC (0.986 in training and 0.950 in testing), indicating that the prediction accuracy of the MLP model is 98.6% and 95% in the training and testing datasets. The RBF and BP models performed consistently in terms of ACC, with a prediction accuracy of 96.6% and 93.3% in the training and testing datasets. The results show that the MLP model has higher fitting and prediction capabilities than RBF and BP. The R value of the MLP model (0.968 in training and 0.920 in testing) is significantly higher than that of the RBF and BP models, indicating that more frequently observed flooding sites can be correctly identified in the MLP model. The F value of the MLP model is the highest (0.984 in training and 0.939 in

testing), followed by the RBF model (0.960 in training and 0.923 in testing) and the BP model (0.960 in training and 0.920 in testing). The MLP model has the smallest RSME value (0.117 in training and 0.223 in testing), and the largest K value is the MLP model (0.972 in training and 0.901 in testing). The ROC curves and AUC values of the three models are shown in Figure 4. The AUC value of the MLP model is the highest, 0.988 in the training dataset and 0.951 in the testing dataset, followed by the RBF model (0.965 in training and 0.944 in testing), and the AUC value of the BP model (0.965 in training and 0.935 in testing). The results demonstrate that the order of model performance is MLP>RBF≈BP. The MLP model showed the best performance on the testing dataset with a prediction accuracy of 95% and AUC value of 0.951. Therefore, the MLP model was selected for flood susceptibility analysis, and the analysis results are used in flood risk assessment.

Table 3. Model performance for the training and testing dataset

Index	Training			Testing		
	MLP	RBF	BP	MLP	RBF	BP
TP	60	60	60	23	24	23
TN	83	80	80	34	32	33
FP	0	0	0	1	0	1
FN	2	5	5	2	4	3
ACC	0.986	0.966	0.966	0.950	0.933	0.933
P	1.000	1.000	1.000	0.958	1.000	0.958
R	0.968	0.923	0.923	0.920	0.857	0.885
F	0.984	0.960	0.960	0.939	0.923	0.920
RSME	0.117	0.203	0.203	0.223	0.258	0.258
K	0.972	0.932	0.932	0.901	0.872	0.871

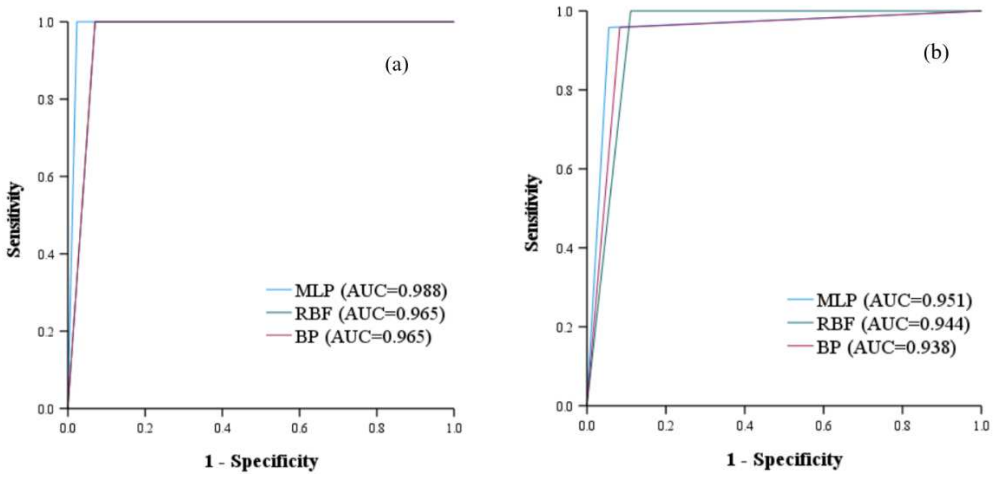


Figure 4. ROC curve: (a)training dataset; (b)testing dataset.

4.2 Flood susceptibility mapping

The simulation results of machine learning were input into GIS software, and the flood sensitivity map of the study area was obtained. In Figure. 5, the study area is divided into five categories (highest, high, medium, low and lowest) using the natural discontinuity method. As shown in Figure 5(a) and Table 4, the highest and high areas in the entire study

area are 83.01 and 82.32 km^2, accounting for 24.71% and 24.50%, respectively. The highest and high risk areas in the Qianshan River Basin are mainly distributed in the central and northern parts of the basin. The terrain is flat and the altitude is low, which is easily affected by water accumulation. And the high degree of urbanization, dense buildings and high impermeability in this area have increased surface runoff and further aggravated the flood crisis. The southern part of the Qianshan River Basin has a well-developed water system, and the areas close to the river are vulnerable to floods and have high flood sensitivity. This conclusion is consistent with previous research results[58, 59].Compared with the simulation results of the MLP model, the RBF and BP models predicted extremely highest and high risk areas, and the ability to predict medium and low risk areas was insufficient. The medium risk and low risk areas obtained by the RBF simulation were 1.46 and 1.31 km^2, and the medium risk and low risk areas obtained by the BP simulation were 1.70 and 0.64 km^2.

The MLP model shows the best prediction performance among the three models, so the article calculates the contribution of different index factors (elevation, slope, CN, manning coefficient, land use, distance from the rivers and TWI) when the MLP model is calculated (Figure. 6). The contribution values of elevation, land use type and slope rank among the top three, with 27.3%, 19.9% and 18% respectively. The contribution of the distance from the river and the CN are 11% and 10.9%, respectively, and the contribution of TWI and manning coefficient are the smallest, which are 7.3% and 5.6%, respectively. This conclusion is consistent with the previous research conclusion[23, 53, 60, 61].

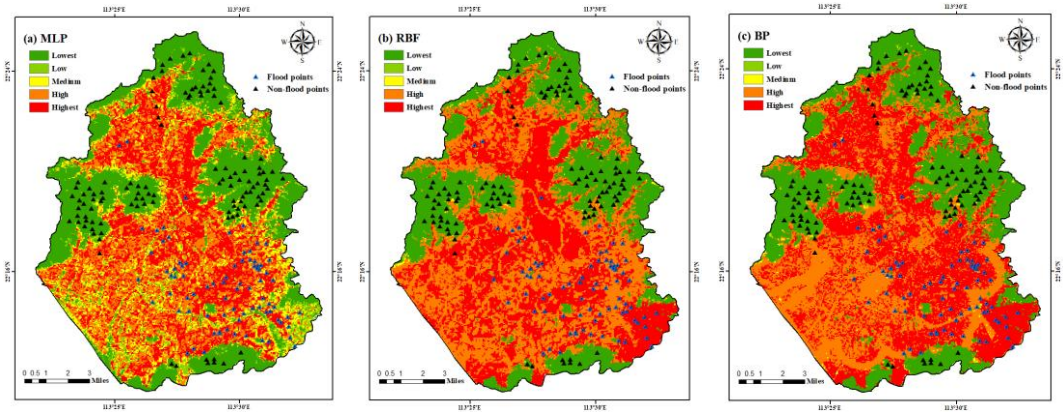


Figure 5. Flood susceptibility maps: (a)MLP; (b)RBF; (c)BP.

Table 4. Flood susceptibility maps from the three compared models

Model	Level	Area(km^2)	Percentage (%)
MLP	Highest	83.01	24.71
	High	82.32	24.50
	Medium	40.17	11.96
	Low	25.37	7.55
	Lowest	105.10	31.28
RBF	Highest	114.66	35.11
	High	119.96	36.73
	Medium	1.46	0.45

	Low	1.31	0.40
	Lowest	89.22	27.32
BP	Highest	124.61	37.08
	High	107.82	32.09
	Medium	1.70	0.51
	Low	0.64	0.19
	Lowest	101.26	30.13

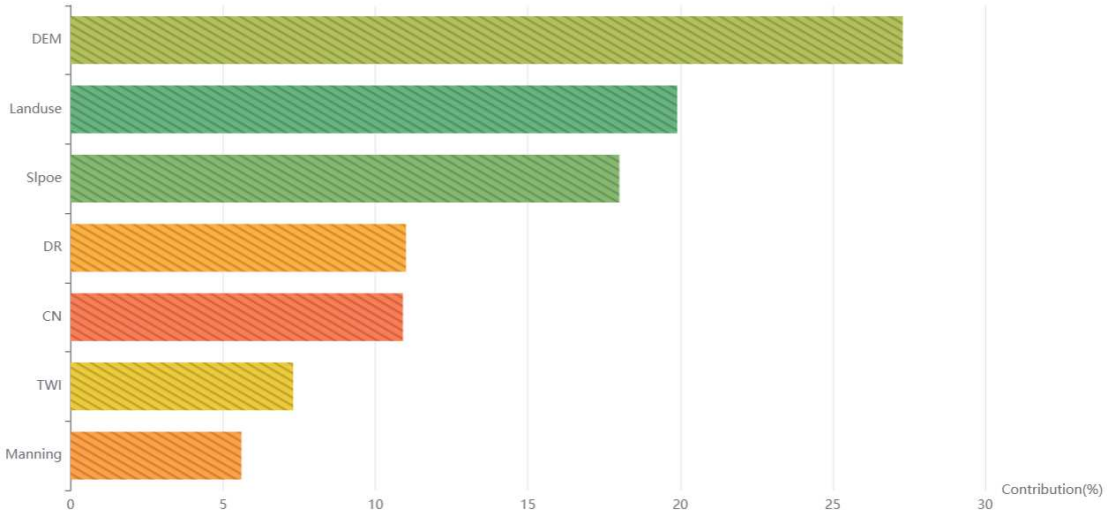


Figure 6. Contribution of the indices to flood susceptibility.

4.3 Flood risk assessment

Finally, the flood risk assessment of Qianshan River Basin is carried out based on the MCDA method. The flood susceptibility map obtained by the MLP model was superimposed with the hazard and exposure. Using ArcGIS powerful data analysis function, different index factors are superimposed and calculated, and the flood risk is divided into five categories (highest, high, medium, low, and lowest) by the natural discontinuity method (Figure. 7). The analysis shows that the areas of the highest and high risk are 12.78 and 48.75 km^2, accounting for 3.9% and 15.1%, and they are mainly distributed in the southeast of the study area. The rainfall in this area is relatively high, and the terrain is flat and prone to water accumulation. And the economy is developed and the population is dense, if flood disaster occurs, it will cause a greater social impact. The northern region has higher terrain, mostly forest land, low population density, and low risk of hazard and exposure, so its flood risk level is low.

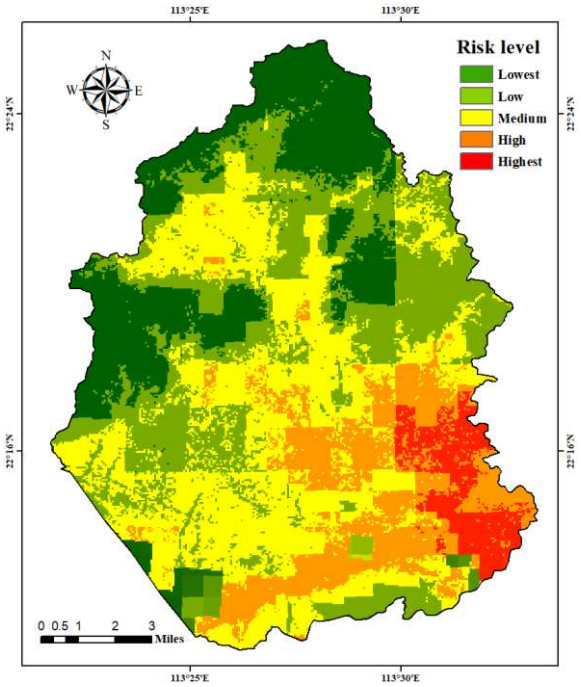


Figure 7. Flood risk assessment result of the Qianshan River Basin.

5.Discussion

5.1 Adaptation of machine learning in flood risk assessment

The accuracy of the machine learning model in flood risk prediction is related to the amount of data in the sample set (flood points and non-flood points) in the study area[26, 34]. The more abundant and representative the sample data, the better predictions of the machine learning model. With the continuous acceleration of urbanization, urban land use types and underlying surface conditions change, which will affect the effect of machine learning. In this paper, a comprehensive MCDA based on machine learning is proposed for urban flood risk assessment, in which machine learning is used to analyze the flood susceptibility of the study area. Through model evaluation, it is found that the performance of the MLP model is better than the RBF and BP models in flood prediction. The flood susceptibility map of the study area derived from the MLP model is superimposed with the hazard and exposure, and the flood risk zoning map was obtained.

With the deepening of computational principles and the improvement of algorithms, neural network models have made great progress, and more advanced machine learning models have emerged, including Random Forest, Boosted-Tree (RFBT), Semi-supervised learning, Decision Trees (DT) and Naïve Bayes (NB) etc. There are many machine learning algorithms, each with its own advantages and disadvantages, and different researchers have introduced different machine models to conduct flood risk assessment and achieved fruitful results[56, 57]. Different scholars hold different opinions on which machine learning model has the highest prediction accuracy and the most reliable results. The three machine learning models MLP, RBF and BP used in this paper are the most basic neural network models, and they all show excellent performance in prediction. The MLP model has higher prediction accuracy, and the MLP method is used to analyze the flood

susceptibility of the study area.

5.2 limitations

Through the research and review of previous research results in this paper, the shortcomings of machine learning in current flood risk assessment research mainly include: The multitude of factors influencing flood risk assessment, which undergo continuous temporal variation, necessitates dynamic flood risk evaluation as a future research direction. When employing machine learning for flood vulnerability analysis, the selected indicator factors may yield divergent assessment outcomes. Even when identical indicator factors are utilized, employing different machine learning models for prediction produces markedly distinct results. This primarily stems from variations in the contribution rates of indicator factors across different machine learning models. Flood risk analyses based on machine learning algorithms must be cross-validated against results from conventional flood risk assessment methods (such as mathematical-statistical approaches, indicator-based systems, and scenario simulation techniques) to ensure mutual corroboration of assessment outcomes.

6.Conclusions

In this study, machine learning models (MLP, RBF and BP) are used to analyze the flood susceptibility of the highly urbanized Qianshan River Basin, and the MCDA method is used to evaluate the flood risk. The main conclusions are as follows:

- Through training, testing and evaluation, the MLP model better in prediction accuracy and performance compared with the RBF and BP model, and is more suitable for flood susceptibility analysis. Elevation, land use, and slope are the top three factors that affect the model's prediction results.
- Based on the MCDA, the flood susceptibility map obtained by MLP model and the hazard (rainfall) and exposure (population density, GDP distribution) were superimposed and analyzed, and the flood risk zoning map of the study area was drawn. The highest and high risk areas are mainly located in the southeastern part of the basin and require more stringent flood prevention measures.
- In this paper, it is proposed to integrate the machine learning algorithm with the MCDA for flood assessment. Flood risk assessment results obtained by the method proposed in this paper can provide a scientific basis for urban risk management decisions.

Declarations

Ethical Approval

All authors kept the ‘Ethical Responsibilities of Authors’.

Consent to Participate

All authors gave explicit consent to participate in this work

Consent to Publish

All authors gave explicit consent to publish this manuscript.

Acknowledgments

The authors wish to gratefully acknowledge the significant contribution by the editor of this journal, the reviewers for their painstaking efforts in evaluating the manuscripts and the editorial team for their unstinting support during the review of the submitted manuscripts.

Authors Contributions

G.L. wrote the manuscript and discussion and analyzed the data; W.S. conceptualized and performed the analysis; Y.Z. supervised the funding acquisition; acquired the data; Y.L. provided technical sights, as well as edited, restructured; Y.Z. provided professionally optimized the manuscript.

Funding

This work was supported by Shandong Agricultural Engineering College High-level Personnel Research Autonomous Project (2024GCCZR-10).

Competing Interest

The authors declare that they have no conflict of interest.

Availability of data and material

The datasets used and analyzed during the current study are available from the corresponding author on request.

Reference

1. Stefanos, et al., Assessment of flood hazard based on natural and anthropogenic factors using analytic hierarchy process (AHP). 2013.
2. Chowdhuri, I., et al., Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India. 2019. **65**(5).
3. Flood disaster monitoring and evaluation in China. Global Environmental Change Part B Environmental Hazards. 2002. **4**(2-3): p. 33-43.
4. Xu, Z.G. Zhao. Impact of urbanization on rainfall-runoff processes: case study in the Liangshui River Basin in Beijing, China. 2016.
5. Barbosa, A.E., et al., Key issues for sustainable urban stormwater management. 2012. **46**(20): p. 6787-6798.
6. A, S.M., et al., Flood risk and adaptation strategies under climate change and urban expansion: A probabilistic analysis using global data. 2015. **538**(4): p. 445-457.
7. Ceola, et al., Topography- and nightlight-based national flood risk assessment in Canada. 2017. **21**(4): p. 2219-2232.
8. Bingyan, M., et al., Process-oriented SWMM real-time correction and urban flood dynamic simulation. Journal of Hydrology. 2022. **605**.
9. Sahar, B., et al., Urban flood simulation and prioritization of critical urban sub-

catchments using SWMM model and PROMETHEE II approach (vol 105, pg 3, 2018). PHYSICS AND CHEMISTRY OF THE EARTH. 2020. **116**.
10. Applicability of SWMM for Semi Urban Catchment Flood modeling using Extreme Rainfall Events. International Journal of Recent Technology and Engineering. 2019. **8**(2).
11. Lee, K.S., et al., Numerical Simulation for Reducing the Flood Damage of Green Park Using MIKE URBAN. 2015. **8**(1): p. 37-54.
12. Jiahong, L., et al., Urban flood analysis for different design storm hyetographs in Xiamen Island based on TELEMAC-2D. Chinese Science Bulletin. 2019. **64**(19).
13. Vu, T.T., et al., Two-Dimensional Hydrodynamic Modelling of Flood Inundation for a Part of the Mekong River with TELEMAC-2D. British Journal of Environment and Climate Change. 2015. **5**(2).
14. Cheng, T., et al., Flood Risk Zoning by Using 2D Hydrodynamic Modeling: A Case Study in Jinan City. 2017. **2017**(pt.9): p. 5659197.5659191-5659197.5659198.
15. Zhang, S.B.J.J.o.H. Pan, An urban storm-inundation simulation method based on GIS. 2014. **517**: p. 260-268.
16. Jinru, W., et al., Assessment of long and short-term flood risk using the multi-criteria analysis model with the AHP-Entropy method in Poyang Lake basin. International Journal of Disaster Risk Reduction. 2022. **75**.
17. Constantin, B., et al., Flood Risk Assessment Based on Flood Hazard and Vulnerability Indexes. Ovidius University Annals of Constanta - Series Civil Engineering. 2020. **22**(1).
18. Tang, Z., et al., Incorporating probabilistic approach into local multi-criteria decision analysis for flood susceptibility assessment. 2018. **32**(3): p. 701-714.
19. Chen, X., et al., Urbanization and climate change impacts on future flood risk in the Pearl River Delta under shared socioeconomic pathways. 2020. **762**: p. 143144.
20. Shereif, et al., Multi-criteria approach to develop flood susceptibility maps in arid regions of Middle East. 2018.
21. Chowdary, V.M., et al., Multi-Criteria Decision Making Approach for Watershed Prioritization Using Analytic Hierarchy Process Technique and GIS. 2013. **27**(10): p. 3555-3571.
22. Kia, M.B., et al., An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. 2012. **67**(1): p. 251-264.
23. Btp, A., et al., Flood risk assessment using deep learning integrated with multi-criteria decision analysis. **219**.
24. Tehrany, et al., Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. 2014.
25. Tehrany, M.S., et al., Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method. 2015. **29**(4): p. 1149-1165.
26. Li, X., et al., Flood Risk Assessment of Global Watersheds Based on Multiple

Machine Learning Models. 2019. **11**(8): p. 1654.
27. Nandi, A., et al., Flood hazard mapping in Jamaica using principal component analysis and logistic regression. 2016. **75**(6): p. 1-16.
28. A, Z.W., et al., Flood hazard risk assessment model based on random forest. 2015. **527**: p. 1130-1141.
29. Victoriano, J.M., et al., Predicting Pollution Level Using Random Forest: A Case Study of Marilao River in Bulacan Province, Philippines. 2019. **3**.
30. Chen, J., et al., Towards better flood risk management: Assessing flood risk and investigating the potential mechanism based on machine learning models. 2021. **293**(1): p. 112810.
31. Guoyi, L.L. Jiahong, Urban Flood Risk Assessment Based on TELEMAC-2D Model. Water Resources Protection, 2022: p. 1-13.
32. Rahmati, O., et al., Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. 2016. **31**(1): p. 42-70.
33. Pourali, S.H., et al., Topography Wetness Index Application in Flood-Risk-Based Land Use Planning. 2016. **9**(1): p. 39-54.
34. Tehrany, et al., Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. 2015.
35. Khosravi, K., et al., A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. 2018. **627**: p. 744–755.
36. Pham, B.T., et al., Bagging based Support Vector Machines for spatial prediction of landslides. 2018. **77**(4): p. 146.141-146.117.
37. Del, I.V.A.V.J.J.o.P.C.S. Starchenko, Forecast of the near ground air temperature based on the multilayer perceptron model. 2021. **1989**(1): p. 012025-.
38. Voyant, C., et al., Uncertainties in global radiation time series forecasting using machine learning: The multilayer perceptron case. 2017. **125**: p. 248-257.
39. El Yacoubi, S., et al., A multilayer perceptron model for the correlation between satellite data and soil vulnerability in the Ferlo, Senegal. 2018: p. 1-10.
40. Altunkaynak, A.T.A.J.U.w.j. Nigussie, Monthly water demand prediction using wavelet transform, first-order differencing and linear detrending techniques based on multilayer perceptron models. 2018. **15**(1-2): p. 177-181.
41. Widiasari, I.R., et al. Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network based hydrology time series data mining. in 2017 International Conference on Innovative and Creative Information Technology (ICITech). 2017.
42. LU, et al. Research on Control Optimization of Flood Discharge Gate at Navigation and Electricity Junction Based on RBF Neural Network. 2020.
43. Ji-Hong, Q.U., et al., Groundwater level prediction based on hybrid hierarchy genetic algorithm and RBF neural network. 2007. **13**(2): p. p.170-174.
44. Rumelhart, D.E.J.L.J.l. Mcclelland, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. 1986.

45. Zhangbao, L., Research on Geological Disaster Forecast Method Based on PCA Structure BP Model. Journal of Physics: Conference Series. 2021. **2083**(4).
46. Zhu, C., et al., Comparison of GA-BP and PSO-BP neural network models with initial BP model for rainfall-induced landslides risk assessment in regional scale: a case study in Sichuan, China. Natural Hazards. 2020. **100**(1).
47. Wen, L.Y. Liu, A research about Beijing's carbon emissions based on the IPSO-BP model. Environmental Progress & Sustainable Energy. 2017. **36**(2).
48. Yuehua, Z., et al., Research progress on risk assessment of heavy rainfall and flood disasters in China. Torrential Rain and Disasters, 2019. **38**(5): p. 8.
49. Flood risk assessment for urban water system in a changing climate using artificial neural network %J Natural Hazards. 2015. **79**(2): p. 1059-1077.
50. Chen, C.S., et al., Development and application of a decision group Back-Propagation Neural Network for flood forecasting. 2010. **385**(1-4): p. 173-182.
51. Thai, P.B., et al., Machine Learning Methods of Kernel Logistic Regression and Classification and Regression Trees for Landslide Susceptibility Assessment at Part of Himalayan Area, India. 2018. **11**(12): p. 1-10.
52. Khosravi, K., et al., A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods %J Journal of Hydrology . 2019. **573**: p. 311-323.
53. Pham, B.T., et al., Flood risk assessment using hybrid artificial intelligence models integrated with multi-criteria decision analysis in Quang Nam Province, Vietnam. 2020. **592**.
54. Chen, et al., Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. 2018.
55. Mohammady, M., et al., Landslide susceptibility mapping at Golestan Province, Iran: A comparison between frequency ratio, Dempster–Shafer, and weights-of-evidence models. 2012. **61**: p. 221-236.
56. Chang, L.-C., et al., Building an Intelligent Hydroinformatics Integration Platform for Regional Flood Inundation Warning Systems. 2019. **11**(1): p. 9.
57. Chang, L.-C., et al., Building ANN-Based Regional Multi-Step-Ahead Flood Inundation Forecast Models. 2018. **10**(9): p. 1283.
58. Rui, S., et al., Impact of various flood scenarios on ruban emergency response times based on the TELEMAC-2D model. J Tsinghua Univ (Sci & Technol), 2022. **62**(01): p. 60-69.
59. Wang, K.J.R.S., Dynamic Assessment of the Impact of Flood Disaster on Economy and Population under Extreme Rainstorm Events. 2021. **13**.
60. Hosseini, F.S., et al., Flash-flood hazard assessment using Ensembles and Bayesian-based machine learning models: application of the simulated annealing feature selection method. 2019. **711**: p. 135161.
61. Degiorgis, M., et al., Classifiers for the detection of flood-prone areas using remote sensed elevation data. 2012. **470-471**(none).