Advancing flood susceptibility modeling using stacking ensemble machine learning:
A multi-model approach

YANG Huilin[1], *YAO Rui[1], DONG Linyao[2], SUN Peng[1], ZHANG Qiang[3], WEI Yongqiang[4], SUN Shao[5], AGHAKOUCHAK Amir[6]

1. School of Geography and Tourism, Anhui Normal University, Wuhu 241002, Anhui, China;
2. Changjiang River Scientific Research Institute, Wuhan 430015, China;
3. Advanced Interdisciplinary Institute of Environment and Ecology, Beijing Normal University, Zhuhai 519087, Guangdong, China;
4. Hunan Institute of Water Resources and Hydropower Research, Changsha 410007, China;
5. State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China;
6. Department of Civil and Environmental Engineering, University of California Irvine, Irvine, CA 92697, USA

Abstract: Flood susceptibility modeling is crucial for rapid flood forecasting, disaster reduction strategies, evacuation planning, and decision-making. Machine learning (ML) models have proven to be effective tools for assessing flood susceptibility. However, most previous studies have focused on individual models or comparative performance, underscoring the unique strengths and weaknesses of each model. In this study, we propose a stacking ensemble learning algorithm that harnesses the strengths of a diverse range of machine learning models. The findings reveal the following: (1) The stacking ensemble learning, using RF-XGB-CB-LR model, significantly enhances flood susceptibility simulation. (2) In addition to rainfall, key flood drivers in the study area include NDVI, and impervious surfaces. Over 40% of the study area, primarily in the northeast and southeast, exhibits high flood susceptibility, with higher risks for populations compared to cropland. (3) In the northeast of the study area, heavy precipitation, low terrain, and NDVI values are key indicators contributing to high flood susceptibility, while long-duration precipitation, mountainous topography, and upper reach vegetation are the main drivers in the southeast. This study underscores the effectiveness of ML, particularly ensemble learning, in flood modeling. It identifies vulnerable areas and contributes to improved flood risk management.

Keywords: flood susceptibility assessment; machine learning; stacking ensemble learning; flood drivers; Xiangjiang River Basin

1   Introduction

Floods stand as one of the most devastating natural disasters globally, leaving long-term impacts on ecosystems and human societies (Li *et al*., 2020; Silva *et al*., 2020). Simultaneously, floods impose a heavy toll on economies and the environment, causing widespread destruction in populated areas, agriculture, and water resources (Wu *et al*., 2019). Flood occurrences can stem from natural causes such as prolonged precipitation, heavy rainfall, and snowmelt, or they can manifest unnaturally due to urbanization, population expansion, and deforestation (Chan *et al*., 2018). In the wake of a 1.2℃ increase in global temperatures since the pre-industrial era (1850–1900), the frequency and intensity of rainfall events have escalated (IPCC, 2019). Alarmingly, projections suggest that over 2 billion individuals could be adversely impacted by floods by 2050 (Siegert *et al*., 2020). China, in particular, grapples frequently with flood and waterlogging disasters, notable for both the scale of affected populations and the ensuing economic losses (Zhao *et al*., 2018). Consequently, the imperative to mitigate flood-induced damages in susceptible regions has emerged as a paramount strategy (Chen *et al*., 2019).

Prior methodologies employed for assessing flood susceptibility have predominantly encompassed statistical techniques and hydrological models. Statistical approaches, such as regression analysis, bivariate statistics, and multi-criteria decision analysis (Youssef *et al*., 2016; Costache and Bui, 2020), often rely on expert insights, consequently introducing an element of uncertainty into their outcomes. Conversely, hydrological models aim to decode the intricate interplay between floods and influential factors by emulating hydrological processes (Rozalis *et al*., 2010; Li *et al*., 2019; Gao *et al*., 2024). However, the implementation of hydrological modeling mandates meticulously curated long-term hydrologic, meteorological, and topographic data, which can be elusive in less documented or ungauged basins. Additionally, the site-specific nature of hydrological models restricts their adaptability to dissimilar basins, rendering them less suitable for expansive regional investigations.

In recent years, the advent of big data collected through diverse sensors for weather monitoring, coupled with new computational paradigms and significant strides in high-performance computing technology, has fostered the era of big data. Within this landscape, the application of machine learning (ML) methods has gained traction in modeling flood susceptibility. At its core, flood susceptibility modeling hinges on constructing predictive models grounded in the correlation between flood occurrences and influencing factors. This endeavor results in the deduction of flood susceptibility, manifested through spatial distribution maps (Liu *et al*., 2022). Notably, the ambit of flood susceptibility modeling spans varied geographical scales, ranging from local basins and regions to the broader national context (Shahabi *et al*., 2020).

Prominent ML techniques fueling these modeling endeavors encompass a spectrum of methodologies. These include the acclaimed Deep transfer learning based on transformer (Xu *et al*., 2023), LSTM (Zhang *et al*., 2018), Random Forest (RF) (Chapi *et al*., 2017; Chen *et al*., 2020), Support Vector Machine (SVM) (Tehrany *et al*., 2015; Zhang *et al*., 2023), Artificial Neural Networks (ANN) (Zhao *et al*., 2018; Li *et al*., 2022), Generalized Linear Models (GLM) (Vandenberg-Rodes *et al*., 2016), Logistic Regression (LR) (Tien *et al*., 2019) and Naive Bayes (NB) (Khosravi *et al*., 2019). These advanced techniques not only under-

score the versatility of ML but also highlight its burgeoning importance in addressing intricate challenges within flood susceptibility modeling.

Given the inherent limitations of individual machine learning (ML) models, such as overfitting and instability, various ensemble learning methodologies have been adopted to improve flood susceptibility assessment. Ensemble learning combines predictions from multiple ML models using techniques like bagging, boosting, average weighted approaches, and stacking. Bagging and boosting unite models of the same type to mitigate singular model instability or bias, such as decision tree-based bagging (Ha *et al*., 2021), Gradient Boosting Decision Tree (GBDT) (Chen *et al*., 2021), eXtreme Gradient Boosting (XGBoost) (Ma *et al*., 2021). The weighted averaging method assigns a weight to each base model that is used to compute the weighted average. These weights can be determined based on the performance of the model on the validation set, and usually better performing models are assigned larger weights (Mahato *et al*., 2021). Stacking typically fuses diverse model types to leverage their respective strengths and mitigate weaknesses (Kuhn and Johnson, 2013), while stacking reduces the risk of overfitting and has better generalization capabilities. The stacking method has been evaluated as a high-performance model in existing studies, with better accuracy and stability compared to the first three methods (Lv *et al*., 2022; Li *et al*., 2023; Wu and Wang, 2023). However, the application of stacking for flood susceptibility assessment remains under-explored, thus forming the primary focus of this study. It is paramount to emphasize that a theoretically optimal model may not inevitably yield superior outcomes in practical implementations. For instance, in evaluating flash flood susceptibility in Markazi, Iran, Pham *et al*. (2020) observed AdaBoost's superior accuracy compared to bagging. Similarly, in a flood risk investigation within China's Pearl River Delta, Chen *et al*. (2021) identified GBDT's superior performance compared to XGBoost, typically regarded as possessing superior learning capabilities. Consequently, the applicability or generalizability of diverse ML methods under distinct geographical contexts remains elusive.

The Xiangjiang River Basin (XRB), spanning the largest drainage area and runoff in Hunan province, China, lies within a subtropical monsoon climate and is acutely susceptible to summer floods from June to August. In July 2006, the region suffered heavy rainfall due to the severe tropical storm "Bilis", triggering floods that impacted 7.29 million people and incurred economic losses exceeding 780 billion yuan (Du *et al*., 2006). The summer of 2017 witnessed an extreme rainfall event, coupled with heavy reliance on riverbank levees for flood control, leading to record water levels in the XRB. This accentuates the pressing need for comprehensive flood risk assessment within the XRB and the identification of primary contributing factors.

Hence, this study endeavors to delineate flood-prone zones within the XRB utilizing diverse models, comparing their outcomes, and identifying flood occurrence determinants (Figure 1). The primary contributions of this study encompass: (1) developing a stacking ensemble model for quantitative flood susceptibility assessment within the XRB, evaluating its performance through climatic, topographical, and environmental lenses; (2) identifying and quantitatively analyzing key influencing factors across distinct flood-prone areas; and (3) quantitatively examining population, construction land, and cropland exposure to floods within the XRB.
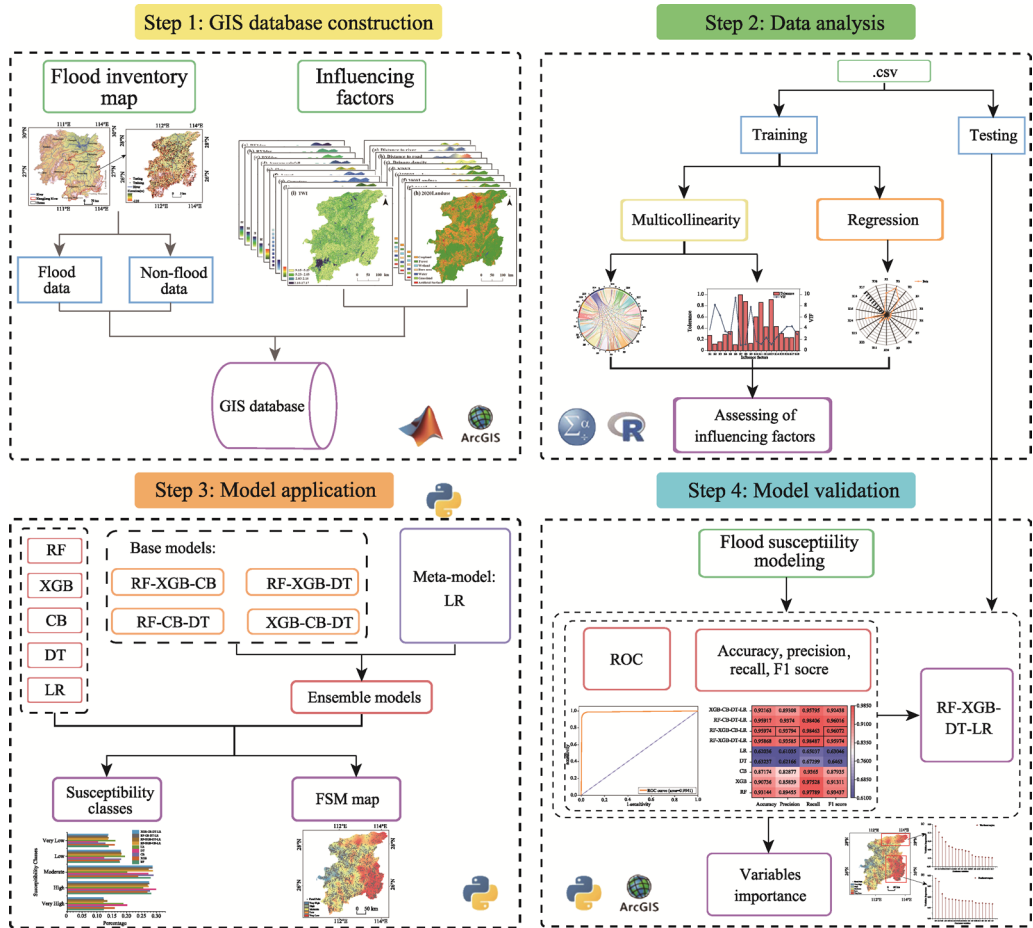
Figure 1 The modeling framework outlined in this article

2   Data source

In this study, the flood inventory map data were provided by the Hunan Provincial Meteorological Bureau (hn.cma.gov.cn), and a total of 5989 flood events occurred in the XRB (Figure 2). Among these, 70% of the flood records were used for modeling purposes, while the remaining 30% were used to assess the model's predictive ability. Based on the literature review, data availability, and insights from previous studies (Fang *et al.*, 2022; Zeng *et al.*, 2022), 18 flood susceptibility factors were selected (Figures 3 and 4): maximum 1-day precipitation (RX1day), maximum 3-day precipitation (RX3day), maximum 5-day precipitation (RX5day), average precipitation during the flood season, elevation, slope, aspect, curvature, Topographic Roughness Index (TRI), Topographic Wetness Index (TWI), distance to rivers, distance to roads, drainage density, NDVI, 1990 landuse, 2000 land use, 2010 land use, and 2020 land use.

   The RX1day dataset contains daily precipitation data from 195 meteorological stations in the XRB. It covers the period from January 1, 1960 to December 31, 2019. This dataset is provided by the National Climate Center of the China Meteorological Administration. RX3day, RX5day, and average precipitation during the flood season are computed from the
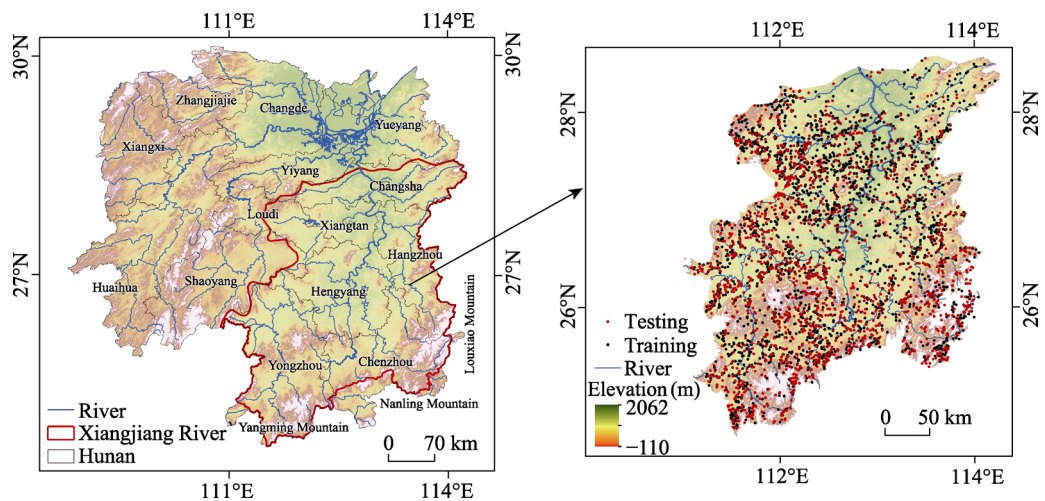
Figure 2 Geographical location of the Xiangjiang River Basin (The red points are the test set for flooding and the black points are the training set for flooding)
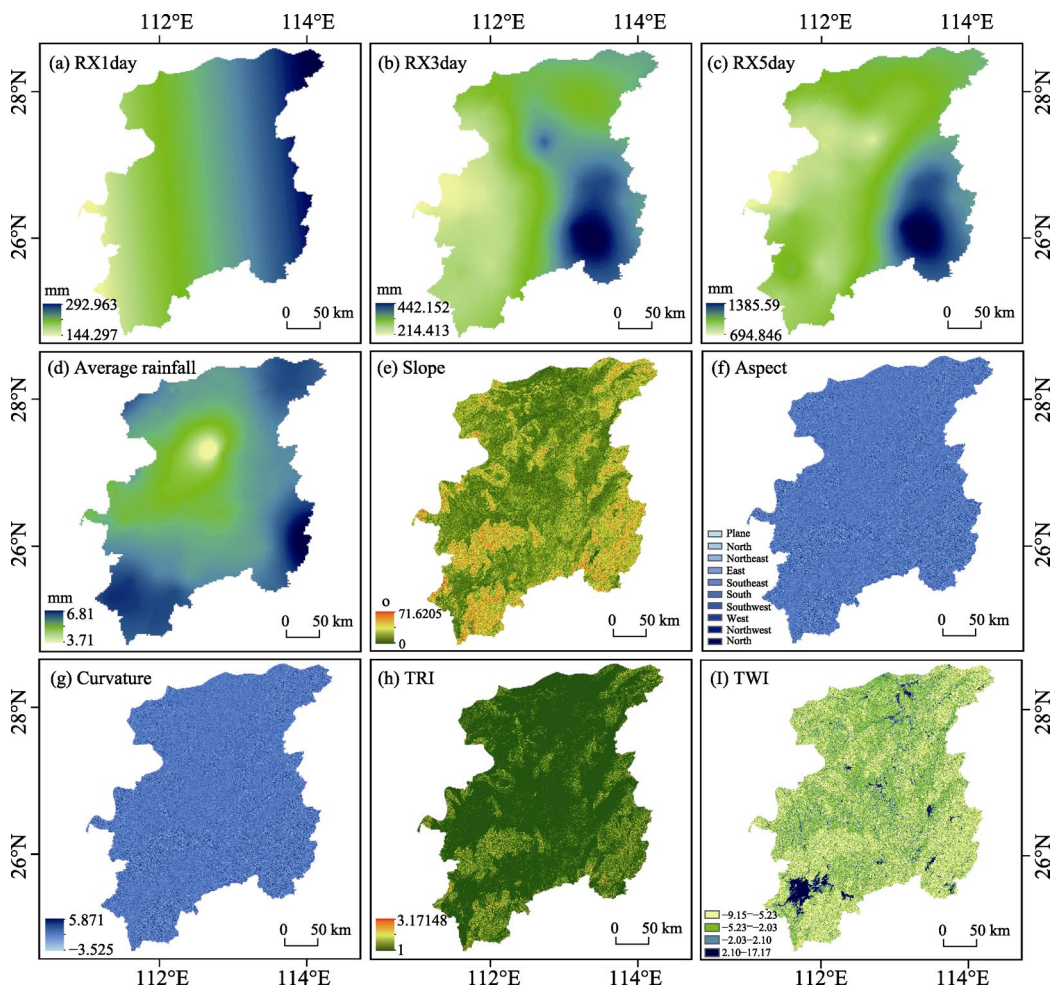

Figure 3 Climatic and topographic information used for flood susceptibility in the Xiangjiang River Basin
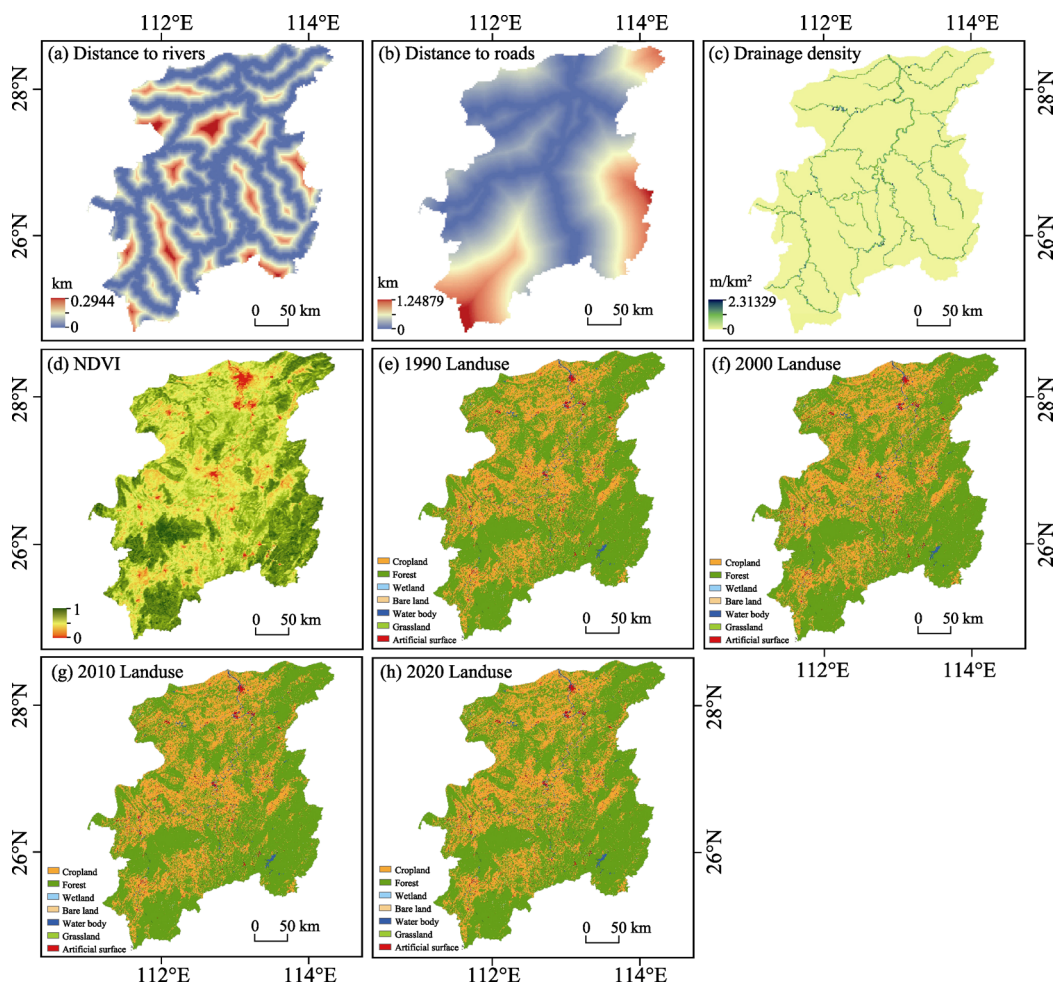
Figure 4   Environmental information used for flood susceptibility in the Xiangjiang River Basin

RX1day. The DEM data are derived from the Geospatial Data Cloud (https://www.gscloud. cn/) for extracting slope, aspect, curvature, TRI, and TWI. The spatial resolution of this data is 90 m. Water system data originated from the 1:250,000 national basic geographic database in the National Geographic Information Resource Catalog Service System (http://www.webmap.cn). Road data originated from the Geographic Information Expertise Service system (kmap.ckcest.cn) in a 1:250,000 dataset. NDVI data were obtained from Landsat-8 image processing. Land use data originated from the Resource and Environment Science and Data Center of the Chinese Academy of Sciences (https://www.resdc.cn). The data are described in Table 1.

3   Methods

3.1   GIS database development

Based on the flood data obtained from the flood inventory maps, a random sampling method was used to generate non-flood data. The GIS database was then constructed by combining

Table 1 Information of flood susceptibility factors

Factor type	Susceptibility factors	Data description	Abbreviation
Climatic factors	RX1day	Maximum rainfall in 1 day	X1
	RX3day	Maximum rainfall in 3 days	X2
	RX5day	Maximum rainfall in 5 days	X3
	Average precipitation during the flood season	Average rainfall April-August	X4
Topographic and geomorphologic factors	Elevation	Water usually runs off from higher elevations and collects at lower elevations	X5
	Slope	Measuring the depth of the ground surface, which directly affects the rate of surface runoff	X6
	Aspect	Identifying the direction of the steepest downslope at a location on the surface	X7
	Curvature	Measurement of unevenness	X8
	TRI	The lower the surface roughness, the more prone to flooding	X9
	TWI	Indicating the extent of waterlogging in the catchment	X10
Environmental factors	Distance to rivers	Areas near rivers are more prone to flooding	X11
	Distance to roads	Areas near roads are more prone to flooding	X12
	Drainage density	Water abundance in the region	X13
	NDVI	The state of the vegetation	X14
	1990 Landuse	The way natural attributes of land were utilized in 1990	X15
	2000 Landuse	The way natural attributes of land were utilized in 2000	X16
	2010 Landuse	The way natural attributes of land were utilized in 2010	X17
	2020 Landuse	The way natural attributes of land were utilized in 2020	X18

the flood susceptibility factors. In this database, all flood data are given a value of "1", while all non-flood ones are given a value of "0". The impact factors were reclassified using the natural breakpoint classification method. To further analyze the data, frequency ratios (FR) are adapted to weighting the data during the preprocessing step. The formula used for FR is as follows:

$$FR = \frac{A / A'}{B / B'} \tag{1}$$

where A represents the number of floods in each influencing factor, A' represents the number of all floods, B is the number of pixels in a particular class, and B' is the total number of pixels (Wang *et al*., 2021). In addition, the FR value of each influencing factor was normalized, and the formula of normalization frequency ratio (NFR) is:

$$U' = \frac{u - Min(u)}{Max(u) - Min(u)}(E - F) + F \tag{2}$$

where U' is the normalized value, u is the original value, E and F are the upper and lower normalization boundary (Tien *et al*., 2012).

3.2   Data processing and analysis

3.2.1   Tests for multicollinearity

Spearman's rank correlation coefficient, tolerance, and variance inflation factor were utilized

to judge the multicollinearity between the influencing factors (Islam *et al*., 2021). Spearman>0.7 indicates a high degree of dependence. The tolerance value < 0.1 indicates the presence of covariance between this independent variable and other independent variables. When VIF < 10, there is no multicollinearity. While VIF≥10, there is strong multicollinearity.

3.2.2 Multiple linear regression (MLR)

Standardized coefficients (Beta values) are used in multiple regressions to compare the significance between variables. A higher Beta value indicates that the variable has a greater influence on the dependent variable.

3.3 Flood susceptibility modeling

3.3.1 Machine learning model

In this paper, five ML models, namely RF, XGB, CB, DT and LR, are selected to assess flood susceptibility and testing a multi-model approach in the XRB. RF, XGB, and CB were chosen in this paper mainly because these three models are themselves integrated learning algorithms, and it has been demonstrated that integrated models outperform other algorithms (Zhao *et al*., 2018; Seydi *et al*., 2023; Ren *et al*., 2024). DT was chosen mainly because DT modeling offers several processing methods such as squared automatic interaction detection (CHAID), classification and regression trees (CRT), and fast, unbiased, and efficient statistical trees (QUEST) (Roe *et al*., 2005). CHAID is the most suitable for the purpose of our modeling because the conditional (predictor) factor selected at each step is the one that has the strongest relationship with the dependent variable factor. In decision tree analysis, creating each new branch of the tree is considered a step. Conditional factor categories are merged if they show significant differences relative to the dependent variable. In terms of its speed and ability to split multiplexed nodes, CHAID in DT is best suited for susceptibility modeling (Kusiak *et al*., 2010). The rationale for choosing LR is that metamodel selection is usually simple during stacking modeling, and therefore it provides a smooth interpretation of the predictions made by the underlying classifiers. Therefore, it is recommended that the linear model be used as a meta-classifier for linear regression, and for the task of classifying predictive class labels, DT should be used (Yaseen *et al*., 2022).

The fundamentals of flood vulnerability modeling begin with dividing the dataset into two subsets: the training set (70% of the total observations) and the test set (30%). Random Forest (RF) is a machine learning algorithm that does not rely on any assumptions about the statistical distribution of the data (Breiman, 2001). It is capable of balancing datasets that are not evenly distributed. Even if a majority of the characteristics are absent, it is still possible to preserve the precision (Zhu *et al*., 2020). For the classification problem, RF consists of a set of classification trees. Each tree is constructed using a subset of the entire training set and a random portion of the input features. The classification tree divides the sample data into smaller homogeneous groups, where homogeneity can be defined by various metrics. When making predictions about new samples, each tree votes on a class. The class probability of a new sample is defined by the proportion of votes for each class, and the class with the most votes is the predicted class, the voting process is explained below:

$$sH(x) = \arg_z^{\max} \sum_{i=1}^{k} I(h_i(x) = Z) \tag{3}$$

where $H(x)$ is the model, h_i is the individual tree structure, I is the indicative function, and Z is the final prediction.

XGBoost (XGB) is a complex ensemble learning algorithm based on classification or regression trees (Chen and Guestrin, 2016). XGB aims to minimize computational complexity and improve accuracy (Merghadi *et al.*, 2020). XGB considers a dataset containing m features and n instances.

$$s\hat{y}_i = \phi(x_i) = \sum_{a=1}^{A} f_a(x_i), f_a \in F \tag{4}$$

$$\hat{y}_i = \phi(x_i) = \sum_{a=1}^{A} f_a(x_i), f_a \in F \tag{5}$$

where F denotes the space of the tree, which can be computed as:

$$F = \left\{ f(x) = w_{q(x)} \right\} (q : R^m \to T, w \in R^T) \tag{6}$$

where q is the structure of a single tree and T denotes a leaf tree within a single tree. In XGB algorithm, in order to optimize the integration tree and minimize the error, the objective function is calculated as follows:

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{7}$$

where l is a loss function that quantifies the error between the measurements and the estimates, and t denotes the number of iterations used to minimize the error. Ω denotes the penalty for model complexity using the regression tree function as follows:

$$f_k = \gamma T + \frac{1}{2} \lambda \| w^2 \| \tag{8}$$

CatBoost (CB) is a Gradient Boosting Decision Tree (GBDT) framework based on symmetric decision trees. It has the advantages of fewer parameters, support for categorical variables, and high accuracy. The main pain point it addresses is the efficient and reasonable handling of categorical features. In addition, the issue of gradient bias and prediction bias is effectively tackled by CB. This helps to minimize overfitting and enhances the accuracy and generalization of the algorithm (Saber *et al.*, 2023). CB depends mainly on the use of gradient boosting, which uses a binary tree classification scheme.

Assume a dataset:

$$D = \left\{ (X_J, Y_J) \right\} \qquad (J = 1, ..., m) \tag{9}$$

where X_J is the combination of attributes and $Y_J \in R$ denotes the desired goal. The input and output datasets are equally and independently dispersed according to the unknown function, and the goal of the learning technique is to train and check that the function $H:R_n \to R$ reduces the loss of information, i.e., $L(H)=E_L(y,H(X))$, where L is the smoothing error function, and (X,y) denotes the test samples from D. The learning technique is based on a series of approximations, $H_t=H(t–1)=gt$, which are derived from a priori approximations. The gradient enhancement method creates a series of approximations H_t, with $H_t=H(t–1)=gt$ being the final function generated from the a priori approximations (Saber *et al.*, 2023).

Decision Tree (DT) is a model for decision making judgment based on a tree structure, which categorises the dataset through multiple conditional discriminative processes and fi-

nally obtains the desired result. There is no need for pre-established associations between input and target variables (Saito *et al*., 2009). DT grades and uniformly categorizes conditioning factors according to the level of susceptibility. The purpose of constructing a tree is to establish a set of decision rules that can form the basis for predicting the outcome of the set of input variables (Debeljak and Džeroski, 2011). Thus, rules are generated by analyzing a set of factors with the aim of predicting outcomes from a set of similar variables (Myles *et al*., 2004).

The use of Logistic Regression (LR) was due to the nature of the data used in this study, which included categorical and continuous data as predictor variables and the presence or absence of flood data variables as outcomes. LR analyses of the training data using forward stepwise regression aim to find the most prudent set of predictors. These predictors were effective in estimating the outcome variables (Kleinbaum and Klein, 2010). LR is a commonly used ML algorithm for binary classification problems. Assuming the probability of flood occurrence is *p*, LR models the logarithm of flood occurrence as a linear function, as shown in (10):

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k \tag{10}$$

where X_i (*i*=1, 2, …, *k*) is a factor associated with flood occurrence, with a total of *k* factors. β_i is the regression coefficient and β_0 is the intercept.

3.3.2 Stacking ensemble modeling

Stacking, first proposed by David (1992), is an ensemble algorithm designed to reduce the generalization error. It accomplishes this by integrating multiple regression or classification algorithms during the training phase. Stacking can enhance overall prediction accuracy by synthesizing the outputs of multiple models. In the first stage, the original dataset is sliced and divided into training sets and test sets based on specific ratios. Suitable base learners are then chosen through cross-validation to train the training set. In the case of stacking K-Fold, the dataset is divided into K parts. One part is designated as the test set, and the remaining parts are used as the training set in K iterations. The final result's accuracy or value is determined by averaging the outcomes obtained from these K iterations. Each model generates a K-Fold prediction, and the data gathered from the K-Fold training of the first layer of models are combined to form a training set for base models.

In the second phase, the predictions from the base models serve as feature data for training and predicting the meta-learner, respectively. The meta-learner is then merged with the labels of the original dataset, along with the feature and sample data acquired in the previous phase. This amalgamated information is utilized to construct the model and generate the final stacking prediction results (Figure 5).

The selection of the base model layer and the meta-model layer is the most significant part of building stacking (Dou *et al*., 2019). The base model layer is more complex due to its numerous classifiers, while the meta-model is chosen to be a relatively simple LR. (Adeli *et al*., 2020). According to the classifier, the choice of base model should be strong and multiple (Pourghasemi *et al*., 2017). Therefore, the base model is selected out of four schemes: RF-XGB-CB, RF-XGB-DT, RF-CB-DT, and XGB-CB-DT.
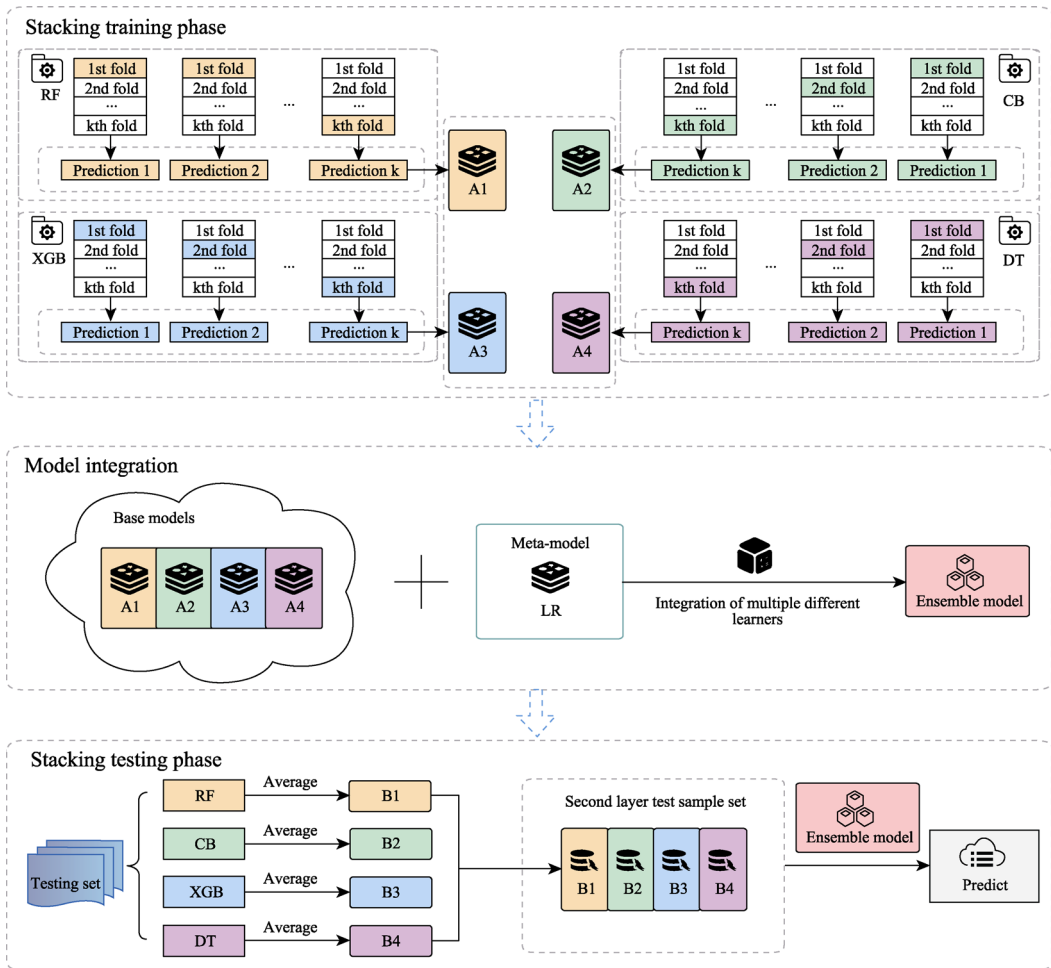
Figure 5   Procedure of stacking for ensemble learning

3.3.3   Validation and performance of the models

To evaluate the flood sensitivity effect, this paper used several statistical metrics of Precision, Recall, Accuracy, and F1 Score to evaluate the model's performance (Chen *et al*., 2018; Zhao *et al*., 2020). Additionally, the predictive ability of the model was evaluated using the ROC curve (Pham *et al*., 2016). The higher the ROC value indicates better model performance, with a range of 0.5(less accurate) to 1(highly accurate) (Bui *et al*., 2015). A ROC value less than 0.6 indicates model failure, while a value above 0.8 suggests better model performance.

4   Results

4.1   Identifying flood susceptibility factors

Figure 6a displays the Spearman correlation among flood susceptibility factors. The correlation between curvature, distance to rivers, drainage density, and NDVI appeared relatively low, with the lowest correlation observed between curvature and NDVI at 0.0018. On the

other hand, RX1day and RX3day, 1990 landuse and 2010 landuse, and RX3day and RX5day demonstrated higher correlations, with the highest correlation being 0.6915 between RX1day and RX3day. However, as this value falls short of 0.7, it suggests that the flood susceptibility factors were mutually independent. A linear regression test for collinearity among 18 factors revealed high VIF values for RX3day, slope, and TRI, with slope exhibiting the highest value at 9.401. In contrast, low tolerance values were found for RX3day, slope, and TRI, with slope showing the lowest value (0.106). Thus, all flood susceptibility factors meet the requirements of VIF<10 and tolerance>0.1 (Figure 6b).

Figure 7 illustrates the contributions of individual flood susceptibility factors, assessed using the multiple linear regression method. The analysis highlighted RX3day as the most
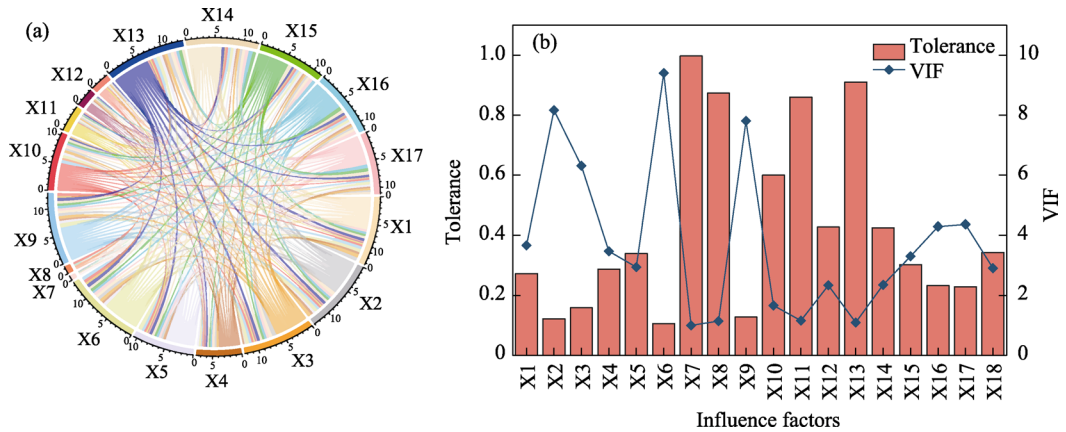

Figure 6 Multicollinearity test between flood susceptibility factors (X1: RX1day; X2: RX3day; X3: RX5day; X4: Average precipitation during the flood season; X5: Elevation; X6: Slope; X7: Aspect; X8: Curvature; X9: TRI; X10: TWI; X11: Distance to rivers; X12: Distance to roads; X13: Drainage density; X14: NDVI; X15: 1990 Landuse; X16: 2000 Landuse; X17: 2010 Landuse; X18: 2020 Landuse)
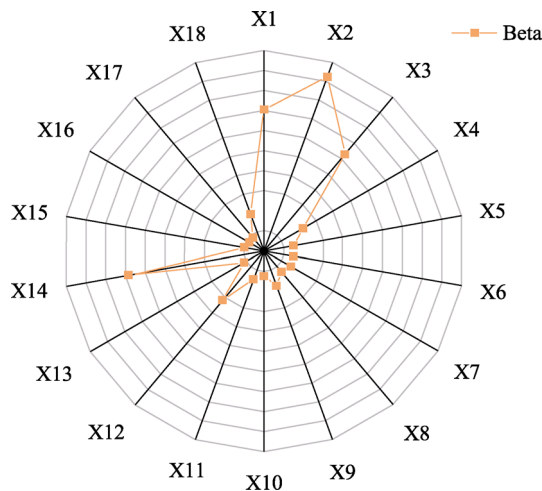

Figure 7 Contribution values of various flood susceptibility factors to floods (X1: RX1day; X2: RX3day; X3: RX5day; X4: Average precipitation during the flood season; X5: Elevation; X6: Slope; X7: Aspect; X8: Curvature; X9: TRI; X10: TWI; X11: Distance to rivers; X12: Distance to roads; X13: Drainage density; X14: NDVI; X15: 1990 Landuse; X16: 2000 Landuse; X17: 2010 Landuse; X18: 2020 Landuse)

influential factor with a contribution value of 0.185, followed by RX1day (0.141) and NDVI (0.137), emphasizing their primary roles in flood occurrence. Other factors such as RX5day, the average precipitation during the flood season, elevation, aspect, TRI, distance to roads, and land use in various years (1990, 2000, 2010, and 2020) exhibited contribution values ranging from 0.01 to 0.126. Notably, all recorded contribution values for flood susceptibility factors exceeded zero, indicating their significant impact on flood dynamics. Consequently, all identified factors were included in subsequent modeling endeavors.

4.2   Construction of flood susceptibility maps

Following the validation of flood impact factors, five ML models and four stacking ensemble learning models were employed to compute flood susceptibility values. The natural breakpoint classification method was utilized to divide flood susceptibility values into five classes: very high, high, moderate, low, and very low. Figure 9 presents the proportion of different flood susceptibility categories. Broadly, across all models, the high susceptibility category predominated over the other categories, accounting for 27.5%. For the high and very high categories, the area proportion in the DT model exceeded those of other models, standing at 29.82% and 20.11% respectively. In the LR model, the categories with substantial area proportions were the low (19.37%) and very low (16.11%) categories. The CB model predominantly predicted the moderate category, constituting 28.95%. Even though the performance of the models varies, the spatial distribution rendered by these nine models exhibited similar characteristics, all indicating that the high susceptibility areas were situated in the northeast and southeast of the XRB, while the low susceptibility areas primarily occupied the western region of the XRB (Figure 8). Collectively, over 40% of the XRB was marked by high and very high flood susceptibility.

4.3   Validation of model performance

The performance of the models was validated using ROC curve values (Figure 10), alongside Accuracy, Precision, Recall, and F1 score indicators (Figure 11). The results indicated that the stacking ensemble models generally outperformed single models in terms of both goodness of fit and generalization ability. The RF-XGB-CB-LR model displayed the highest ROC curve value (0.9941), and the overall values for Accuracy, Precision, Recall, and F1 score indicators were also the highest, implying optimal model performance. Conversely, the ROC curve values for the DT and LR models stood at 0.6760 and 0.6359 respectively, and the values of Accuracy, Precision, Recall, and F1 score indicators were relatively low, suggesting poorer model performance. The ROC curve values of the remaining models surpassed 0.94, denoting generally good performance.

4.4   Importance of flood impact factors

Based on the evaluations presented in Figures 10 and 11, the RF-XGB-CB-LR model was identified as the optimal model. Subsequently, Python was utilized to evaluate the feature importance of flood impact factors within this model, as depicted in Figure 12. The results revealed that X14 (NDVI) was the most critical factor, boasting an importance value of 0.22. This was followed by X3 (average precipitation during the flood season), X12 (distance to roads), and X1 (RX1day). In contrast, other factors demonstrated relatively low feature im-

portance, with land use in different years and drainage density appearing as the least important variables. Due to the complex relationship between impact factors and the model, each impact factor contributed differently to the model.
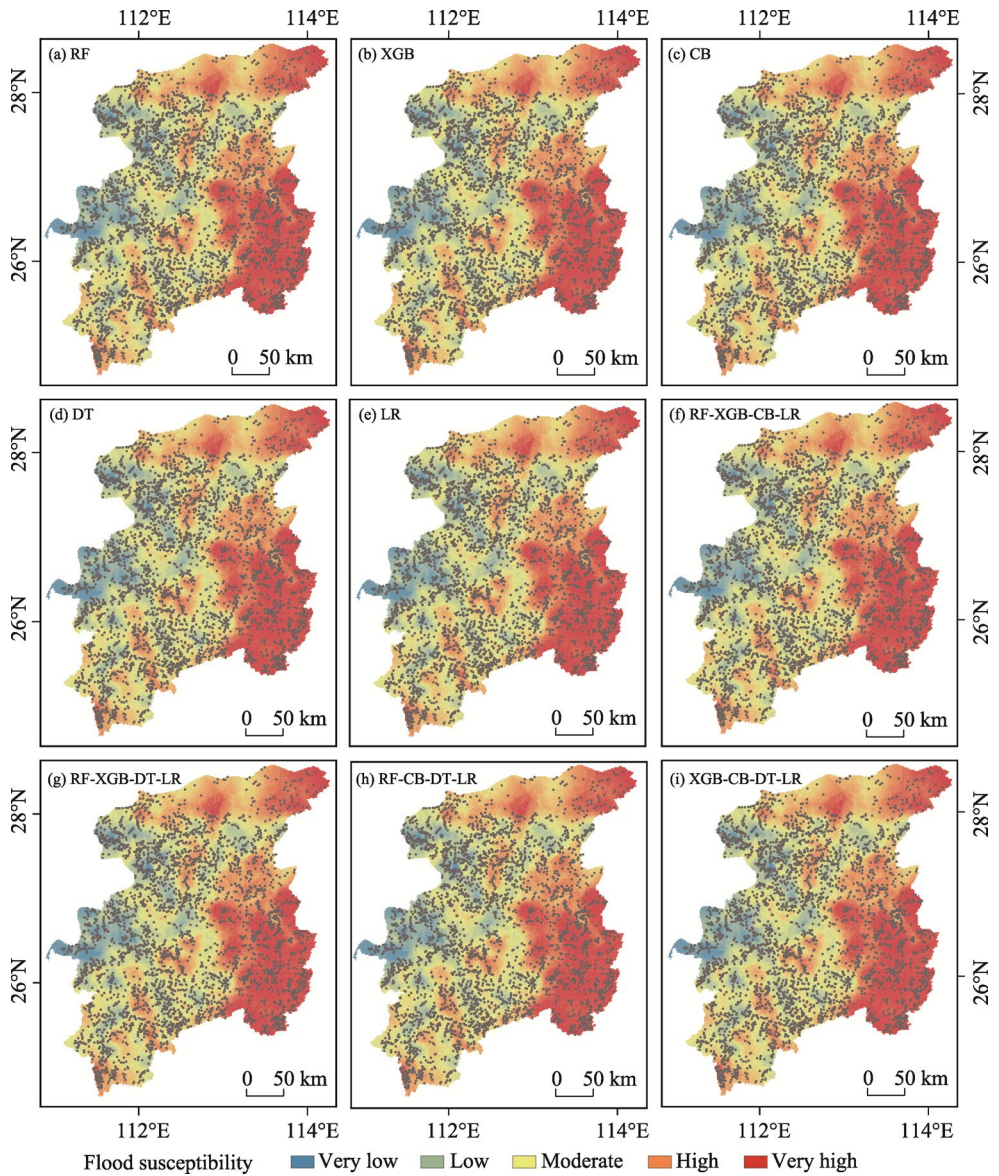

Figure 8   Flood susceptibility in the Xiangjiang River Basin

4.5   Statistical analysis of population, built environment, and cropland exposed to floods

The analysis of the percentage of population, built environment, and cropland exposed to floods is depicted in Figure 13. A significant number of people in the XRB were at risk due to high flood susceptibility areas. To be precise, 156,000 individuals, which made up 34% of the total basin population, reside in these areas (Figure 13a). When considering the exposure of buildings, the area of built environment affected by floods serves as an accurate indicator (Figure 13b). High flood susceptibility areas encompassed 870.872 km^2 of built environ-
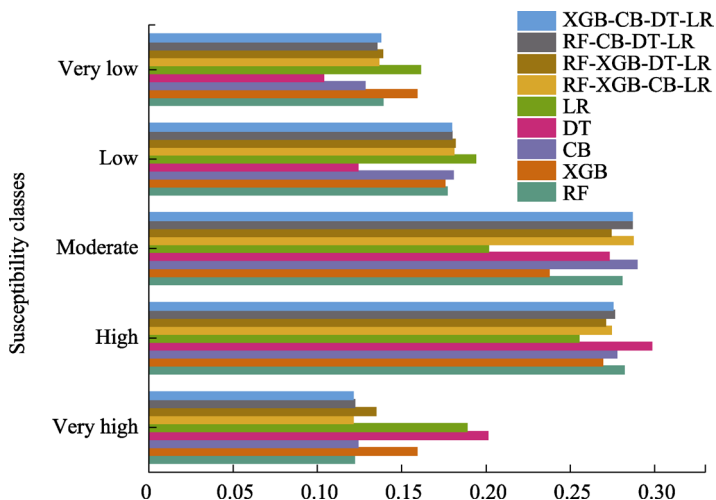
Figure 9   Percentage of different flood susceptibility categories
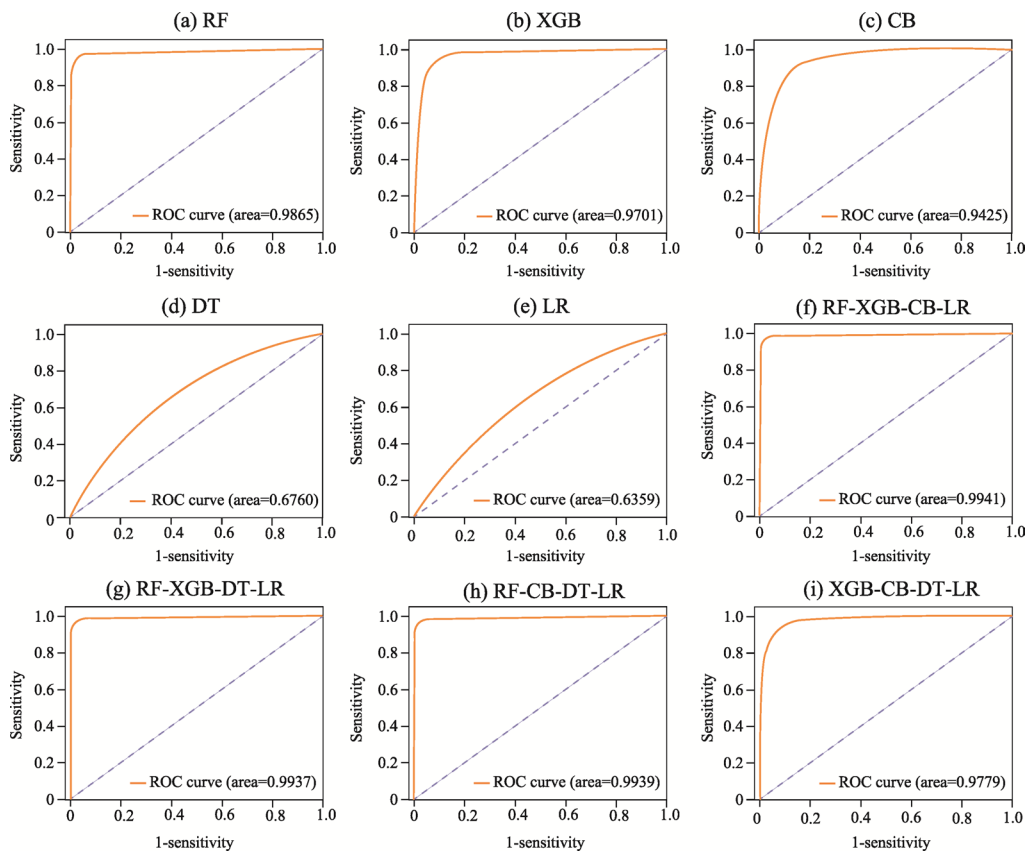

Figure 10   ROC curves for model validation

ment, accounting for 24% of the total. It is worth noting that the Changsha-Zhuzhou-Xiangtan (Changzhutan) urban agglomeration, one of the most industrialized, urbanized, and densely populated regions in the XRB, falls within these high susceptibility zones. This geographical coincidence results in a higher exposure of both population and built environ-
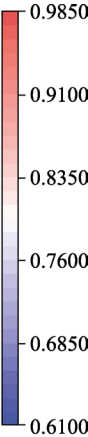
Figure 11 Performance indicators of different models (The black wireframe represents the highest value of each evaluation indicator)
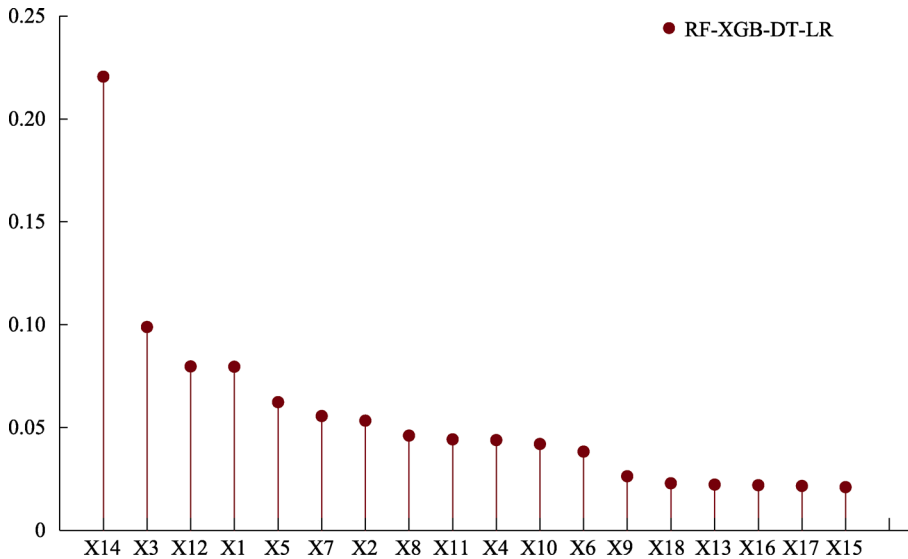

Figure 12 The importance of different features in the optimal model (X1: RX1day; X2: RX3day; X3: RX5day; X4: Average precipitation during the flood season; X5: Elevation; X6: Slope; X7: Aspect; X8: Curvature; X9: TRI; X10: TWI; X11: Distance to rivers; X12: Distance to roads; X13: Drainage density; X14: NDVI; X15: 1990 Landuse; X16: 2000 Landuse; X17: 2010 Landuse; X18: 2020 Landuse)

ment. Cropland exposure was estimated by considering the area of agricultural land that was prone to floods (Figure 13c). Of the total crop area, 8898 km^2 (or 20%) was found within high susceptibility zones. The agricultural belt around the Nanling and Luoxiao mountains is particularly vulnerable, as it falls within a high flood susceptibility area, and thus the crops there suffer severely from disasters. Despite this, the largest proportion of cropland exposure was at a low level, accounting for 33% of the total. This is attributed to regions such as the outskirts of the Changzhutan urban agglomeration and the hilly areas in central and southern Hunan, which have been intensifying agricultural development and are situated in low susceptibility areas. Therefore, cropland exposure in these regions was comparatively lower.
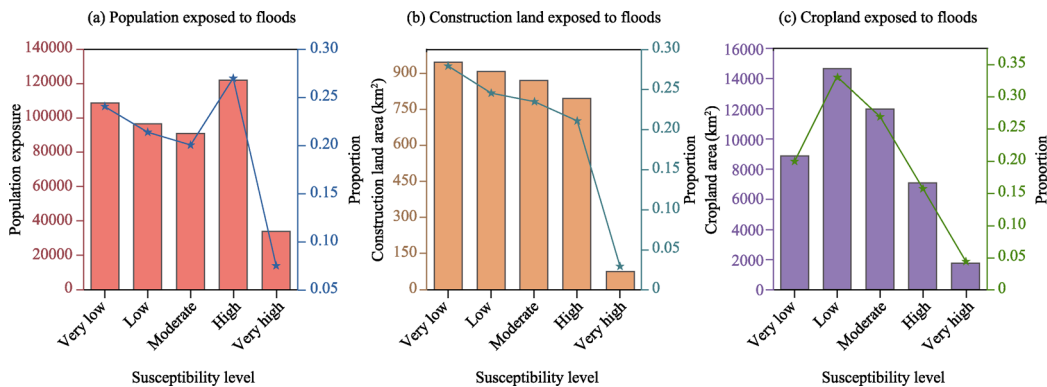
Figure 13   Statistical analysis of population, built environment, and cropland exposure to floods

5   Discussion

Identifying flood-prone areas assumes paramount importance in devising effective disaster preparedness strategies and judicious resource allocation for impending flood events. Despite the widespread utilization of ML techniques in flood prediction, crafting highly precise flood susceptibility maps remains an intricate undertaking. In this research, we harnessed five ML methods and employed four stacking ensemble models to formulate flood susceptibility maps, leveraging climate, topographic, and environmental variables. Subsequently, we scrutinized their predictive capabilities within the XRB and evaluated the factors influencing flood susceptibility within the region. Across all models, a consistent observation emerged: the northeastern and southeastern sectors of the XRB exhibited heightened vulnerability to flooding. Remarkably, among the array of models, the RF-XGB-CB-LR ensemble model emerged as the exemplar, boasting an impressive ROC score of 0.9941.

5.1   Performance of stacking ensemble model

The realm of ML models has made remarkable strides across diverse domains, especially in the context of spatial flood modeling. In this study, a suite of five ML models was harnessed alongside stacking ensemble techniques to meticulously assess flood-prone regions within the XRB. Through the comparative lens, ensemble algorithms outshone conventional ML approaches, solidifying ensemble methodologies as the avant-garde in data mining classification. The supremacy of ensemble models, which amalgamate the strengths of individual models for predictive precision, has been corroborated by antecedent research endeavors (Arabameri *et al*., 2019; Prasad *et al*., 2022).

Among the evaluated models, the RF-XGB-CB-LR ensemble model distinguished itself with a commendable assessment of flood-prone areas (ROC=0.9941). An integration of the three most adept standalone models—RF, XGB, and CB—yielded even superior outcomes. As the performance of the DT model lagged, its inclusion slightly moderated the ensemble model's accuracy. Notably, the top-performing individual learners, RF, XGB, and CB, are in themselves ensemble learning methods, orchestrating resilient learners by synergizing multiple weaker components. However, less proficient base learners like DT might have exerted a diluting influence on overall accuracy (Pourghasemi *et al*., 2017).

While ensemble models have primarily found utilization within the XRB for tasks such as

water level forecasting, their application in assessing flood susceptibility remains relatively nascent. Thus, this study pioneers the application of stacking ensemble models for flood susceptibility evaluation within the XRB. Notably, stacking ensembles outperformed other ensemble techniques employed (Table 2). The model's efficacy is intrinsically intertwined with factors such as data quality, biophysical features of the study area, and selection of appropriate flood susceptibility determinants. Notably, a more comprehensive inclusion of factors augments the model's potential accuracy (Donati and Turrini, 2002). Importantly, the methodology undertaken here holds broader applicability, extending beyond flood susceptibility to encompass other environmental calamities such as landslides, debris flows, and avalanches. Furthermore, its potential extends to exploring the prediction accuracy of models within diverse topographical settings.

Table 2   Comparison between stacking ensemble model and other ensemble models

Region	Impact factor	Ensemble model	ROC	Reference
Haraz watershed	Slope, curvature, TWI, elevation, NDVI, rainfall, lithology, stream density, distance to rivers, land use	Bagging-LMT	0.95	Chapi *et al.*, 2017
Khiyav-Chai watershed	Altitude, slope, aspect, drainage density, land use, curvature, distance to rivers, TWI, soil depth, soil hydrological groups SHG, land use, lithology	Weighted and unweighted averaging	0.91	Choubin *et al.*, 2019
Teesta Sub-catchment	Elevation, curvature, SPI, aspect, slope, TRI, TWI, STI, LULC, distance to rivers, soil type, rainfall	Dagging	0.87	Islam *et al.*, 2021
Xiangjiang River Basin	RX1day, RX3day, aspect, RX5day, slope, TRI, TWI, average precipitation during the flood season, elevation, distance to rivers, distance to roads, drainage density, NDVI, curvature, 1990 landuse, 2000 landuse, 2010 landuse, 2020 land use	Stacking	0.99	This work

Stacking is a powerful ensemble learning technique that enhances model generalizability by leveraging the strengths of multiple individual models. Unlike single machine learning (ML) models. Stacking combines the performance of various base models, each potentially excelling on different data subsets and feature sets. Through cross-validation, the diverse strengths of these base models are harnessed to complement one another, resulting in improved overall prediction accuracy. This ensemble approach fosters both competition and collaboration among the base models, which mitigates the risk of overfitting. Specifically, the meta-model, which serves as the final decision-making layer, learns from the predictions generated by the base models across the training data. By focusing on the averaged outputs of the base models rather than directly on the training data, the meta-model reduces the likelihood of overfitting to any single model's biases or noise. Stacking not only lowers the variance of the model but also increases its robustness to noise, outliers, and unknown inputs. This integration-based learning strategy enhances the model's ability to generalize across different regions and similar geographic environments, thereby offering superior performance in diverse and complex scenarios.

5.2   Analysis of major influencing factors of floods in the XRB

In deciphering the optimal model outcomes, it was evident that NDVI played a paramount role in influencing floods within the XRB, with an importance score of 0.22. NDVI, commonly used to assess vegetation attributes in an area, demonstrates an inverse relationship

between vegetation density and flood occurrence (Kumar and Acharya, 2016). Vegetation density aids water infiltration, curbing surface water volume and subsequently diminishing the likelihood of flood incidents (Turoğlu and Dölek, 2011). Amidst the rapid urbanization in central cities like Changzhutan, Hengyang, Yongzhou, and Loudi, which harbor intensive development activities in manufacturing, services, transportation, and residential sectors, an expansion of deforested regions has ensued. This, coupled with similar trends in peripheral areas of Changzhutan urban agglomeration, the hilly tracts in central and southern Hunan, and the agricultural belt of the Nanling and Luoxiao mountains, has led to decreased afforestation, consequently exacerbating flood occurrences in the XRB. Notably, precipitation patterns emerged as the predominant influencer of flood occurrences in the XRB. The spatial distribution of precipitation generally mirrored an east-high and west-low trend, akin to the spatial variability exhibited by floods. Regions characterized by low flood susceptibility were predominantly concentrated in the western territory. This can be attributed to the protective influence of Yangming Mountain on the southern part of the western region, mitigating the effects of typhoon cloud systems and reducing the likelihood of intense rainfall events (Chen *et al*., 2011).

The XRB's subtropical monsoon humid climate, marked by abundant yet unevenly distributed rainfall throughout the year, engenders prolonged periods of elevated temperature and humidity within the basin (Hang *et al*., 2022). Impervious surfaces such as roads, sidewalks, and parking lots intensify rainfall runoff (Mukherjee and Singh, 2020), making areas adjacent to roads more susceptible to flooding. Across the XRB, the majority of regions are located near roads, significantly increasing their vulnerability to floods.

High flood susceptibility regions predominantly cluster in the XRB's northeastern and southeastern sectors. The optimum RF-XGB-CB-LR model was employed to dissect the feature importance within these regions (Figure 14). In the northeast, the chief influencing factor was the highest daily rainfall (RX1day), bearing an importance score of 0.19 (Figure 14a). There exists a robust positive correlation between flood events and RX1day (Ávila *et al*., 2016). Impacted by substantial rainfall, RX1day in the northeastern XRB ascends to its peak value of 292.963 mm, culminating in heightened water levels for rivers such as the Daxi River. The downstream nature of the northeast, coupled with lower elevation compared to upstream regions, fosters substantial downstream flow due to heavy upstream rainfall, resulting in elevated water levels in lower-lying regions. In the absence of large-scale dam construction due to the region's economic significance, flood control in the Changzhutan metropolitan area primarily relies on dikes, complicated by intricate river-lake dynamics and the pressure from Lake Dongting (Hang *et al*., 2022).

Conversely, the southeastern sector sees the average precipitation during the flood season reigning supreme as the key flood influencer (Figure 14b). The average flood season rainfall, RX1day, and RX5day in this region soar to 6.81 mm, 442.152 mm, and 1385.59 mm respectively. Persistent and intense rainfall, particularly torrential downpours, render this region markedly susceptible to floods (Bui *et al*., 2017). Torrential rain generation and distribution are steered by factors like the Pacific subtropical high pressure, frontal rainfall, shear lines, cyclones, and typhoon climates (Luo, 2006). Nestled between the middle section of the Nanling Mountains and the southern span of the Luoxiao Mountains, the area's mountainous terrain augments rainfall intensity (Gao *et al*., 2006). With subpar vegetation and severe soil

erosion, the upper reaches of this flood-prone zone primarily encompass agricultural terrain. Intense human activity has triggered deforestation, leading to extensive soil and water loss. Degraded vegetation and soil, compounded by water depletion, have eroded the natural storage capacity of forests and soil, truncating runoff confluence time, exacerbating river siltation, and reducing flood passage cross-sections (Chenzhou Flood Control Office, 2020). Cumulatively, these factors magnify flood susceptibility in the region.
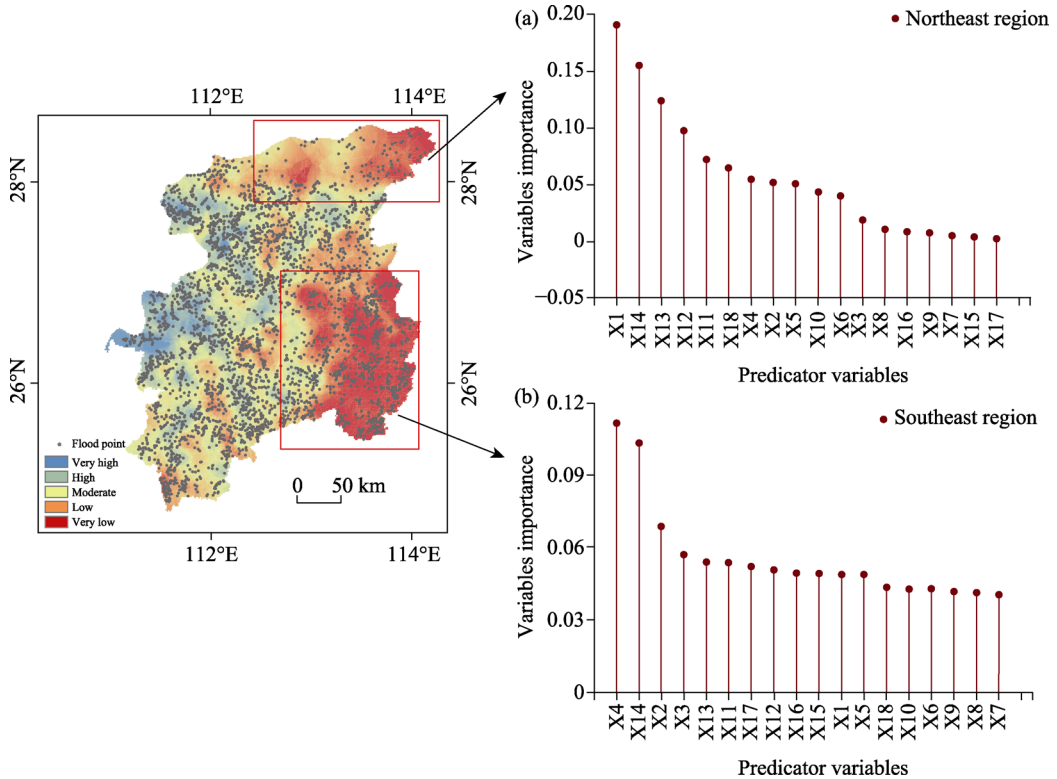

Figure 14 The importance of characteristics in high susceptibility areas of the Xiangjiang River Basin (X1: RX1day; X2: RX3day; X3: RX5day; X4: Average precipitation during the flood season; X5: Elevation; X6: Slope; X7: Aspect; X8: Curvature; X9: TRI; X10: TWI; X11: Distance to rivers; X12: Distance to roads; X13: Drainage density; X14: NDVI; X15: 1990 Landuse; X16: 2000 Landuse; X17: 2010 Landuse; X18: 2020 Landuse)

6 Conclusions

This study employed machine learning models in conjunction with stacking ensemble learning techniques to develop a flood susceptibility assessment for the XRB, while also examining the contributing factors. The main findings are:

(1) Factors such as rainfall, low NDVI values, and impervious surfaces contribute to floods. Stacking models, especially RF-XGB-CB-LR, outperformed individual machine learning approaches, yielding an ROC of 0.99 and high evaluation metrics.

(2) Precipitation patterns matched flood spatial distribution, with the east-high, west-low trend driven by Pacific subtropical high pressure.

(3) Over 40% of the XRB is highly susceptible to flooding, mainly in the northeast and southeast. Other than heavy rainfall, factors such as low terrain, and human activities am-

plify the flood risk.

(4) High susceptibility regions affect 156,000 people (34% of the basin), with urban areas most exposed. The built environment and cropland exposure in these areas are 24% and 20% respectively.

Nevertheless, this study has its limitations, stemming from its exclusive reliance on ML models and the absence of time series data. To enhance future research, incorporating a range of diverse deep learning models and integrating meticulously acquired high-precision time-series flood data from multiple sources could pave the way for more comprehensive insights.

References

Adeli E, Li X, Kwon D *et al*., 2020. Logistic regression confined by cardinality-constrained sample and feature selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(7): 1713–1728.

Arabameri A, Chen W, Loche M *et al*., 2019. Comparison of machine learning models for gully erosion susceptibility mapping. *Geoscience Frontiers*, 11(5): 1609–1620.

Ávila A, Justino F, Wilson A *et al*., 2016. Recent precipitation trends, flash floods and landslides in southern Brazil. *Environmental Research Letters*, 11(11): 114029.

Breiman L, 2001. Random forests. *Machine Learning*, 45: 5–32.

Bui D T, Bui Q T, Nguyen Q P *et al*., 2017. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agricultural and Forest Meteorology*, 233: 32–44.

Bui D T, Khosravi K, Shahabi H *et al*., 2019. Flood spatial modeling in northern Iran using remote sensing and GIS: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. *Remote Sensing*, 11(13): 1589.

Bui D T, Pradhan B, Lofman O *et al*., 2012. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *Catena*, 96: 28–40.

Bui D T, Tuan T A, Klempe H *et al*., 2015. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, 13: 361–378.

Chan F K S, Griffiths J A, Higgitt D *et al*., 2018. "Sponge City" in China: A breakthrough of planning and flood risk management in the urban context. *Land Use Policy*, 76: 772–778.

Chapi K, Singh V P, Shirzadi A *et al*., 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. *Environmental Modelling and Software*, 95: 229–245.

Chen J L, Huang G R, Chen W J, 2021. Towards better flood risk management: Assessing flood risk and investigating the potential mechanism based on machine learning models. *Journal of Environmental Management*, 293: 112810.

Chen T, Zhao Y L, Peng G *et al*., 2011. The climatic characteristics and variation of rainstorm in Hengyang during recent five years. *Journal Disaster*: *Prevention Science and Technology*, 13(1): 27–30. (in Chinese)

Chen T Q, Guestrin C, 2016. XGBoost: A scalable tree boosting system. *Machine Learning*, 9: 785–794.

Chen W, Hong H, Li S J *et al*., 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, Li Y, Xue W *et al*., 2020. Modeling flood susceptibility using data-driven approaches of nave bayes tree, alternating decision tree, and random forest methods. *Science of The Total Environment*, 701: 134979.

Chen W, Zhang S, Li R W *et al*., 2018. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. *Science of The Total Environment*, 644: 1006–1018.

Chenzhou Flood Control Office, 2020. Causes of floods in the Chenjiang River Basin and disaster prevention measures. *Hunan Hydro&Power*, (3): 44–45. (in Chinese)

Choubin B, Moradi E, Golshan M *et al.*, 2019. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of The Total Environment*, 651: 2087–2096.

Costache R, Bui D T, 2020. Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. *Science of The Total Environment*, 712: 136492.

David H W, 1992. Stacked generalization. *Neural Networks*, 5: 241–259.

Debeljak M, Džeroski S, 2011. Decision trees in ecological modelling. In: Fred J, Hauke R, Broder B. Modelling Complex Ecological Dynamics. Springer: London Press, 197–209.

Donati L, Turrini M C, 2002. An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology: Application to an area of the Apennines (Valnerina; Perugia, Italy). *Engineering Geology*, 63(3/4): 277–289.

Dou J, Yunus A P, Bui D T *et al.*, 2019. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. *Landslides*, 17: 641–658.

Du J, He F, Shi P J, 2006. Integrated flood risk assessment of Xiangjiang River Basin in China. *Journal of Natural Disasters*, 15: 8–44. (in Chinese)

Fang L, Huang J L, Cai J T *et al.*, 2022. Hybrid approach for flood susceptibility assessment in a flood-prone mountainous catchment in China. *Journal Hydrology*, 612: 128091.

Gao D, Yin J, Wang D *et al.*, 2024. Modelling and validation of flash flood inundation in drylands. *Journal of Geographical Sciences*, 34(1): 185–200.

Gao Y Z, Xing J J, Wang C L *et al.*, 2006. Cause and forecast of mountain flood from rainstorm. *Journal of Natural Disasters*, 4: 65–70. (in Chinese)

Ha H, Luu C, Bui Q D *et al.*, 2021. Flash flood susceptibility prediction mapping for a road network using hybrid machine learning models. *Natural Hazards*, 109: 1247–1270.

Hang Z, Huang J Q, Li Z Z *et al.*, 2022. Nonstationary Bayesian modeling of extreme flood risk and return period affected by climate variables for Xiangjiang River Basin, in South-Central China. *Water*, 14(1): 66.

Huang C X, Hu S S, Huang Y, 2023. Analysis on spatiotemporal variation characteristics and influencing factors of NDVI in Hunan province. *Ecological Indicators*, 42: 114–126.

IPCC, 2019. IPCC Special Report: Climate Change and Land. Paris, France.

Islam A R M T, Talukdar S, Mahato S *et al.*, 2021. Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, 12: 101075.

Khosravi K, Shahabi H, Pham B T *et al.*, 2019. A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *Journal Hydrology*, 573: 311–323.

Kleinbaum D G, Klein M, 2010. Modeling strategy for assessing interaction and confounding. In: Logistic Regression: Statistics for Biology and Health. New York, NY: Springer.

Kuhn M, Johnson K, 2013. Applied Predictive Modeling. New York, NY: Springer.

Kumar R, Acharya P, 2016. Flood hazard and risk assessment of 2014 floods in Kashmir Valley: A space-based multisensor approach. *Natural Hazards*, 84: 437–464.

Kusiak A, Li M, Zhang Z, 2010. A data-driven approach for steam load prediction in buildings. *Applied Energy*, 87(3): 925–933.

Li K, Zhao J, Lin Y, 2023. Debris-flow susceptibility assessment in Dongchuan using stacking ensemble learning including multiple heterogeneous learners with RFE for factor optimization. *Natural Hazards*, 118: 2477–2511.

Li S S, Wang Z L, Lai C G *et al.*, 2020. Quantitative assessment of the relative impacts of climate change and human activity on flood susceptibility based on a cloud model. *Journal of Hydrology*, 588: 125051.

Li W J, Lin K R, Zhao T T G *et al.*, 2019. Risk assessment and sensitivity analysis of flash floods in ungauged

basins using coupled hydrologic and hydrodynamic models. *Journal Hydrology*, 572: 108–120.

Li Z, Zhang Y, Wu Q *et al*., 2022. Study on flood forecasting ensemble correction based on hierarchical optimization and LSTM. *Water Resources and Hydropower Engineering*, 53(8): 41–49.

Liu J, Wang J Y, Xiong J N *et al*., 2022. Assessment of flood susceptibility mapping using support vector machine, logistic regression and their ensemble techniques in the Belt and Road region. *Geocarto International*, 37: 1–30.

Luo Z L, 2006. Discussion on the prevention and control of mountain flood disasters in the border region between Hunan, Jiangxi, Guangdong, and Fujian provinces: Taking Chenzhou city as an example. *Land Resource Guide*, 6: 45–47. (in Chinese)

Lv L, Chen T, Dou J *et al*., 2022. A hybrid ensemble-based deep-learning framework for landslide susceptibility mapping. *International Journal of Applied Earth Observation and Geoinformation*, 108: 102713.

Ma M H, Zhao G, He B S *et al*., 2021. XGBoost-based method for flash flood risk assessment. *Journal Hydrology*, 598: 126382.

Mahato S, Pal S, Talukdar S *et al*., 2021. Field based index of flood vulnerability (IFV): A new validation technique for flood susceptible models. *Geoscience Frontiers*, 12(5): 101175.

Merghadi A, Yunus A P, Dou J *et al*., 2020. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207: 103225.

Mukherjee F, Singh D, 2020. Detecting flood prone areas in Harris County: A GIS based analysis. *Geological Journal*, 85: 647–663.

Myles A J, Feudale R N, Liu Y *et al*., 2004. An introduction to decision tree modeling. *Journal Chemometrics*, 18(6): 275–285.

Pham B T, Avand M, Janizadeh S *et al*., 2020. GIS based hybrid computational approaches for flash flood susceptibility assessment. *Water*, 12: 683.

Pham B T, Bui D T, Dholakia M B *et al*., 2016. A comparative study of least square support vector machines and multiclass alternating decision trees for spatial prediction of rainfall-induced landslides in a tropical cyclones area. *Geotechnical and Geological Engineering*, 34: 1807–1824.

Pourghasemi H R, Yousefi S, Kornejady A *et al*., 2017. Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling. *Science of The Total Environment*, 609: 764–775.

Prasad P, Loveson V J, Das B *et al*., 2022. Novel ensemble machine learning models in flood susceptibility mapping. *Geocarto International*, 37: 16.

Ren H, Pang B, Ping B *et al*., 2024. Flood susceptibility assessment with random sampling strategy in ensemble learning (RF and XGBoost). *Remote Sensing*, 16(2): 320.

Roe B P, Yang H J, Zhu J *et al*., 2005. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nuclear Instruments and Methods in Physics Research*, 543(2): 577–584.

Rozalis S, Morin E, Yair Y *et al*., 2010. Flash flood prediction using an uncalibrated hydrological model and radar rainfall data in a Mediterranean watershed under changing hydrological conditions. *Journal of Hydrology*, 394: 245–255.

Saber M, Boulmaiz T, Guermoui M *et al*., 2023. Enhancing flood risk assessment through integration of ensemble learning approaches and physical-based hydrological modeling. *Geomatics, Natural Hazards and Risk*, 14: 1.

Saito H, Nakayama D, Matsuyama H, 2009. Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi Mountains, Japan. *Geomorphology*, 109: 108–121.

Seydi S T, Kanani-Sadat Y, Hasanlou M *et al*., 2023. Comparison of machine learning algorithms for flood susceptibility mapping. *Remote Sensing*, 15(1): 192.

Shahabi H, Shirzadi A, Ghaderi K *et al*., 2020. Flood detection and susceptibility mapping using Sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on k-nearest neighbor classifier. *Remote Sensing*, 12: 266.

Siegert M, Alley R B, Rignot E *et al*., 2020. Twenty-first century sea-level rise could exceed IPCC projections for strong-warming futures. *One Earth*, 3: 691–703.

Silva M M G T D, Kawasaki A, 2020. A local-scale analysis to understand differences in socioeconomic factors

affecting economic loss due to floods among different communities. *International Journal of Disaster Risk Reduction*, 47: 101526.

Tehrany M S, Pradhan B, Mansor S *et al.*, 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena*, 125: 91–101.

Tien B D, Khosravi K, Shahabi H *et al.*, 2019. Flood spatial modeling in northern Iran using remote sensing and GIS: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. *Remote Sensing*, 11(13): 1589.

Tien B D, Pradhan B, Lofman O *et al.*, 2012. Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *Catena*, 96: 28–40.

Turoğlu H, Dölek İ, 2011. Floods and their likely impacts on ecological environment in Bolaman River basin (Ordu, Turkey). *The Journal of Agricultural Science*, 43: 167–173.

Vandenberg-Rodes A, Moftakhari H R, AghaKouchak A *et al.*, 2016. Projecting nuisance flooding in a warming climate using generalized linear models and Gaussian processes. *Journal of Geophysical Research: Oceans*, 121: 8008–8020.

Wang Y, Fang Z C, Hong H Y *et al.*, 2021. Flood susceptibility mapping by integrating frequency ratio and index of entropy with multilayer perceptron and classification and regression tree. *Journal of Environmental Management*, 289: 112449.

Wu F, Sun Y F, Sun Z X *et al.*, 2019. Assessing agricultural system vulnerability to floods: A hybrid approach using emergy and a landscape fragmentation index. *Ecological Indicators*, 105: 337–346.

Wu X, Wang J, 2023. Application of bagging, boosting and stacking ensemble and easy ensemble methods for landslide susceptibility mapping in the Three Gorges Reservoir area of China. *International Journal of Environmental Research and Public Health*, 20(6): 4977.

Xu Y, Lin K, Hu C *et al.*, 2023. Deep transfer learning based on transformer for flood forecasting in data-sparse basins. *Journal Hydrology*, 625: 129956.

Yang Y, Du J, Cheng L L *et al.*, 2017. Applicability of TRMM satellite precipitation in driving hydrological model for identifying flood events: A case study in the Xiangjiang River Basin, China. *Natural Hazards*, 87: 1489–1505.

Yaseen A, Lu J, Chen X, 2022. Flood susceptibility mapping in an arid region of Pakistan through ensemble machine learning model. *Stochastic Environmental Research and Risk Assessment*, 36: 3041–3061.

Youssef A M, Pradhan B, Sefry S A, 2016. Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. *Environmental Earth Sciences*, 75: 12.

Zeng H, Huang J Q, Li Z Z *et al.*, 2022. Nonstationary Bayesian modeling of extreme flood risk and return period affected by climate variables for Xiangjiang River Basin, in South-Central China. *Water*, 14: 66.

Zhang J, Zhu Y, Zhang X *et al.*, 2018. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal Hydrology*, 561: 918–929.

Zhang R, Chai Z, Zhang T *et al.*, 2023. Research progress of flood forecasting based on machine learning models. *Water Resources and Hydropower Engineering*, 54(11): 89–101. (in Chinese)

Zhao G, Pang B, Xu Z *et al.*, 2018. Mapping flood susceptibility in mountainous areas on a national scale in China. *Science of The Total Environment*, 615: 1133–1142.

Zhao G, Pang B, Xu Z X *et al.*, 2020. Urban flood susceptibility assessment based on convolutional neural networks. *Journal Hydrology*, 590: 125235.

Zhu R C, Hu X F, Hou J Q *et al.*, 2020. Application of machine learning techniques for predicting the consequences of construction accidents in China. *Process Safety and Environmental Protection*, 145: 293–302.