Flash Flood Disaster Risk Simulation Technology Based on Classified Recursive Feature Elimination-Random Forest Optimization Algorithm

Xiaolei Zhang[1*], Ruihua Qin[2], Changqi Dong[3], Xuecheng Xing[1], Cheng Zhang[1], Jing Xiao[4], Ronghua Liu[1]

Research Center on Flood & Drought Disaster Prevention and Reduction of the Ministry of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing, 100038, China[1,*]
Email: zhangxl@iwhr.com

Xi’an University of Architecture and Technology, Xi’an, 710311, China[2]
Email: 1106236218@qq.com

Sun Yat-sen University, Guangzhou, 510275, China[3]
Email: dongchq5@mail2.sysu.edu.cn

China Institute of Water Resources and Hydropower Research, Beijing, 100038, China[1]
Email: xingxch@iwhr.com

Research Center on Flood & Drought Disaster Prevention and Reduction of the Ministry of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing, 100038, China[1]
Email: zhangc@iwhr.com

Hydrology and Water Resources Survey Bureau of Hainan Province, Haikou, 570100, China[4]
Email: xiaojing2012tju@163.com

State Key Laboratory of Water Cycle and Water Security, China Institute of Water Resources and Hydropower Research, Beijing, 100038, Research Center on Flood & Drought Disaster Prevention and Reduction of the Ministry of Water Resources, Beijing, 100038, China[1]
Email: liurh@iwhr.com

ABSTRACT

Flash flood disasters inflict severe economic losses and casualties on human society, making the scientific identification and assessment of these risks an urgent priority in disaster management research. This study aims to improve the accuracy of flash flood risk prediction by coupling feature selection and the Random Forest algorithm, thereby establishing a scientific basis for disaster early warning. Seventeen features influencing flash flood occurrence were initially identified, and a feature selection technique coupling Recursive Feature Elimination (RFE) with the Random Forest-based optimization algorithm was developed, which effectively determines the optimal feature combination for flash flood risk simulation. The research results indicate that the optimal feature combination obtained using classification-based RFE significantly improves the predictive performance of the Random Forest model, achieving an ROC value of 0.947. High-risk areas for flash flood disasters in Fujian Province are mainly distributed in the Wuyi Mountains, Daiyun Mountain, and Damao Mountain regions. These areas cover 49,000 km^2 and affect 27 million people.

KEYWORDS: Random forest; Recursive feature selection; Flash flood disaster risk; Small watershed scale

1 INTRODUCTION

With continuous socio-economic development and worsening climate change, flood disasters have gradually increased, causing heavy casualties and economic losses. Flash floods are particularly prominent due to their characteristics of intense short-duration rainfall, rapid outbursts, and high risk to human safety and property (Fu et al., 2024). Flash floods account for approximately 70% of flood-related fatalities. Therefore, Strengthening prevention measures and accurately identifying high-risk areas is vital for reducing casualties and property losses. To achieve this goal, scientific and systematic risk assessment and prediction of flash flood are essential.

Current flash flood risk assessment primarily include mathematical models, physical models, and hybrid approaches that integrate machine learning with statistical models. The first category comprises deterministic mathematical models, which use coefficient of determination method to evaluate the sensitivity of various influencing factors across different intervals and classify flash flood risks using the relative entropy index (Yang et al., 2023). Raillani et al. employed the Dirac delta function approximation to estimate the probability density function, describing disaster occurrence rates in different regions to identify high-risk areas (Rayllani et al., 2023). Nugraheni et al. integrated 15 primary indicators (e.g., social, economic, policy, and land use factors) with 65 secondary indicators (e.g., working-age population, land-use change trends, and policy objectives) in combination with the CLEAR model, CLUE-S model, and DROP model to identify flash flood risk areas (Nugraheni et al., 2018). Chen et al. established a flash flood risk assessment system for Qingyuan City based on GIS and the Analytic Hierarchy Process (AHP), delineating risk zones into different levels. Their results delineated the spatial distribution of flash flood risks in Qingyuan City and provided technical support for flood risk management planning (Chen et al., 2017). Zhang et al. investigated flash floods in Chongqing, identifying key influencing factors and calculating risk indices to support decision-making in risk management (Zhang et al., 2019). Zhang et al. employed GIS, geographic detectors, and AHP, using nine indicators such as snow water equivalent, slope, and the maximum 1-hour precipitation to identify key drivers of flash flood risk in Xinjiang. Their study comprehensively assessed the risks based on disaster-causing factors, environmental conditions, and vulnerability (Zhang et al., 2024). Wibowo et al. used a composite flood disaster index to assess flash flood risk (Wibowo et al., 2020).

The second category comprises physically based models solving shallow water equations. One-dimensional and two-dimensional hydraulic models have been employed to simulate flood dynamics in developing mountainous areas, which supports precise risk zoning and hazard mapping. (Zhang et al., 2020;Hu et al., 2018). Liu et al. established a bidirectional coupled hydrological-hydrodynamic model (CNFF-IFMS) to simulate flash floods in the Zhaigang River Basin, validating the performance of the coupled model in mountainous flood simulations (Liu et al., 2025). Waleed et al. integrated ArcGIS, ERDAS, and WMS with HEC hydrological model and HEC-RAS two-dimensional hydraulic modeling software to assess the impact of flash floods on the holy sites and eastern urban areas of Mecca (Waleed et al., 2022).. Salih et al. applied the HEC-HMS model and IMERG data to evaluate flood risks in the Wadi Hail basin in southwestern Saudi Arabia, providing valuable insights for flood management strategies (Salih et al., 2024). Li et al. simulated flash flood processes by integrating hydrological and hydrodynamic models and developed flood risk maps (Li et al., 2019). Al-Kuisi et al. applied a hydrological model coupled with GIS-based AHP analysis to generate flash flood risk zoning maps for Petra and Wadi Musa, Jordan, supporting the local early warning system(Al-Kuisi et al., 2023).

Furthermore, studies demonstrate that single-model approaches (e.g., hydrological and hydrodynamic) are inherently limited in quantifying cumulative effects of riverine structures (e.g., bridges, weirs, and dams) on flood propagation dynamics, particularly their backwater interactions. This can lead to an underestimation of flash flood susceptibility and potential hazards. Therefore, the development of coupled hydrological-hydrodynamic models (such as CNFF-IFMS) enables more accurate and refined flash flood simulation studies (Liu et al., 2025).

The third category comprises hybrid statistical-physical models that integrate machine learning and physical mechanisms. In particular, machine learning techniques are increasingly applied in flash flood risk assessment and management. Fan et al. employed Light Gradient Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost) algorithms to construct flash flood risk maps, providing more accurate data support for disaster early warning systems (Fan et al., 2023). Chu et al. used the Bayesian Model Averaging (BMA) approach combined with three classical machine learning algorithms—Gradient Boosting Decision Trees (GBDT), Backpropagation Neural Networks (BP), and Random Forest (RF)—to simulate flash flood risk (Chu et al., 2023). Ma et al. applied Least Squares Support Vector Machines (LS-SVM) and other artificial intelligence techniques to assess flash flood risk in Yunnan, generating risk maps of flash flood occurrences (Ma et al., 2019). Jing et al. adopted a combination of Fuzzy Comprehensive Evaluation, Analytical Hierarchy Process (AHP), and GIS to evaluate flash flood risk in Datong City (Jing et al., 2018). Ruidas et al. proposed a bivariate logistic regression method to map flash flood risk in the Gandheswari River Basin, India (Ruidas et al., 2022). Ju et al. employed five different machine learning regression models, including Random Forest, Support Vector Regression (SVR), XGBoost, and LightGBM, to assess the risk of four types of natural disasters in Changzhou, demonstrating the advantages of machine learning algorithms in risk assessment(Ju et al., 2024). These studies establish physical principles and quantitative tools (e.g., risk assessment matrices) for flash flood management, directly informing the development of mitigation strategies in operational contexts.

Flash flood risk simulation is requires multi-physics coupling due to the nonlinear interactions of multi-factor, including meteorological conditions, topography, and land-use changes, making it difficult to predict. Previous studies have shown that: (i) deterministic mathematical models and traditional physical models have played a significant role in simulating and predicting flash floods. However they have many limitations—they require specific types of data and lack flexibility, while mathematical models may be influenced by the designer's experience and biases, potentially introducing additional errors. Flash floods often involve multiple nonlinear interactions; (ii) compared to physical models, machine learning models are data-driven and effectively capture complex nonlinear relationships, resulting in more accurate predictions. Furthermore, they leverage historical historical data and identify complex nonlinear patterns; and (iii) Compared to Deterministic mathematical models, machine learning algorithms reduce subjective errors. However, they may be prone to overfitting in certain scenarios, where the model performs well on training data but lacks generalization ability when applied to new data, preventing it from achieving the expected outcomes (Bu et al,, 2020).

To improve flash flood risk risk simulations and support flood forecasting and early warning, this study proposes a coupling technique based on Recursive Feature Elimination and Random Forest (RFE-RF). A classification-based RFE method is applied to remove insignificant features, which reduces data dimensionality, mitigates overfitting, and simplifies the model requirements. Historical flash flood data are then used for risk analysis and prediction. The results show that compared to using all features, reducing dimensionality not only lowers the degree of overfitting but also improves prediction accuracy to 0.88 and increases the ROC value to 0.947.

2 METHODOLOGY

This study establishes a flash flood disaster risk identification methodology combining the categorical feature elimination method and a random forest coupling algorithm to identify flash flood disaster risk areas and summarize their risk characteristics. The specific process includes eigenvalue classification, feature combination screening, model training, and risk simulation.

a. Eigenvalue Classification: In this study, to analyze the causes of flash flood disasters, the influencing factors are categorized into three main groups: rainfall datasets, underlying surface datasets, and socio-economic datasets.

b. Feature Combination Screening: In this study, to optimize the feature selection process and accurately identify the most significant factors influencing flash flood disasters, Recursive Feature Elimination (RFE) is used to process each category of data (i.e., rainfall datasets, underlying surface datasets, and socio-economic datasets) separately. At least one factor from each category is retained to reflect the causes of flash flood disasters.

c. Model Training: The optimal feature combinations obtained from the RFE process are used to train a random forest model. 80% of the historical flash flood disaster dataset is allocated for model training, while the remaining 20% reserved for testing the model's accuracy.

d. Risk Simulation: Using the trained random forest model, the likelihood of flash flood disasters is predicted and categorized into five distinct groups: extremely high, high, medium, low, and extremely low.

Recursive Feature Elimination (RFE) iteratively removes the least important features, gradually refining the feature set. Initially, all available features are used to train the model, after which the model is repeatedly retrained while eliminating insignificant features until the feature count meets a predetermined condition. Through RFE, redundancy in the feature space can be effectively reduced, improving both the efficiency and performance of the model (Han et al., 2021). On large datasets, selecting an optimal subset of features is essential for machine learning algorithms, such as Random Forest. Too many features can degrade algorithm performance and increase the risk of overfitting (Jeon et al., 2020).

Random Forest is a classifier that that constructs multiple decision trees to predict sample outcomes. Introduced by Breiman in 2001, this algorithm has been widely adopted for both classification and regression tasks, particularly when the number of variables far exceeds the number of observations (Blau-G et al., 2016). It improves predictive performance by averaging the outcomes of individual trees for regression tasks or by applying majority voting for classification tasks. Its advantages include high accuracy, the capability to handle large datasets and high-dimensional features, and the capability to assess feature importance.

This study employs a coupled RFE-Machine Learning algorithm, leveraging RFE to reduce excessive feature dimensionality and mitigate overfitting, thereby enhancing the performance of the Random Forest model. This methodology is used to identify key features influencing flash flood disasters and simulate flash flood disaster risks in Fujian Province.

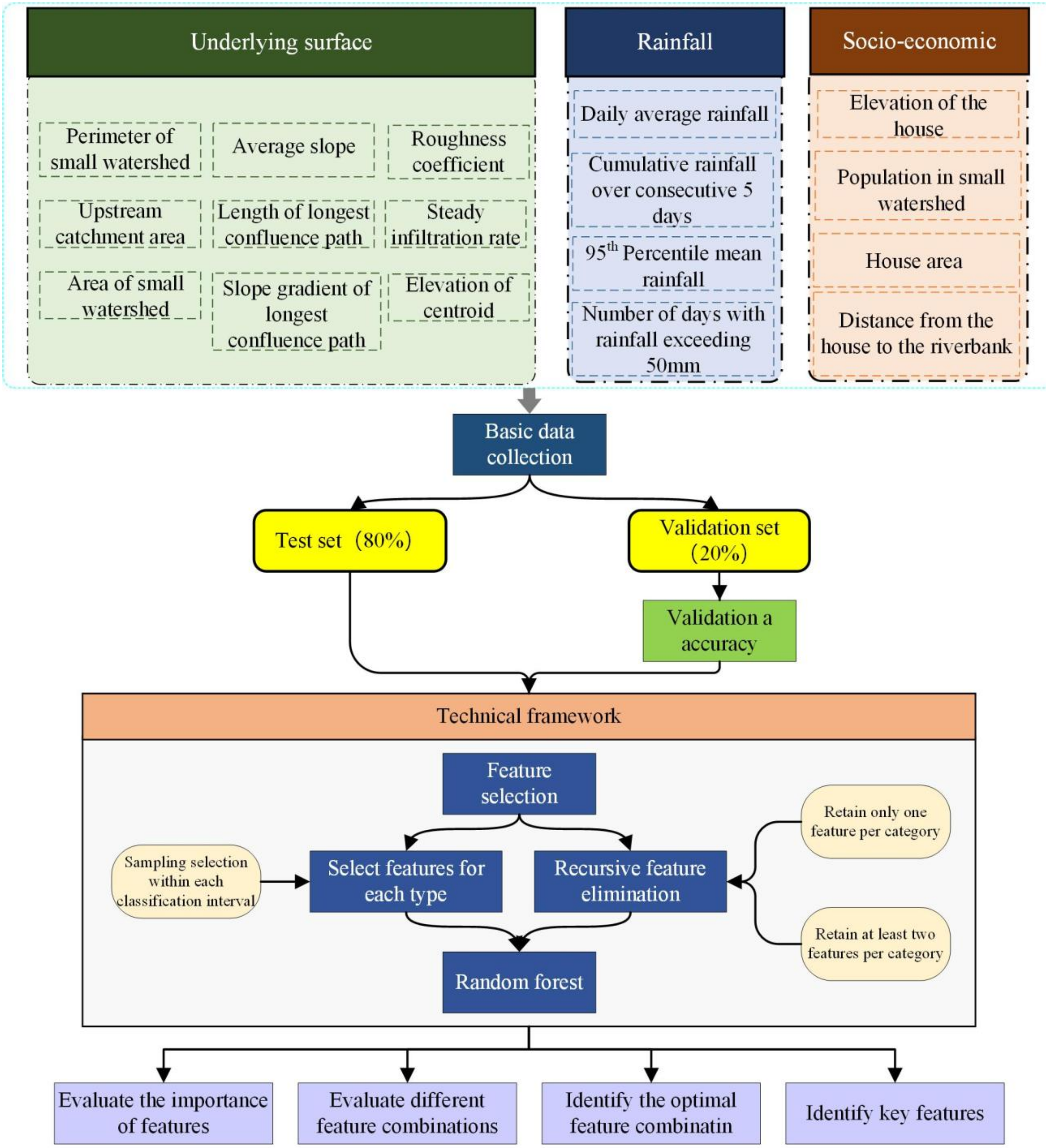


Figure 1: Technical Framework

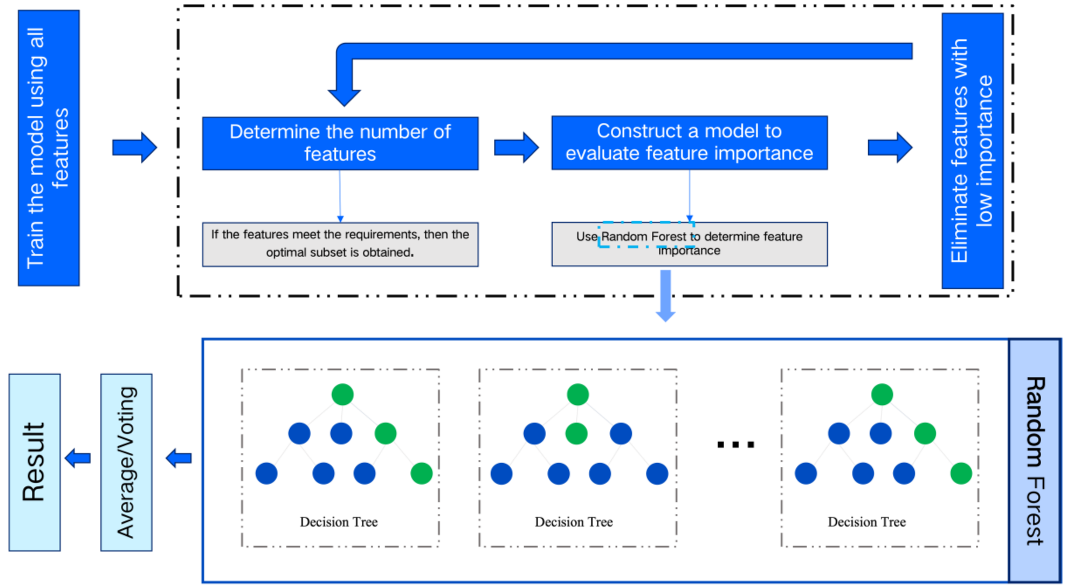


Figure 2: Classification based recursive feature selection method

3 CASE STUDY

3.1 Study Area

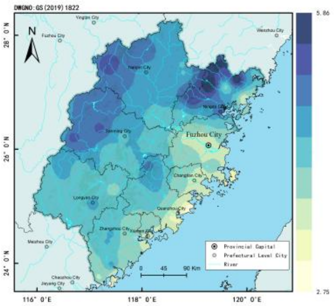
Using Fujian Province as a case study, this paper aims to identify key contributing factors to flash flood disasters, assess high-risk zones across the province, visualize risk distribution patterns, and investigate the underlying mechanisms of hazard concentration areas. Fujian Province, situated along China's southeastern coast, is dominated by mountainous terrain where 90% of its land exhibits undulating topography, 1.35 times the national average. Notably, 66 counties (77.7% of provincial municipalities) fall within these geologically active mountain zones. The unique landform and climate change result in frequent occurrences of flash flood disasters, causing significant casualties and posing a severe challenge to flood control and disaster mitigation (Liang et al., 2020). According to the database of the National flash flood Disaster Survey and Evaluation Results from the China Institute of Water Resources and Hydropower Research, there have been 2,306 recorded historical flash flood disaster events in Fujian Province since the founding of the People's Republic of China, causing severe economic damage and casualties. Notably, the number of flash flood disasters has increased sharply since 1988 (Xiong et al., 2020; Li et al., 2022; Zhang et al., 2019).

3.2 Data

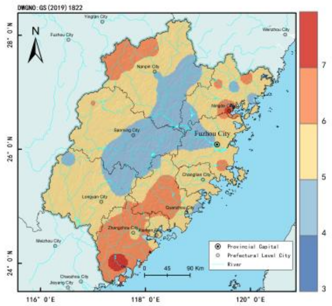
Based on data availability, this study identified 17 impact factors in Fujian Province, categorized into four types of rainfall factors, nine underlying surface factors, and four socio-economic factors. Departing from grid-scale methodologies, this study uses small watersheds as research unit (ranging from 15 to 50 km^2) as research units and fully considers the topological relationships between upstream and downstream watersheds. The impact factors of flash flood disasters are processed and analyzed at the watershed scale. Historical data on these disasters—encompassing casualties and property losses from various watersheds in Fujian Province—serve as the sample. In this study, data from 2,306 historical disaster sites in Fujian Province were employed. Using a random sampling method, the small watersheds containing these historical disaster sites were divided into a training set (80%) and a validation set (20%).A binary classification method is adopted to quantify the casualties and economic losses from flash flood disasters at the small watershed scale. Small watersheds with casualties and economic losses due to flash flood disasters are labeled as 1, while those without losses are labeled as 0. Twenty percent of the sample is used as the test set, while the remaining eighty percent is used for training.

Table 1: Flash flood influencing features of Fujian province

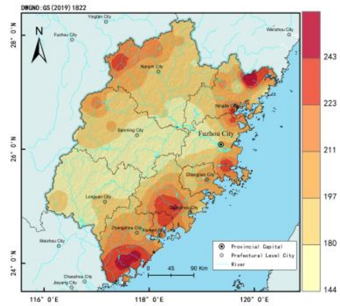
Data type	Characteristic factor	Data source	Abbreviation in English
Rainfall indicators (4 indicators)	a.Daily average rainfall(1951-2013)	China Meteorological Data Network Rainfall Dataset	ADP
	b. Days with rainfall exceeding 50mm (1951-2013)		RF>50mm
	c. Cumulative rainfall over five consecutive days (1951-2013)		CAVD
	d. 95th percentile rainfall (1951-2013)		P95R
Underlying surface type indicators (9 items)	e. Perimeter of a small watershed	National Flash Flood Disaster Prevention Project Team National Flash Flood Disaster Investigation and Assessment Database	SWP
	f. Outlet drainage area		UCA
	g. Small watershed area		SAW
	h. Average slope		AvSl
	i. Longest flow path length		LCP
	j. Longitudinal slope of the longest flow path		LSPG
	k. Roughness coefficient		N
	l. Steady infiltration rate		SIR
	m. Centroid elevation		CE
Socio-economic indicators (4 items)	n. Building elevation	National Flash Flood Disaster Prevention Project Team National Flash Flood Disaster Investigation and Assessment Database	BE
	o. Distance of the building from the riverbank		NH50SW
	p. Building area		BA
	q. Population of the small watershed		PSW
Sample data	r. Historical flash flood disaster data	National Flash Flood Disaster Prevention Project Team National Flash Flood Disaster Investigation and Assessment Database	HFFD



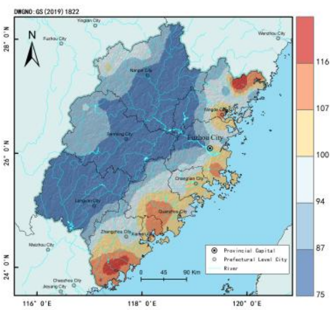
(a) Daily average rainfall



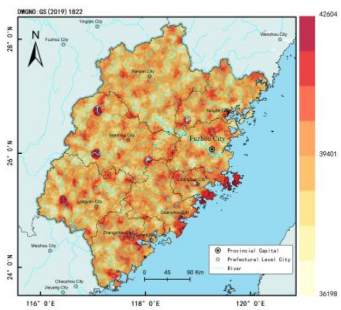
(b) Days with rainfall exceeding 50mm



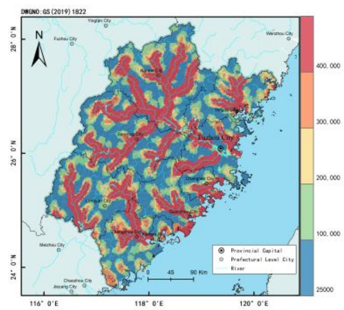
(c) Cumulative rainfall over five consecutive days



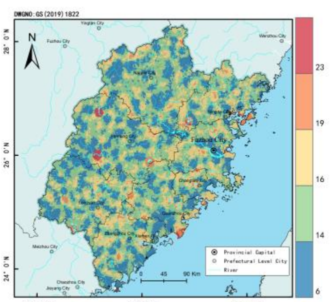
(d) 95th percentile rainfall



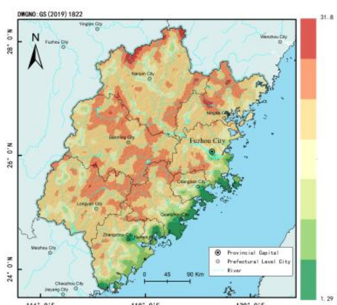
(e) Perimeter of small watershed



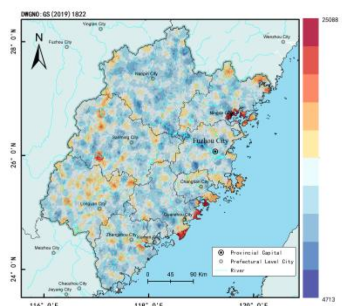
(f) Outlet drainage area



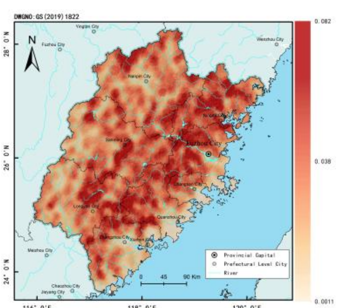
(g) Small watershed area



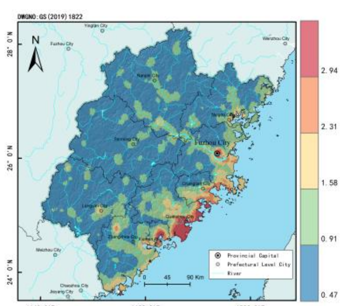
(h) Average slope



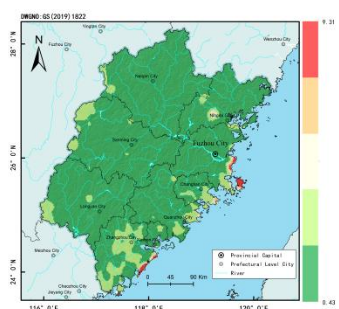
(i) Longest flow path length



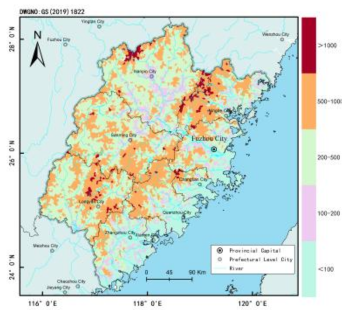
(j) Longitudinal slope of the longest flow path



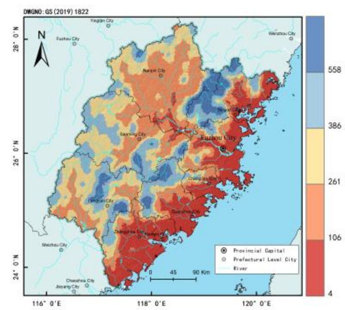
(k) Roughness coefficient



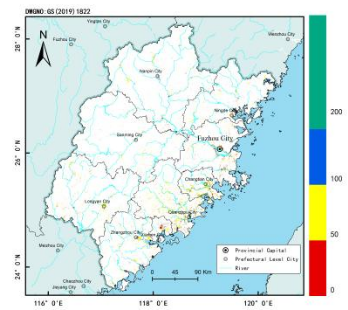
(l) Steady infiltration rate



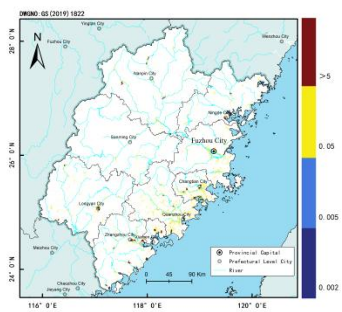
(m) Centroid elevation



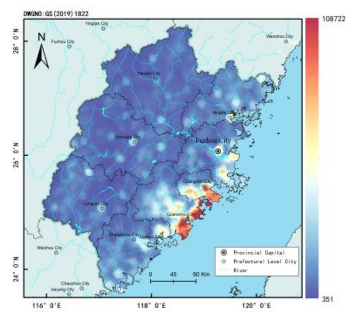
(n) Building elevation



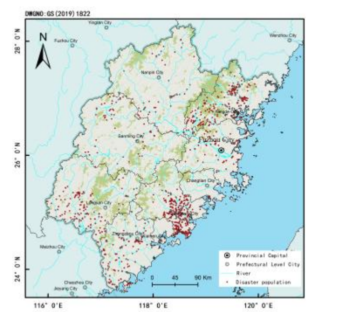
(o) Distance of the building from riverbank



(p) Building area



(i) Population of the small watershed



(r) Historical flash flood disaster data

Figure 3: Feature section datasets and historical flash flood disaster data

4 RESULT ANALYSIS

4.1 Results

After 14 rounds of iterative training, the optimal feature combination was determined and used to simulate flash flood risk levels in Fujian Province. The results indicate that high-risk areas encompass a population of approximately 27 million, and the total area of small watersheds in the medium-high risk zones reaches 49,000 km². The extremely high-risk and high-risk areas for flash floods are mainly concentrated in mountainous regions, particularly in the Wuyi Mountains, Daiyun Mountains, and Daimai Mountains. These regions are characterized by steep slopes and highly variable terrain, with complex topographical conditions and sharp elevation changes. Combined with the abundant daily rainfall of the subtropical monsoon climate, these factors markedly increase the likelihood of flash floods.

As shown in Fig. 4, high-risk areas in Fujian Province are primarily distributed in regions closely connected to mountain ranges. This distribution pattern is closely related to the province's complex topography, particularly in mountainous areas where the risk is most pronounced. Notably, the spatial distribution of these high-risk areas does not strictly follow watershed boundaries but is more significantly influenced by the orientation of the mountain ranges. Due to the unique topographical features of these areas, flash floods are more likely to occur during heavy rainfall events. Additionally, the study indicates that the risk of flash floods is relatively low in coastal regions and low-altitude areas in the southwest. This phenomenon further validates the significant influence of high-altitude regions on the probability and severity of flood disasters.

The study's findings not only provide critical support for disaster prevention and mitigation in Fujian Province but also offer a scientific basis for developing future flash flood early warning and prevention strategies.

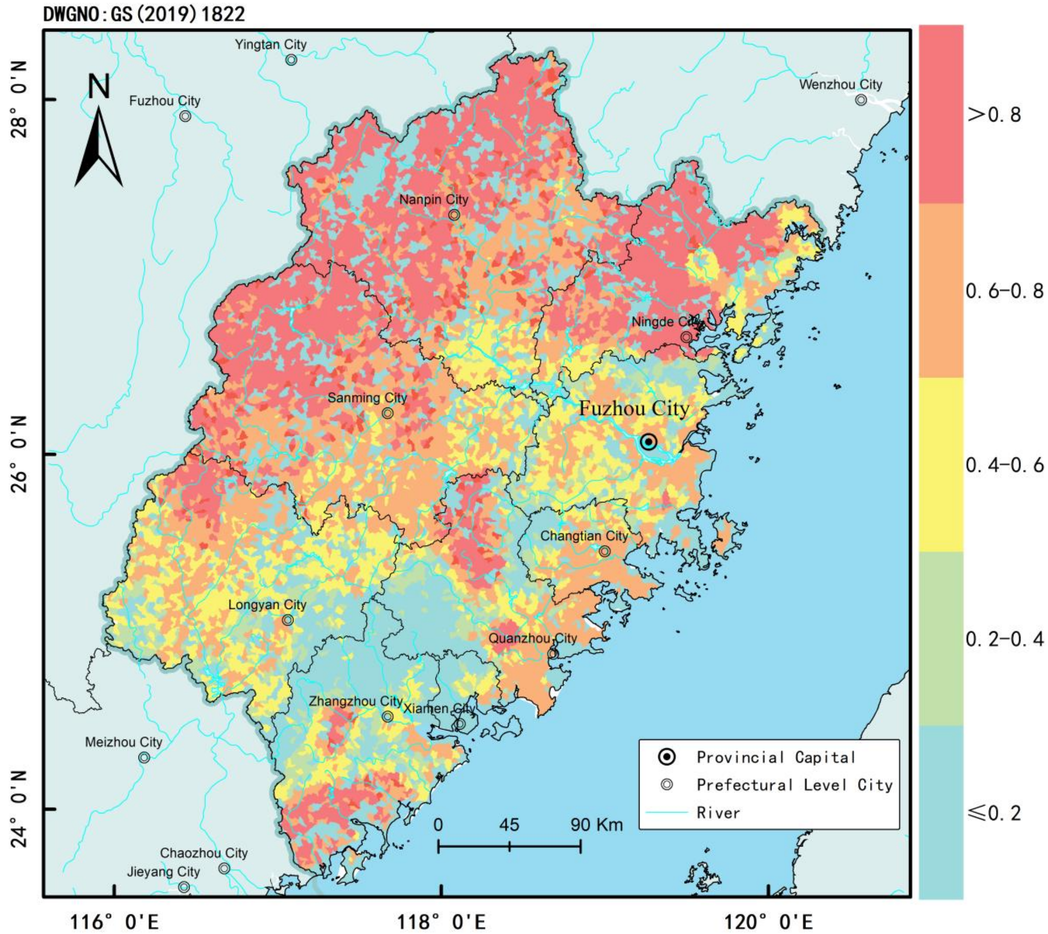


Figure 4: Risk Distribution Map of Flash Flood Disasters

4.2 Discussions

4.2.1 Accuracy Analysis Using Classification RFE

This study employed the classification recursive feature elimination (RFE) method, which progressively removes factors with low correlation to flash flood occurrences while retaining key features that contribute significantly to flash flood events (Table 2). The Receiver Operating Characteristic (ROC) value initially increased as insignificant features were eliminated; however, when the number of selected features became too small, the ROC value began to decline. After multiple iterations, optimal model performance was achieved with eight selected features. The eight key features identified were: Number of days with rainfall exceeding 50mm, Daily average rainfall, Five-day cumulative rainfall, 95th percentile rainfall, Slope, Longitudinal slope of the longest flow path, Building elevation, Population of the small watershed. The model achieved an ROC value of 0.947. The model comprehensively considered three major categories of key features: rainfall, underlying surface conditions, and socio-economic factors. The optimized feature combination not only improved the predictive performance of the model but also more accurately reflected the primary influencing factors of flash flood occurrences in Fujian Province. This approach provides a reliable and comprehensive basis for subsequent flood prevention and mitigation decision-making.

Table 2: Risk indicators of flash floods in mountainous areas

Indicator combination	Remaining indicators after feature selection	ROC value	Indicator number	Eliminated features
(4，4，2)	Number of days with rainfall exceeding 50mm, Daily average rainfall, Five-day cumulative rainfall, 95th percentile rainfall, Slope, Longitudinal slope of the longest flow path, Roughness coefficient, Centroid elevation, Building elevation, Population of the small watershed	0.946	10	Steady infiltration rate, Perimeter, Outlet drainage area, Area, Longest flow path length, Centroid elevation, Number of buildings, Distance of buildings from the riverbank
(4，3，2)	Number of days with rainfall exceeding 50mm, Daily average rainfall, Five-day cumulative rainfall, 95th percentile rainfall, Slope, Longitudinal slope of the longest flow path, Roughness coefficient, Building elevation, Population of the small watershed	0.945	9	Steady infiltration rate, Perimeter, Outlet drainage area, Area, Longest flow path length, Centroid elevation, Number of buildings, Distance of buildings from the riverbank, Centroid elevation
(4，2，2) Optimal combination	**Number of days with rainfall exceeding 50mm, Daily average rainfall, Five-day cumulative rainfall, 95th percentile rainfall, Slope, Longitudinal slope of the longest flow path, Building elevation, Population of the small watershed**	**0.947**	**8**	**Steady infiltration rate, Perimeter,Outlet drainage area, Area, Longest flow path length, Centroid elevation, Number of buildings, Distance of buildings from the riverbank, Centroid elevation**
(3，2，2)	Number of days with rainfall exceeding 50mm, Daily average rainfall, Five-day cumulative rainfall, Slope, Longitudinal slope of the longest flow path, Building elevation, Population of the small watershed	0.937	7	Steady infiltration rate, Perimeter,Outlet drainage area, Area, Longest flow path length, Centroid elevation, Number of buildings, Distance of buildings from the riverbank, Centroid elevation, Roughness coefficient, 95th percentile rainfall

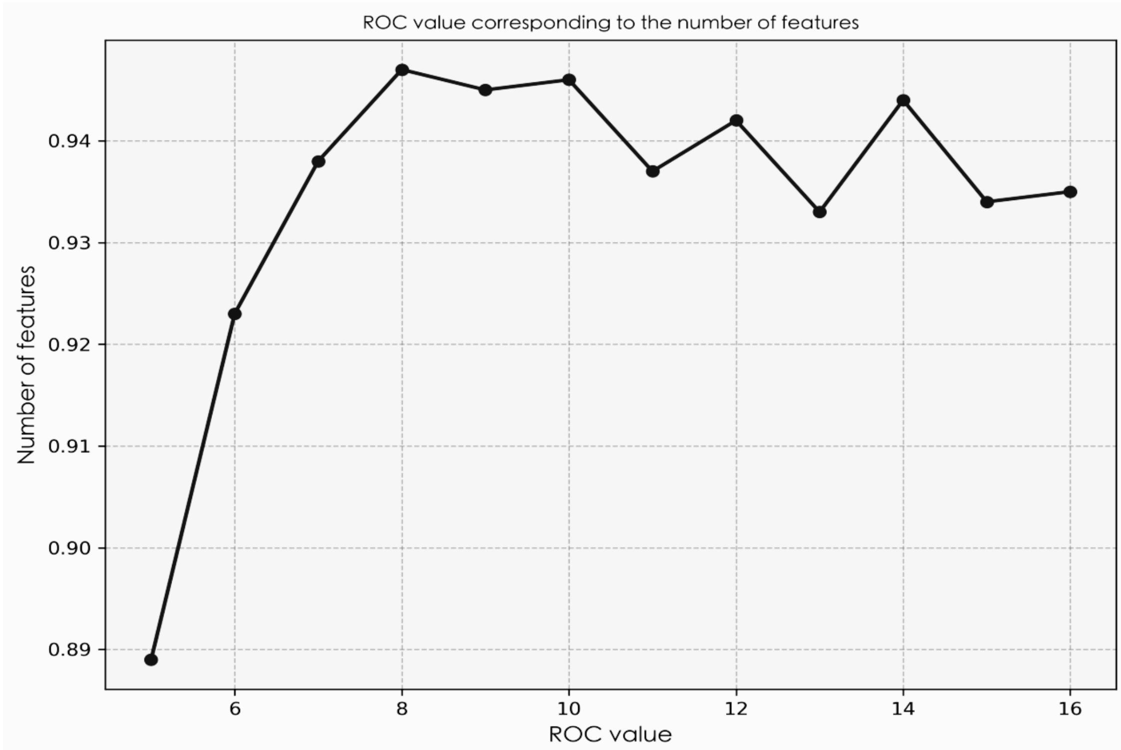


Figure 5: The corresponding ROC values for different numbers of features

4.2.2 Analysis of Flash Flood Disaster Leading Factors

Fig. 6 presents the importance assessment results of various factors. Daily average rainfall is identified as the most critical, followed by the number of days with rainfall exceeding 50 mm, indicating that rainfall-related indicators predominantly influence the final outcome. Among the three indicator categories, underlying surface factors rank second, with the longitudinal slope of the longest flow path exerting the most significant effect within this group. Among socio-economic factors, the small watershed population is the most influential. Overall, the analysis in Fig. 7 underscores the central role of rainfall indicators in prediction and analysis, while also emphasizing the significance of underlying surface characteristics and socio-economic factors. This findings provides scientific basis for further optimizing disaster risk assessments.

The study results further demonstrate that rainfall factors, socio-economic factors, and terrain factors all play key roles in the model. Validation using the classification feature elimination method reveals that the predictive accuracy decreases when only one type of indicator is retained thus substantiating the necessity of a multi-factor joint analysis to enhance model performance. This conclusion highlights the interdependence of different indicator types and offers theoretical support for optimizing future disaster risk assessment and prediction models.

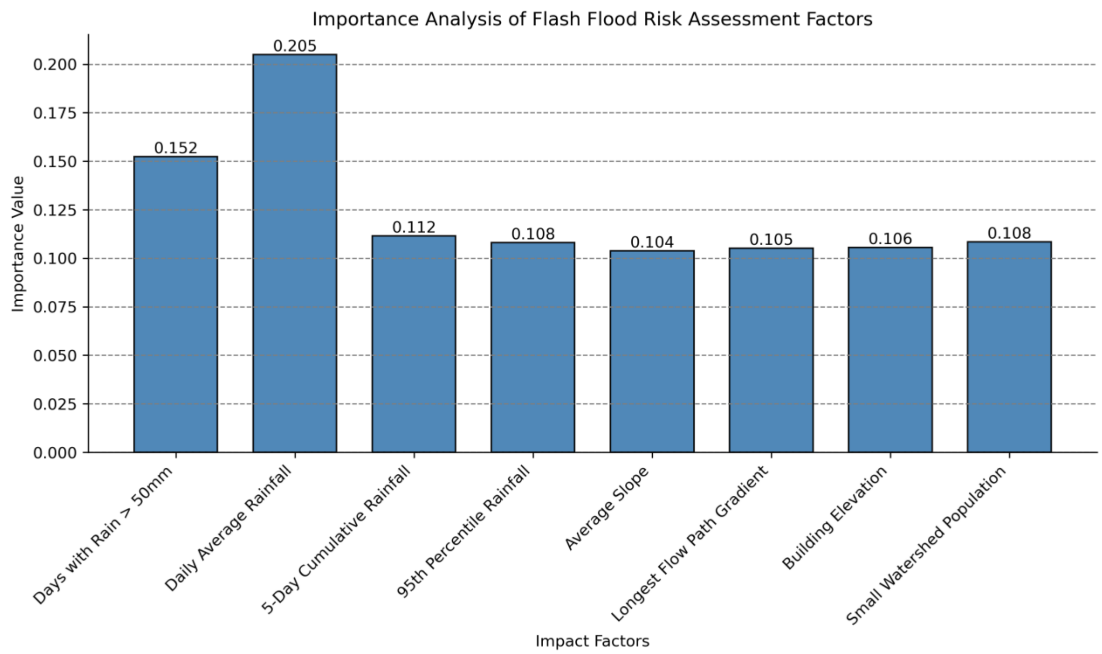


Figure 6: The importance of each feature in the optimal feature combination

Figure captions and table headings should be sufficient to explain the figure or table without needing to refer to the text. Figures and tables not cited in the text should not be presented. The following is an example for Table 1.

Table 2 Title of the Table

Type of nanoparticles	Average size (nm)	Variance (nm)
CuO	47	4.2
NiO	35	6.4
Al_2O_3	42	2.1
SnO_2	27	3.9

Tables and figures should be placed close after their first reference in the text. All figures and tables should be numbered with Arabic numerals. Table headings should be centred above the tables. Figure captions should be centred below the figures as shown in Figure 1.

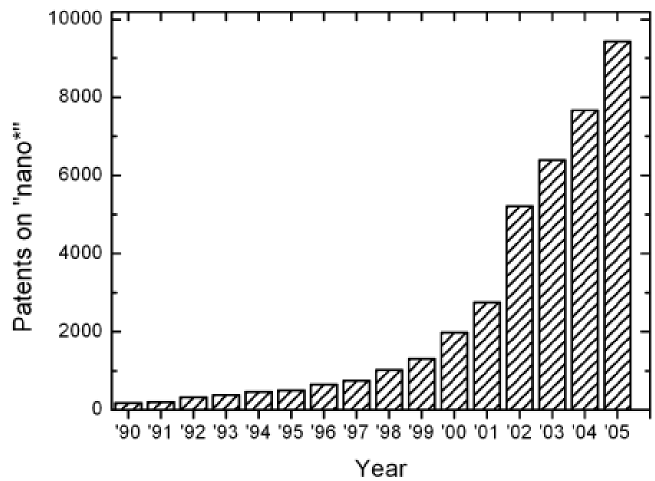


Figure 1: Number of patents on nanotechnology with time

5 CONCLUSIONS

This study developed a flash flood disaster risk prediction model by coupling Random Forest (RF) with Recursive Feature Elimination (RFE), systematically revealing the spatial distribution patterns of flash flood disaster risk levels and providing a scientific basis for disaster prevention. The study focuses on the southeastern coastal region of China, Fujian Province, and uses a classification recursive feature elimination method to optimize feature selection, identify key influencing factors, and build a high-precision risk assessment model, thereby improving the accuracy of risk prediction. The main conclusions are as follows:

(i)A flash flood risk assessment model based on the RFE-RF optimization algorithm was developed, achieving an ROC value exceeding 0.947, which demonstrates its high predictive capability.

(ii)Dominant predictive factors were identified, including the number of days with rainfall exceeding 50mm, daily average rainfall, cumulative rainfall over five consecutive days, 95th percentile rainfall, slope, longest flow path gradient, building elevation, and population in small watersheds.

(iii)Spatial analysis revealed that high-risk flash flood areas in Fujian Province are primarily concentrated in the Wuyi Mountains, Daiyun Mountains, and Hainan Mountains, where steep terrain and abundant rainfall increase the likelihood of flash floods.

The study provides technical support for the precise identification of high-risk flash flood areas and offers a scientific foundation for local governments and relevant agencies to develop disaster prevention and mitigation policies. The findings can guide the enhancement of flash flood monitoring, forecasting, and early warning systems, particularly in formulating detailed emergency response plans and raising public awareness of disaster prevention measures to minimize human and economic losses.

6 ACKNOWLEDGEMENTS

This work was supported by the Jing-Jin-Ji Regional Integrated Environmental Improvement-National Science and Technology Major Project (2025ZD1208302) and National Natural Science Foundation of China (NSFC) [Grant number: 52239006, 42201093].

REFERENCES

Al Kuisi, M., Al Azzam, N., Hyarat, T., Al-Qudah, K., Al-Mughaid, H., & Alshawamreh, B. (2023). Flood hazard and risk assessment of flash floods for Petra catchment area using hydrological and analytical hierarchy (AHP) modeling. Water, 15(16), 2910. https://doi.org/10.3390/w15162910

Biau, G., & Scornet, E. (2016). A random forest guided tour. TEST, 25(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bu, C., & Zhang, Z. (2020). Research on overfitting problem and correction in machine learning. Journal of Physics: Conference Series, 1693(1), 012100. https://doi.org/10.1088/1742-6596/1693/1/012100

Chen, G., Li, W., & Lin, K. (2017). GIS-based risk assessment of flash flood disasters in Qingyuan City. Pearl River, 38(10), 55–59.

Chu, Y., Song, W., & Chen, D. (2023). Risk identification of flash flood hazard using machine learning and Bayesian model averaging techniques. Water, 15(11), 1890–1897. https://doi.org/10.3390/w15111890

El-Saoud, W. A., & Othman, A. (2022). An integrated hydrological and hydraulic modelling approach for flash flood hazard assessment in eastern Makkah city, Saudi Arabia. Journal of Hydrology, 33(4), 102045. https://doi.org/10.1016/j.jhydrol.2022.102045

Fan, G., Tang, J., Gao, X., Wang, R., & Li, X. (2023). Risk assessment of flash flood disasters in the Western Region of Nanyang City based on LightGBM. China Rural Water and Hydropower, 8, 135–141.

Fu, X., Wang, Z., Qi, S., Song, L., & Han, X. (2024). Review and response process analysis of flash flood disasters from an emergency perspective: A case study of six typical events in 2022. Water Resources and Hydropower Engineering (in Chinese and English), 55(11), 15–25.

Han, Y., Lan, H., & Zhang, F. (2021). A dynamic recursive feature elimination framework (dRFE) to further refine a set of omic biomarkers. Bioinformatics, 37(18), 2845–2852. https://doi.org/10.1093/bioinformatics/btab151

Hu, X., & Song, L. (2018). Hydrodynamic modeling of flash flood in mountain watersheds based on high-performance GPU computing. Natural Hazards, 91(3), 1265–1284. https://doi.org/10.1007/s11069-018-3185-3

Jeon, H., & Oh, S. (2020). Hybrid-recursive feature elimination for efficient feature selection. Applied Sciences, 10(9), 3211. https://doi.org/10.3390/app10093211

Ju, W., Xing, Z., & Wu, J. (2024). Comprehensive risk assessment of natural disasters based on machine learning in Changzhou City, China. Environment, Development and Sustainability. Advance online publication. https://doi.org/10.1007/s10668-024-04485-w

Li, W., Cao, W., Wang, Y., & Zhao, J. (2019). Risk assessment and sensitivity analysis for flash floods in ungauged basins using coupled hydrological and hydrodynamic models. Journal of Hydrology, 574, 884–896. https://doi.org/10.1016/j.jhydrol.2019.04.093

Liang, L., & Bi, Q. (2020). Fujian Province innovates defense means to enhance the level of flash flood disaster prevention. China Flood Control & Drought Relief, 30(Z1), 105–108.

Liu, Z., Wang, H., & Zou, Y. (2025). Multi-scenario simulation study of flash flood disasters in the Zhaigang River Basin under the influence of water-related engineering. Water Resources and Hydropower Engineering (in Chinese and English), 56(1), 1–21.

Ma, M., Liu, C., Zhao, G., Li, J., & Wang, H. (2019). Flash flood risk analysis based on machine learning techniques in the Yunnan Province, China. Remote Sensing, 11(2), 197–203. https://doi.org/10.3390/rs11020197

Nugraheni, I. L., Widayati, A., Supriharyono, S., Suman, A., & Mahmudah, I. (2018). Flood disaster mitigation modeling using participation community based on land conversion and disaster resilience. International Journal of Environmental Science and Development, 9(4), 98–102.

Raillani, H., Chen, W., Ye, Z., & Zhang, Q. (2023). Uncertainty quantification for disaster modelling: Flooding as a case study. Stochastic Environmental Research and Risk Assessment, 5(1), 1–7. https://doi.org/10.1007/s00477-023-02448-7

Ruida, S., Saha, A., Islam, A. R. M. T., Kundu, B., & Matin, A. (2022). Development of geo-environmental factors controlled flash flood hazard map for emergency relief operation in complex hydro-geomorphic environment of tropical river, India. Environmental Science and Pollution Research, 29, 1–16. https://doi.org/10.1007/s11356-022-19220-3

Salih, A., & Hassablla, A. (2024). Flash flood risk assessment in the Asir region, southwestern Saudi Arabia, using a physically-based distributed hydrological model and GPM IMERG satellite rainfall data. Atmosphere, 15(6), 624. https://doi.org/10.3390/atmos15060624

Shuang, Y., Xin, M., & Yu, L. (2018). Risk assessment of flash flood disasters in Datong City based on fuzzy comprehensive evaluation method. China Water Resources and Hydropower Engineering, 49(5), 84–89.

Sun, L., Renberth, K. E., Ren, Z., & Li, C. (2022). Development of an ensemble Bayesian inference-based copula approach for bivariate risk evaluation of extreme precipitation under climate change. International Journal of Climatology, 42(14), 7417–7436. https://doi.org/10.1002/joc.7668

Wibowo, R. C., Sarkowi, M., Setiawan, A. F., & Prasetyo, Y. (2020). Flash flood hazard areas assessment in Bandar Negeri Suoh (BNS) region using an index-based approach and analytical hierarchy process. Journal of Physics: Conference Series, 1434(1), 012006. https://doi.org/10.1088/1742-6596/1434/1/012006

Xiong, J., Lu, D., Tian, F., & Liu, H. (2020). Spatiotemporal characteristics and driving force analysis of flash floods in Fujian Province. International Journal of Geo-Information, 9(2), 133–137. https://doi.org/10.3390/ijgi9020133

Yang, L., Fang, X., Yuan, L., & Wang, X. (2023). Sensitivity analysis of flash flood factors and disaster risk assessment in small watersheds of Jiangxi Province. Science Technology and Engineering, 23(10), 4448–4456.

Zhang, J., Dong, L., Zhang, N., & Li, Y. (2024). GIS-based risk assessment of flash flood disasters in Xinjiang Uyghur Autonomous Region. Yangtze River, 55(11), 81–88, 95.

Zhang, Q., Zhang, H., Li, Y., & Wang, S. (2019). Research on Chongqing flash flood disaster risk assessment system based on AHP-GIS. In IOP Conference Series: Earth and Environmental Science (Vol. 330, No. 3, p. 032028). IOP Publishing. https://doi.org/10.1088/1755-1315/330/3/032028

Zhang, W., Ma, G., Wu, J., & Li, X. (2020). Assessment of flood inundation by coupled 1D/2D hydrodynamic modeling: A case study in mountainous watersheds along the coast of Southeast China. Water, 12(9), 2550. https://doi.org/10.3390/w12092550

Zhang, X., Liu, R., Liu, Q., Li, C., & Guo, L. (2019). Risk identification and quantitative analysis of flash flood disasters in Fujian Province. Journal of China Institute of Water Resources and Hydropower Research, 17(4), 6–12.