Dear Editor and Reviewers,

We sincerely appreciate the time and effort that the editor and both reviewers have dedicated to evaluating our manuscript. We are grateful for the insightful comments and constructive suggestions, which have significantly helped us improve the quality and clarity of our work. We have carefully considered each comment and have made the necessary revisions accordingly. Below, we provide detailed responses to each of the reviewers' suggestions and explain the corresponding modifications made in the manuscript.

Thank you again for your valuable feedback and support.

Reply on RC1

1. Thank you for your valuable feedback. In response to your comment that the datasets were not sufficiently described, we have revised and improved Section 2.2.1 (Data Sources) to provide clearer and more detailed descriptions of the datasets used in this study.

Specifically, we have modified the table format and content to:

(1) Clarify dataset details (e.g., data processing, resolution, and coverage).

(2) Ensure consistency in formatting and descriptions across all dataset categories.

(3) Improve readability by structuring the information in a more accessible way.

The updated Table 1 now provides a more comprehensive overview of the datasets used, including spatial resolution, data preprocessing methods, and specific sources.

2. Thank you for your valuable feedback. In response to your comment regarding the level of detail in the results, we have carefully revised the manuscript and incorporated detailed modifications based on the specific comments you provided.

Each of your specific concerns has been addressed individually, with revisions made to enhance the clarity, completeness, and validity of the results. These modifications ensure that the findings are now presented with a greater level of detail, making them more transparent and easier to evaluate.

We appreciate your insightful suggestions, which have significantly contributed to improving the manuscript.

3. The combination of machine learning and AHP methods for flood risk assessment is already quite common. However, using only a single machine learning algorithm tends to result in poor interpretability of flood risk, leading to uncertainty in the model's flood risk results. Additionally, further research is needed to efficiently and accurately select the optimal machine learning algorithm for the region.

Automatic machine learning tries to automatize the steps of feature extraction, model and algorithm selection, parameter optimization, and so on so that it needs no human assistance and avoids man-made bias. This approach only requires the configuration of different run times, allowing the algorithm to explore a wider array of model and parameter combinations within the allocated time, ultimately leading to the identification of the best-performing model.

This paper selects the Auto-Sklearn framework to address the binary classification problem of flooded point identification and to calculate flood risk. By utilizing the characteristics of automated machine learning, the efficiency of machine learning is improved, and the importance ranking of flood hazard factors is obtained. The next step is to use the AHP to calculate the relevant weights by combining the flood risk results with exposure, vulnerability, and resilience indicators. AHP aims to quantify the decision-making process by scoring weights according to their level of importance, ultimately yielding the flood risk results.

4. Line 57 - Replace the sentence in line 57 with: " Through continuous improvement and development of machine learning algorithms, ensemble methods have been addressed the limitations of traditional machine learning models. "

5. Line 76 - The meaning expressed in line 76 is inaccurate; a more precise phrasing would be: "The effectiveness of machine learning "automatically improves with experience," and a key challenge in the research is how to integrate the data processing capabilities and feature selection strengths of hybrid models with ensemble models."

6. Line 178 - The reason for choosing uneducated individuals as one of the indicators is that they are often part of socially vulnerable groups. People with lower levels of education may not fully understand warning information, disaster prevention measures, or have access to sufficient disaster

preparedness resources, which increases their vulnerability during disasters. Additionally, those with higher education levels typically have access to more information channels, while individuals with lower education levels may be unfamiliar with new technologies or information channels (such as mobile apps and internet alerts). At the same time, lower-educated groups may live in areas with less developed infrastructure, making it difficult for them to receive timely social aid and support. Uneducated individuals may also find it harder to regain economic independence after a disaster, as they may lack access to technical training or knowledge updates, leading to slower recovery. Although income level is an important factor in assessing vulnerability, average income distribution can sometimes obscure individual differences. For example, a region may have a high average income level, but low-income groups (such as uneducated individuals) can still be in a highly vulnerable state. Additionally, income level may not directly reflect an individual's awareness of disaster, knowledge reserves, or ability to take action.

7. Line 179- Urbanization rate refers to the proportion of the urban population to the total permanent population in a given region, and it reflects the level of urbanization in that area. This indicator has an inverse relationship with flood vulnerability. Generally speaking, the higher the urbanization rate of a region, the higher the level of social development and the capacity for protection, which can reduce flood vulnerability to some extent.

8. Line 187- The flood inventory map in this paper was developed using inundation data from the Global Flood Database and flood disaster data from the EM-DAT database, supplemented by satellite and Google image interpretation and verified against existing historical flood records. The actual flood-affected areas were delineated based on flood traces from the inundation datasets and image interpretations. For this study, 278 flood inundation points were randomly selected within the inundation data range during the study period, and the location of each point was used as the basis for subsequent statistical analysis of flood events.

9. Line 188- Thank you for your valuable feedback. To enhance clarity and better address your comment regarding Line 188, we have revised Section 2.2 in the manuscript.

The original Section 2.2 "Data Sources" has been renamed to "Section 2.2 Data Sources and

Processing" and is now divided into two subsections:

Section 2.2.1 Data Sources – Provides details on the datasets used in the study, including their sources and resolutions.

Section 2.2.2 Data Standardization and Preprocessing – Specifically addresses the issue raised in Line 188, explaining how features with different resolutions were mapped to meteorological stations to construct the training dataset.

In Section 2.2.2, we have added a detailed explanation of the (1) Unification of Spatial Scale and (2) Normalization of the Numerical Range:

(1) Unification of spatial scale means aligning data within the same coordinate range and resolution. The research data is standardized through projection transformation, converting all datasets into the same geographic and projected coordinate systems. The Kriging interpolation method is used to spatially process all discrete data. Finally, if the spatial data has different resolutions, resampling is performed to standardize all data to the same resolution, which in this study is unified to 30m×30m.

(2) Normalization of the numerical range can be achieved using a normalization process. Through a linear transformation, the values of the data are mapped to the range [0, 1], thus eliminating the influence of differing dimensions among the data indicators. In this study, the Min-Max Normalization method is used for normalization, and the formula is as follows:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

These modifications ensure that datasets from different sources and resolutions are properly standardized for flood risk assessment.

10. Line 194 - Data balancing was necessary in this study because an imbalanced flood and non-flood sample ratio led to biased model performance, where the classifier tended to favor the majority class. Through experiments, we found that using a 1:1 ratio for flooded and non-flooded points in the training dataset significantly improved the model's predictive performance, compared to 1:2 and 1:3 ratios, which resulted in decreased recall for the minority class. Therefore, we adopted a 1:1 sampling strategy to ensure a more balanced representation of flood and non-flood samples during training.

11. Line 201- Each label consists of 278 points representing entire dataset.

12. Line 201- The overview map of the study area has been revised to display the spatial distribution of flooded and non-flooded points.

13. Line 231- The text here contains a definition error, which has been corrected in the main body of the paper. Thank you to the reviewer for pointing this out.

14. Line 290- The definition of consistency has been added to the paper: In a pairwise comparison matrix, the decision-maker's judgments must exhibit logical coherence and transitivity. This means that if option A is considered more important than option B, and option B is considered more important than option C, consistency requires that option A must also be judged more important than option C.

15. The λ in Eq. 11 should be λmax as defined in Eq. 10, and this has been corrected in the paper.

16. Line 299 - This sentence has been revised: Where average Random Consistency Index (RI) represents the average random consistency which depends only on the order of the judgment matrix. The RI values for judgment matrices of order 1 to 10 are shown in Table 3.

17. Line 316 - Thank you for your valuable feedback. In response to your comment on Line 316, we have removed accuracy as an evaluation metric. Instead, we have adjusted the relevant sections throughout the manuscript to focus on Precision, Recall, F1-score, and the Kappa coefficient, which provide a more reliable evaluation of model performance in an imbalanced dataset.
We have carefully revised all occurrences of accuracy in the text to ensure consistency and clarity in our evaluation methodology. Please let us know if further modifications are needed.

18. Line 319 - The probability thresholds for accuracy, precision, recall, and F1-score range from [0, 1], while the Kappa coefficient ranges from [-1, 1].

19. Line 327 - We used 5-fold cross-validation to assess overfitting by comparing the performance of the training and testing sets. The experimental results indicate that the performance of the training set and testing set is relatively close, suggesting that the model does not exhibit overfitting.

20. Line 368 - Thank you for your valuable suggestion. We have adopted your recommendation and have cited Lundberg and Lee (2017) instead of Wang et al. (2023a) for the description of SHAP. Additionally, we have revised the sentence for improved clarity as follows:
"SHAP is an explanation method based on game theory and belongs to post-hoc model interpretation methods (Lundberg and Lee, 2017)."

21. Figure 8 - Thank you for your question regarding Figure 8. The figures were generated based on a flood risk assessment model that combines CatBoost under the AutoML framework and AHP to calculate flood hazard, exposure, vulnerability, and resilience in the YRDUA.
The flood hazard level (a), flood exposure level (b), flood vulnerability level (c), flood resilience level (d), and flood risk spatial distribution (e) were derived through natural breaks classification in ArcGIS software, using weights determined by the model. The final result is a flood risk zoning map for the YRDUA.

22. Line 450 - Thank you for your comment. We have revised the original text and incorporated the following content to clarify the causes of differences in flood risk over the past decades:
"The differences in flood risk among cities in the YRDUA over the past few decades are primarily due to a complex interplay of various factors, including geographic and climatic conditions, urbanization processes, socio-economic factors, ecological changes, and historical flood events. The topography and precipitation patterns of different cities affect their capacity for rainwater drainage and accumulation, while urbanization leads to an increase in impervious surfaces and variations in infrastructure development, impacting flood management capabilities. Additionally, differences in population density, economic development levels, and flood management policies can exacerbate flood risk. Furthermore, the increasing frequency of extreme weather events due to climate change further elevates flood risk. These factors determine the varying levels of flood risk among cities

within the YRDUA."

Reply on RC2

1. Line 13 - Thank you for your careful review. We appreciate your attention to detail. We have adopted your suggestion and revised "remains" to "remain" in Line 13 to ensure grammatical accuracy.

We appreciate your insightful comments and your help in improving the clarity and correctness of our manuscript.

2. Line 19 - Thank you for your suggestion. We have introduced CatBoost properly in the revised manuscript and have modified the sentence as follows:

"Results indicate that, among different assessment models, the Categorical Boosting (CatBoost) model achieves the highest accuracy, precision, F1-score, and kappa coefficient for flooded point identification."

This ensures that CatBoost is properly introduced before being referenced.

3. Thank you for your careful review. We have corrected all missing spaces before citations as mentioned and have thoroughly checked the entire manuscript to ensure consistency in citation formatting.

We appreciate your attention to detail and your valuable feedback in improving the clarity and presentation of our manuscript.

4. Line 78 - The point has been removed. Thank you for your correction.

5. Line 111-114 Thank you for your valuable suggestion. We have carefully revised Lines 108–114 to remove detailed results from the introduction and instead provide an overview of the manuscript structure. The revised section now reads:

"The comparative analysis of superimposed flooded points data shows a strong alignment between the distribution of flooded points in the study area and the high to medium-high risk areas, highlighting the reliability and applicability of the proposed model. The remainder of this paper is structured as follows: Section 2 describes the study area, data sources, and methodology; Section 3 presents the results and analysis; Section 4 discusses the findings and their implications; and Section 5 concludes the study with key insights and recommendations."

6. Figure 2 - Thank you for your constructive feedback on Figure 2. I have updated the color palette to avoid any confusion with rivers and lakes, and have also inserted an inset map to illustrate the position of the YRDUA relative to the entire China. The Digital Elevation Model (DEM) units of measure, which are indeed meters above sea level (m asl), have been clearly stated. The revised figure is attached in the submission documents for your review. I appreciate your detailed suggestions and hope these revisions meet your expectations.

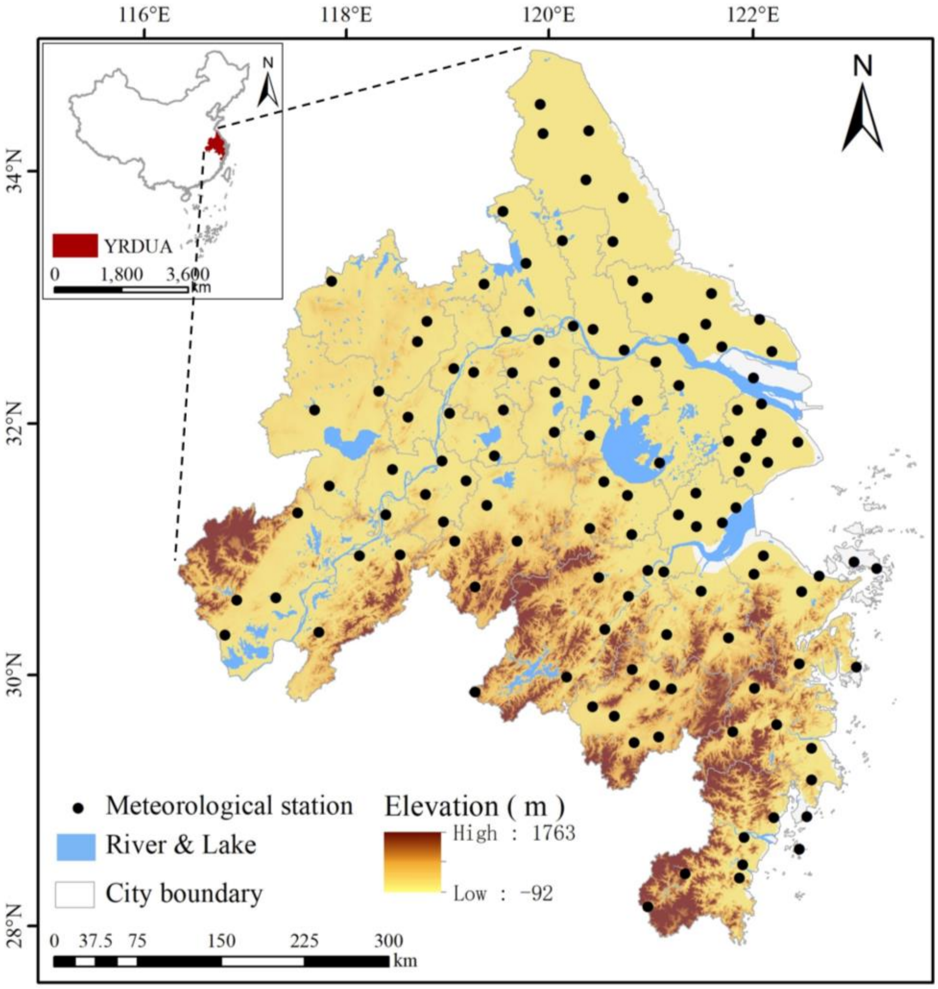

Figure 1: The schematic map of the YRDUA.

7. Table 1- Thank you for your valuable suggestions regarding Table 1. We have carefully considered your feedback and made the following improvements:

(1) Enhanced Dataset Descriptions – We have provided more detailed explanations for basic geographic information data to ensure clarity and completeness.

(2) Improved Table Readability – To enhance readability, we have inserted a horizontal dividing line after each data category, making it easier to distinguish between different dataset categories.

(3) Revised Table Caption – The table title has been updated to "Table 1: Description of the Datasets Used for Flood Risk Assessment, Their Characteristics, and Data Sources", better reflecting its comprehensive content beyond just a list of data sources.

These modifications ensure that dataset details are clearer, formatting is more structured, and the table is easier to interpret. We appreciate your detailed review and insightful recommendations, which have significantly improved the clarity and presentation of the dataset information.

Table 1: Description of the Datasets Used for Flood Risk Assessment, Their Characteristics, and Data Sources.

<table>
<tr><th>Category</th><th>Details</th><th>Resolution</th><th>Data Source</th></tr>
<tr><td rowspan="4">Basic Geographic Information Data</td><td>Administrative boundaries and river network density data.</td><td rowspan="4">30m</td><td rowspan="4">-Resources and Environmental Science and Data Center, CAS (https://www.resdc.cn/).
-USGS (https://earthexplorer.usgs.gov/).
-Wuhan University CLCD dataset (https://zenodo.org/records/8176941).
- National Ecosystem Science Data Center (nesdc.org.cn).</td></tr>
<tr><td>Digital Elevation Model (DEM) based on SRTM1 (30m), mosaicked and clipped to the study area (27 core cities).</td></tr>
<tr><td>Land use data from CLCD (30m), includes 7 types: farmland, forest, shrubland, grassland, water, bare land, and impervious surfaces.</td></tr>
<tr><td>NDVI data (2000–2020) calculated using the GEE platform.</td></tr>
<tr><td>Meteorological Data</td><td>Hourly precipitation data from 120 meteorological stations. Data preprocessed for outlier removal and missing value handling.</td><td>Station data</td><td>National Meteorological Information Center, China Meteorological Administration</td></tr>
<tr><td>Social Statistics</td><td>Population, unemployment, GDP, and healthcare statistics at the prefecture level.</td><td>Prefecture-level</td><td>Provincial and municipal statistical yearbooks and bulletins</td></tr>
</table>

	Urbanization rate calculated using urban population proportion.		
	GDP density and per capita GDP derived from total GDP and land area/population.		
Historical Disaster Data	Flood inundation data from the MODIS-based Global Flood Database (2000–2018), processed to focus on the YRDUA region. To ensure comprehensive selection of inundation points, the inundated areas within the time frame were overlaid to produce a historical flood map.	250m	Global Flood Database (https://www.emdat.be/).

8. Lines 168-183 Thank you for your insightful comment regarding the definition of heavy rainfall in Line 172. Considering its importance as a key indicator in our analysis, we have revised Lines 168-183 to provide a clearer and more precise definition.

The revised section now explicitly states:

(1) A heavy rainstorm event is defined according to the Meteorological Bureau's criteria as rainfall of 50mm or more within 24 hours.

(2) Annual Cumulative Heavy Rainfall Duration (DURA) is defined as the total number of days in a year when heavy rainstorm events occur at meteorological stations within the study area.

(3) The explanation now clarifies the relationship between prolonged heavy rainfall duration and flood risk, reinforcing why DURA is a more crucial factor than total annual precipitation in flood hazard assessment.

The revisions are as follows:

The hazard indicators consisted of six indices: Average annual precipitation (PREC), Annual Cumulative Heavy Rainfall Duration (DURA), Digital Elevation Model (DEM), SLOPE, Drainage Density (DD), and Normalized Difference Vegetation Index (NDVI). Rainfall is the primary factor leading to flooding, particularly extreme rainstorms caused by climate change. According to the Meteorological Bureau's definition, a heavy rainstorm event is characterized by rainfall of 50mm or more within 24 hours. DURA is defined as the total number of days with heavy rainstorm events occurring at all meteorological stations within the study area each year. The more days heavy rainstorms accumulate and the longer their duration, the greater the likelihood of flooding and other disaster events. DEM and SLOPE are important topographical indicators. Areas with low DEM and SLOPE values are generally more susceptible to flood threats. DD refers to the area of rivers or lakes per unit of land surface area and is a crucial indicator of a watershed's structural characteristics. It determines the watershed's capacity to withstand flooding. The higher the DD, the greater the likelihood of flooding and the higher the potential flood risk. Vegetation plays a role in water storage, retention, and infiltration. The lower the vegetation coverage, the weaker the buffering capacity, making it more likely for surface water to accumulate. The NDVI index measures the relative abundance of green vegetation, with values ranging from -1 to 1. The higher the value, the greater the vegetation coverage, and the lower the risk of flooding.

Land area (AREA), Population Density (DPOP), GDP Density (DGDP), and Building Density (DBUI) were selected as exposure indicators to assess the degree of vulnerability of both the natural environment and social systems to flooding. The land area for each administrative unit at the prefecture-level city is calculated individually. A larger land area corresponds to a greater extent exposed to flooding. DPOP and DGDP represent the concentration of population and assets, respectively. Areas with higher DPOP and DGDP are more susceptible to potential threats from pluvial flooding. DBUI, the ratio of total building area to total land area in a region, indicates the building density. A higher DBUI reflects greater exposure to flooding.

Vulnerability indicators focus more on the social aspects of flood disasters. This study selects four vulnerability indicators: Proportion of Child Population (PPOP_CHI), Proportion of Elderly Population (PPOP_ELD), Proportion of Uneducated Population (PPOP_UEDU), and Urbanization Rate (UR). Age is a key feature of social vulnerability, and both the population aged 0-14 and those over 65 are considered vulnerable groups, as these age groups are more susceptible to flood damage. The uneducated population generally has a weaker awareness of disaster risks and lower self-protection capacity, which makes this group more vulnerable to flooding. The urbanization rate refers to the proportion of the urban population in the total resident population of a region. This indicator is inversely related to flood vulnerability. In general, a higher urbanization rate indicates greater social development and stronger protective capacities, which can reduce vulnerability to flooding to some extent.

The resilience indicators selected in this study include Gross Domestic Product (GDP) per capita, Unemployment Rate (UEMP), Number of Doctors (DOCS), Number of Medical Institutions

(INSTS), and Number of Hospital Beds (BEDS). GDP per capita is the ratio of a region's GDP to its total resident population over a specified period, reflecting the region's economic condition. A higher GDP per capita indicates a more developed economy, which is associated with a greater capacity to recover quickly after a flooding event. The Unemployment Rate (UEMP) measures the proportion of the idle labor force, indirectly reflecting the stability of urban development. A high unemployment rate signals economic instability, which weakens the capacity to cope with floods and extends the time required for post-disaster recovery, thus impeding disaster response efforts. The indicators of DOCS, INSTS, and BEDS provide insights into a region's healthcare and medical support capabilities. Areas with stronger healthcare systems are better positioned to manage flood risks and recover more effectively from such disasters.

9. Thank you for your insightful question regarding the inclusion of three indicators related to the sanitary sector. These indicators—doctors, medical institutions, and hospital beds—were selected to comprehensively capture the region's healthcare capacity, which plays a crucial role in resilience during and after disasters.

While all three indicators pertain to healthcare, each represents a distinct aspect of flood resilience:
(1) Doctors reflect the availability of medical personnel to provide immediate care.
(2) Medical institutions indicate the infrastructure of healthcare facilities, which is essential for disaster response.
(3) Hospital beds measure the capacity to accommodate affected individuals, particularly in large-scale flood events.

Together, these indicators provide a balanced and multidimensional assessment of how the healthcare sector contributes to flood resilience.

Regarding the suggested inclusion of civil protection forces, law enforcement, and firefighters, we acknowledge their importance in disaster response. However, data availability constraints prevent us from including these indicators in our analysis. The relevant data for civil protection and emergency response forces are only available after 2012, whereas our study covers the years 1990, 2000, 2010, and 2020. Due to this limitation, we prioritized indicators with consistent data availability across all study periods.

Furthermore, previous research has demonstrated the importance of healthcare-related indicators in flood risk assessments. Ekmekcioğlu et al., (2021) proposed a hierarchical procedure that incorporates multiple flood vulnerability and hazard criteria, including healthcare capacity, to generate district-based flood risk maps. The study highlights that integrating healthcare infrastructure in flood risk assessments improves the ability to quantify social vulnerability and disaster mitigation capacity.

Additionally, research by Ahmed et al., (2022) in Geocarto International emphasized that mitigation capacity is a critical component in spatial flood risk mapping. The study found that regions with better healthcare infrastructure exhibit enhanced resilience and faster recovery from flood disasters. The inclusion of doctors, medical institutions, and hospital beds aligns with these findings, as they directly contribute to flood preparedness, emergency response efficiency, and post-disaster recovery. Based on these studies, we believe that healthcare-related indicators are essential for evaluating community resilience in flood risk assessments. While we recognize the role of emergency response units, our choice of indicators ensures consistency across different time periods and provides a comprehensive understanding of flood resilience.

10. Thank you for your insightful question regarding the division of the training and testing datasets. Based on your feedback, we have revised the manuscript to provide a clearer explanation of this process.

（1）Clarified Historical Disaster Data in Section 2.2

In Section 2.2, we have provided a more detailed explanation of the historical disaster data, specifically the flood inundation data from the MODIS-based Global Flood Database (2000–2018). This dataset was processed and cropped to focus on the Yangtze River Delta Urban Agglomeration (YRDUA).

（2）New Subsection 2.3: "Extraction of Historical Flood Inundation Points"

To explicitly address the division of the training and validation datasets, we have added a new subsection (2.3) titled "Extraction of Historical Flood Inundation Points", positioned before the original "2.3 Establishment of a Flood Risk Assessment Indicator System".This section explains how the historical flood map for the study area was generated and the criteria used to extract and separate the flooded and non-flooded points.To further illustrate this process, we have included Figure 3:

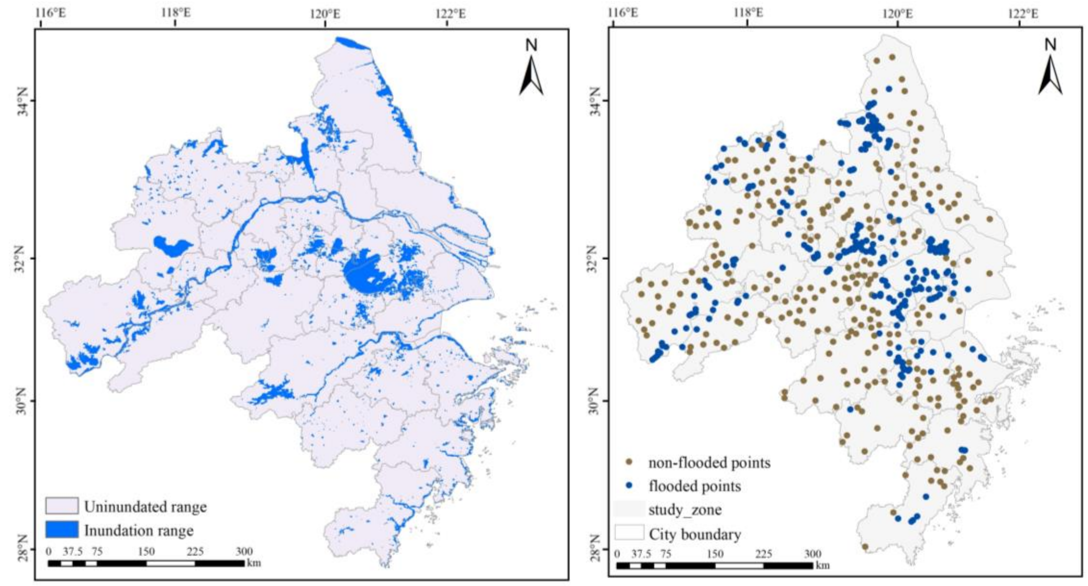

Figure 3: (a) Flood inundation map of the study area. (b) Spatial Distribution of Flooded and Non-Flooded Points in the YRDUA.

（3）Revised Section Numbering for Better Flow

The original Section 2.3 has been renumbered as Section 2.4 ("Establishment of a Flood Risk Assessment Indicator System"), and the order of the subsequent sections has been adjusted accordingly to maintain logical coherence in the manuscript.

11. Thank you for your insightful comments regarding the model names used in the manuscript. Based on your feedback, we have revised the model names to ensure clarity and accuracy.

(1) "Linear" has been updated to "Linear Regression" to explicitly specify that this refers to the linear regression model.

(2) "Neural Network" has been updated to "Multi-layer Perceptron (MLP) Neural Network" to clarify the specific type of neural network used in the study.

12. Section 2.4: Thank you for your insightful comments regarding data processing and spatial resolution. To enhance clarity and better address your concerns, we have made the following revisions in Section 2.2 of the manuscript.

(1) Renaming and Restructuring Section 2.2

(a) The original "Section 2.2 Data Sources" has been renamed "Section 2.2 Data Sources and Processing" to better reflect both the datasets and the processing steps.

(b)This section is now divided into two subsections:

Section 2.2.1 Data Sources – Provides details on the datasets used in the study, including their sources and resolutions.

Section 2.2.2 Data Standardization and Preprocessing – Addresses how features with different resolutions were mapped and standardized for analysis.

(2) Clarification on Raster Data and Resolution Standardization

(a) The analysis was performed using raster data, and we ensured that all datasets were standardized to a uniform resolution of 30m × 30m before model training.

(b) The preprocessing workflow involved projection transformation, spatial interpolation, and resampling to align all datasets within the same coordinate reference system and spatial scale.

(3) Detailed Explanation in Section 2.2.2

In Section 2.2.2, we have explicitly explained the two key preprocessing steps:

(a) Unification of Spatial Scale:

Data were standardized through projection transformation to ensure that all datasets were aligned within the same geographic and projected coordinate system. To process discrete spatial data, the Kriging interpolation method was applied, ensuring a smooth and continuous representation. Additionally, for datasets with different resolutions, resampling was performed to standardize them to a uniform 30m × 30m resolution for consistency in analysis

(b) Normalization of the Numerical range

A Min-Max Normalization process was applied to scale all values to [0,1], ensuring that different feature dimensions do not introduce bias into the model. The formula is as follows:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

13. Line 231-232 The text here contains a definition error, which has been corrected in the main body of the paper. Thank you to the reviewer for pointing this out.

14. Thank you for your suggestion regarding Equation 5. Based on your feedback, we have modified the notation to subscript "score" as recommended. We appreciate your careful review and constructive input, which have helped improve the clarity and consistency of the manuscript.

15. Thank you for your attention to detail in reviewing the manuscript. I have updated "judgements" to "judgments" in Table 2 as you suggested. I appreciate your guidance on this matter.

16.Line 299: This sentence has been revised: Where average Random Consistency Index (RI) represents the average random consistency which depends only on the order of the judgment matrix. The RI values for judgment matrices of order 1 to 10 are shown in Table 3.

17. Section 3.1.1: Thank you for your comment. We have added the training phase results in Section 3.1.1 and updated **Table 4** to include both training and test set performances. The results show that most models performed well on the training set but experienced a decline on the test set, indicating variations in generalization ability. CatBoost demonstrated strong robustness, while models like Decision Tree and Nearest Neighbors showed a more significant drop, suggesting higher sensitivity to overfitting.

18. Thank you for your comment. We have made the requested changes and added an explanation of the micro- and macro-average ROC curves in the text to clarify their meaning and relevance in our analysis.

19. Thank you for your valuable feedback. Based on your suggestion, we have made revisions to Section 3.1.2 to improve clarity and coherence. The two subsections—"Ranking of Importance" and "SHAP Interpretability Analysis"—have now been integrated into a single, more cohesive section titled "Importance and Interpretability Analysis of Hazard Factors".

In this revised section, we first present the importance ranking of the key flood hazard indicators using the CatBoost model and then follow with an in-depth explanation of the SHAP analysis. We also clarified the use of SHAP interaction values to capture the interaction effects between key features, specifically DEM and SLOPE, which was highlighted through the SHAP dependency plot. These revisions aim to ensure that the content flows more logically and provides a more integrated discussion of the analysis. We hope this addresses your concern and improves the overall clarity of the manuscript.

20. Thank you for your suggestion regarding the sentence in lines 354-355. I have rewritten the sentence for clarity and conciseness in the manuscript.

21. Figure 7 (b): Thank you for your helpful comment. The unit is now included in the figure.

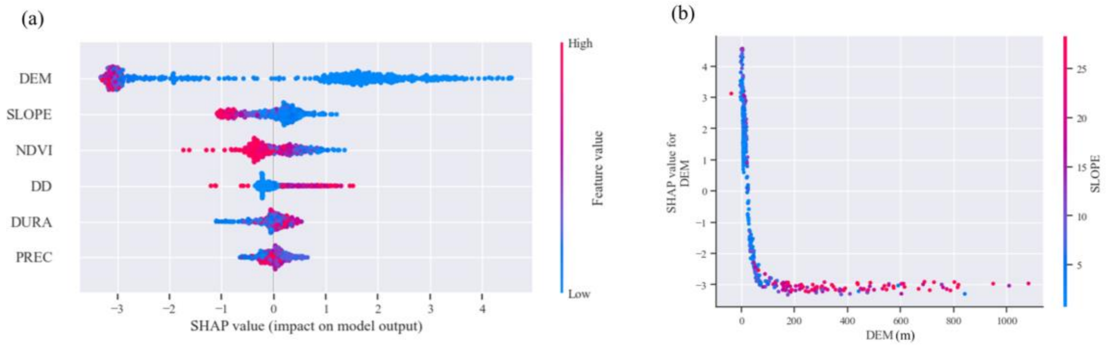

Figure 2: (a) Scatter Plot of Hazard Indicators from SHAP Analysis. (b) SHAP Dual Dependence Analysis of Elevation and Slope Factors

22. L404 and Table 5: Thank you for your valuable feedback. To clarify the meaning of the "Attribute," we have made the following revisions:

The specific indicator weights and their corresponding impacts on flood risk are shown in Table 5. The "Attribute" column represents the impact of each indicator on flood risk, with "+" indicating a positive impact on flood risk and "-" indicating a negative impact on flood risk.

23. Thank you for your valuable feedback. In response to your suggestion, we have added a new section 2.7 "Determination of Flood Risk Levels" in the methodology, where we provide a detailed explanation of the natural breakpoint classification method.

24. Thank you for your inquiry regarding the resolution of the rasters used in Figures 8 and 9. The raster data for these figures were obtained at a resolution of 30m x 30m.

Additionally, to clarify the data preprocessing steps, we have added an explanation in Section 2.2.2, "Data Standardization and Preprocessing."

25. The following content has been added to Section 3.2: In this analysis, variables such as PREC

and DURA exhibit clear temporal variability, as they change year by year due to weather patterns. However, other factors like DEM, SLOPE, NDVI, and urbanization indicators such as DPOP and GDP are spatial variables that do not exhibit direct temporal changes, but their effects on flood risk are influenced by changing socio-economic and ecological conditions.

26. Line 483 - Thank you for your suggestion. We have revised the percentage change values to retain two decimal to improve readability and clarity.

27. Thank you for your valuable suggestion. We have revised **Table 6** to explicitly indicate that the change rates are expressed as percentages. Additionally, we have updated the table caption and added percentage signs (%) to all values for clarity.

28. Thank you for your valuable suggestion. To address your comment, we have added point (5) in the conclusion section, highlighting the practical implications of our findings for flood risk management and policy-making. This addition discusses how the flood risk maps generated in this study can serve as a scientific basis for urban planning, infrastructure development, emergency response, and disaster prevention. It also emphasizes the potential of integrating these assessments into decision-making processes to enhance flood prevention measures and regional resilience. Furthermore, the AutoML framework used in this study can be applied to other regions and incorporated with future climate change scenarios for long-term forecasting and proactive disaster mitigation planning.

Reference

Ahmed, N., Hoque, M. A.-A., Howlader, N., and Pradhan, B.: Flood risk assessment: role of mitigation capacity in spatial flood risk mapping, Geocarto International, 37, 8394–8416, https://doi.org/10.1080/10106049.2021.2002422, 2022.

Anon: A weighted metric scalarization approach for multiobjective BOHB hyperparameter optimization in LSTM model for sentiment analysis, Information Sciences, 644, 119282, https://doi.org/10.1016/j.ins.2023.119282, 2023a.

Anon: Efficient LBP-GLCM texture analysis for asphalt pavement raveling detection using eXtreme Gradient Boost, Construction and Building Materials, 401, 132731, https://doi.org/10.1016/j.conbuildmat.2023.132731, 2023b.

Ekmekcioğlu, Ö., Koc, K., and Özger, M.: District based flood risk assessment in Istanbul using fuzzy analytical hierarchy process, Stoch Environ Res Risk Assess, 35, 617–637, https://doi.org/10.1007/s00477-020-01924-8, 2021.

Omar, A. and Abd El-Hafeez, T.: Quantum computing and machine learning for Arabic language sentiment classification in social media, Sci Rep, 13, 17305, https://doi.org/10.1038/s41598-023-44113-7, 2023.