26 (2025) 100490

Contents lists available at ScienceDirect

The Journal of Climate Change and Health

journal homepage: www.elsevier.com/locate/joclim

Research article

Flood risk prediction and modeling in Bauchi: Leveraging machine learning models and explainable AI for urban resilience

Kamil Muhammad Kafi [a,b,*], Zakiah Ponrahono [a], Zulfa Hanan Ash'aari [a], Aliyu Salisu Barau [b]

[a] Department of Environment, Faculty of Forestry and Environment, Universiti Putra Malaysia 43400 Serdang Selangor, Malaysia
[b] Department of Urban and Regional Planning, Faculty of Earth and Environmental Sciences, Bayero University Kano, Nigeria

ARTICLE INFO

Keywords:
Climate change: Extreme events: Disaster: Africa: Public health: Urban planning

ABSTRACT

Introduction: Floods are amongst the most destructive weather and climate-related disasters, causing significant loss of life and property globally. Accurate flood risk prediction is crucial for improving disaster resilience and urban planning.
Methods: This study employed artificial intelligence (AI) techniques, specifically Random Forest (RF), XGBoost (XGB), and Support Vector Machine (SVM) models, to predict and model flood risk in Bauchi, Nigeria. Additionally, Explainable AI analysis was utilized to interpret the model outcomes.
Results: The study revealed that high-risk areas have a history of frequent and severe flooding based on RF and XGBoost predictions. Settlement formality, elevation, population, and rainfall were the most influential factors in exacerbating flood risk. The RF model outperformed both XGBoost and SVM, with a precision of 0.857 and ROC-AUC of 0.93, while SVM performed the least, with a precision of 0.757 and ROC-AUC of 0.84.
Conclusion: The findings provide valuable insights for climate action, particularly in flood risk and exposure, and emphasize the role of urban planning and effective disaster risk reduction strategies in enhancing urban resilience.

1. Introduction

Extreme weather and climate events are becoming more frequent and impactful disasters affecting communities worldwide [1–4]. Among these, floods are particularly devastating due to their recurring nature and their ability to cause widespread damage. This makes flooding the most catastrophic of extreme events globally, with immense consequences for both the environment and human populations [5–7]. The frequency and intensity of flooding have been on the rise, driven primarily by climate change, increasing urbanization, the expansion of economic activities in flood-prone regions, and the absence of adequate flood management systems [7–11]. As one of the most globally recognized destructive hazards, floods account for a significant proportion of property damage and loss of life, with annual costs running into tens of billions of dollars [12–14], making floods the most pervasive of all-natural hazards, responsible for over 50 % of disaster-related fatalities [10,14,15]. In addition to these direct impacts, floods also pose serious public health risks, including outbreaks of infectious diseases, waterborne illnesses, and dermatological conditions caused by exposure to contaminated water [16–18].

The impacts of climate-induced flooding disproportionately affect both high- and low-income countries [19–22]. This is evident in the rising incidences of severe floods across Africa, Asia, the United States, and Europe, underscoring the global threat posed by weather and climate extremes [11]. Countries with high exposure, such as Bangladesh, Egypt, Vietnam, Indonesia, India, and Japan, are especially vulnerable due to high population concentrations along major rivers and low-lying coastal zones [23]. In Africa and Asia, flood vulnerability is further heightened by rapid, unchecked urbanization and low coping capacity amidst escalating effects of climate change [24–27]. Consequently, Nigeria grapples with significant vulnerability and exposure to frequent floods and other climate-related disasters [28,29]. According to the global disaster database EM-DAT, compiled by the Centre for Research on the Epidemiology of Disasters (CRED), Nigeria ranks as the 7th most impacted country in terms of mortality due to disasters, particularly floods (EM-DAT, 2023). In 2022 alone, floods caused 603 deaths, the highest ever recorded in the country from a single extreme weather event [30]. Additionally, Nigeria ranks among the top nations for economic losses caused by disasters, positioned as the 8th most affected globally, with flood-related damages amounting to an

* Corresponding author.
E-mail address: kmkafi.urp@buk.edu.ng (K.M. Kafi).

https://doi.org/10.1016/j.joclim.2025.100490
Received 7 January 2025; Accepted 5 July 2025
Available online 8 October 2025

unprecedented $4.2 billion USD in 2022 [31]. This marked the first time Nigeria surpassed India in flood-related economic destruction, underscoring the country's escalating weather and climate vulnerability.

Bauchi, a rapidly urbanizing city in Nigeria, faces significant risks from weather and climate extremes, especially floods and windstorms [30,32]. Flooding is a recurring issue, with over 20 major incidents affecting at least ten neighborhoods annually. Despite growing evidence of increasing flood events in Bauchi and Nigeria [33–35], predicting the locations and degree of future risk remains uncertain. This highlights the urgent need for comprehensive disaster risk prediction models to identify vulnerable areas, estimate potential impacts, and improve preparedness and risk reduction efforts. Accurate flood risk modeling is crucial for preventing future disasters and safeguarding at-risk communities [36].

The deployment of artificial intelligence (AI) has been transformative in disaster predictions and modeling, significantly enhancing the accuracy and efficiency of risk assessments [37,38]. In particular, GeoAI, which integrates geospatial data with AI techniques such as machine learning (ML) and deep learning (DL), has become a powerful tool for predicting extreme weather events, including floods [36]. These methods have simplified the complexities of disaster risk modeling, especially in data-scarce regions like Nigeria [39].

This study employed ML models to predict and assess flood risks in Bauchi. By leveraging AI's capability to analyze hydrological, topographical, and environmental data, it highlights how flood prediction accuracy can be improved, even in data-scarce regions. The findings offer valuable insights for climate action, particularly in disaster risk reduction and enhancing urban resilience.

2. Materials and methods

2.1. Study area

Bauchi city, the capital of Bauchi State, is located between latitudes 9° and 12° North and longitudes 8° and 11° East [40]. As a State, Bauchi borders several other states, including Kano to the northwest, Jigawa to the north, Yobe to the northeast, Gombe to the east, Taraba to the southeast, and Plateau to the west [41] (see Fig. 1). With a population of over 600,000, Bauchi is one of the most populous cities in the region. It experiences a tropical climate, with distinct wet and dry seasons. The wet season lasts from May to October, marked by heavy rainfall, while the dry season spans from November to April, characterized by hot and dry conditions [42].

2.2. Flood prediction and risk modeling

In this study, a multi-method approach was employed to model and predict flood risk, focusing on the use of ML and XAI algorithms for enhanced accuracy and interpretability. While many previous studies have applied ML or DL models for flood prediction [43–45], only a limited number have utilized XAI to interpret the influence of key predictor factors on risk outcomes [1,46]. This study undertook a comparative approach of three ML models and leveraged the XAI algorithm to provide deeper insights into how specific hazard conditioning factors contribute to flood risk prediction during disaster scenarios (see Fig. 2).

2.2.1. Sample inventory and acquisition

The inventory data consisted of flood events recorded over the past

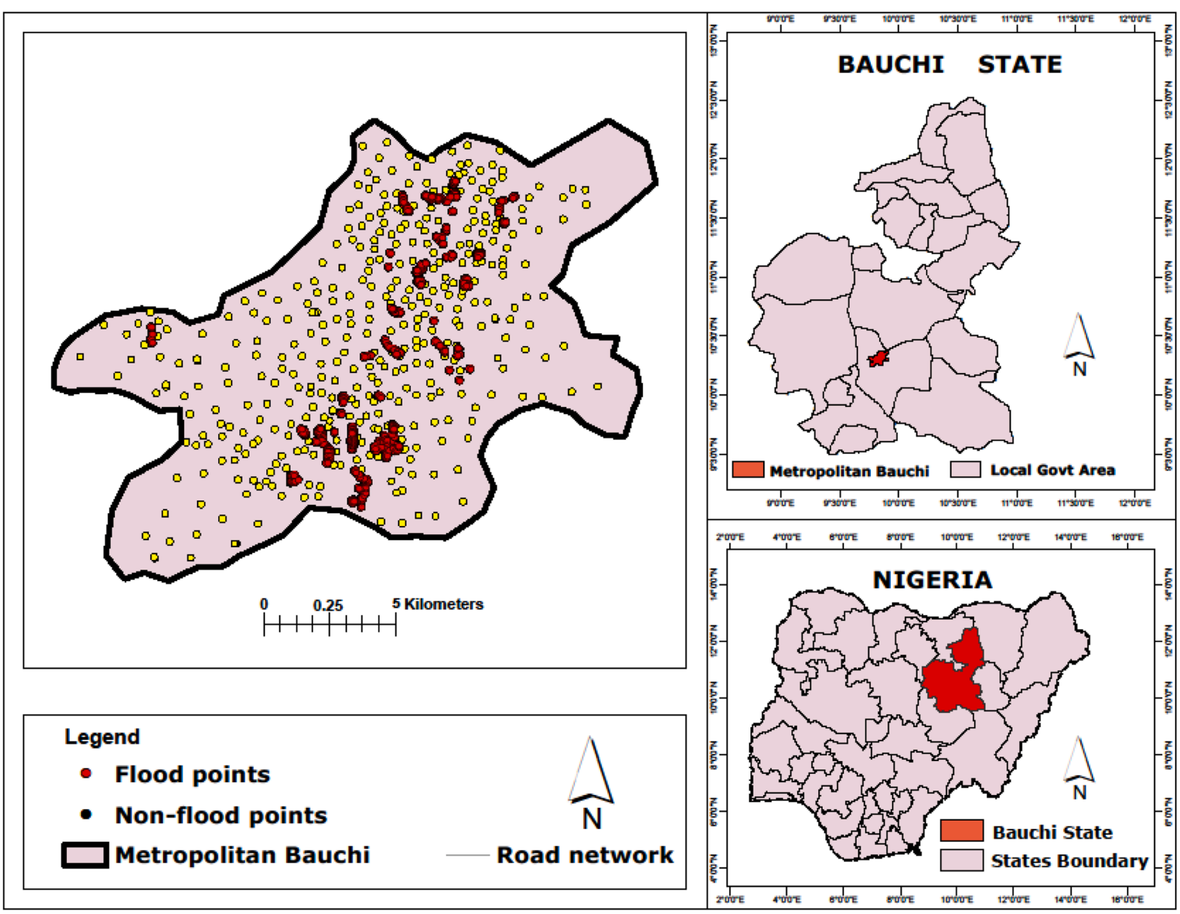

Fig. 1. Study area showing locations with histories of floods and non-flood experiences.

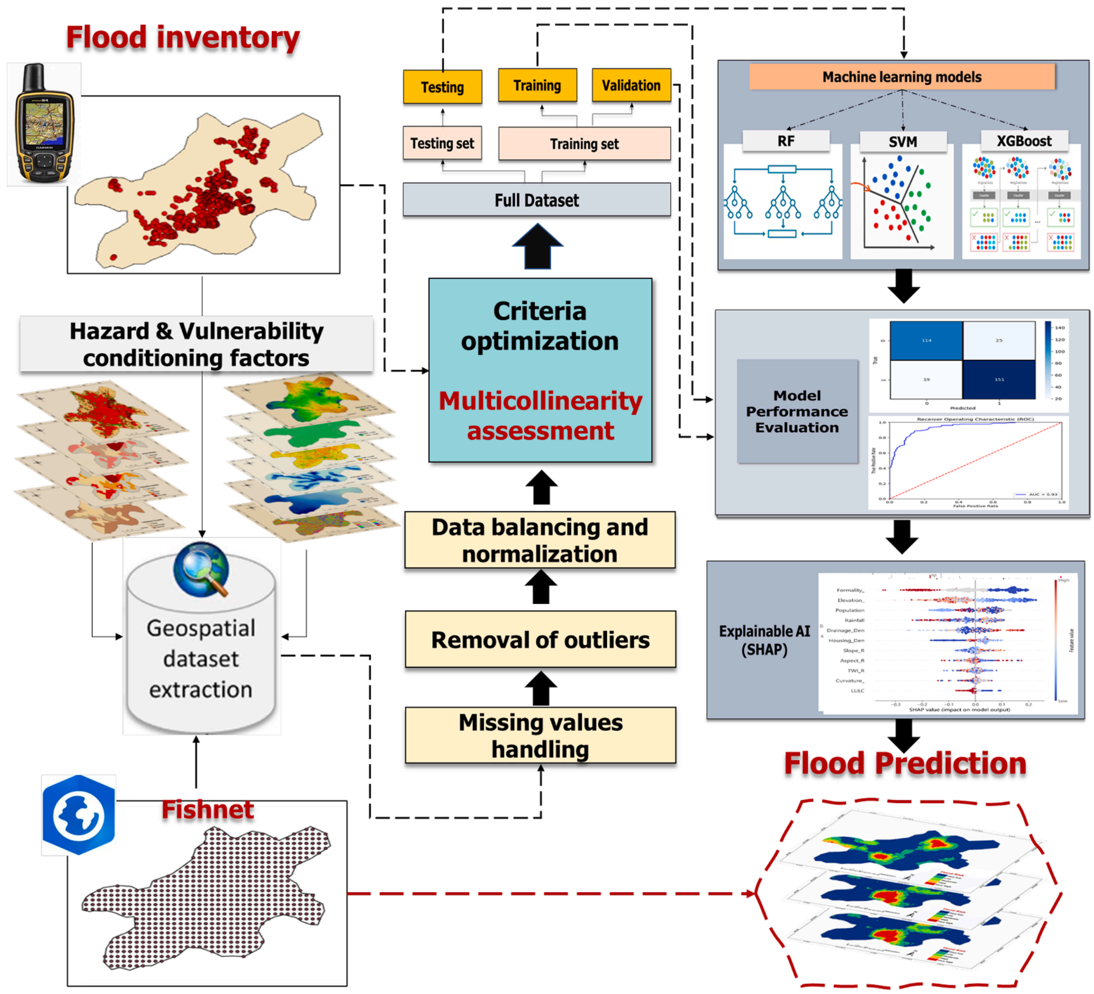

Fig. 2. Methodological workflow for flood risk prediction and modeling, integrating GPS and satellite-derived datasets with machine learning algorithms and explainable AI techniques, implemented in Python using the Jupyter Notebook environment.

five years (2019–2024), where areas affected by these events were labeled as 1, while non-flood locations were randomly selected throughout the study area and labeled as 0 [1]. The study utilized a GPS device to record the coordinates of locations with histories of flood damage [47,48]. A comprehensive census was conducted to document affected locations between 2018 and 2023, capturing their accompanying impact information. In total, 524 flood-affected locations and 505 non-affected locations were recorded. To ensure a more even distribution of non-affected areas, Google Earth Pro was employed in randomly selecting some of the non-affected flood locations, aiding in the collection of sample points across both affected and non-affected locations using GPS coordinates [43] (see Fig. 1).

2.2.2. Hazard and vulnerability conditioning factors

Hazard conditioning factors are environmental and physical characteristics, such as hydrological, meteorological, geological, and terrain variables, that contribute to the occurrence and severity of disasters. These factors influence the likelihood and intensity of flood events by shaping interactions between natural processes and the landscape [12, 49]. For flood risk prediction and modeling, ten hazard datasets were generated (see Table 1).

Vulnerability factors encompass the social, physical, and economic susceptibility of communities, infrastructure, and assets to disasters [7, 50]. They reflect the extent to which these elements are exposed to disaster impacts and their capacity to cope or recover [7]. The vulnerability dimension provides insights into an area's susceptibility to disasters and its resilience in the face of hazards [10]. In this study, four vulnerability datasets were generated for flood risk modeling (see Table 1). The study utilized a widely recognized equation that incorporates both hazard and vulnerability predictors [39]. The equation is expressed as:

Table 1
Hazard conditioning factors.

Major Factor	Conditioning Factors	Hazard	Vulnerability
Geological	Curvature	√	
	Aspect	√	
	Elevation	√	
	Slope	√	
Hydrological	Rainfall	√	
	Drainage Density	√	
	TWI	√	
Socio-physical	Housing density		√
	Settlement formality		√
Social	Population density		√
Land use	LULC		√

$$\text{Risk} = \text{Hazard} \times \text{Vulnerabilty} \tag{1}$$

Where;

Hazard is defined as the triggering factor, while vulnerability quantifies the social, physical, and economic susceptibility of the people, infrastructure, and asset elements to hazard occurrences.

2.2.3. Multicollinearity test

The study performed a multicollinearity test using IBM-SPSS-25 to evaluate the correlation between independent variables (conditioning factors) and dependent variables (flood inventories) [12]. Multicollinearity testing is crucial to ensure that the independent variables do not have high correlations with each other, as this can undermine the model's predictive accuracy [51]. By identifying and addressing multicollinearity, the model's reliability and interpretability are improved, providing clearer insights into the relationship between predictors and target variables [52]. The values for each conditioning factor were extracted using the feature extraction tool in GIS (ArcMap 10.8), aligning the flood inventory sample locations with their corresponding values for all conditioning factors.

2.2.4. Machine learning models application

To model and predict flood risk areas, three different ML models were applied, compared, and evaluated [1,12,46]. They are RF, XGBoost, and SVM models. Each model was employed to predict and map flood risk areas based on certain identified risk predictors. By utilizing multiple models, the study aimed to enhance predictive accuracy and identify the most reliable method for assessing the spatial distribution of flood risk areas [36]. Additionally, these models have been widely used for disaster predictions in previous studies and have consistently produced reliable spatial outputs with excellent prediction accuracy [53–55]. The three machine learning models were implemented using Python programming language. Python offers a robust ecosystem of libraries such as scikit-learn for ML algorithms and specialized libraries like XGBoost for gradient boosting, making it an ideal platform for predictive modeling [51].

2.2.4.4. Normalization and data splitting

Data normalization scales the features to a common range, typically between 0 and 1, ensuring that features with large numerical values do not dominate the model [36]. This helps improve the performance and convergence speed of machine learning algorithms. The experimental setup for this study involved dividing the inventory and extracted conditioning factor dataset into training and testing sets [36,56]. The training set was used to train the model, while the testing set evaluated the model's performance on unseen data, ensuring that the model would generalize well to new data and wasn't overfitting [57,58]. The study allocates 60 % of the dataset for training and 40 % for testing [51].

2.2.4.5. Model evaluation. The performance of the machine learning models was evaluated using key metrics: Accuracy, Kappa coefficient, and ROC-AUC [36,51]. Accuracy provides an overall measure of prediction correctness, while the Kappa coefficient adjusts for random chance, offering more insight in imbalanced datasets. ROC-AUC assesses the model's ability to differentiate between classes, reflecting its discriminatory power [12,46]. The confusion matrix, consisting of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), provides a tabular representation of model performance. ROC, a graphical tool in binary classification, shows the trade-off between sensitivity and false positive rate across different thresholds. Equations 2 through 5 describe the model evaluation metrics.

$$Sensitivity = \frac{TN}{TN + FP} \tag{2}$$

$$Specificity = \frac{TN}{TN + FN} \tag{3}$$

The ROC-AUC value ranges between 0 and 1, where a higher AUC indicates better model performance. It provides a single scalar value to compare different classifiers, with an AUC of 1 suggesting a perfect classifier, while an AUC of 0.5 represents a classifier performing randomly.

$$A \cup C = \sum TP \sum \frac{TN}{P} + N \tag{4}$$

Accuracy: It measures the overall correctness of predictions and is computed as the ratio of correctly predicted instances to the total number of instances within the dataset [2].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

2.2.5. Explainable AI (XIA)

XAI has emerged as a set of techniques that make ML models more transparent and understandable by providing insights into how models make decisions [59]. As a result, XAI has been increasingly utilized to interpret ML prediction outcomes [46]. In this study, the use of XAI is crucial for interpreting the results of machine learning models used to predict flood risks. SHAP, a popular XAI method, is applied to explain the contribution of each hazard and vulnerability conditioning factor to the model's predictions. SHAP assigns importance values to features, showing how they influence the model's output, which helps in better understanding the key drivers of flood risks [1].

3. Results

3.1. Flood risk predictors

The integration of hazard and vulnerability conditioning factors is fundamental to any comprehensive disaster risk assessment [20]. These factors serve as key predictors (see Figs. 3 and 4), collectively shaping the spatial patterns that inform risk modeling and enhance the accuracy of predicting flood occurrence. The inclusion of a broader set of conditioning factors typically improves the robustness and precision of the model, enabling more accurate identification of high-risk areas and contributing to more effective disaster risk reduction [59].

3.2. Multi collinearity assessment

Prior to implementing ML models for flood predictions, a multicollinearity analysis was performed to ensure there was no collinearity among the predictor variables [52], as multicollinearity can affect the accuracy of the training results [51]. The analysis revealed no substantial collinearity issues among the independent flood-related variables (Table 2). The lowest tolerance (TOL) value was found for slope, at 0.348, which is above the critical threshold of 0.10. Furthermore, all variance inflation factor (VIF) values are below the critical threshold of 10.0, indicating no evidence of multicollinearity [60].

3.3. Model performance

Metrics computation was conducted to assess the performance of the three ML models—RF, XGBoost, and SVM—applied for flood predictions in the study [57,61]. Key metrics, including accuracy, precision, sensitivity, specificity, Kappa coefficient, and ROC-AUC score, were used to evaluate both the effectiveness of flood predictions and model comparisons [12,52]. To further optimize model performance, the study applied grid search hyperparameter tuning using the *GridSearchCV* library. Since Random Forest (RF) is generally less prone to overfitting compared to other models [59], it was trained using the default parameters, while hyperparameter tuning was conducted on the XGBoost and SVM models to mitigate overfitting [46]. The optimized XGBoost parameters—max_depth = 7, learning_rate = 0.1, and n_estimators = 100—yielded the best performance with an ROC-AUC score of 0.92. Similarly, the SVM model with the optimized parameters—C = 10, kernel = rbf, gamma = 0.01, and degree = 2—produced an ROC-AUC score of 0.84. RF achieved the highest accuracy (0.857) for flood prediction, slightly outperforming XGBoost (0.847), indicating superior

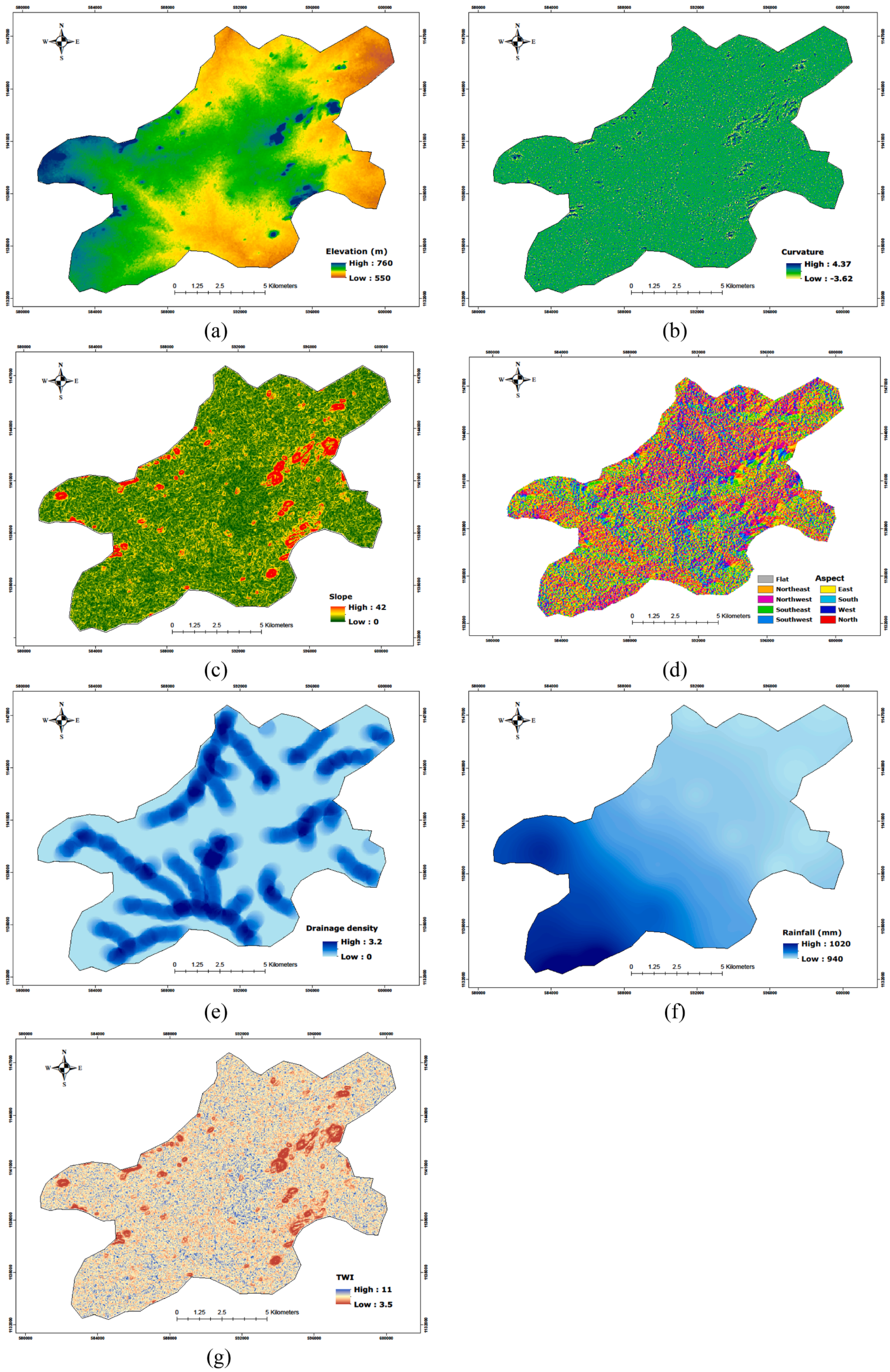

Fig. 3. Flood risk predictor variables incorporated as hazard factor layers in the GIS Environment (a) Elevation, (b) Curvature, (c) Slope, (d) Aspect, (e) Drainage density, (f) Rainfall, (g) TWI.

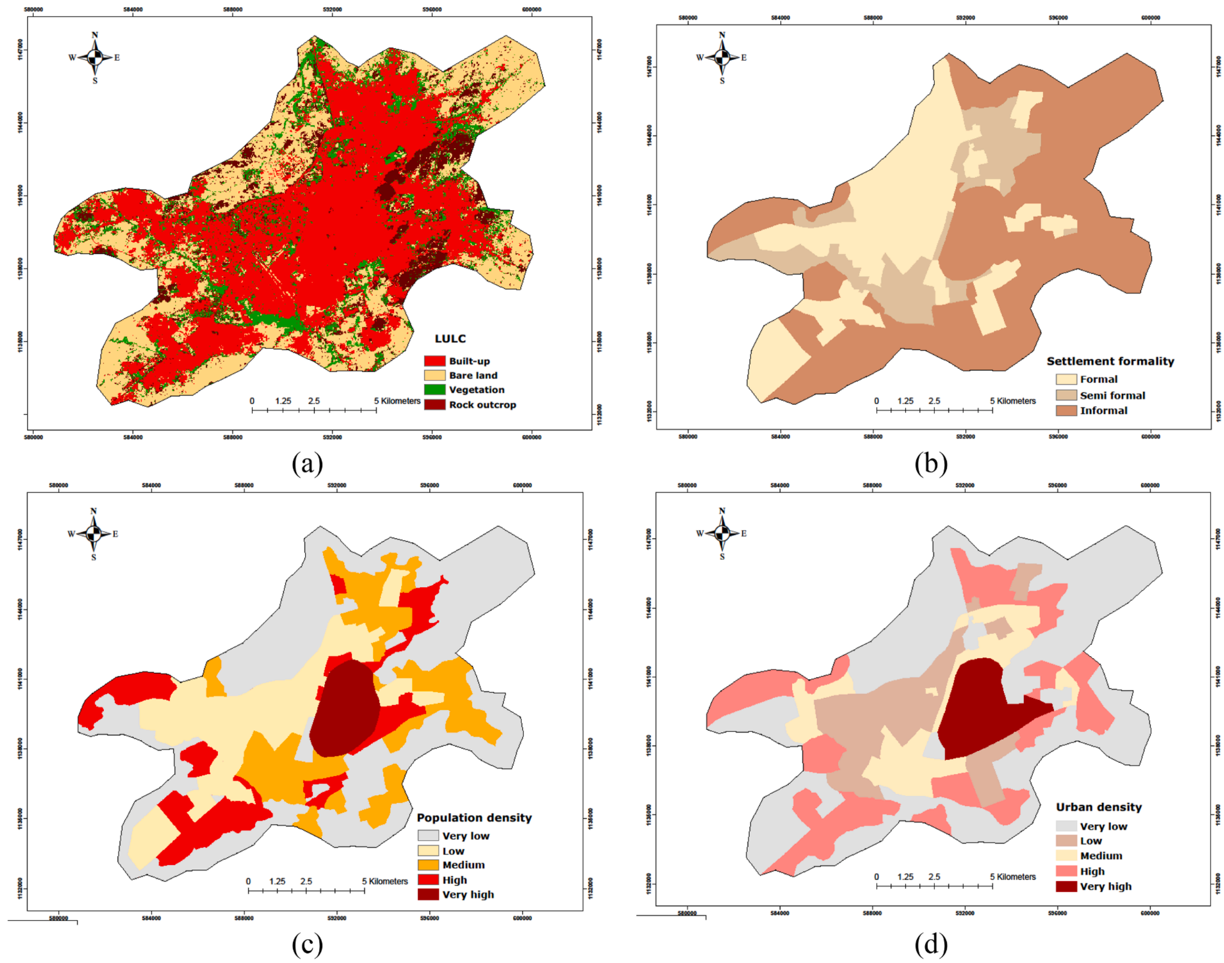

Fig. 4. Flood risk predictor variables incorporated as vulnerability factor layers in the GIS Environment (a) LULC, (b) Settlement formality, (c) Population density, and (d) Urban density.

Table 2
Multi collinearity assessment for flood risk predictors.

Factor	Flood	
	Tolerance	VIF
Aspect	0.989	1.011
Drainage Density	0.642	1.558
TWI	0.447	2.236
Rainfall	0.915	1.093
Curvature	0.938	1.066
Elevation	0.448	2.234
Slope	0.384	2.603
Population density	0.523	1.913
Formality	0.740	1.352
Housing Density	0.729	1.372
LULC	0.918	1.089

Table 3
. Performance Evaluation Metrics.

Metrics	Flood		
	RF	XGBoost	SVM
Accuracy	0.857	0.847	0.757
Precision	0.857	0.848	0.760
Sensitivity	0.888	0.883	0.781
Specificity	0.820	0.810	0.742
Kappa coefficient	0.714	0.70	0.515
ROC-AUC	0.93	0.92	0.84

overall predictive performance. In contrast, SVM recorded a lower accuracy of 0.757, reflecting its reduced ability to correctly classify risk areas (see Table 3). RF also excelled in precision with a score of 0.857, effectively identifying flood-prone areas with minimal errors, followed by XGBoost at 0.848. SVM's precision was lower at 0.757, indicating less accuracy in predicting high-risk areas [46]. In terms of sensitivity, RF led with 0.888, closely followed by XGBoost at 0.883, both effectively detecting flood-prone areas. SVM had a lower sensitivity of 0.781, suggesting reduced effectiveness in identifying high-risk zones. For specificity, RF again performed best at 0.820, with XGBoost at 0.810, while SVM's specificity was weaker at 0.742, highlighting its difficulty in correctly classifying non-risk areas [51].

The Kappa coefficient showed RF as the top performer at 0.71, indicating strong agreement between predicted and actual outcomes, followed closely by XGBoost at 0.70, both within the "good" performance range ($0.55 < k < 0.85$). SVM had a lower Kappa value of 0.52, reflecting only moderate agreement. The ROC-AUC score, which measures the model's ability to distinguish between high-risk and low-risk areas, further supported the performance patterns. RF led with a score of 0.93, closely outperforming XGBoost with 0.92, while SVM achieved a score of 0.84 (see Fig. 5a–c). RF and XGBoost demonstrated superior efficiency in classifying flood-prone areas accurately. Overall, RF emerged as the top-performing model for flood risk prediction, consistently demonstrating high accuracy and minimizing misclassification across all metrics.

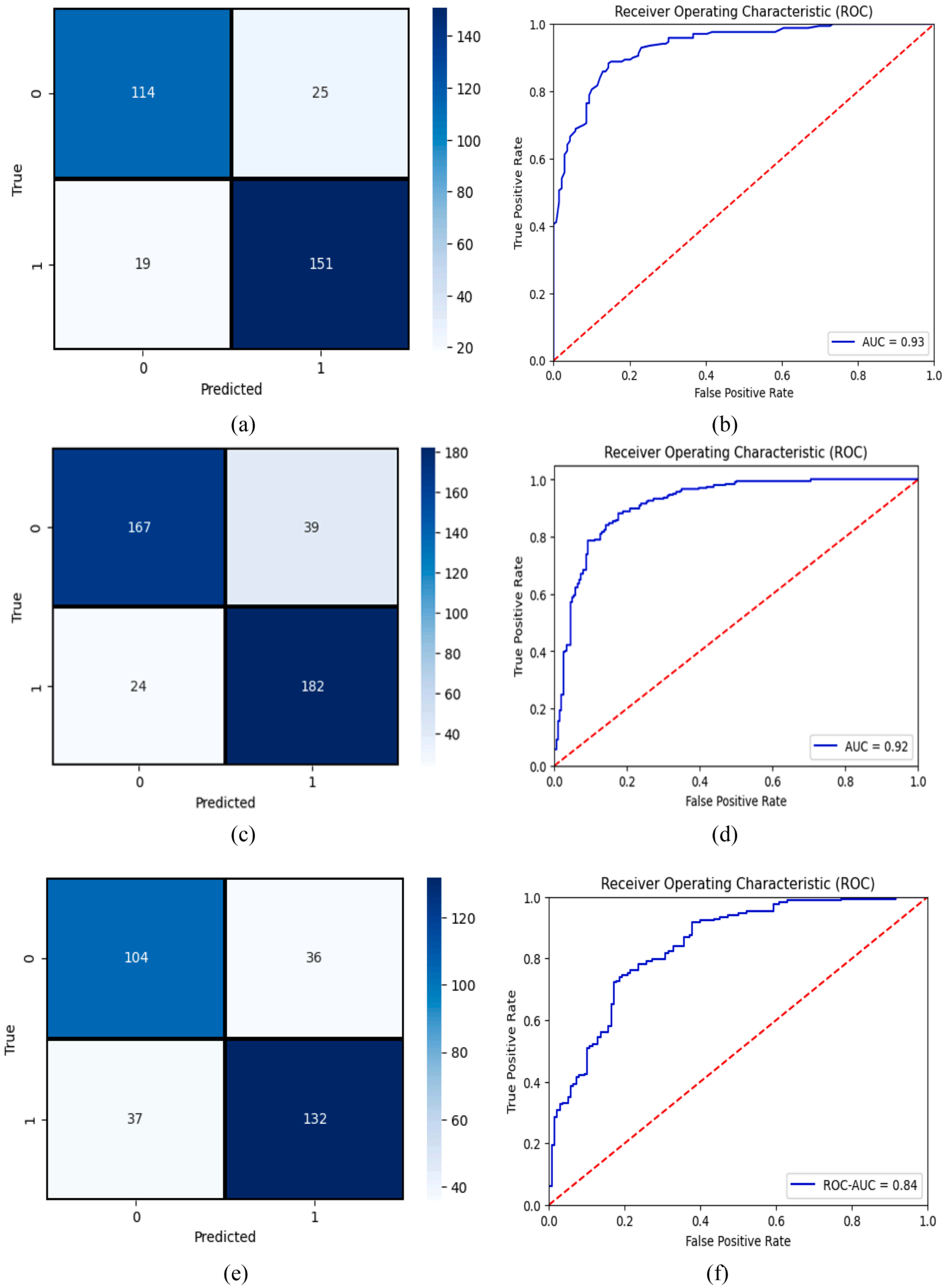

Fig. 5. Confusion matrices and ROC-AUC curves for flood prediction (a, b) Random Forest (c, d) XGBoost and (e, f) Support Vector Machine.

3.4. Factor importance and explainable AI

The SHAP summary plot highlights settlement formality and elevation as the key drivers of flood risk in both RF and XGBoost models, with similar weight values (see Fig. 7 and Table 5). Higher formality, indicating planned settlements, is associated with reduced flood risk, while higher informal settlements face significantly greater risk due to the concentration of buildings in either hazard-prone areas or areas with inadequate drainage infrastructure (Fig. 6a–d). Elevation is also crucial, with lower areas being more prone to flooding as revealed by the SHAP plot (see Fig. 7a and 7b).

Both models consistently identify informal, low-lying regions as the most vulnerable, underscoring the strong influence of these factors on flood risk. Rainfall is ranked as the third most important factor in XGBoost and fourth in RF, with higher rainfall significantly increasing flood risk. Population density, ranked third in RF and fourth in XGBoost, also contributes positively to flood risk, suggesting that the densely populated areas are more susceptible to flooding (see Fig. 7a and 7b). Both models identify curvature and LULC as the least influential factors, though an increase in LULC still raises flood risk, as indicated by positive SHAP values. This underscores how urban expansion without adequate planning amplifies flood vulnerability.

3.5. Flood risk mapping

Risk mapping reveals key insights into flood-prone areas, with XGBoost and RF exhibiting similar modeling patterns for flood inundation prediction [39], (see Fig. 8a and 8b). Both models effectively highlight regions that frequently experience severe flood events, placing them in the "high" and “very high” risk categories [1]. Areas such as Bayan-Kotu, Ungwan-Duhu, Zango, Fadaman-Mada, Korama, Magaji-Quarters, and Gwallagan-Mayaka are identified as very high-risk zones, where flood occurrence is frequent and severe. While the SVM model successfully identified some frequently flood-affected areas, such as Fadaman-Mada, Ungwan-Duhu, Magaji-Quarters, and Korama, aligning with historical flood patterns [54]. However, it failed to accurately predict other critical flood-prone neighborhoods like Gwallagan-Mayaka, Zango, Bakin-Kura, and Bayan-Kotu, which have a history of frequent and severe flooding, misclassifying them into lower-risk categories. The inability of SVM to capture these key high-risk zones highlights the model’s weaker predictive performance compared to more robust models like RF and XGBoost, which demonstrated a more comprehensive mapping of flood-prone regions.

Having achieved a similar prediction pattern between RF and XGB, the assessed spatial swath of these risk classes across different risk levels ranges from "very low" to "very high." For areas classified as very low risk, RF predicts 64.61 % of the area, while XGB closely follows with 64.1 %, indicating strong agreement between the models. In the low-risk category, RF estimates 14.88 %, and XGB predicts 13.5 %, showing that RF identifies a slightly larger portion of the area as low risk compared to XGB. For moderate risk, RF predicts 9.02 %, and XGB predicts 8.8 %, indicating a close alignment between the models in this category. However, the models differ a little significantly in the “high risk” category. RF estimates 6.26 % of the area under high risk, while XGB predicts slightly more, at 7.1 %, suggesting that XGB identifies more high-risk areas. This trend continues in the very high-risk category, where RF predicts 5.21 % and XGB predicts 6.6 %, indicating that XGB classifies a larger portion of the area as being at very high risk. While there are slight differences in the spatial coverage of some risk classes between RF and XGB, the models produced similar spatial patterns in flood risk mapping, reinforcing the reliability of their predictions for identifying flood-prone areas. The general findings reveal that the most flood-prone areas are often high-density informal neighborhoods located in low-lying water channels [39,62,63]. These areas, characterized by poor drainage and less regulated urban planning, face significantly higher flood risks [26].

Table 5
Weight of each predictor variable in flood predictions.

Predictors	Flood	
	RF	XGBoost
Aspect	0.04	0.07
Drainage Density	0.09	0.09
TWI	0.03	0.06
Rainfall	0.12	0.11
Curvature	0.02	0.04
Elevation	0.20	0.18
Slope	0.04	0.06
Population density	0.13	0.12
Formality	0.22	0.21
Housing density	0.09	0.05
LULC	0.02	0.01

4. Discussion

The results of the ML analysis indicate that the majority of neighborhoods identified as very high- and high-flood-risk areas in Bauchi have a history of frequent flooding events. These areas, including Bayan-Kotu, Ungwan-Duhu, Magaji-Quarters, and Gwallgan-Mayaka, are particularly prone to flooding due to their proximity to water channels and their elevation in low-lying regions. The recurrence of flooding in these neighborhoods underscores the persistent vulnerability of urban spaces, especially those that have been subjected to unchecked urbanization and poor spatial planning. These findings highlight a critical issue in Bauchi, where flood risks are not just theoretical but are visibly exacerbated by historical patterns of flooding.

The study also reveals that settlement formality and elevation are the most influential factors contributing to the heightened flood risk in Bauchi. Using XAI techniques, it was found that informal settlements, which have proliferated rapidly in low-lying hazard-prone areas, are significant drivers of flood vulnerability. The densification of these informal settlements has increased the pressure on flood-prone regions, with little consideration given to flood risks during the expansion of housing. As a result, areas like Bayan-Kotu and Magaji-Quarters, which are situated near water channels, are not only identified as very high-risk flood zones but also face heightened public health risks of disease outbreaks due to the contamination of wells and the overflow of sewage systems during floods [16]. Similar patterns have been observed in Ghana, where flood-prone slum areas experienced higher incidences of malaria, diarrhea, cholera, and skin diseases [18].

This pattern of urban growth is a direct consequence of unchecked urbanization and unregulated spatial planning in Bauchi. Despite the increasing frequency of floods, there has been limited effort to control the spread of settlements in vulnerable areas [11]. The lack of effective urban planning, coupled with inadequate infrastructure and flood management systems, has contributed to the worsening of flood risks. These findings suggest that the expansion of urban settlements without proper consideration of environmental factors such as elevation and proximity to water bodies is a major contributor to the high flood risk in the city.

Ultimately, the study emphasizes the urgent need for improved urban planning and regulation in Bauchi to mitigate flood risks. By addressing the flood-exacerbating factors such as settlement formality, building in low elevation and flood-prone areas, and controlling urban sprawl, policymakers could significantly reduce urban vulnerability through policy decisions and implementation that prioritize urban resilience.

Finally, this study demonstrates the effectiveness of ML and Explainable AI in predicting and modeling flood risk, as well as unraveling the underlying anthropogenic factors that exacerbate these risks amid the changing climate. This approach can also be applied in future research to predict public health risks in areas like Bauchi, where

Fig. 6. Flood incidences (a) commercial development encroaching on a water way along Federal Low-cost Bye-pass, (b) the same commercial building submerged by the 4th September 2024 flooding, (c) Bayan Kotu completely submerged by the 4th September 2024 flooding and (d) Bayan kotu showing arrangement of buildings along water channel.

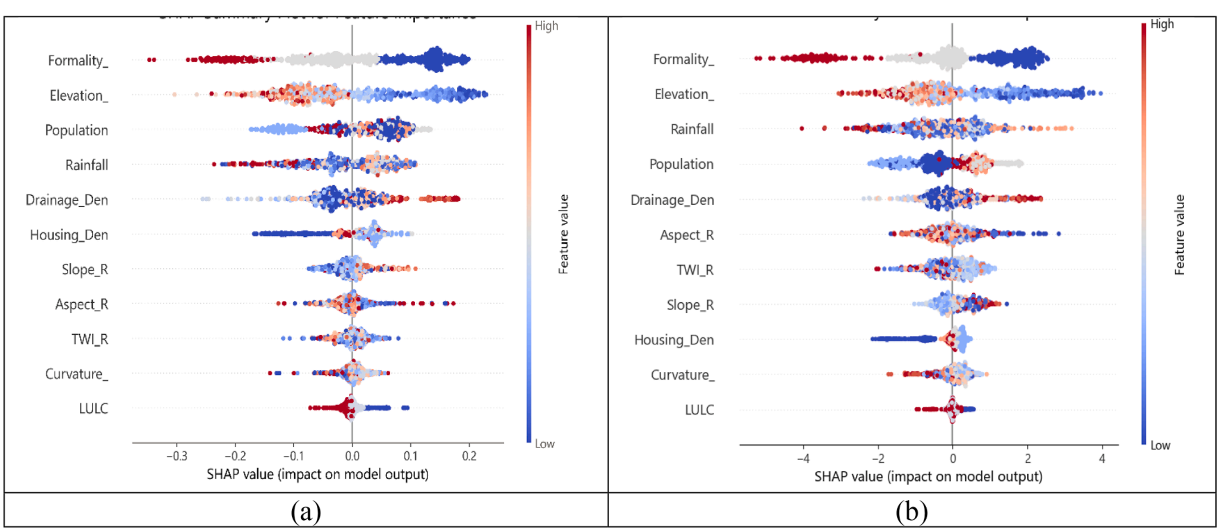

Fig. 7. SHAP summary plot showing the importance and effect of conditioning factors on individual predictions for (a) RF flood, (b) XGBoost flood. The plots demonstrate how each factor's value influences the model outputs across the dataset.

repeated severe flooding and water contamination increase the threat of disease outbreaks [16].

CRediT authorship contribution statement

Kamil Muhammad Kafi: Writing – review & editing, Writing – original draft, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Zakiah Ponrahono:** Visualization, Validation, Supervision, Conceptualization. **Zulfa Hanan Ash'aari:** Visualization, Supervision, Conceptualization. **Aliyu Salisu Barau:** Visualization, Validation, Supervision, Conceptualization.

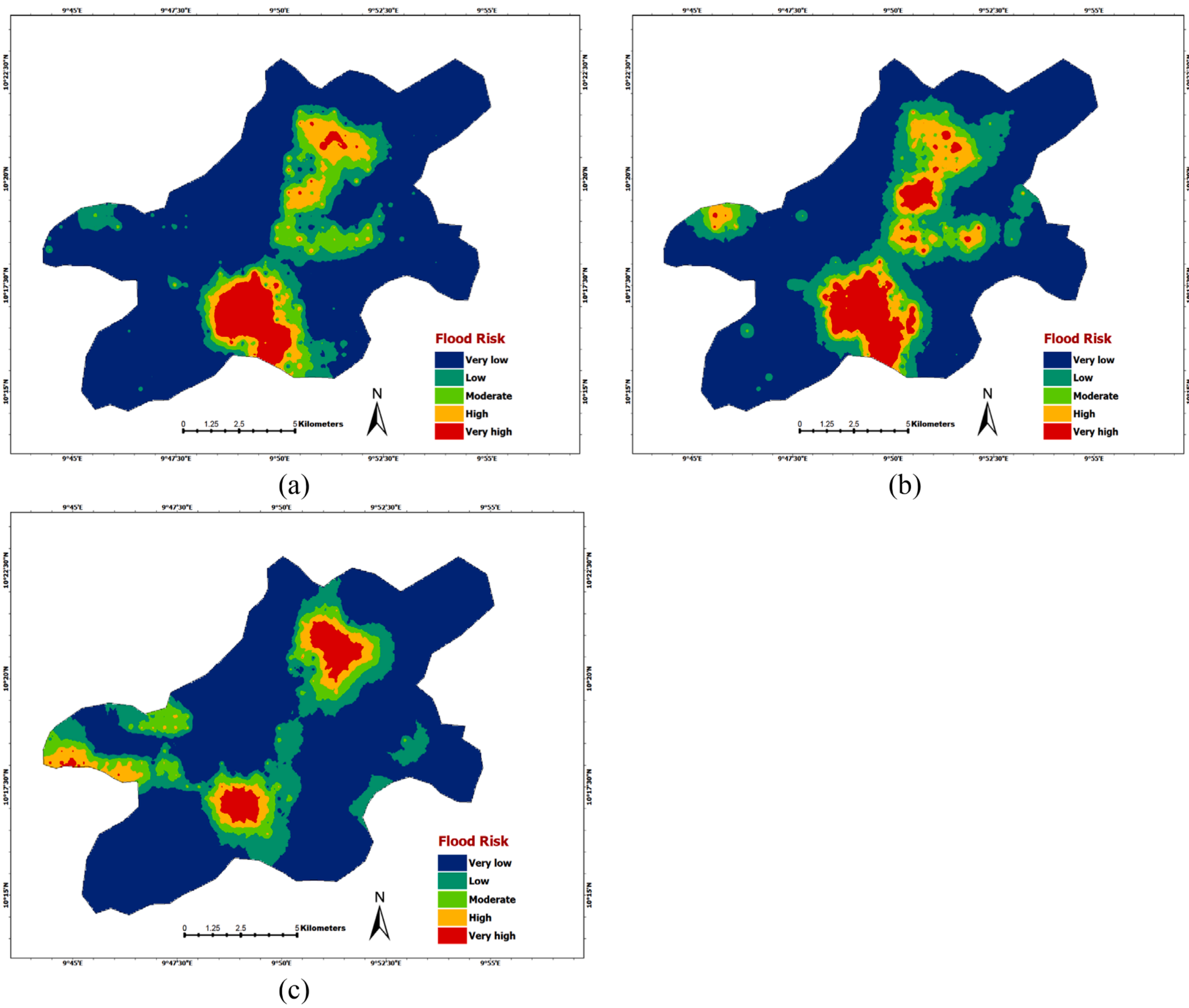

Fig. 8. Flood risk prediction model using (a) Random Forest (b) XGBoost, and (c) Support Vector Machine.

Declaration of competing interest

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

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.joclim.2025.100490.

References

[1] Al-Ruzouq R, Shanableh A, Jena R, Gibril MBA, Hammouri NA, Lamghari F. Flood susceptibility mapping using a novel integration of multi-temporal sentinel-1 data and eXtreme deep learning model. Geosci Front 2024;15:101780. https://doi.org/10.1016/j.gsf.2024.101780.
[2] Avand M, Kuriqi A, Khazaei M, Ghorbanzadeh O. DEM resolution effects on machine learning performance for flood probability mapping. J Hydro-Environ Res 2022;40:1–16. https://doi.org/10.1016/j.jher.2021.10.002.
[3] McDermott TKJ. Global exposure to flood risk and poverty. Nat Commun 2022;13:3529. https://doi.org/10.1038/s41467-022-30725-6.
[4] Ahmed A, Mohamed NS, Siddig EE, Algaily T, Sulaiman S, Ali Y. The impacts of climate change on displaced populations: a call for action. J Clim Change Health 2021;3:100057. https://doi.org/10.1016/j.joclim.2021.100057.
[5] Tate E, Rahman MA, Emrich CT, Sampson CC. Flood exposure and social vulnerability in the United States. Nat Hazards 2021;106:435–57. https://doi.org/10.1007/s11069-020-04470-2.
[6] Vojinovic Z, Alves A, Gómez JP, Weesakul S, Keerakamolchai W, Meesuk V, et al. Effectiveness of small- and large-scale nature-based solutions for flood mitigation: the case of Ayutthaya. Thailand. Sci Total Environ 2021;789:147725. https://doi.org/10.1016/j.scitotenv.2021.147725.
[7] De Silva M, Kawasaki A. Socioeconomic vulnerability to disaster risk: a case study of flood and drought impact in a rural Sri Lankan community. Ecol Econ 2018;152:131–40.
[8] Aerts JCJH, Botzen WJ, Clarke KC, Cutter SL, Hall JW, Merz B, et al. Integrating human behaviour dynamics into flood disaster risk assessment. Nat Clim Change 2018;8:193–9. https://doi.org/10.1038/s41558-018-0085-1.
[9] Dutal H. Determining the effect of urbanization on flood hazard zones in Kahramanmaras, Turkey, using flood hazard index and multi-criteria decision analysis. Environ Monit Assess 2023;195:92. https://doi.org/10.1007/s10661-022-10693-6.
[10] Membele GM, Naidu M, Mutanga O. Examining flood vulnerability mapping approaches in developing countries: a scoping review. Int J Disaster Risk Reduct 2022;69:102766.
[11] Kafi KM, Ponrahono Z. Weather and climate extremes in Nigeria: modeling the perceived impact of spatial planning and community practices on windstorm and flood exposure using PLS-SEM and correlogram. Int J Disaster Risk Reduct 2025;124:105554. https://doi.org/10.1016/j.ijdrr.2025.105554.
[12] Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B, et al. A comparative assessment of flood susceptibility modeling using Multi-Criteria decision-making analysis and machine learning methods. J Hydrol 2019;573:311–23. https://doi.org/10.1016/j.jhydrol.2019.03.073.
[13] Moriyama K, Sasaki D, Ono Y. Comparison of global databases for disaster loss and damage data. J Disaster Res 2018;13:1007–14.
[14] Winsemius HC, Jongman B, Veldkamp TI, Hallegatte S, Bangalore M, Ward PJ. Disaster risk, climate change, and poverty: assessing the global exposure of poor people to floods and droughts. Environ Dev Econ 2018;23:328–48.
[15] Dottori F, Mentaschi L, Bianchi A, Alfieri L, Feyen L. Cost-effective adaptation strategies to rising river flood risk in Europe. Nat Clim Change 2023;13:196–202. https://doi.org/10.1038/s41558-022-01540-0.
[16] Kafi KM, Ponrahono Z, Salisu Barau A. Addressing knowledge gaps on emerging issues in weather and climate extreme events: a systematic review. Clim Change 2024;177:56. https://doi.org/10.1007/s10584-024-03714-5.

[17] Parker ER, Mo J, Goodman RS. The dermatological manifestations of extreme weather events: a comprehensive review of skin disease and vulnerability. J Clim Change Health 2022;8:100162. https://doi.org/10.1016/j.joclim.2022.100162.
[18] Damte E, Manteaw BO. Wrigley-Asante C. Urbanization, climate change and health vulnerabilities in slum communities in Ghana. J Clim Change Health 2023;10: 100189. https://doi.org/10.1016/j.joclim.2022.100189.
[19] Ahmad Bukhari SI, Hassan Rizvi S. Impact of floods on women: with special reference to flooding experience of 2010 flood in Pakistan. J Geogr Nat Disasters 2015;05. https://doi.org/10.4172/2167-0587.1000140.
[20] Aladejana OO, Ebijuoworih EJ. Flood risk assessment in Kogi State Nigeria through the integration of hazard and vulnerability factors. Discov Geosci 2024;2:31. https://doi.org/10.1007/s44288-024-00036-y.
[21] Kemter M, Marwan N, Villarini G, Merz B. Controls on flood trends across the United States. Water Resour Res 2023;59:e2021WR031673. https://doi.org/10.1029/2021WR031673.
[22] Rodríguez-Giralt I, Arenas M, Gómez DL. Children, participation and disasters in Europe: a poor record. Child. young people's particip. Disaster Risk Reduct., Policy Press; 2020. p. 15–36.
[23] Rentschler J, Salhab M, Jafino BA. Flood exposure and poverty in 188 countries. Nat Commun 2022;13:3527. https://doi.org/10.1038/s41467-022-30727-4.
[24] MdA Chowdhury, MdK Hasan, Islam SLU. Climate change adaptation in Bangladesh: current practices, challenges and the way forward. J Clim Change Health 2022;6:100108. https://doi.org/10.1016/j.joclim.2021.100108.
[25] Douglas I. Flooding in African cities, scales of causes, teleconnections, risks, vulnerability and impacts. Int J Disaster Risk Reduct 2017;26:34–42.
[26] Gaisie E, Cobbinah PB. Planning for context-based climate adaptation: flood management inquiry in Accra. Environ Sci Policy 2023;141:97–108. https://doi.org/10.1016/j.envsci.2023.01.002.
[27] Yucel I, Onen A, Yilmaz KK, Gochis DJ. Calibration and evaluation of a flood forecasting system: utility of numerical weather prediction model, data assimilation and satellite-based rainfall. J Hydrol 2015;523:49–66.
[28] Idrees MO, Yusuf A, Mokhtar ES, Yao K. Urban flood susceptibility mapping in Ilorin, Nigeria, using GIS and multi-criteria decision analysis. Model Earth Syst Environ 2022;8:5779–91. https://doi.org/10.1007/s40808-022-01479-3.
[29] Eboreime E, Anjorin O, Obi-Jeff C, Ojo TM, Hertelendy A. From drought to displacement: assessing the impacts of climate change on conflict and forced migration in West Africa's Sahel Region. J Clim Change Health 2025;23:100448. https://doi.org/10.1016/j.joclim.2025.100448.
[30] Umar N, Gray A. Flooding in Nigeria: a review of its occurrence and impacts and approaches to modelling flood data. Int J Environ Stud 2023;80:540–61. https://doi.org/10.1080/00207233.2022.2081471.
[31] Egbueri JC, Ighalo JO, Pande CB. Climate change impacts on nigeria: environment and sustainable development. Cham: Springer International Publishing; 2023. https://doi.org/10.1007/978-3-031-21007-5.
[32] Kafi KM, Barau AS, Aliyu A. The effects of windstorm in African medium-sized cities: an analysis of the degree of damage using KDE hotspots and EF-scale matrix. Int J Disaster Risk Reduct 2021;55:102070.
[33] Adekola O, Lamond J. A media framing analysis of urban flooding in Nigeria: current narratives and implications for policy. Reg Environ Change 2018;18: 1145–59. https://doi.org/10.1007/s10113-017-1253-y.
[34] Barau A, Wada AS. Do-it-yourself flood risk adaptation strategies in the neighborhoods of Kano City. Nigeria. Afr. Handb. Clim. Change Adapt. 2021: 1353–80.
[35] Oe A, Ka I, Ec O. Climate change and coastal floods: the susceptibility of coastal areas of Nigeria. J Coast Zone Manag 2017;20. https://doi.org/10.4172/2473-3350.1000443.
[36] Kafi KM, Ponrahono Z. Advances in weather and climate extreme studies: a systematic comparative review. Discov Geosci 2024;2:66. https://doi.org/10.1007/s44288-024-00079-1.
[37] Pandey M. Artificial intelligence algorithms in flood prediction: a general overviewPandey PC, Kumar R, Pandey M, Giuliani G, Sharma RK, Srivastava PK, editors. Artificial intelligence algorithms in flood prediction: a general overview. Geo-Inf. Disaster Monit. Manag. 2024:243–96. https://doi.org/10.1007/978-3-031-51053-3_12.
[38] Yigitcanlar T, Butler L, Windle E, Desouza KC, Mehmood R, Corchado JM. Can building "artificially intelligent cities" safeguard humanity from natural disasters, pandemics, and other catastrophes? an urban scholar's perspective. Sensors 2020; 20:2988.
[39] Darabi H, Choubin B, Rahmati O, Torabi Haghighi A, Pradhan B, Kløve B. Urban flood risk mapping using the GARP and QUEST models: a comparative study of machine learning techniques. J Hydrol 2019;569:142–54. https://doi.org/10.1016/j.jhydrol.2018.12.002.
[40] Sadiq AA, Amin SA, Ahmad D, Umara BG. Characteristics of irrigation tube wells on major river flood plains in Bauchi State. Nigeria. Ambiente E Agua - Interdiscip J Appl Sci 2014;9:602–9. https://doi.org/10.4136/ambi-agua.1314.
[41] Kafi KM, Shafri HZM, Shariff ABM. An analysis of LULC change detection using remotely sensed data; A Case study of Bauchi City. IOP Conf Ser Earth Environ Sci 2014;20:012056. https://doi.org/10.1088/1755-1315/20/1/012056.
[42] Alkali AN, Yadima SG, Ibrahim UA. Relationship between rainfall and sediment yield in Gubi Dam reservoir, Bauchi State, Nigeria. Int J Eng Res Dev 2016;12: 19517.
[43] Bouramtane T, Kacimi I, Bouramtane K, Aziz M, Abraham S, Omari K, et al. Multivariate analysis and machine learning approach for mapping the variability and vulnerability of urban flooding: the case of Tangier City, Morocco. Hydrology 2021;8:182. https://doi.org/10.3390/hydrology8040182.
[44] Jiang J, Huang Z-G, Grebogi C, Lai Y-C. Predicting extreme events from data using deep machine learning: when and where. Phys Rev Res 2022;4:023028.
[45] Costache R, Arabameri A, Moayedi H, Pham QB, Santosh M, Nguyen H, et al. Flash-flood potential index estimation using fuzzy logic combined with deep learning neural network, naïve Bayes, XGBoost and classification and regression tree. Geocarto Int 2022;37:6780–807. https://doi.org/10.1080/10106049.2021.1948109.
[46] Almansi KY, Ujang U, Azri S, Wickramathilaka N. Traffic noise prediction model using GIS and ensemble machine learning: a case study at Universiti Teknologi Malaysia (UTM) Campus. Environ Sci Pollut Res 2024. https://doi.org/10.1007/s11356-024-35243-0.
[47] Kafi KM, Gibril MBA. GPS application in disaster management: a review. Asian J Appl Sci 2016;4.
[48] Tempa K, Yuden K. Multi-hazard zoning for national scale population risk mapping: a pilot study in Bhutan Himalaya. Geoenvironmental Disasters 2023;10: 7. https://doi.org/10.1186/s40677-023-00239-4.
[49] Chukwu M, Huang X, Peng B, Oloruntimilehin I. Flood risk evaluation using AHP-based model and GIS technique: a case study of Ethiope East and West local government areas. Nigeria. J Indian Soc Remote Sens 2023;51:2561–76. https://doi.org/10.1007/s12524-023-01786-x.
[50] Gautam D, Adhikari R, Jha P, Rupakhety R, Yadav M. Windstorm vulnerability of residential buildings and infrastructures in south-central Nepal. J Wind Eng Ind Aerodyn 2020;198:104113. https://doi.org/10.1016/j.jweia.2020.104113.
[51] Sachit MS, Shafri HZM, Abdullah AF, Rafie ASM, Gibril MBA. Global spatial suitability mapping of wind and solar systems using an explainable AI-based approach. ISPRS Int J Geo-Inf 2022;11:422. https://doi.org/10.3390/ijgi11080422.
[52] Khosravi K, Pham BT, Chapi K, Shirzadi A, Shahabi H, Revhaug I, et al. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. Sci Total Environ 2018;627:744–55. https://doi.org/10.1016/j.scitotenv.2018.01.266.
[53] Bochenek B., Ustrnul Z. Machine learning in weather prediction and climate analyses—Applications and perspectives 2022.
[54] Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Sci Total Environ 2019;651:2087–96.
[55] Li B, Yang G, Wan R, Dai X, Zhang Y. Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China. Hydrol Res 2016;47:69–83.
[56] Moon S-H, Kim Y-H, Lee YH, Moon B-R. Application of machine learning to an early warning system for very short-term heavy rainfall. J Hydrol 2019;568: 1042–54.
[57] Jiang Z, Yang S, Liu Z, Xu Y, Xiong Y, Qi S, et al. Coupling machine learning and weather forecast to predict farmland flood disaster: a case study in Yangtze River basin. Environ Model Softw 2022;155:105436. https://doi.org/10.1016/j.envsoft.2022.105436.
[58] Zhou Y, Wu W, Nathan R, Wang QJ. A rapid flood inundation modelling framework using deep learning with spatial reduction and reconstruction. Environ Model Softw 2021;143:105112. https://doi.org/10.1016/j.envsoft.2021.105112.
[59] HAH Al-Najjar, Pradhan B, He X, Sheng D, Alamri A, Gite S, et al. Integrating physical and machine learning models for enhanced landslide prediction in data-scarce environments. Earth Syst Environ 2024. https://doi.org/10.1007/s41748-024-00508-8.
[60] Hong H, Tsangaratos P, Ilia I, Liu J, Zhu A-X, Chen W. Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. Sci Total Environ 2018;625:575–88. https://doi.org/10.1016/j.scitotenv.2017.12.256.
[61] Gaitán CF. Machine learning applications for agricultural impacts under extreme events. Clim. Extrem. Their Implic. Impact Risk Assess. 2020:119–38.
[62] Ali H, Modi P, Mishra V. Increased flood risk in Indian sub-continent under the warming climate. Weather Clim Extrem 2019;25:100212.
[63] Birkmann J, Jamshed A, McMillan JM, Feldmeyer D, Totin E, Solecki W, et al. Understanding human vulnerability to climate change: a global perspective on index validation for adaptation planning. Sci Total Environ 2022;803:150065.