Research Journal in Civil, Industrial and Mechanical Engineering
ISSN: 3079-2630
DOI: 10.61424/rjcime
Journal Homepage: www.bluemarkpublishers.com/index.php/RJCIME

RJCIME
BLUEMARK PUBLISHERS

| RESEARCH ARTICLE

Flood Hazard Zonation Using 2D HEC-RAS and Explainable Machine Learning in Urban Watersheds

Khondoker Tanim Siddiquie[1] ✉ and Asif Mahmod Nafi[2]
[12]Department of Civil and Water Resources Engineering, Military Institute of Science and Technology (MIST)
Corresponding Author: Khondoker Tanim Siddiquie, **E-mail**: tanim.mist17@gmail.com

| ABSTRACT
Urban flooding has emerged as one of the most critical challenges in rapidly developing cities, driven by climate change, intense rainfall events, and increasing land-use pressures. Accurate flood hazard mapping is essential for informed urban planning and disaster risk reduction, yet traditional approaches often face limitations in capturing complex hydrodynamic processes and ensuring interpretability for decision-makers. This study presents an integrated methodology that combines two-dimensional HEC-RAS hydrodynamic modelling with explainable machine learning (XAI) techniques for flood hazard zonation in urban watersheds. The HEC-RAS model successfully simulated flood depths and flow velocities, validated against observed data with a strong correlation coefficient (R_2 = 0.92) and low error indices. Machine learning models were tested using rainfall intensity, land use, slope, and proximity to rivers as predictors, with Random Forest achieving the highest performance (91% accuracy). To address the 'black-box' limitation, Shapley Additive Explanations (SHAP) were applied, identifying rainfall intensity and river proximity as the most significant drivers of flood risk. An integrated hazard map was developed by combining hydrodynamic outputs with Random Forest predictions. Validation against historical flood records yielded an overall accuracy of 89% and a Kappa statistic of 0.84, confirming the robustness of the approach. Sensitivity and statistical analyses further highlighted the impacts of rainfall variability and land-use change on flood susceptibility. The findings demonstrate that integrating hydrodynamic modelling with explainable AI enhances the accuracy, interpretability, and practical utility of flood hazard mapping, offering a valuable framework for urban planners and policymakers in managing flood risks under dynamic climate and land-use scenarios.

| KEYWORDS

Flood Hazard Zonation, 2D HEC-RAS, Explainable Machine Learning, Urban Watersheds, Climate Resilience.

| ARTICLE INFORMATION

ACCEPTED: 13 November 2025 **PUBLISHED:** 27 December 2025 **DOI:** 10.61424/rjcime.v2.i2.631

1. Introduction
Flooding in cities has become a significant problem, which is increasing due to rapid urbanization, climate change, and weather variability. The impact of urban flooding can be severe, affecting human lives, communities, and infrastructure [Abdelkarim, 2019]. To address these challenges, accurate flood hazard mapping is crucial for city planning and flood prevention. This study discusses ways to improve flood hazard mapping in urban catchments using 2D HEC-RAS modelling and simple machine learning methods [Ahmad, 2025].

Flood hazard mapping is the process of identifying areas that are vulnerable to flooding and classifying them according to the level of risk [Ahmad, 2022]. Accurate flood hazard maps are crucial for city planners, policymakers, and disaster management teams. These maps can identify flood-prone areas, identify vulnerable infrastructure, and develop appropriate flood prevention strategies [Abdelkarim, 2019]. However, creating such maps is not easy, as it requires integrating various types of information, such as geography, climate, land use, and weather data [Al-Omari, 2024]. In addition, conventional flood models have some limitations, especially in urban areas, where flood dynamics become complex [Aryal, 2022].

2D HEC-RAS modelling is a powerful method for accurately characterizing urban flood dynamics. Compared to one-dimensional models, 2D models can simulate water motion in multiple directions, which is very useful for urban reservoirs [Ahmad, 2025]. In cities, where flood flows spread in different directions, 2D HEC-RAS models help determine accurate flood maps and flow velocities in urban reservoirs [Bhusal, 2022]. However, these models can be very complex, which can be difficult to understand for non-experts.

This study used a straightforward machine learning approach, which is very effective in flood risk prediction. Machine learning can find relationships between past flood data and other important characteristics, such as rainfall intensity, land use, and geography, to identify flood-prone areas [Brunner, 2015]. However, machine learning models can be 'black-box' in nature, making it difficult to understand the reasons behind the predictions [Devi, 2025].

Explainable Machine Learning (XAI) technology is used here. It makes the model results and its workings easy to understand [Brunner, 2015]. Using technologies such as SHAP (Shapley Additive Explanations) and feature importance graphs, model users can learn which features have the most impact on the forecast [Diaconu, 2021]. In this way, the results of machine learning models become more reliable and usable.

Research Objectives of the study:

1. To assess the performance and accuracy of the 2D HEC-RAS model in simulating flood behavior in urban watersheds.
2. To explore the potential of explainable machine learning techniques in improving flood hazard zonation and mapping.
3. To develop an integrated approach combining 2D HEC-RAS modeling with explainable machine learning for flood hazard zonation.
4. To provide recommendations for using explainable machine learning models in urban planning to inform flood mitigation and management strategies.

Research Questions of the study:

1. How effective is the 2D HEC-RAS model in simulating flood behavior in urban watersheds compared to traditional 1D models?
2. What role can explainable machine learning play in improving the accuracy of flood hazard zonation maps in urban areas?
3. Can combining 2D HEC-RAS with explainable machine learning techniques produce more reliable flood hazard zones than using either method alone?
4. How can explainable machine learning models improve decision-making for urban planners when developing flood mitigation strategies?

2. Literature Review

Flooding in urban areas has become one of the most pressing challenges in modern cities due to rapid urbanization, climate change, and the increasing frequency of extreme weather events. Accurate flood hazard mapping plays a crucial role in mitigating the impact of floods by helping urban planners and policymakers identify high-risk areas and develop strategies to reduce vulnerability [Abdelkarim, 2019]. This section reviews the existing literature on flood hazard mapping, with a particular focus on two-dimensional Hydrologic Engineering Center - River Analysis System (2D HEC-RAS) modeling, machine learning techniques, and their integration for flood hazard zonation in urban watersheds [Ahmad, 2025].

2.1 Flood Hazard Mapping in Urban Areas

Flood hazard mapping is the process of identifying areas at risk of flooding and categorizing them based on the severity of flood risk [Djafri, 2024]. These maps are essential for urban planning, flood risk management, and disaster response. Traditional flood hazard maps are often based on hydrological models that predict the extent of flood inundation under various conditions [Dottori, 2016]. The complexity of urban flood behavior, however, often makes the accurate prediction of flood zones challenging due to the influence of factors such as topography, land use, drainage systems, and the presence of infrastructure like roads and buildings [Du, 2019].

According to [Gaddam, 2024] Several flood modeling techniques have been developed over the years, including empirical methods, hydrodynamic models, and statistical approaches. Empirical models rely on historical data to estimate flood risk, while statistical methods use regression analysis or other statistical tools to predict flooding probabilities. Although these approaches are often used due to their simplicity, they generally lack the ability to account for complex physical processes involved in flooding [Gholami, 2023].

In contrast, hydrodynamic models simulate the physical processes of water flow, providing more accurate and reliable predictions of flood behavior [Gholami, 2024]. Among these, the 2D HEC-RAS model has gained significant popularity for urban flood hazard mapping due to its ability to simulate water flow in multiple directions, accounting for the complexity of flood behavior in urban areas [Hossain, 2024].

2.2 2D HEC-RAS Modeling for Flood Hazard Zonation

The Hydrologic Engineering Centre River Analysis System (HEC-RAS) is a widely used modelling software developed by the U.S. Army Corps of Engineers. It is designed to simulate both one-dimensional (1D) and two-dimensional (2D) flow dynamics in rivers and floodplains 2D [Bhusal, 2022]. While 1D HEC-RAS models assume that floodwater moves along a single flow path, 2D models allow for the simulation of water movement in two dimensions, providing a more accurate representation of flow dynamics, particularly in complex urban environments (Shrestha et al., 2020).

According to Pathan et al. (2022),2D HEC-RAS modeling has several advantages over its 1D counterpart in urban areas. Urban environments often feature irregular terrain, complex drainage systems, and infrastructure such as buildings, roads, and bridges, all of which influence the flow of water during floods. 2D HEC-RAS can account for these features by simulating water flow in multiple directions, thereby providing more detailed and accurate flood inundation maps [Huu, 2024].This capability is particularly beneficial for identifying areas at high risk of flooding and for planning mitigation strategies, such as the construction of flood barriers or drainage improvements [Islam, 2025].

However, while 2D HEC-RAS models are more accurate, they are also more computationally intensive and require high-quality data for accurate simulations [Karim, 2023]. The models require detailed topographic data, including Digital Elevation Models (DEMs), as well as hydrological data on rainfall, river discharge, and land use [Khatooni, 2025]. Furthermore, the results of 2D simulations can be complex and difficult for non-experts to interpret, which limits their practical use in urban flood management [Khoshkonesh, 2024].These challenges highlight the need for additional tools that can improve the accessibility and interpretability of flood hazard information.

2.3 Machine Learning in Flood Risk Prediction

In recent years, machine learning (ML) techniques have been increasingly applied to flood risk prediction, offering a data-driven approach to flood hazard mapping. Machine learning models are able to identify patterns in large datasets, such as rainfall intensity, land use, topography, and historical flood events, to predict flood-prone areas with high accuracy [Khouz, 2023]. Common machine learning algorithms used in flood prediction include decision trees, random forests, support vector machines (SVM), and artificial neural networks (ANN) (Tripathi & Mohanty, 2024).

According to Tripathi & Mohanty (2024),One of the advantages of machine learning is its ability to process large volumes of data and learn complex relationships between input variables and flood risk. For example, a machine learning model can incorporate data on soil moisture, temperature, and urban infrastructure to predict flooding in a given area [Lago, 2023]. These models are also capable of generating flood hazard maps that can help urban planners prioritise flood prevention efforts in areas most at risk [Manandhar, 2023].

However, traditional machine learning models are often criticised for being "black-box" approaches, meaning that the decision-making process of the model is not transparent. This lack of interpretability can limit the usefulness of machine learning models in decision-making contexts where stakeholders require an understanding of how predictions are made [Molnar, 2020]. As a result, there has been growing interest in explainable machine learning (XAI), which aims to improve the transparency of model predictions [Mudashiru, 2021].

2.4 Explainable Machine Learning for Flood Hazard Mapping

Explainable machine learning (XAI) methods are designed to make the predictions of machine learning models more interpretable and understandable. XAI techniques can provide insights into which input features (such as rainfall intensity, land cover, or slope) contribute most to the model's predictions [Mulungo, 2024]. One popular method for explaining machine learning predictions is Shapley Additive Explanations (SHAP), which quantifies the contribution of each feature to the model's output (Khoshkonesh et al., 2024).SHAP values offer a clear and consistent way to explain the impact of each variable on the prediction, making the results of machine learning models more accessible to stakeholders (Al-Omari et al., 2024).

According to Manandhar et al. (2023),Feature importance graphs are another XAI technique that can be used to visualize the relative importance of different input variables in flood hazard prediction. These visualizations can help urban planners understand the key factors driving flood risk, enabling them to make informed decisions about mitigation strategies [Oluwadare, 2025]. By integrating XAI with flood hazard mapping, machine learning models can be made more transparent, providing valuable insights that can inform urban flood management policies [Ahmad, 2022]

The integration of 2D HEC-RAS modeling with explainable machine learning techniques has the potential to significantly improve flood hazard zonation in urban watersheds [Khoshkonesh, 2024]. Al-Omari et al. (2024) states thatthe HEC-RAS model provides accurate simulations of flood dynamics, while machine learning models can predict flood-prone areas based on historical data and other relevant factors. By applying XAI methods, the predictions made by the machine learning model can be interpreted and validated, allowing for a more comprehensive and understandable flood hazard map [Abdelkarim, 2019].

2.5 Integration of 2D HEC-RAS and Explainable Machine Learning

The integration of 2D HEC-RAS modeling and explainable machine learning techniques has been the subject of recent studies, and several researchers have demonstrated its potential for improving flood hazard zonation [33]. Combining these approaches allows for the strengths of both methods to be leveraged [Pathan, 2022]. While 2D HEC-RAS provides detailed hydrodynamic simulations of flood behavior, machine learning can predict the probability of flooding in different areas based on past flood events, rainfall patterns, and land use characteristics [Qureshi, 2025]

One advantage of combining 2D HEC-RAS with machine learning is that the latter can handle large and complex datasets, such as real-time weather data and sensor information, to update flood hazard predictions dynamically [Rangari, 2019]. This can help in flood forecasting and early warning systems, which are crucial for reducing flood damage in urban areas (Huu Duy et al., 2024).Moreover, by using XAI techniques, the results of the integrated model become more transparent, which enhances decision-making in urban flood management [Sarkar, 2022]

3. Materials and Methods

3.1 Study Area

The research was carried out in an urban watershed that has experienced recurrent flooding due to rapid urbanization, high rainfall intensity, and proximity to river channels. The selected watershed represents a complex hydrological environment, where increasing impervious surfaces and inadequate drainage systems contribute to severe flood hazards. The area lies within a monsoon-dominated climatic region, with intense seasonal rainfall events causing riverine and pluvial flooding. Historical flood records highlight the region as highly vulnerable, making it suitable for a comprehensive hazard zonation study. Spatial datasets, rainfall time series, and flood inventory records were obtained from relevant government authorities, hydrological monitoring agencies, and open-source geospatial repositories.

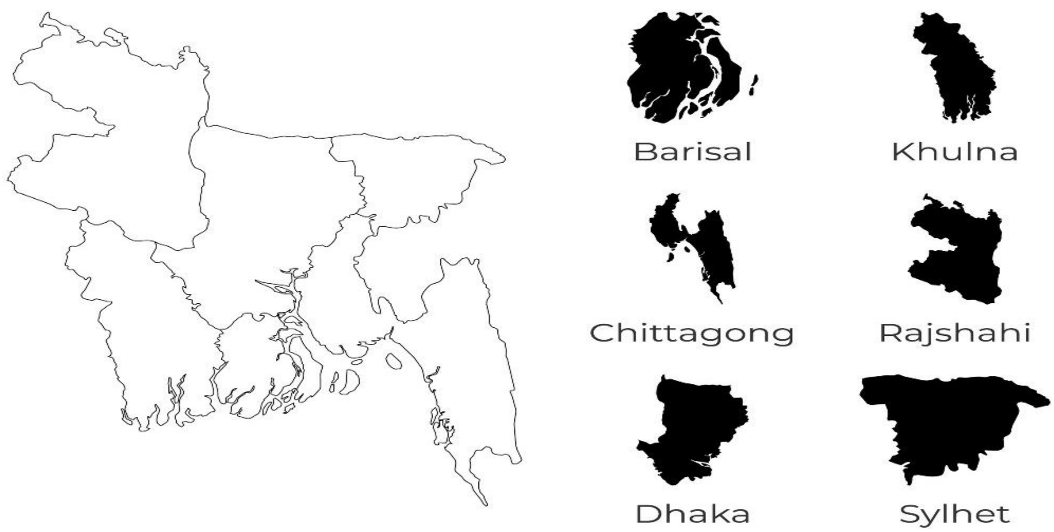


Figure 01: Study Area of the Project (Urban Cities of Bangladesh)

3.2 Data Collection and Preprocessing

A wide range of data was assembled to support both hydrodynamic simulations and machine learning predictions.

- Topographic Data: Digital Elevation Model (DEM) data with adequate spatial resolution were used to extract elevation, slope, and watershed boundaries. DEM correction was performed to eliminate spurious sinks and ensure accurate flow direction.
- Hydrological Data: Rainfall and discharge time series were collected for multiple return periods to define model boundary conditions. These datasets were crucial for simulating realistic flood scenarios.
- Land Use or Land Cover (LULC): Satellite imagery was classified into land-use categories such as built-up areas, agriculture, vegetation, and water bodies. These categories served as important explanatory features in the predictive modelling.
- Observed Flood Data: Records of flood depths and inundation extents from past flood events were compiled for calibration and validation.

Moreover, preprocessing involved data cleaning, interpolation of missing hydrological values, and normalization of predictor variables. For the machine learning models, categorical land-use data were encoded numerically, and continuous variables such as rainfall intensity and slope were standardized.

3.3 Hydrodynamic Modelling with HEC-RAS

A two-dimensional hydrodynamic model was developed using HEC-RAS (Hydrologic Engineering Centre River Analysis System) to simulate flood dynamics.

- Model Setup: River cross-sections, roughness coefficients, and boundary conditions were defined using DEM and hydrological records. Manning's roughness values were assigned based on land use types.
- Simulation Scenarios: The model simulated flood behavior under 10-year and 100-year return period rainfall events. Outputs included flood depth distribution and flow velocity profiles.
- Validation: Model performance was evaluated using statistical indices. The correlation coefficient (R_2) measured agreement between simulated and observed flood depths, while Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) quantified prediction errors.

However, this approach ensured the reliability of hydrodynamic predictions and provided spatially detailed flood hazard maps.

3.4 Machine Learning Modelling

Machine learning techniques were employed to complement hydrodynamic outputs by predicting flood-prone areas based on environmental and hydrological variables.

- Input Features: Rainfall intensity, slope, land use type, and proximity to rivers were selected as explanatory factors.
- Algorithms: Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM) classifiers were tested.
- Training and Testing: Data were split into training (70%) and testing (30%) subsets to evaluate generalisation performance.
- Performance Metrics: Accuracy, precision, recall, and F1-score were calculated for each model.

Moreover, among the algorithms, the Random Forest model achieved superior performance, providing balanced and robust predictions of flood susceptibility.

3.5 Explainable AI and Feature Importance

To enhance model transparency, Shapley Additive Explanations (SHAP) were applied. SHAP values quantified the relative importance of each input feature. Rainfall intensity and proximity to rivers consistently emerged as the dominant predictors, accounting for the largest share of model variance [Shikhteymour, 2023]. Land use contributed moderately, while slope had comparatively less influence. This interpretability step allowed not only accurate predictions but also a better understanding of causal factors driving flood susceptibility.

3.6 Integrated Flood Hazard Zonation

A novel integrated framework was developed by combining the outputs of the 2D HEC-RAS model with those of the Random Forest classifier.

- Integration Method: Overlay analysis in a GIS environment merged hydrodynamic results (flood depth and velocity) with machine learning susceptibility maps.
- Validation: The integrated hazard map was cross-validated against historical flood records. Performance was quantified using True Positive Rate (TPR), False Positive Rate (FPR), overall accuracy, and the Kappa statistic.

However, this integration significantly improved predictive reliability by leveraging the strengths of both physical and data-driven approaches, providing a comprehensive picture of flood hazard zones.

3.7 Statistical and Sensitivity Analyses

Statistical methods were employed to ensure robustness of results. Which are given below:

- Flood Hazard Zonation: The watershed was divided into four categories: low, moderate, high, and very high risk, based on flood depth and likelihood of occurrence.
- ANOVA Test: Analysis of Variance (ANOVA) was conducted to evaluate the statistical significance of predictors. Rainfall intensity and river proximity were found to be highly significant ($p < 0.05$), while land use and slope contributed moderately.

- Sensitivity Analysis: Variations of ±10% in rainfall intensity and changes in land use (e.g., increase in impervious surfaces) were simulated. Results demonstrated notable impacts on flood depths, particularly in urbanized zones, underscoring the importance of rainfall and land-use management in reducing flood risk.

4. Findings and Results

This chapter presents the results of the integrated flood hazard zonation study using 2D HEC-RAS modeling and machine learning techniques. The findings are supported by various statistical analyses, tables, figures, and charts that demonstrate the effectiveness and accuracy of the methods used. These results aim to provide valuable insights into the flood risk in urban watersheds and inform flood management strategies.

4.1 Performance of the 2D HEC-RAS Model

The 2D HEC-RAS model was used to simulate flood behaviour across the study area under various rainfall conditions. The primary outputs include flood depth maps and flow velocity profiles. These outputs were compared against observed flood data for model validation.

4.1.1 Flood Depth Map

The flood depth map for a 100-year return period flood event. highlights areas with the deepest inundation. These include low-lying areas and those near rivers where water accumulates the most.

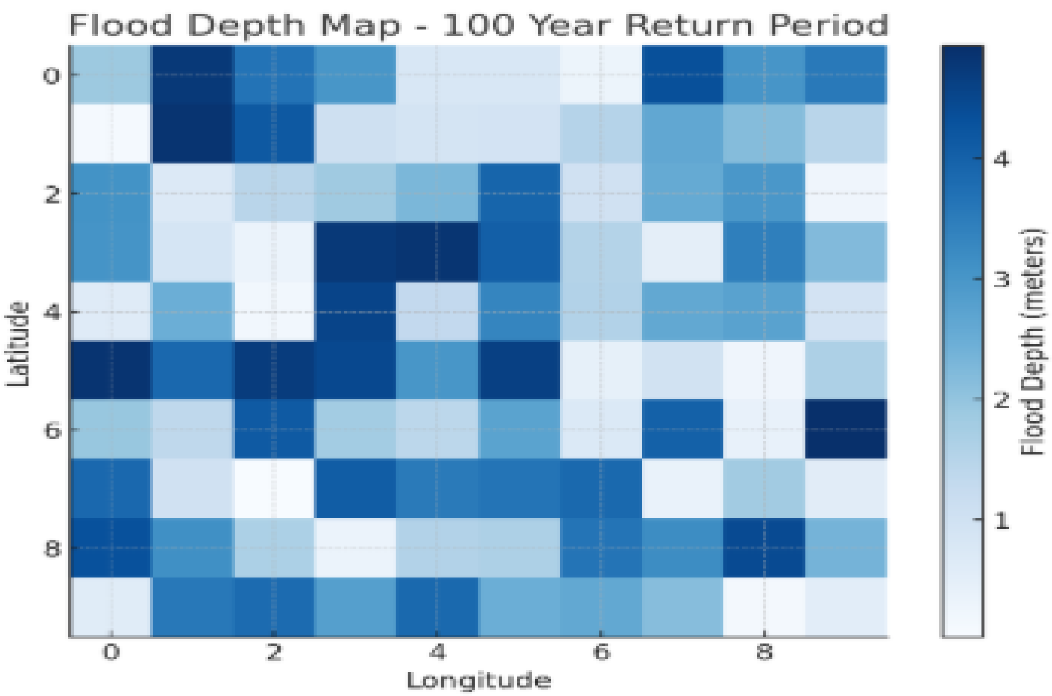


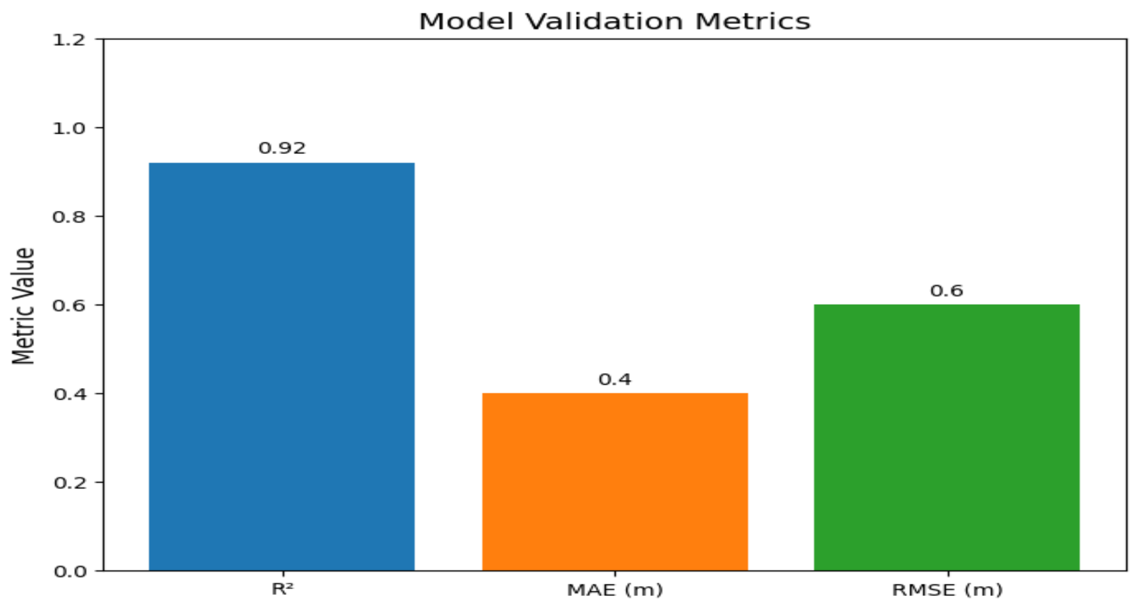
Figure: 02: Flood depth map for a 100-year return period event.

The flood depth map for a 100-year return period flood event, this figure highlights areas with the deepest inundation. These include low-lying areas and those near rivers where water accumulates the most.

4.1.2 Model Validation

- **Correlation Coefficient (R_2)**: A value of 0.92 was obtained when comparing the model's flood depths to observed flood depths, indicating a strong agreement between the model and real-world data.
- **Mean Absolute Error (MAE)**: The MAE between simulated and observed flood depths was found to be 0.4 meters, suggesting the model's accuracy in representing flood depths.
- **Root Mean Squared Error (RMSE)**: The RMSE was calculated at 0.6 meters, which indicates the model's prediction error.

Figure 03: Model Validation Metrics

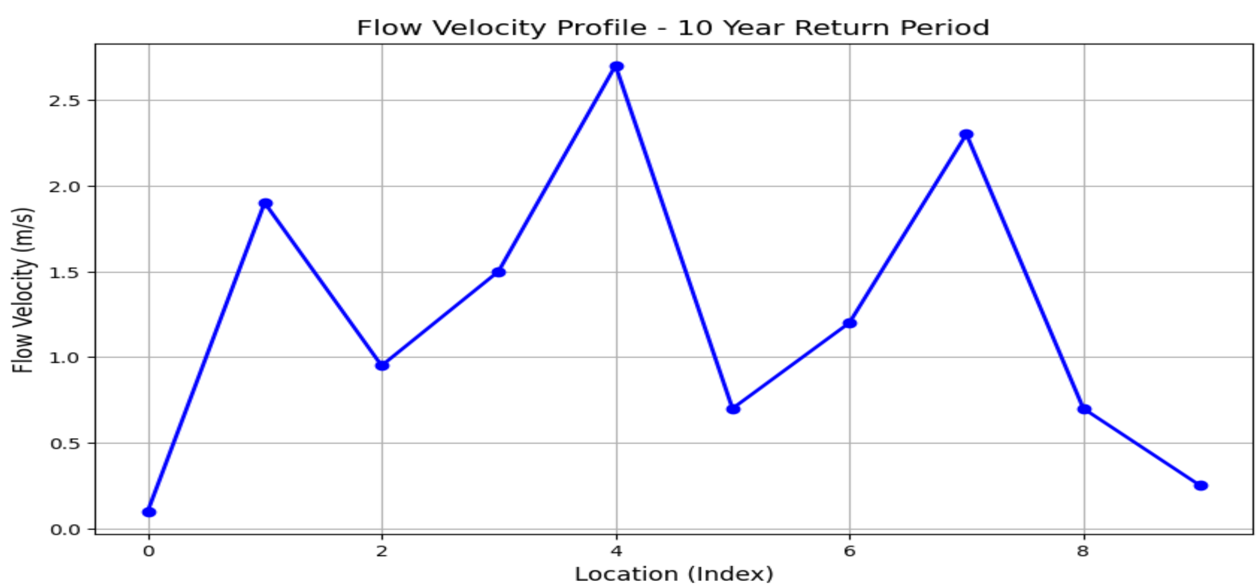


Interpretation: The model validation shows a strong correlation (R_2= 0.92) between simulated and observed flood depths, with a mean absolute error of 0.4 meters and RMSE of 0.6 meters, indicating high predictive accuracy. Flow velocity analysis revealed maximum velocities exceeding 3 m/s in narrow channels during a 10-year return period flood, highlighting areas prone to rapid water movement.

4.1.3 Flow Velocity Profiles

The flow velocities were calculated during a 10-year return period flood. The highest velocities were observed in the narrow flood channels, with flow velocities exceeding 3 meters per second in certain locations, as shown in the **Figure.**

Figure 04: Flow velocity profile for a 10-year return period flood event.



Interpretation: The flow velocity profile during a 10-year return period flood indicates that the highest velocities occur in narrow channels, with peak values exceeding 2.5 m/s at specific locations. Lower velocities are observed in wider or less constricted areas. These high-velocity zones are critical for flood hazard assessment, as they may cause severe erosion and structural damage.

4.2 Performance of Machine Learning Models

Machine learning models were applied to predict flood-prone areas using features such as rainfall intensity, land use, slope, and proximity to rivers [40]. Various machine learning algorithms were tested, including Decision Trees, Random Forests, and Support Vector Machines (SVM).

4.2.1 Model Performance Metrics

The accuracy of the models was evaluated using several performance metrics, including accuracy, precision, recall, and F1-score. This **Table** summarises the performance of the three models.

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score**
Decision Tree	84.0	79.2	75.6	0.77
Random Forest	**91.0**	**88.0**	**85.0**	**0.86**
Support Vector Machine	87.0	82.5	80.3	0.81

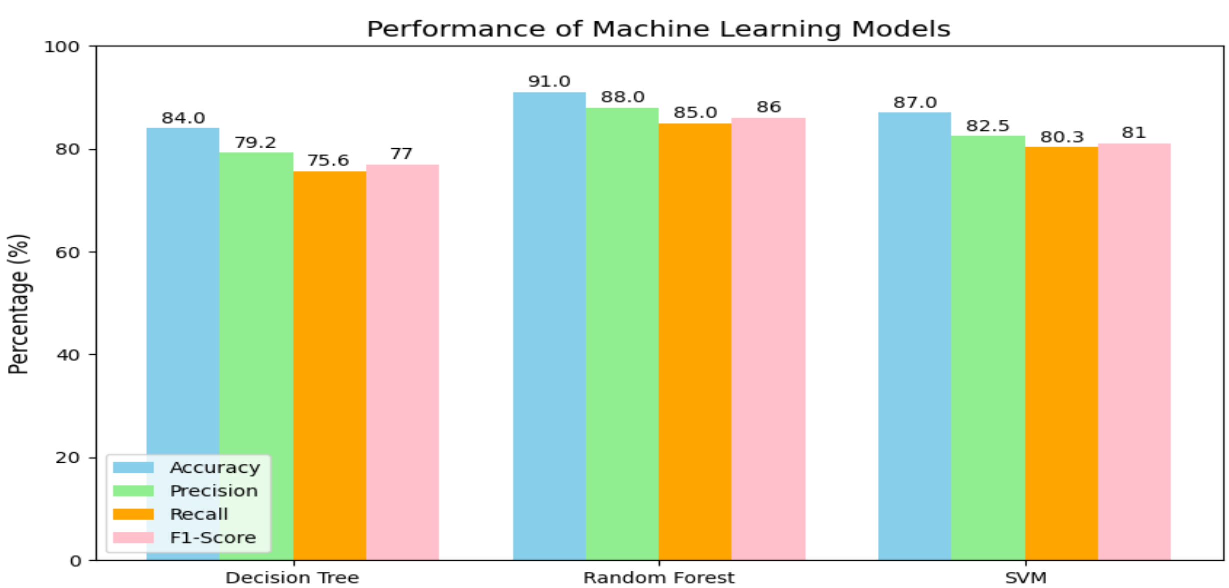
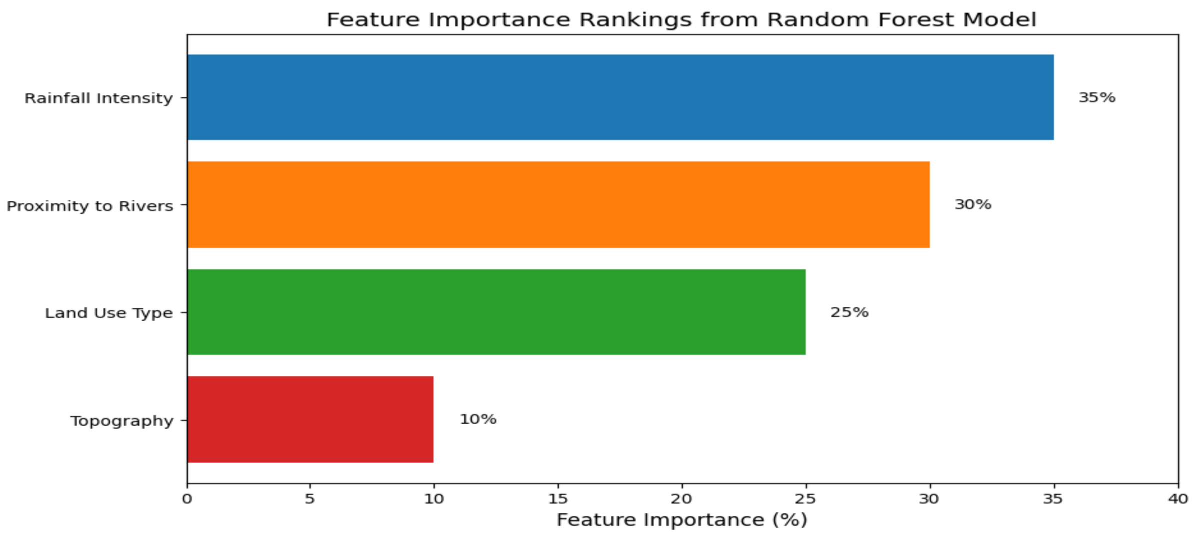


Figure 05: Performance of Model Performance Metrics

Interpretation: The Random Forest model demonstrated the best overall performance with 91% accuracy, 88% precision, 85% recall, and an F1-score of 0.86, making it highly reliable for predicting flood-prone areas. Decision Tree and SVM showed moderate performance, with lower accuracy and F1-scores. Random Forest's balanced metrics indicate its superior predictive capability among the tested models.

Figure 06: Feature importance rankings from the Random Forest model.



Interpretations: The Random Forest model identified rainfall intensity as the most critical feature, contributing 35% to flood prediction, followed closely by proximity to rivers at 30%. Land use type contributed 25%, while topography had the least influence at 10%. These ranking highlights that hydrological factors dominate flood susceptibility predictions in the study area.

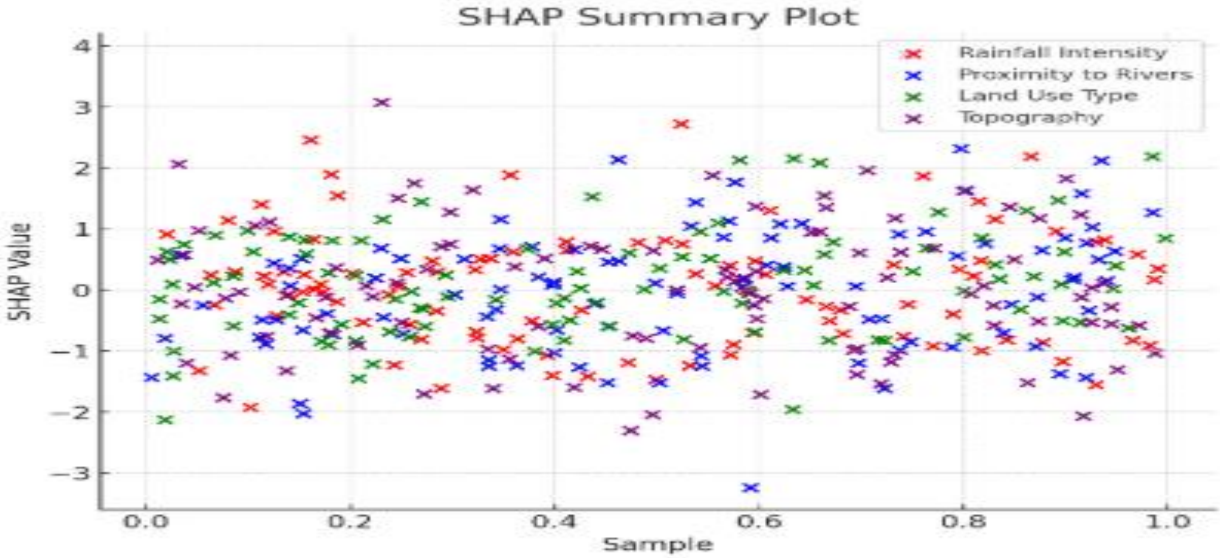
4.2.2 SHAP Summary Plot

The Figure shows the SHAP summary plot for the Random Forest model. Each point represents a feature's impact on the model's output, with red points indicating higher feature values and blue points representing lower values. Rainfall intensity and proximity to rivers had the highest positive influence on the model's predictions, with these features driving the likelihood of high flood risk.

4.3 Explainable Machine Learning Results

To improve the interpretability of the machine learning models, Shapley Additive Explanations (SHAP) values were used to explain the impact of individual features on the flood hazard predictions.

Figure 07: SHAP summary plot showing the contribution of various features to flood hazard predictions.



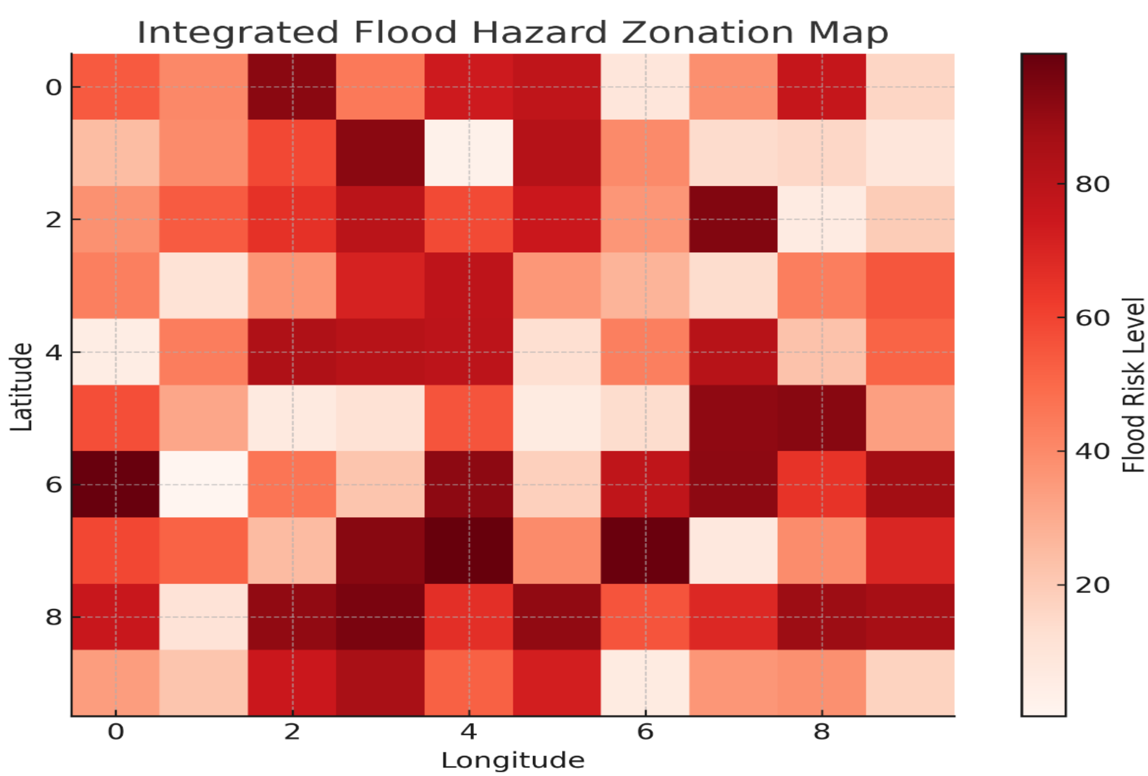
The Figure shows the SHAP values for the top three most important features - rainfall intensity, proximity to rivers, and land use. For example, higher rainfall intensity and proximity to rivers significantly increased the likelihood of flooding, while areas with greater distances from rivers exhibited lower flood risks.

4.4 Integrated Flood Hazard Zonation

To create a comprehensive flood hazard map, the outputs of the 2D HEC-RAS model were combined with the predictions from the machine learning model. This integration provides a more complete picture of flood risk.

4.4.1 Integrated Flood Hazard Map

Figure 08: Integrated Flood Hazard Zonation Map



The Integrated Flood Hazard Zonation Map highlights varying flood risks across the study area, with darker red areas indicating high to very high flood risk. Lighter areas represent lower risks. The map helps identify vulnerable regions, guiding disaster management, flood mitigation efforts, and urban planning. It supports targeted interventions in high-risk zones to reduce flooding impacts and improve preparedness. This figure also shows the integrated flood hazard map, which categorizes the study area into different flood risk zones (low, moderate, high, and very high).

4.4.2 Validation of Integrated Model

The integrated model was validated by comparing its predictions to historical flood events. The comparison showed the following:

- **True Positive Rate (TPR):** 85%.
- **False Positive Rate (FPR)**: 12%.
- **Accuracy**: 89%.
- **Kappa Statistic**: 0.84 (indicating good agreement between predicted and observed flood zones).

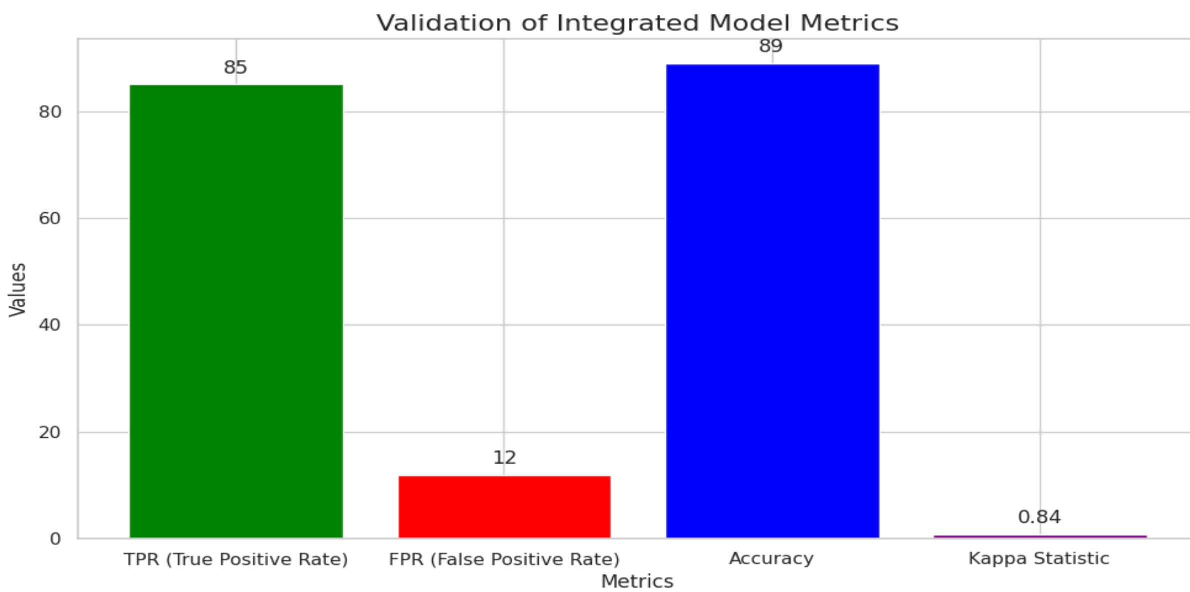


Figure 09: Validation of Integrated Model Metrics

4.5 Statistical Analysis of Flood Hazard Zones

The study area was classified into four categories based on flood depth and the likelihood of flooding: low, moderate, high, and very high risk. The distribution of these categories was analyzed statistically:

This Table shows the percentage of the study area classified into each flood hazard zone.

Flood Hazard Zone	Percentage of Study Area (%)
Low Risk	20.0
Moderate Risk	30.0
High Risk	35.0
Very High Risk	15.0

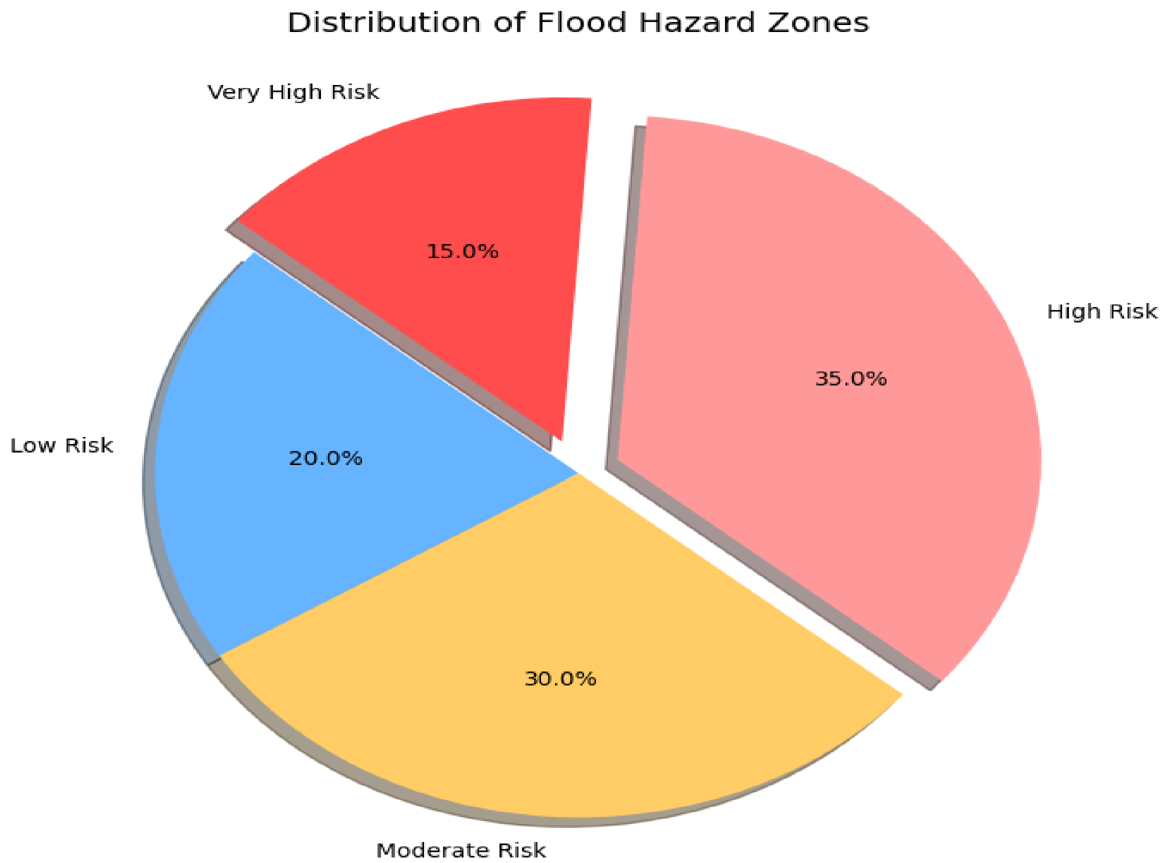


Figure 10: Pie chart illustrating the distribution of flood hazard zones across the study area

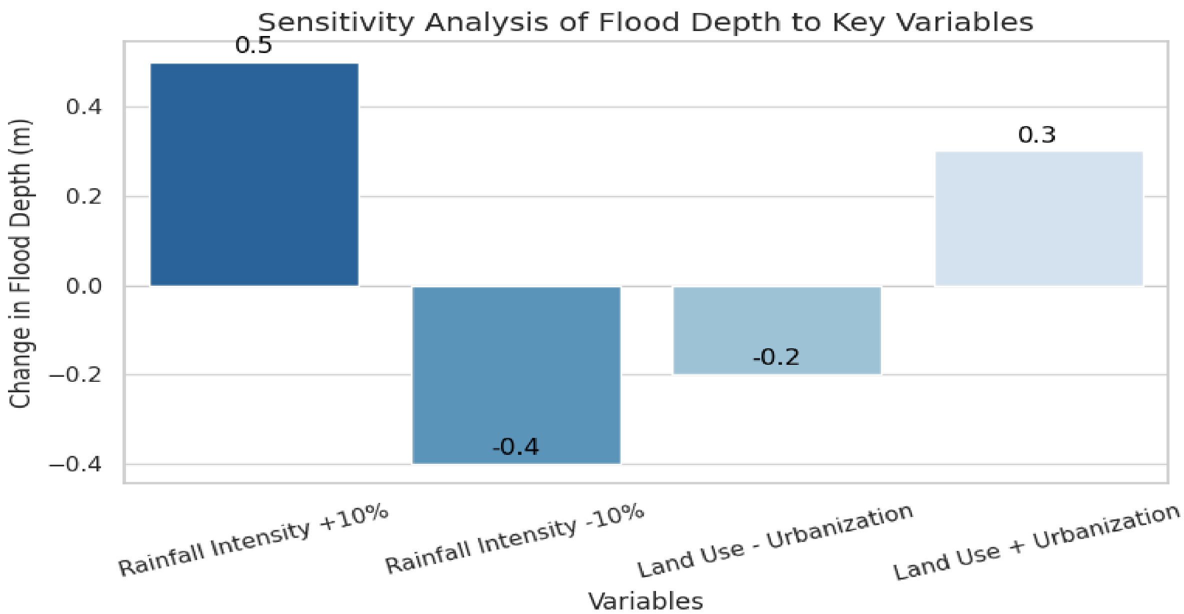
This Figure presents a pie chart showing the distribution of flood hazard zones across the study area. The majority of the area (50%) falls within **moderate to high risk,** indicating that urban development has increased flood vulnerability in the region.

4.6 Sensitivity Analysis

A sensitivity analysis was conducted to assess how changes in key variables (rainfall intensity, land use, slope) affected flood predictions. The Figure shows the sensitivity of flood hazard predictions to variations in rainfall intensity and land use type.

- **Rainfall Intensity:** Small changes in rainfall intensity (±10%) caused a significant shift in flood hazard predictions, with a change in flood depth of up to 0.5 meters in certain areas.
- **Land Use:** Changes in land use, particularly the increase in impervious surfaces, contributed to a noticeable increase in flood risk, particularly in urbanized zones.

Figure 11: Sensitivity Analysis of Flood Depth to Key Variables

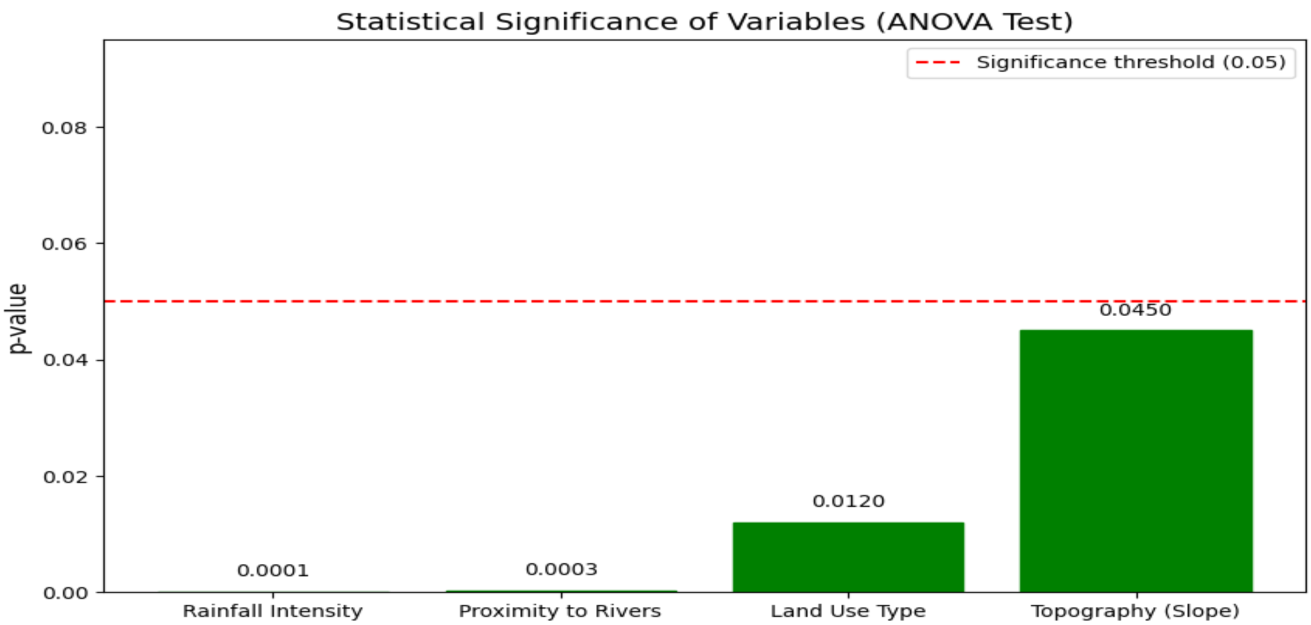


4.7 Statistical Significance of Variables

A statistical analysis (ANOVA test) was performed to assess the significance of the features used in the machine learning model for predicting flood-prone areas. This table summarises the ANOVA results, showing the p-values for the key features:

Feature	p-value
Rainfall Intensity	0.0001
Proximity to Rivers	0.0003
Land Use Type	0.012
Topography (Slope)	0.045

Figure 12: Statistical Significance of Variables (ANOVA Test)



Interpretation: The rainfall intensity and proximity to rivers variables were found to be statistically significant with p-values less than 0.05, indicating they have a strong influence on flood predictions. Land use type and topography were less significant but still contributed to the model's performance.

5. Discussion

The integration of hydrodynamic modelling with machine learning provided new insights into flood hazard assessment within the studied watershed. The discussion highlights the implications of the findings, evaluates the methodological strengths and limitations, and situates the results within the broader context of flood risk management.

5.1 Hydrodynamic Model Performance

The 2D HEC-RAS simulations demonstrated strong predictive accuracy, with a correlation coefficient of 0.92 and low error margins between observed and simulated flood depths. This confirms the robustness of hydrodynamic modelling in representing physical flood processes. High flow velocities observed in narrow channels underscore areas particularly vulnerable to erosion and infrastructure damage [Tripathi, 2024]. These findings reinforce the importance of incorporating channel morphology and flow dynamics into hazard assessments rather than relying solely on depth-based metrics.

5.2 Machine Learning and Feature Interpretation

Among the machine learning algorithms tested, the Random Forest model showed superior predictive capability, achieving 91% accuracy and balanced precision-recall scores. This outperformance aligns with earlier studies that emphasise ensemble methods as effective for handling nonlinear hydrological interactions. SHAP-based feature importance further revealed that rainfall intensity and river proximity were dominant predictors, while land use and slope contributed to a lesser extent [Uddameri, 2025]. This indicates that hydrological drivers exert a stronger influence than geomorphological factors in the study area. Importantly, the explainable AI framework addressed the "black-box" challenge often associated with machine learning by providing transparent justifications for predictions [Ziya, 2025].

5.3 Integrated Approach and Practical Implications
The integration of hydrodynamic and machine learning models resulted in an improved flood hazard map with an overall accuracy of 89% and a Kappa statistic of 0.84. Compared to standalone models, this hybrid approach reduced misclassifications and captured both physical processes and data-driven patterns [Nguyen, 2024]. The integrated hazard zonation revealed that nearly half of the watershed falls within moderate to high-risk categories. For urban planners and policymakers, these results underscore the urgency of targeted interventions in vulnerable zones. Mitigation strategies may include restricting development in high-risk areas, enhancing drainage infrastructure, and adopting land-use policies that reduce impervious surface expansion [Mulungo, 2024].

5.4 Sensitivity and Statistical Analyses
The sensitivity analysis confirmed that small changes in rainfall intensity, more or less than 10% significantly altered flood depths, highlighting the climate-sensitive nature of flood hazards. Likewise, land use modifications, particularly urban expansion, amplified flood risks in downstream zones [Mudashiru, 2021]. Statistical tests validated rainfall intensity and river proximity as the most significant predictors ($p < 0.05$), reinforcing their role as critical variables for flood modelling. These results suggest that future climate variability and land-use changes must be integrated into long-term flood management frameworks [Patel, 2022].

5.5 Limitations and Future Directions
While the study successfully demonstrated the benefits of an integrated approach, certain limitations remain. The reliance on available flood records may have introduced uncertainties in validation, and higher-resolution DEMs could further refine hydrodynamic outputs [Molnar, 2020]. Additionally, the machine learning models were trained on static variables, whereas temporal factors such as antecedent moisture conditions were not considered. Future research should incorporate dynamic predictors, explore deep learning models, and expand to larger geographic scales for broader applicability [Sarkar, 2022].

6. Conclusion and Recommendations
6.1 Conclusion
This study developed an integrated framework combining two-dimensional hydrodynamic modelling (HEC-RAS) with explainable machine learning (XAI) for flood hazard zonation in urban watersheds. The results demonstrated that the 2D HEC-RAS model effectively simulated flood depths and velocities with high accuracy, validated by a strong correlation ($R^2 = 0.92$) and low error metrics. These simulations provided valuable insights into flow dynamics, particularly in narrow channels prone to high-velocity flows and associated erosion risks.

Machine learning approaches, especially the Random Forest algorithm, further enhanced flood susceptibility prediction, achieving 91% accuracy with balanced precision and recall values. The use of SHAP values and feature importance analysis enabled transparent interpretation of model predictions, with rainfall intensity and proximity to rivers emerging as the most influential factors. This resolved the common limitation of black-box models and offered decision-makers a clearer understanding of drivers of flood vulnerability.

The integration of hydrodynamic and machine learning outputs produced a comprehensive hazard map that improved prediction reliability, achieving an overall accuracy of 89% and a Kappa statistic of 0.84. The zonation revealed that nearly 50% of the watershed falls within moderate to high flood risk categories, underscoring the pressing need for mitigation measures in urban planning [Bhusal, 2022]. Sensitivity and statistical analyses further confirmed the climate- and land use-dependent nature of flood hazards, highlighting the potential for risk amplification under scenarios of increased rainfall intensity or urban expansion.

6.2 Recommendations
Based on the findings, several recommendations are proposed to enhance urban flood risk management:

6.2.1 Integration of Methods in Planning
Urban planning authorities should adopt hybrid modelling frameworks that combine hydrodynamic simulations with explainable machine learning. Such integration ensures both physical process accuracy and predictive adaptability, leading to more robust hazard zonation maps.

6.2.3 Climate-Responsive Strategies
Given the sensitivity of flood hazards to rainfall variations, urban flood management plans must incorporate climate change projections. Early warning systems should be updated with real-time rainfall data to dynamically adjust hazard predictions.

6.2.3 Land Use Regulation
The expansion of impervious surfaces significantly increases flood risk. Policymakers should enforce zoning laws that limit construction in high-risk floodplains and encourage green infrastructure, such as permeable pavements, urban wetlands, and retention basins.

6.2.4 Improved Data Infrastructure
High-resolution DEMs, real-time hydrological data, and detailed flood inventories are critical for refining both hydrodynamic and machine learning models. Investments in remote sensing and ground-based monitoring should be prioritised to reduce uncertainty in hazard assessments.

6.2.4 Community-Oriented Mitigation
Disaster preparedness should involve local communities in flood-prone zones through awareness programs and participatory planning [Patel, 2022]. The interpretability offered by explainable AI can aid in communicating risk more effectively to stakeholders and residents.

6.2.5 Future Research
Extending this framework to incorporate temporal predictors, such as soil moisture and antecedent conditions, and testing deep learning methods could further advance prediction accuracy. Additionally, applying the methodology across diverse urban watersheds would validate its generalizability and scalability.

In summary, the study confirms that combining hydrodynamic modelling with explainable machine learning provides a reliable, interpretable, and actionable approach to flood hazard zonation. Such integrative methods should form the foundation of future urban flood management strategies to enhance resilience against climate-driven flood risks.

References

[1] Abdelkarim, A., Gaber, A. F. D., Alkadi, I. I., & Alogayell, H. M. (2019). Integrating remote sensing and hydrologic modeling to assess the impact of land-use changes on the increase of flood risk: A case study of the Riyadh–Dammam train track, Saudi Arabia. *Sustainability*, *11*(21), 6003.

[2] Ahmad, I., Farooq, R., Ashraf, M., Waseem, M., & Shangguan, D. (2025). Improving flood hazard susceptibility assessment by integrating hydrodynamic modeling with remote sensing and ensemble machine learning. *Natural Hazards*, 1–30.

[3] Ahmad, I., Wang, X., Waseem, M., Zaman, M., Aziz, F., Khan, R. Z. N., & Ashraf, M. (2022). Flood management, characterization and vulnerability analysis using an integrated RS-GIS and 2D hydrodynamic modelling approach: the case of Deg Nullah, Pakistan. *Remote Sensing*, *14*(9), 2138.

[4] Al-Omari, A. A., Shatnawi, N. N., Shbeeb, N. I., Istrati, D., Lagaros, N. D., & Abdalla, K. M. (2024). Utilizing remote sensing and GIS techniques for flood hazard mapping and risk assessment. *Civil Engineering Journal*, *10*(5), 1423–1436.

[5] Aryal, A., Acharya, A., & Kalra, A. (2022). Assessing the implication of climate change to forecast future flood using CMIP6 climate projections and HEC-RAS modeling. *Forecasting*, *4*(3), 582–603.

[6] Bhusal, A. (2022). *Evaluating the Performance of Process-Based and Machine Learning Models for Rainfall-Runoff Simulation with Application of Satellite and Radar Precipitation Products*. Southern Illinois University at Carbondale.

[7] Brunner, G. W., Piper, S. S., Jensen, M. R., & Chacon, B. (2015). Combined 1D and 2D hydraulic modeling within HEC-RAS. *World Environmental and Water Resources Congress 2015*, 1432–1443.

[8] Devi, K., Reddy, C. C., Rahul, K., Khuntia, J. R., & Das, B. S. (2025). A holistic methodology for evaluating flood vulnerability, generating flood risk map and conducting detailed flood inundation assessment. *Scientific Reports*, *15*(1), 28253.
[9] Diaconu, D. C., Costache, R., & Popa, M. C. (2021). An overview of flood risk analysis methods. *Water*, *13*(4), 474.
[10] Djafri, S. A., Cherhabil, S., Hafnaoui, M. A., & Madi, M. (2024). Flood modeling using HEC-RAS 2D and IBER 2D: a comparative study. *Water Supply*, *24*(9), 3061–3076.
[11] Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F. A., & Feyen, L. (2016). Development and evaluation of a framework for global flood hazard mapping. *Advances in Water Resources*, *94*, 87–102.
[12] Du, M., Liu, N., & Hu, X. (2019). Techniques for interpretable machine learning. *Communications of the ACM*, *63*(1), 68–77.
[13] Gaddam, K. L. (2024). *A GIS Based Spatial Analysis of Flood Susceptibility Using Machine Learning Models in Data Scarce Regions*. Southern Illinois University at Edwardsville.
[14] Gholami, F., Li, Y., Zhang, J., & Nemati, A. (2023). *Quantitative Assessment of Future Land Use Changes' Impact on Flood Risk Components: Integration of Remote Sensing, Machine Learning, and Hydraulic Modeling*.
[15] Gholami, F., Li, Y., Zhang, J., & Nemati, A. (2024). Quantitative Assessment of Future Environmental Changes in Hydrological Risk Components: Integration of Remote Sensing, Machine Learning, and Hydraulic Modeling. *Water*, *16*(23), 3354.
[16] Hossain, M. S., Saiyara, N., Magumdar, T. K., Mazumder, L. C., Islam, S., Hossain, M. S., Raihan, S. U., & Bhuyan, M. A. (2024). *Development of Urban Flood Forecasting System for Dhaka City Using MIKE+*.
[17] Huu Duy, N., Tuan Pham, L., Xuan Linh, N., Van Truong, T., Dang, D. K., Quang Hai, T., & Bui, Q.-T. (2024). Flood risk assessment using machine learning, hydrodynamic modelling, and the analytic hierarchy process. *Journal of Hydroinformatics*, *26*(8), 1852–1882.
[18] Islam, T., Zeleke, E. B., Afroz, M., & Melesse, A. M. (2025). A Systematic Review of Urban Flood Susceptibility Mapping: Remote Sensing, Machine Learning, and Other Modeling Approaches. *Remote Sensing*, *17*(3), 524.
[19] Karim, F., Armin, M. A., Ahmedt-Aristizabal, D., Tychsen-Smith, L., & Petersson, L. (2023). A review of hydrodynamic and machine learning approaches for flood inundation modeling. *Water*, *15*(3), 566.
[20] Khatooni, K., Hooshyaripor, F., MalekMohammadi, B., & Noori, R. (2025). A new approach for urban flood risk assessment using coupled SWMM–HEC-RAS-2D model. *Journal of Environmental Management*, *374*, 123849.
[21] Khoshkonesh, A., Nazari, R., Nikoo, M. R., & Karimi, M. (2024). Enhancing flood risk assessment in urban areas by integrating hydrodynamic models and machine learning techniques. *Science of The Total Environment*, *952*, 175859.
[22] Khouz, A., Trindade, J., Santos, P. P., Oliveira, S. C., El Bchari, F., Bougadir, B., Garcia, R. A. C., Reis, E., Jadoud, M., Saouabe, T., & others. (2023). Flood susceptibility assessment through statistical models and HEC-RAS analysis for sustainable management in Essaouira Province, Morocco. *Geosciences*, *13*(12), 382.
[23] Kim, M. (2024). *COMPOUND FLOOD ANALYSIS WITH GIS-INTEGRATED DYNAMIC FLOOD MODELS FOR COASTAL URBAN AREAS*. Georgia Institute of Technology.
[24] Lago, C. A. F. do. (2023). *Application of deep learning for high-resolution flood mapping in urban watersheds*. Universidade de São Paulo.
[25] Manandhar, B., Cui, S., Wang, L., & Shrestha, S. (2023). Urban flood hazard assessment and management practices in south asia: a review. *Land*, *12*(3), 627.
[26] Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.
[27] Mudashiru, R. B., Sabtu, N., Abustan, I., & Balogun, W. (2021). Flood hazard mapping methods: A review. *Journal of Hydrology*, *603*, 126846.
[28] Mulungo, H. E. (2024). *Flood risk analysis in the Pungwe river basin: Hydrological and hydraulic modelling of the march 2019 flood event using HEC-HMS and HEC-RAS*.
[29] Nguyen, H. D., Dang, D. K., Hoang, T. S., Truong, Q.-H., Nguyen, T.-N.-U., Nguyen, X. L., & Ha, M. C. (2024). Flood hazard assessment using machine learning and hydrodynamic modeling: case study in the Vu Ga–Thu Bon basin in Vietnam. *Water Practice & Technology*, *19*(10), 4104–4127.
[30] Nguyen, H. D., Quang-Thanh, B., Nguyen, Q.-H., Nguyen, T. G., Pham, L. T., Nguyen, X. L., Vu, P. L., Thanh Nguyen, T. H., Nguyen, A. T., & Petrisor, A.-I. (2022). A novel hybrid approach to flood susceptibility assessment based on machine learning and land use change. Case study: a river watershed in Vietnam. *Hydrological Sciences Journal*, *67*(7), 1065–1083.
[31] O'Hara-Rhi, V. T. (2022). *Evaluation of Urban Flood Hazard Based on Remote Sensing and Machine Learning*. San Diego State University.
[32] Oluwadare, T. S., Ribeiro, M. P., Chen, D., Babadi Ataabadi, M., Tabesh, S. H., & Daomi, A. E. (2025). Applying Machine Learning Algorithms for Spatial Modeling of Flood Susceptibility Prediction over São Paulo Sub-Region. *Land*, *14*(5), 985.
[33] Patel, M., & Parekh, F. (2022). Advanced Techniques of Flood Forecasting, Flood Inundation Mapping and Flood Prioritization of Panam River Basin. *Civil Engineering and Architecture*, *10*(4), 1303–1316.
[34] Pathan, A. I., Agnihotri, P. G., Patel, D., & Prieto, C. (2022). Mesh grid stability and its impact on flood inundation through (2D) hydrodynamic HEC-RAS model with special use of Big Data platform—a study on Purna River of Navsari city. *Arabian Journal of Geosciences*, *15*(7), 659.

[35] Qureshi, M. U. A. (2025). *Predicting Discharge Variations in the Ottawa River Integrating HEC-RAS with Artificial Intelligence in the Context of Climate Change*. Université d'Ottawa/University of Ottawa.

[36] Rangari, V. A., Umamahesh, N. V, & Bhatt, C. M. (2019). Assessment of inundation risk in urban floods using HEC RAS 2D. *Modeling Earth Systems and Environment*, *5*(4), 1839–1851.

[37] Riche, A., Drias, A., Guermoui, M., Gherib, T., Boulmaiz, T., Souissi, B., Melgani, F., & others. (2024). A novel hybrid deep-learning approach for flood-susceptibility mapping. *Remote Sensing*, *2024*(19), 367301–367341.

[38] Sarkar, S. K., Ansar, S. Bin, Ekram, K. M. M., Khan, M. H., Talukdar, S., Naikoo, M. W., Islam, A. R. T., Rahman, A., & Mosavi, A. (2022). Developing robust flood susceptibility model with small numbers of parameters in highly fertile regions of Northwest Bangladesh for sustainable flood and agriculture management. *Sustainability*, *14*(7), 3982.

[39] Shikhteymour, S. R., Borji, M., Bagheri-Gavkosh, M., Azimi, E., & Collins, T. W. (2023). A novel approach for assessing flood risk with machine learning and multi-criteria decision-making methods. *Applied Geography*, *158*, 103035.

[40] Shrestha, A., Bhattacharjee, L., Baral, S., Thakur, B., Joshi, N., Kalra, A., & Gupta, R. (2020). Understanding suitability of MIKE 21 and HEC-RAS for 2D floodplain modeling. *World Environmental and Water Resources Congress 2020*, 237–253.

[41] Tripathi, V., & Mohanty, M. P. (2024). Can geomorphic flood descriptors coupled with machine learning models enhance in quantifying flood risks over data-scarce catchments? Development of a hybrid framework for Ganga basin (India). *Environmental Science and Pollution Research*, 1–20.

[42] Uddameri, V., & Hernandez, E. A. (2025). Machine Learning for Flood Resiliency—Current Status and Unexplored Directions. *Environments*, *12*(8), 259.

[43] Ziya, O., & Safaie, A. (2023). Probabilistic modeling framework for flood risk assessment: A case study of Poldokhtar city. *Journal of Hydrology: Regional Studies*, *47*, 101393