MACHINE LEARNING BASED FLOOD SUSCEPTIBILITY ASSESSMENT IN THE WEST RAPTI RIVER BASIN USING MULTI SOURCED GEOSPATIAL DATA

Dibikshya Shrestha[1], Rabi Shrestha[2]
[1]*Kathmandu Metropolitan City, jbdibikshyashrestha@gmail.com*
[2]*NEA Engineering Company Ltd, rabi.shrestha@neaec.com.np*

KEY WORDS: Flood Susceptibility Mapping (FSM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Artificial Neural Network (ANN)

ABSTRACT:
Floods are one of the most destructive natural hazards, causing significant human loss. Identification of flood-prone areas is a substantial need for effective planning and risk management. This study aims to generate flood susceptibility maps for the West Rapti River Basin incorporating three machine learning models: Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Artificial Neural Network (ANN). Among the models, XGBoost achieved the highest predictive performance with an accuracy of 0.9706, followed by RF (0.9510) and ANN (0.9412). Variable importance analysis showed that distance from the river is the most influential factor. The resulting flood susceptibility maps indicate that approximately 30% of the basin is highly susceptible to flooding, primarily in low lying areas near river channels. The findings demonstrate the effectiveness of machine learning approaches in flood susceptibility assessment. This study provides valuable insights for disaster risk reduction and early warning systems in the West Rapti River Basin.

1. INTRODUCTION
1.1 Background

Floods are considered as one of the most catastrophic natural disasters that has affected a large population worldwide, causing significant damage to the environment as well as socio-economic systems (Bhattrai & Ghimere, 2023). It ranks among the most devastating natural disaster that severely affect human safety and infrastructure (Nusit et al., 2025). Their increasing occurrence depends on both climatic factors such as heavy rainfall, snowmelt, and extreme weather conditions and human activities, including deforestation, rapid urban growth, and improper land use practices(Asrade et al., 2026) which has resulted in billions of dollar in damages along with disrupting livelihoods especially in vulnerable region of the world "(Rahman et al., 2025).

The assessment of extreme events such as floods is important for mitigating risk and adapting sustainable strategies —(Hussain et al., 2023). In order to access their associated risk, damage and spatial extend, many researchers have focused on estimating and explaining flood hazards for which reliable and accurate Flood Susceptibility Mapping (FSM) is essential (Seydi et al., 2023). FSM quantifies the risk of flooding in a specific region in accordance to its relationship with its causes(Seydi et al., 2023). It reveals flood-prone areas under specific conditions of topography, hydrology, land cover, rainfall, and artificial constructions (Nguyen et al., 2023). Traditionally, FSM has applied quantitative, semi-quantitative and qualitative methodologies which often face uncertainties(Abdo et al., 2025).

In recent years, Machine learning techniques have arisen as powerful alternative to model flood susceptibility due to their ability to capture relationship efficiently (Nusit et al., 2025). High level Machine learning models such as: Artificial Neural Network (ANN), decision trees can predict floods in low computational costs without in-depth knowledge of physical and hydrological processes involved (Kaur et al., 2025; Nusit et al., 2025). For instance, (Nguyen et al., 2023) achieved accuracy of (AUC>0.9) , training ML models: XGBoost, RF, CART and LightGBM using a set of 11 topological and hydrological variables. Despite these studies, the issue of which model performs better in a particular geographical location is still a prevailing issue.

Nepal is extremely vulnerable to flood events due to its complex geology, topography and climatic events due to which flood events occur frequently(Bhattrai & Ghimere, 2023). West Rapti River Basin (WRRB), located in the mid-western part of Nepal, is one of the flood prone river basins that is inundated each year due to heavy rainfall (Bhattrai & Ghimere, 2023). Literature shows that despite being extremely vulnerable to flood events, most previous studies have adapted traditional approach to map flood susceptibility in this river basin. Given this context, this study aims to map flood susceptibility in West Rapti River basin and evaluate the performance of three different machine learning models: Random Forest (RF), Xtreme Gradient Descent (XGBoost) and Artificial Neural Network (ANN) along with 10 different multi-sourced flood inducing variables. The findings of this study can assist in early warning systems, risk management measures and proper infrastructure planning

2. OBJECTIVES
The main objective of this research includes:
1.Generate Flood Susceptibility Maps (FSMs) for Random Forest (RF), Xtreme Gradient Descent (XGBoost) and Artificial Neural Network (ANN) models,
2.Evaluate and validate the models using statistical measures,
3.Identify flood vulnerability in terms of land use type and affected settlements.

3. MATERIALS AND METHODS
3.1 STUDY AREA
a) The study area for this research is West Rapti River Basin (WRR), located in the mid-western part of Nepal with an area of 6971 km² . It is one of the tributaries of Karnali River (known as Ghagra River in India). The boundary of the basin extends from 27°40′35″ N to 28°34′21″ N latitude and 81°39′23″ N to 83°09′39″ E longitude. This river basin stands at a maximum elevation of 3665m and minimum elevation of 125m. (river atlas of nepal) The Rapti River in this basin originates in the central highlands of Nepal and flows southward. The basin includes several tributaries such as: Arun Khola, Jhimruk Khola, Dundhuwa River, Mari Khola and Lungri khola. The river flows downstream from the junction of the Jhimruk and Madi Khola. With an average slope of 16.8%, the runoff is mainly caused by monsoon rainfall and groundwater(Bhattrai & Ghimire, 2023).

b)

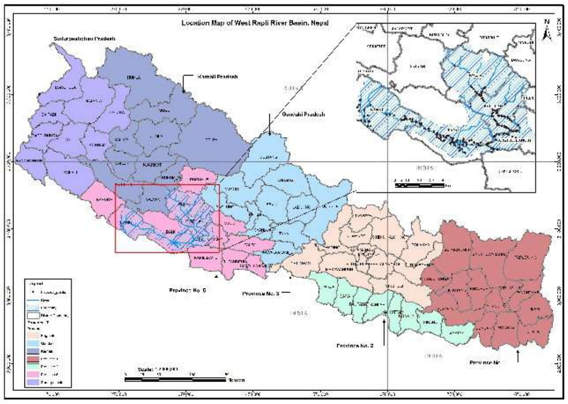


Figure 1: Location Map of West Rapti River Basin

c) The upper portion of the basin (>2000 meter from mean sea level) experiences a temperate climate with dry winters and warm summer and hot summers below 2000 meters. In contrast, the lower portion of the basin has tropical climate. During summer, temperatures in the lower WRRB can exceed 45 degrees Celsius, while in the upper WRRB temperature drop below 2 degrees Celsius in winter, which shows a wide variation of climate across the basin. WRRB experiences four main seasons: Pre monsoon (March–May), Monsoon (June–September), Post-monsoon (October–November), and Winter (December–February) (Talchabhadel et al., 2021). The average annual rainfall of this river basin is 1587mm (Secretariat, 2024). The study area receives summer monsoon rainfall from June to September which accounts for about 80% of the total annual rainfall. The relative humidity ranges from around 60% in May to more than 90% in January (Bhattrai & Ghimire, 2023).

3.2 Data Sources

The different types of datasets used in this study, along with their sources are presented in Table 1. These datasets are analysed using GIS and ESA SNAP Software to evaluate various flood conditioning factors and to identify areas with high flood risk in the WRRB.

Table 1: Datasets used and their sources

S.N	Datasets	Source
1.	SRTM DEM (30m)	USGS
2.	Soil/Soil Moisture (9 km)	FAO/UNESCO, SMAP
3.	Land Use (10m)	Esri/Sentinel 2
4.	Rainfall (5 km)	Chirps
5.	NDVI (30 m)	Landsat 8
6.	Historic Flood Points(10m)	Copernicus/Sentinel 1
7.	Nepal Boundary/ River Network	National Geoportal of Nepal
8.	River basin boundary	HydroSHEDS

3.3 Research Methodology

a) This study employed a remote sensing and machine learning based approach to develop flood susceptibility maps for the West Rapti River Basin (WRRB). The overall methodology integrates satellite image processing, spatial analysis, and machine learning techniques to identify areas vulnerable to flooding in the study arrea.

b) The pre-flood and post-flood Sentinel-1 Synthetic Aperture Radar (SAR) images from the Copernicus Open Access Data Platform was collected. Three major historical flood events that occurred in 2017, 2019, and 2024 were selected for the analysis. Sentinel-1 imagery was chosen because of its ability to acquire data under all weather conditions, including cloudy environments during flood events.

c) The collected Sentinel-1 images were preprocessed using the European Space Agency's Sentinel Application Platform (SNAP). The preprocessing procedure included the application of orbit files, subsetting of the study area, thermal noise removal, radiometric calibration, speckle filtering, terrain correction, and band mathematics for the identification of water surfaces. These preprocessing steps improved the quality and reliability of the satellite data for flood mapping.

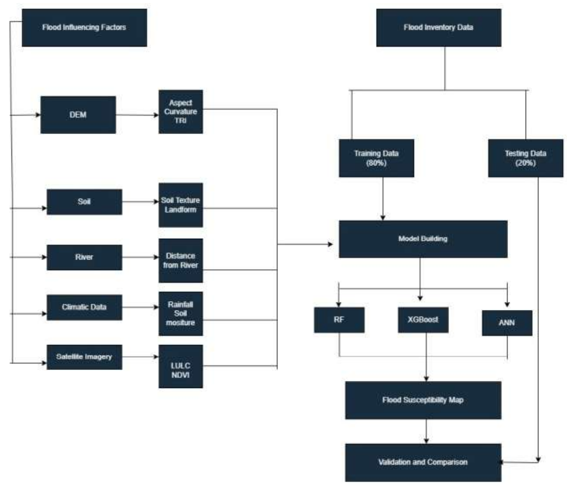


Figure 2: Methodological Flowchart

d) Following preprocessing, flooded areas were identified using thresholding techniques, change detection methods, and manual digitization in QGIS and Google Earth Pro. Flooded points were then extracted from the identified inundated areas for each flood event year (2017, 2019, and 2024). A total of 280 flooded points were collected and labeled as "1" to represent flood occurrence.

e) To prepare the flood-conditioning dataset, the values of all influencing factors were extracted at the flooded point locations using Google Earth Engine and QGIS. The selected flood-conditioning factors included aspect, curvature, Terrain Ruggedness Index (TRI), rainfall, soil moisture, distance from river, Normalized Difference Vegetation Index (NDVI), soil texture, land use/land cover (LULC), and landform. These factors were selected based on their relevance to flood occurrence and previous flood susceptibility studies.

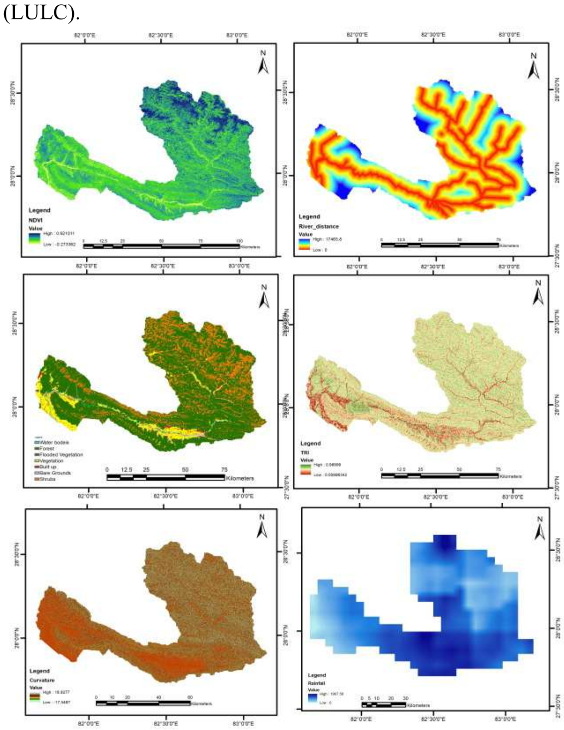
f) Similarly, 280 non-flooded points were generated through manual digitization using post-flood satellite imagery. These points were spatially distributed across the study area to ensure balanced representation of non-flooded locations. To minimize spatial autocorrelation and improve model performance, a buffer distance of 2500 meters was maintained between flooded and non-flooded points. The non-flooded points were labeled as "0".

g) The final dataset consisted of 560 points, including both flooded and non-flooded samples. Prior to model development, the dataset was cleaned by removing null values and applying Interquartile Range (IQR) capping to reduce the influence of outliers. In addition, collinearity analysis was performed among the input variables to examine multicollinearity. The correlation values for all variables were found to be below ±0.5, indicating the absence of significant multicollinearity among the predictors.

h) The prepared dataset was then divided into training and testing datasets using an 80:20 ratio, where 80% of the data were used for model training and 20% were used for model testing and validation. Finally, three machine learning models Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Artificial Neural Network (ANN) were applied to generate flood susceptibility maps for the West Rapti River Basin. The overall methodological framework adopted in this study is illustrated in Figure 2.

3.3 Flood Influencing Factors

Studies have identified four main categories of factors that influence flood occurrence: geomorphological, geological, hydrological and anthropogenic factors (Asrade et al., 2026). Based on these studies, availability of relevant data and avoiding multicollinearity among the input variables, in the study area, nine flood conditioning factors were selected for analysis. These factors include Aspect (direction that a slope faces), Curvature (shape of the terrain surface), Terrain Ruggedness Index (TRI) (quantifies how uneven or rugged the surface is by calculating the average variation in elevation), Rainfall (amount, intensity and duration of rain), Soil Moisture (amount of water present in the soil), Distance from River, Normalized Difference Vegetation Index (NDVI) (remote sensing indices used to measure the density and health of a vegetation), Soil Texture (relative proportion of mineral particles in soil) and Land Use Land Cover (LULC).



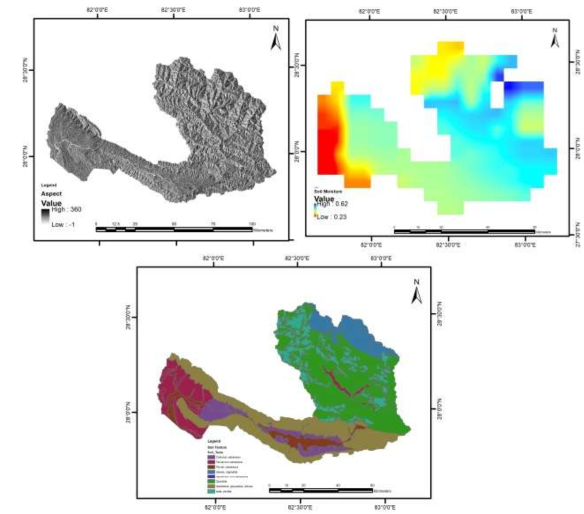


Figure 3: Major factors that influence flood occurrence as represented in maps

3.4 Multicollinearity analysis

Multicollinearity analysis refers to the process of identifying the degree of correlation between independent variables in a machine learning model. It is an important step in flood susceptibility assessment as highly correlated variables can distort model performance. In this study, two commonly used multicollinearity test methods Variance Inflation Factor (VIF) and Pearson Correlation Coefficient (PCC) are employed to evaluate multicollinearity. VIF measures the increment in the variance of predicted regression coefficient due to collinearity with other predictors. Similarly, Pearson Correlation Coefficient (PCC) measures the linear relationship between two independent variables. VIF Equation (1) and PCC Equation (2) is formulated as shown:

$$VIF = 1\frac{1}{1-R^2} \tag{1}$$

$$PCC = \frac{\sum_{i=1}^{n} (X_i - \underline{X})(Y_i - \underline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \underline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \underline{Y})2}} \tag{2}$$

Where, R^2 represents the unadjusted coefficient of determination, X_i and Y_i represents the individual input variables and $\bar{X}$ and $\bar{Y}$ are the mean of X and Y respectively.

3.5 Machine learning Algorithms

3.5.1 Random Forest (RF)

Random Forest (RF), is an ensemble learning algorithm used for both classification and regression tasks. It performs by combining multiple decision trees, which increases stability and reduces variance. RF is based on bootstrap aggregating (bagging), where the final prediction is obtained by averaging the outputs of several trees, as shown in Equation (3):

$$\hat{Y} = \frac{1}{M}\sum_{m=1}^{M} hm(x) \tag{3}$$

where, $\hat{Y}$ represents the predicted output, M is the total number of decision trees, and hm(x) is the prediction from the mthdecision tree for input variables x. In flood susceptibility studies, RF is advantageous because it can handle complex and diverse datasets (Abdo et al., 2025; Rahman et al., 2025).

3.5.2 Extreme Gradient Boost (XGBoost)

Extreme Gradient Boost (XGBoost) is a type of tree-based ensemble machine learning technique that works based on the gradient boosting approach, where decision trees are built one after another. Each new tree is created by correcting the errors made by the previous trees, and this process continues until the overall loss function is minimized.

XGBoost develops and ensemble of decision trees, using the following formulation Equation (4):

$$\hat{y}i = \sum_{k=1}^{K} f_k(x_i)\,, f_k \in F \qquad (4)$$

where, $\hat{y}i$ is the predicted output at location i, K is the number of trees, and $f_k(x_i)$ is the prediction from kth decision tree for the ith location based on input variables. F is the space of all decision trees.

XGBoost is widely used in flood susceptibility mapping because it can effectively analyse complex data such as topography, climate, soil, and vegetation, and identify the most important factors influencing flood risk. It manages overfitting by adding regularization terms "(Abdo et al., 2025; Rahman et al., 2025).

3.5.3 Artificial Neural Network (ANN)

ANN is a machine learning model inspired by the human brain organized into input, hidden and output layers. ANN has the ability of identifying patterns and modelling complex and non-linear relationships. The relationship between input and output in ANN can be expressed as Equation (5):

$$Z = f\left(\sum_{i-n}^{n} w_i x_i + b\right) \qquad (5)$$

where, z is the predicted output by ANN, x_i is the input independent variable, w_i represents the connecting weights, b stands for bias and f is the activation function. In flood susceptibility mapping, ANN possesses a high significance due to its ability to model non-linear relationships, handle large and complex datasets(Abdo et al., 2025; Rahman et al., 2025).

3.6 Accuracy Assessment

After the final cleaning of the dataset, 80% of the data was used for Machine Learning training purposes and the rest 20% of the dataset was used for testing and validation. To ensure unbiased and representative split of the data, random sampling was employed. The accuracy and the performance of the predictive models were assessed using: Accuracy, Recall, F1-score and Area Under the Curve- Receiver Operating Characteristics (AUC-ROC) Curve.

3.6.1 Accuracy

It measures the overall correctly classified flood prone and non-flood prone areas. Accuracy is calculated by Equation (6):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+F} \qquad (6)$$

where, TP, TN, FP and FN stand for True Positive, True Negative, False Positive and False Negative respectively(Abdo et al., 2025).

3.6.2 Recall

Recall measures the ratio of accurately identified flood prone areas to the total number of actual flood prone areas. Recall is calculated by Equation (7):

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (7)$$

where, TP (True Positive) is the correctly predicted flood area and FN (False Negative) is the actual flood area that the model missed (Halder et al., 2024).

3.6.3 F1- Score

F1-Score measures the balance between Recall and Precision particularly beneficial when the data is imbalanced. F1-Score is calculated by Equation (8):

$$\text{F1-Score} = \frac{2\,x\,TP}{2XTP+FP+FN} \qquad (8)$$

Where, TP is True Positive, FP is False Positive and FN is False Negative (Halder et al., 2024).

3.6.4 Area Under the Curve - Receiver Operating Characteristics (AUC-ROC)

AUC measures how accurately the model distinguishes between flooded and non-flooded areas whereas ROC is the graphical representation of AUC value. AUC is calculated by Equation(9):

$$\text{AUC} = \sum TP + \sum TN\,/(K+S) \qquad (9)$$

where, K represents the total number of flooded events and S represents the total number of non-flooded events (Abdo et al., 2025).

4. RESULTS AND DISCUSSION

4.1 Multicollinearity analysis

Multicollinearity among the nine input variables was evaluated using the Pearson Correlation Coefficient (PCC) and Variance Inflation Factor (VIF). The result indicates that all variables are free from multicollinearity, as all PCC values are below 0.6 and all VIF values are less than 5. The highest PCC value was observed between Soil Texture and NDVI (0.39), while the lowest was between Curvature and NDVI (−0.01). Similarly, the highest VIF value was for NDVI (1.58), and the lowest was for Aspect (1.01). These results suggest that there is no significant correlation among the variables, and each variable can be considered independent. The value of PCC and VIF for each variable is shown in Table (2):

Table 3: Value of PCC and VIF for each variable

Variables	PCC	VIF
NDVI	0.151	1.582
TRI	0.116	1.130
Rainfall	0.041	1.180
LULC	0.111	1.201
Curvature	0.077	1.031
Aspect	0.032	1.014
River distance	0.300	1.243
Soil Texture	0.110	1.336
Soil Moisture	0.056	1.327

4.2 Importance of Variables

The importance of the nine input variables was evaluated using feature importance graph in order to understand their contribution to the model. The results indicate that river distance is the most significant variable across all the three models, with an importance score of 0.30, indicating its strong effect on the outcome. In contrast, Aspect has the lowest contribution, with an importance score of 0.03, indicating a weak effect in the model's outcomes.

4.3 Model Performance and Validation

Three Machine Learning models: Random Forest (RF), Extreme gradient Boosting (XGBoost) and Artificial Neural Network (ANN) were employed for flood susceptibility mapping. These models were particularly selected due to their ability to handle complex and non-linear relationships. The optimal parameters of Machine learning algorithms or the hyperparameter tuning parameters was identified using GridSearch CV method which is a technique used in Machine Learning to find the optimal combination of parameters for a model. The final optimal parameter for each model is presented in Table (3):

Table 3: Final optimal parameters for each model

Model Name	Optimal Parameters
RF	n_estimators = 128, max_depth = 16, max_features = 'sqrt', min_sample_split = 2, bootstrap = 'True'
XGBoost	N_estimators = 150, max_depth = 4, learning rate = 0.08, gamma = 0.1, sub_sample = 0.85
ANN	hidden_layer_sizes = (128,), activation = 'relu', batch_size = 64, learning_rate = 0.005, solver= 'adam'

The final susceptibility map was prepared from the predicted probabilities of the best model which was further classified into 5 categories: very low, low, moderate, high and very high susceptibility using the jensky classification method that minimize variance within classes while maximizing differences between them. The final values of the model suggest that 27.64%, 30.87% and 33.01% of the total area of the basin is highly susceptible to the flood. Overall, almost 30% of the basin is highly susceptible to flood which is mainly confined to the western part of the basin. The final flood susceptibility map from three different models is shown inError! Reference source not found.

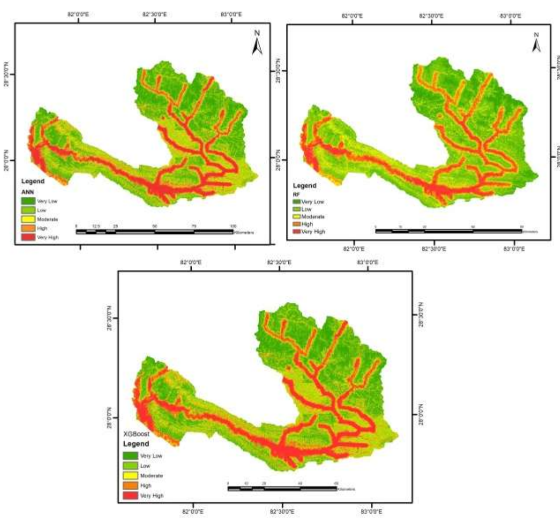


Figure 5: Final Susceptibility Maps as per different models classified under jensky classification method

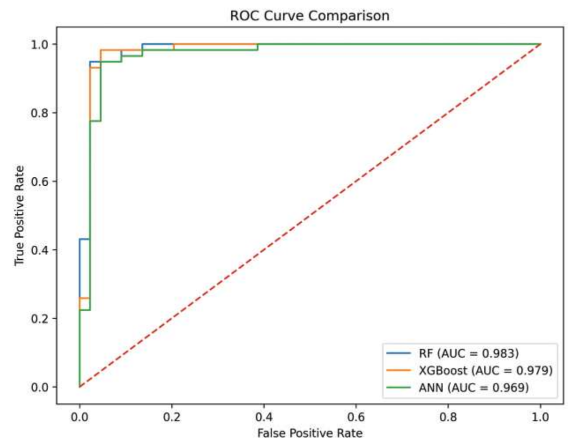
4.3.1 Model Validation

The performance of each model was evaluated using four different statistical measures which include accuracy, recall, f1-score and AUC-ROC Curve. In this study, each model has been trained using a 10-fold cross validation method in order to avoid overfitting of the model. All the performance evaluation metrics are shown in Table (4):

Table 4: Performance evaluation metrics

ML Model	Accuracy	AUC	F1-Score	Recall
RF	0.9510	0.9828	0.9565	0.9483
XGBoost	0.9706	0.9788	0.9744	0.9828
ANN	0.9412	0.9690	0.9474	0.9310

XGBoost model attained the highest accuracy of 0.9706 followed by RF (0.9510) and ANN (0.9412). Regarding, the values of F1-Score and Recall for models they have achieved as follows : XGBoost (F1-Score is 0.9744, Recall is 0.9828), for RF (F1-Score is 0.9565, Recall is 0.9483) and for ANN (F1-Score is 0.9474, Recall is 0.9310).Regarding AUC, RF , XGBoost and ANN has acquired accuracy of 0.9788, 0.9828 and 0.9690 respectively as shown inFigure 6.



1. DISCUSSION

a) The West Rapti River Basin (WRRB) is highly susceptible to recurrent flood events, which significantly affect human life, infrastructure, agriculture, and the regional economy each year. In recent decades, rapid anthropogenic activities and the increasing impacts of climate change have intensified the frequency and severity of flooding events, highlighting the urgent need for accurate flood susceptibility assessment and management strategies (Rahman et al., 2025). In this context, the present study employed data driven machine learning approaches to develop flood susceptibility maps for the WRRB, differing from previous studies that primarily relied on multi-criteria decision making techniques and conventional statistical approaches.

b) A total of nine flood conditioning factors were integrated with flooded and non-flooded sample points to evaluate flood susceptibility across the basin. Previous research has demonstrated the effectiveness of machine learning techniques in flood susceptibility mapping due to their ability to model complex nonlinear relationships among environmental variables (Abdo et al., 2025; Rahman et al., 2025). The flood susceptibility maps generated by all three machine learning models consistently indicated that areas located in close proximity to river channels and water bodies are highly vulnerable to flooding. This finding is consistent with the geomorphological characteristics of the WRRB, which is predominantly composed of low-lying and relatively flat Terai plains that facilitate flood inundation during intense rainfall events.

c) Although precipitation is considered the primary triggering factor for flood occurrence, several additional environmental and topographic variables also play significant roles in influencing flood susceptibility. In the present study, variables such as distance to river, land use/land cover (LULC), soil moisture, and topographic characteristics contributed considerably to model performance. Among these factors, distance to river emerged as the most influential predictor across all models, whereas aspect demonstrated the least influence on flood occurrence. The dominance of river proximity as a controlling factor can be attributed to the hydrological behavior of the basin, where river overflow and lateral expansion during monsoon periods are the major causes of inundation.

d) The performance evaluation of the models demonstrated excellent predictive capability, with overall accuracy values ranging from 0.94 to 0.97. The strong performance of the models can largely be attributed to their ability to capture nonlinear relationships and interactions among flood conditioning variables. Among the evaluated models, XGBoost achieved the highest predictive performance. The superior performance of XGBoost may be associated with its regularization capability, which effectively minimizes overfitting while enhancing model generalization and predictive accuracy. Similar findings were reported by Li et al. (2026), who identified XGBoost as the best-performing model among several machine learning algorithms for flood susceptibility assessment. In contrast, the Artificial Neural Network (ANN) model demonstrated comparatively lower performance, which may be associated with limitations related to training data volume, computational constraints, and model optimization requirements.

e) Despite the satisfactory predictive performance obtained in this study, several limitations remain that should be addressed in future research. One of the primary concerns is the potential risk of overfitting, particularly in tree based ensemble models such as Random Forest (RF) and XGBoost. Although the present study implemented 10 fold cross-validation and hyperparameter tuning to improve model robustness, overfitting remains a persistent challenge in machine learning applications. Future studies may further reduce this issue through the incorporation of higher-resolution field based datasets and more representative hydrological observations, which could improve the spatial accuracy of susceptibility predictions. Additionally, the present study incorporated only three historical flood events due to data limitations. The inclusion of a larger number of historical flood records and long term hydrometeorological datasets could further enhance model reliability.

f) Overall, the findings of this study demonstrate the significant potential of machine learning approaches, particularly XGBoost, for flood susceptibility mapping in data scarce regions such as the West Rapti River Basin. The generated flood susceptibility maps can support policymakers, disaster management authorities, and urban planners in implementing effective flood mitigation, land use planning, and disaster preparedness strategies for reducing future flood risk in the region.

2. CONCLUSION

a) Flooding is one of the most destructive natural hazards affecting the West Rapti River Basin (WRRB), causing significant impacts on human life, infrastructure, agriculture, and regional socio-economic activities. Therefore, flood susceptibility mapping is essential for effective flood risk assessment, disaster preparedness, and the development of early warning systems. The present study aimed to classify the WRRB into different flood susceptibility zones and to evaluate the performance of various machine learning (ML) models for flood susceptibility prediction.For this purpose, a total of 560 flooded and non-flooded sample points were collected from three major historical flood events that occurred in 2017, 2019, and 2024. In addition, nine flood conditioning factors were incorporated to analyze their influence on flood occurrence. These datasets were used to train and evaluate three machine learning models, namely Extreme Gradient Boosting (XGBoost), Random Forest (RF), and Artificial Neural Network (ANN).

b) The results revealed that distance to river was the most influential factor controlling flood susceptibility within the basin, emphasizing the strong hydrological influence of river proximity on flood occurrence in the predominantly flat Terai region. Among the evaluated models, XGBoost demonstrated the highest predictive performance, outperforming RF and ANN in terms of overall accuracy and model reliability. The generated flood susceptibility maps indicated that approximately 30% of the total basin area falls within highly susceptible flood zones, whereas nearly 58% of the basin is categorized under low flood susceptibility. The highly susceptible areas are primarily concentrated near river channels and water bodies.

c) The findings of this study provide valuable insights into the spatial distribution of flood prone areas within the WRRB and can support policymakers, planners, and disaster management authorities in implementing effective flood mitigation and risk management strategies. Furthermore, the proposed machine learning based methodology demonstrates strong potential for application in other flood prone regions for hazard assessment and susceptibility mapping.

REFERENCES

Abdo, H. G., Richi, S. M., Zeng, T., Alqadhi, S., Prasad, P., Nguyen-Huy, T., Alharbi, M. M., & Mallick, J. (2025). Machine learning-based assessment of flood susceptibility in the Eastern Mediterranean: a case study of Baniyas River basin. Geomatics, Natural Hazards and Risk, 16(1). https://doi.org/10.1080/19475705.2025.2524417

Asrade, T. M., Abebe, S. A., Tadesse, K. B., Kerebih, M. S., & Meshesha, T. M. (2026). Flood susceptibility assessment using three machine learning techniques and comparison of their performance. Scientific Reports, 16(1), 1–20. https://doi.org/10.1038/s41598-026-38391-0

Bhattrai, T. N., & Ghimere, S. (2023). Flood Susceptibility Analysis in the West Rapti. 3(1), 1–24.

Halder, K., Ghosh, A., Srivastava, A. K., Pal, S. C., Chatterjee, U., Bisai, D., Ewert, F., Gaiser, T., Islam, A. R. M. T., Alam, E., & Islam, M. K. (2024). SAR-driven flood inventory and multi-factor ensemble susceptibility modelling using machine learning frameworks. Geomatics, Natural Hazards and Risk, 15(1). https://doi.org/10.1080/19475705.2024.2409202

Hussain, M., Tayyab, M., Ullah, K., Ullah, S., Rahman, Z. U., Zhang, J., & Al-Shaibah, B. (2023). Development of a new integrated flood resilience model using machine learning with

GIS-based multi-criteria decision analysis. Urban Climate, 50(June), 101589. https://doi.org/10.1016/j.uclim.2023.101589

Kaur, B., Binns, A., McBean, E., Sandink, D., Castro, K., & Gharabaghi, B. (2025). Urban Flood Susceptibility Mapping for Toronto, Canada, Using Supervised Regression and Machine Learning Models. Journal of Flood Risk Management, 18(2). https://doi.org/10.1111/jfr3.70051

Li, H., Dong, J., Zhou, S., Zhang, Y., Liu, H., Pang, Y., & Du, B. (2026). Application and comparison of multiple machine learning models in flood susceptibility assessment in the Beijing-Tianjin-Hebei region of China. Frontiers in Water, 8(February), 1–18. https://doi.org/10.3389/frwa.2026.1767400

Nguyen, D. L., Chou, T. Y., Hoang, T. V., & Chen, M. H. (2023). Flood Susceptibility Mapping Using Machine Learning Algorithms: A Case Study in Huong Khe District, Ha Tinh Province, Vietnam. International Journal of Geoinformatics, 19(7), 1–15. https://doi.org/10.52939/ijg.v19i7.2739

Nusit, K., Tantanee, S., & Sooraksa, P. (2025). Flood Susceptibility Mapping Using Machine Learning Models with Novel Flood Inventory Sampling Strategies. Sensors and Materials, 37(9), 3829–3839. https://doi.org/10.18494/SAM5586

Rahman, Z. U., Zhang, M., Chen, F., Ullah, S., Ahmad, M., Feroz, A., Danhassan, S. S., & Azzam, A. (2025). Flood susceptibility mapping using supervised machine learning models: insights into predictors' significance and models' performance. Geomatics, Natural Hazards and Risk, 16(1). https://doi.org/10.1080/19475705.2025.2516728

Secretariat, E. C. (2024). River basin atlas of nepal.

Seydi, S. T., Kanani-Sadat, Y., Hasanlou, M., Sahraei, R., Chanussot, J., & Amani, M. (2023). Comparison of Machine Learning Algorithms for Flood Susceptibility Mapping. Remote Sensing, 15(1). https://doi.org/10.3390/rs15010192

Talchabhadel, R., Aryal, A., Kawaike, K., Yamanoi, K., & Nakagawa, H. (2021). A comprehensive analysis of projected changes of extreme precipitation indices in West Rapti River basin, Nepal under changing climate. International Journal of Climatology, 41(S1), E2581–E2599. https://doi.org/10.1002/joc.6866

AUTHOR INFORMATION

Name	:	**Dibikshya Shrestha**
Academic Qualification	:	Geomatics Engineer
Organization	:	Kathmandu Metropolitan City
Current Designation	:	GIS Engineer
Work Experience	:	2 years
Published Article	:	2