GIS and Machine Learning Approaches in Flood Hazard Mapping: A Case Study of Lower Niger River Basin

[*1]Adeyemi, Adedoyin Benson and [1]Komolafe, Akinola Adesuji

[1]Department of Remote Sensing and Geoscience Information System (GIS), Federal University of Technology, Akure, Nigeria

*Corresponding Author's Email: abadeyemi97@gmail.com, Phone No.: +2348141154368

ABSTRACT

Flooding is a recurrent and destructive natural disasterintensified byelements such as extreme rainfall, urbanization, climate change, topography, and human activities. This study primarily aims to integrate Geographic Information System (GIS) and Machine Learning (ML) techniques in flood hazard mapping in the lower Niger River basin in Nigeria. Twenty flood influencing factors including elevation, slope, aspect, flow direction, flow accumulation, drainage density, distance from river, plan curvature, profile curvature, roughness, topographic wetness index (TWI), stream power index (SPI), sediment transport index (STI), normalized difference vegetation index (NDVI), normalized difference moisture index (NDMI), land use/land cover (LULC), soil, geology, temperature,and rainfall, were considered and analyzed within the GIS framework. The Extreme Gradient Boosting (XGBoost) model was applied to generate the flood hazard zones within the study area. Based on historical flood events within the study area, 1164 flooded and non-flooded points were identified and utilized to train and test the model. The ML model achieved high accuracy of 0.905 (90.5%), and an ROC-AUC score of 0.88. The generated flood susceptibility map indicated that 4.67%, 4.98%, 10.31%, 11.13%, and 68.91% of the basin are respectively at very high, high, moderate, low, and very low risk of flooding. The successful integration of GIS with machine learning validates the potential to improve flood hazard prediction and mitigation efforts in the Niger River basin and other similar flooding environments in Nigeria.

Keywords: Flood Hazard Mapping, Geographic Information System, Machine Learning, XGBoost, Niger River Basin

1.0 INTRODUCTION

Globally, flood is acknowledged to be one of the most frequent and devasting natural hazards that endangershuman lives, property, and infrastructure (Ibitoye *et. al.,* 2020). Across the globe, floods inflict unimaginable agony on people and economic hardship (Ghosh *et. al.,* 2023). TheUnited Nations Platform for Space-based Information for Disaster Management and Emergency Response (UN-SPIDER) in 2019, referred flooding to be the presence of water where it is not wanted.It often occurs when a river or water body exceeds its capacity. In 2022, the World Bank accounted for over 1.81 billion people across the globe to be directly exposed to flooding at a depth of over 0.15m. Significant economic and human losses result from the annual increase in the frequency of floods, which is made worse by intense precipitation, climate change, and fast urbanization.Although there are benefits to flooding, such as improved soil fertility, replenished water supplies, and the development or restoration of habitats for a variety of animals and plants, (Aldardasawi and Eren, 2021; Maharjan *et. al.,* 2024), the drawbacks of flooding are arguably greater than the benefits.The physical geography of low-lying coastal areas and river floodplains, which have consistently drawn human settlement over time, forms the basis for the phenomenon of flooding causing economic damage. The movement of people from rural to urban areas, or within cities, often leads them to settle in locations highly prone to flooding, thereby increasing their susceptibility in the absence of adequate flood defense mechanisms (Jha *et. al.,* 2022).

While the occurrence of flood spans across latitudes and longitudes, Nigeria is a prime example of a country facing the tremendous difficulties brought on by frequent floods. Flooding has been shown to have caused millions of deaths, destroyed businesses, poisoned water sources, increased the risk of sickness in several parts of Nigeria (Etuonovbe, 2011), and caused destruction of farmlands thereby having negative impacts on food security. The disruptive impacts of flooding are especially dangerous for Nigeria's agriculture industry. Large swaths of agricultural area are often submerged, resulting in crop failures and lower yields. Recent experiences in the nation indicate that the disastrous floods that occurred in 2012 and 2022 were the worst (Adaji *et. al.,* 2019). Over 14 states in the nation were reported to have been impacted by the 2012 flood (Tokunbo and Ezigbo, 2012).According to EM-DAT, the flood in 2012 was estimated to have impacted 7,000,867 lives, resulted in 363 fatalities, and caused economic damages of roughly $500,000 (Guha-Sapir *et. al.,* 2013; Komolafe *et. al.,* 2015).Nigeria has seen more frequent and severe floods recently, especially in the Niger River Basin. There are noticeable seasonal changes in the water level of the Niger River. Flood dangers are increased by this large floodplain and the tropical environment that is marked by heavy rainfall such as the Niger River basin.This necessitates the need for comprehensive nonstructural measures to assess the potential flood risk associated with the basin.

The utilization of GIS technology, by integrating geospatial datasets helps understand the complex interplay of flood-influencing factors and the invaluable insights it offers in flood hazard assessment and management (Komolafe *et. al.,* 2020). Machine learning which is referred to be a subsection of Artificial Intelligence serves as aninfluential tool in extracting patterns and knowledge from large datasets by learning from the data (Ighile*et. al.,* 2022). An effective way to enhance flood risk assessments is by utilizing GIS and machine learning to integrate several geographic datasets, such as those related to topography, hydrology, climate, environmental,and anthropogenicfeatures (Edamo*et. al.,* 2022). Creating more precise and educational flood hazard maps is feasible when GIS spatial analytic skills are combined with machine learning prediction capacity. This study aims to integrate Geographic Information System (GIS) and Machine Learning (ML) techniques in flood hazard mapping in the lower Niger River basin in Nigeria.

2.0 STUDY AREA

The area considered for this study is the combination of sub-basins of the Niger River in Nigeria (Figure 1). The area covers parts of 15 States of the Country, descending from the North-Central to the South-South. The study area is located between Latitudes $9^0$30’0’’ N and $4^0$28’0’’ N, and Longitudes $5^0$0’0’’ E and $9^0$0’0’’ E.The main Niger River being the largest river basin of western Africa runs in a crescent shape from the Guinea Highlands in Guinea, through Mali, Niger, and then Nigeria where it joins with the Benue River, its main tributary. The vast hydrological system of the Niger River, characterized by its susceptibility to frequent and severe flooding, offers an unparallel opportunity to thoroughly examine patterns of flooding.

The study area has a total area size of approximately 120,197 km^2. The topography of the study area is divided into regions: the coast, north-central plateaus, and the Niger-Benue rivers (Ighile*et. al.,* 2022). According to the Köppen climate classification system, the study area primarily falls under the Af(tropical rainfall) climate type characterized by high temperature across the year, the relative distribution of abundant rainfall, and lush vegetation.The maximum temperature in the south ranges from 30^0 C to 32^0 C, while in the north, the temperature ranges from 33^0 C to 35^0 C.The derived savannah, southern Guinea savannah, and humid forest are among the agroecological zones found in the research region. These zones are arranged from the north to the south.The geological features of the study area are primarily the basement complex, which occupies the northern part of the basin and is composed of rocks like schists, granites, and gneisses, and the sedimentary basins, which cover the southernportion of the basin and are composed of rocks like sandstones, shales, and limestones.

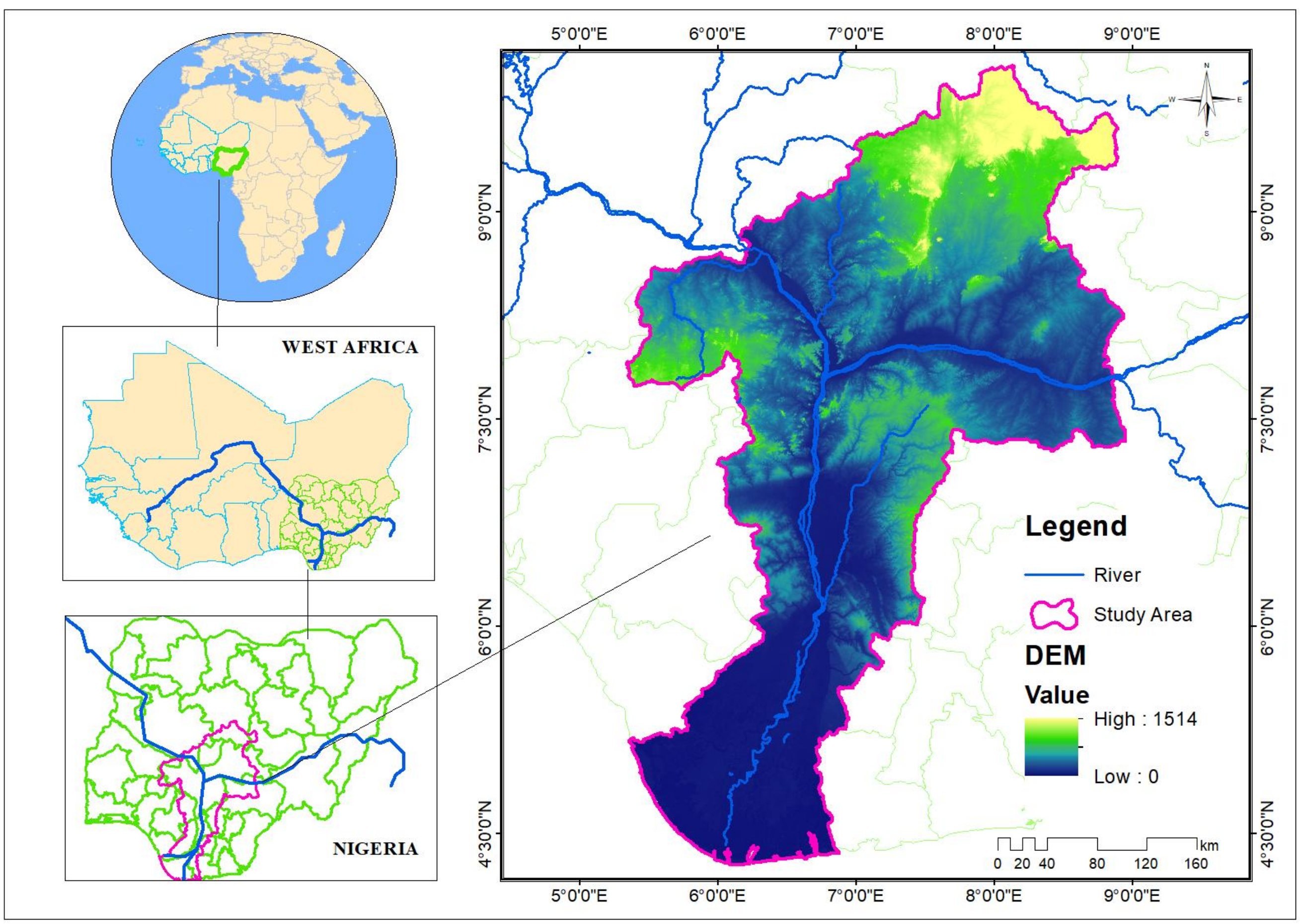


Figure 1: Location map of the study area

3.0 MATERIALS AND METHODS

3.1 Datasets

The study utilized datasets to map the flood-prone area within the sub-lower Niger River basin of Nigeria. The datasets are highlighted alongside their sources in Table 1. These datasets, their resolutions, and their sources are summarized in Table 1.

Table 1: Dataset used and their source

S/N	DATA	RESOLUTION	SOURCE
1.	SRTM DEM Data	30m	USGS Earth Explorer Website
2.	Sentinel-2Satellite Imagery Data	15m	Google Earth Engine
3.	Climate Data	30secs	WorldClim Website
4.	Soil Data	-	FAO Website
5.	Geology Data	-	Nigeria Geological Survey Website

3.2 Methods

The methodology adopted to generate the flood susceptibility map of this study is described in the methodology flowchart (Figure 2). The process involved selection of the flood influencing factors to be considered for the flood susceptibility mapping of the study area, acquisitionofdata for theextraction of flood factors and historical data for the flood inventory map, preparationof the flood influencing factors through geospatial analysis, selection suitable factors through multicollinearity investigation, splitting of the dataset for modeling to training sets and testing set, training the XGBoost machine learning model and evaluating the model, and finally, production ofthe flood susceptibility map.

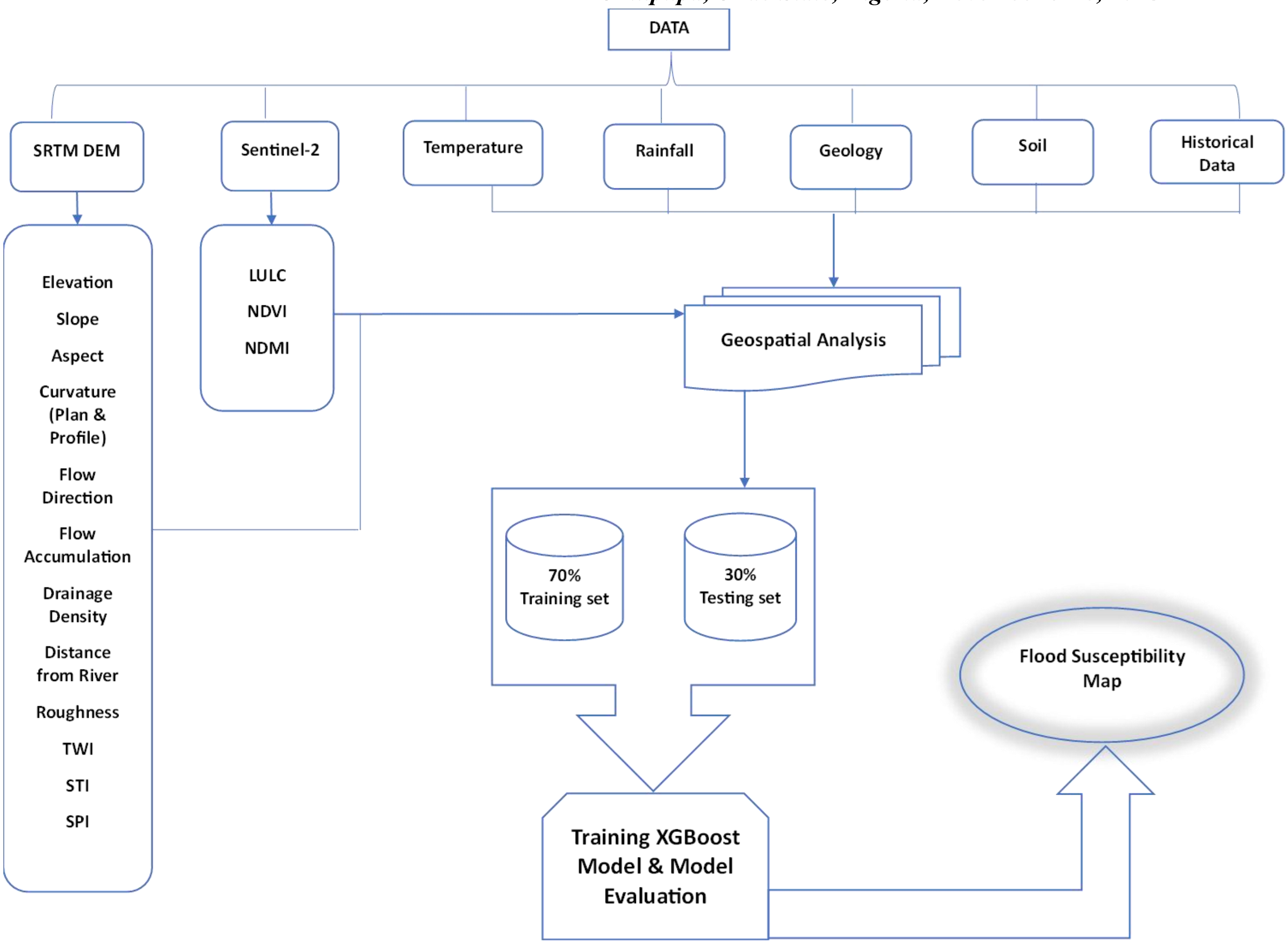


Figure 2: Methodology flowchart of the study

3.2.1 Flood Inventory Map

A flood inventory map is crucial in the comprehensive mapping and evaluation of flood hazards in an area (Ghosh *et. al.,* 2023). The flood inventory map of the study area was created utilizing historical flood occurrence data. The historical flood data were collated from various sources, including, the review of literature, satellite imagery, maps and photos of previous floods, and field surveys. A total of 1164 flooded and non-flooded points were collected within the study area, out of which 70% were considered as a training dataset to train the XGBoost machine learning model while the remaining 30% were considered as a testing dataset to validate the model.

3.2.2 Flood Influencing Factors

The flood-influencing factors considered for the prediction of flood susceptible zones in this research were selected based on existing literature by researchers and experts on flood mapping and modeling. The selected flood-influencing factors in this study are categorized into topographic, hydrologic, climatic, and environmental factors. The acquired remotely sensed and geospatial data were prepared for further image processing and analysis was used in producing each factor map. The digital elevation model (DEM) data was used to generate the topographic and hydrologic factors map such as slope, aspect, curvature (plan and profile), topographic wetness index (TWI), sediment transport index (STI), stream power index (SPI), distance from the river, and drainage density.

Elevation is regarded in many studies as one of the most crucial variables in flood mapping or modeling since a decrease in elevation increases the probability of flooding in a given location. The slope is another significant factor that influences how surface water flows (Edamo*et. al.,* 2022). The degree of slope has an impact on the pace of water infiltration and surface runoff. A region's chance of flooding reduces as its slope decreases. The index of aspects provides a more precise assessment of flood risk mapping (Edamo*et. al.,* 2022). It is well-recognized that low-lying, downslope areas may be more susceptible to floods. Curvature, sometimes referred to as its "slope of slope," (Longley et. al., 2011) was considered for this study.Selecting its two types (plan and profile curvatures), they influence the likelihood of floods by highlighting the divergent and convergent runoff zones. Areas in the study area that are concave and flat have a higher chance of flooding (Ighile*et. al.,* 2022). Using ArcGIS's spatial analyst tools, the slope aspect and both curvature maps were created from the DEM data.Roughness is identified to indicate disparity of elevation between adjoining pixels (Mahdizadeh and Perez, 2022). The map is generated from the DEM data following the equation:

$$Roughness = \frac{(FSmean - FSmin)}{(FSmax - FSmin)}$$

Where, FSmean, FSmin, and FSmax denote the mean, minimum, and maximum focal statistical layer, respectively.

The ability to distinguish between directions of flow is among the essential features of surface hydrology. (Edamo*et. al.,* 2022). The flow direction raster was generated from the fill DEM layer and it was further used to create the flow accumulation map. High flow accumulation indicates areas with a significant volume of water draining through them making the area more susceptible to flooding.The impact of drainage density on the amount of runoff that develops and exits the floodplain area makes it a crucialelementin flood susceptibility mapping (Avand *et. al.,* 2021).Since places that are susceptible to flooding are often located in close proximity to rivers, mapping flood susceptibility also heavily depends on the distance from the river feature.An individual or feature's likelihood of being impacted by flooding decreases with distance from the river (Liu *et al.,* 2021; Edamo*et. al.,* 2022). The drainage density and distance from the river layers were generated from the stream layer using the density tool and distance tool respectively in ArcGIS. TWI provides a concrete component in research on the incidence of floods since it indicates the amount of water present in a region (Ighile*et. al.,* 2022).An increase in the value of TWI in a given location denotes a significant probability of flooding. The STI which describes the particles in water moving due to water flow was also selected being one of the most important parameters used in flood modeling (Ighile*et. al.,* 2022).Similarly, SPI has a major impact on the hydrologic system (Edamo*et. al.,* 2022). The SPI calculates the erosive power of flowing water (Ighile*et. al.,* 2022). The TWI, STI, and SPIwere generated from DEM data using the Raster Calculator tool in ArcGIS according to the following equations respectively:

$$TWI = \ln \frac{\propto}{tan\beta}$$

Where, $\propto$ is the upstream discharge at a certain point, and tan β represents the slope in radians.

$$STI = \left(\frac{As}{22.13}\right)^{0.6}\left(\frac{sin\beta}{0.0896}\right)^{1.3}$$

Where, As is the area of the catchment/flow accumulation and β is the slope.

$$SPI = \propto * tan\beta$$

Where, $\propto$ is the upstream release at a certain point, and tanβ represents the slope in radians. Furthermore, the anthropogenic and environmental factors map including land use/land cover (LULC), normalized difference vegetation (NDVI), and normalized difference moisture index (NDMI) maps were generated from the Sentinel-2 satellite imagery.In addition to being a significant contributing factor to flooding, the LULC was chosen because it clarifies the connection between floods and human activities in the natural environment. The supervised classification method was adopted using Google Earth Engine to classify the Sentinel-2 satellite imagery to generate the LULC map of the study area. The NDVI and NDMI are both vegetation indices used to determine respectively the health and moisture contents of the vegetation in the study area. Using Google Earth Engine, each index map was generated following the equations respectively:

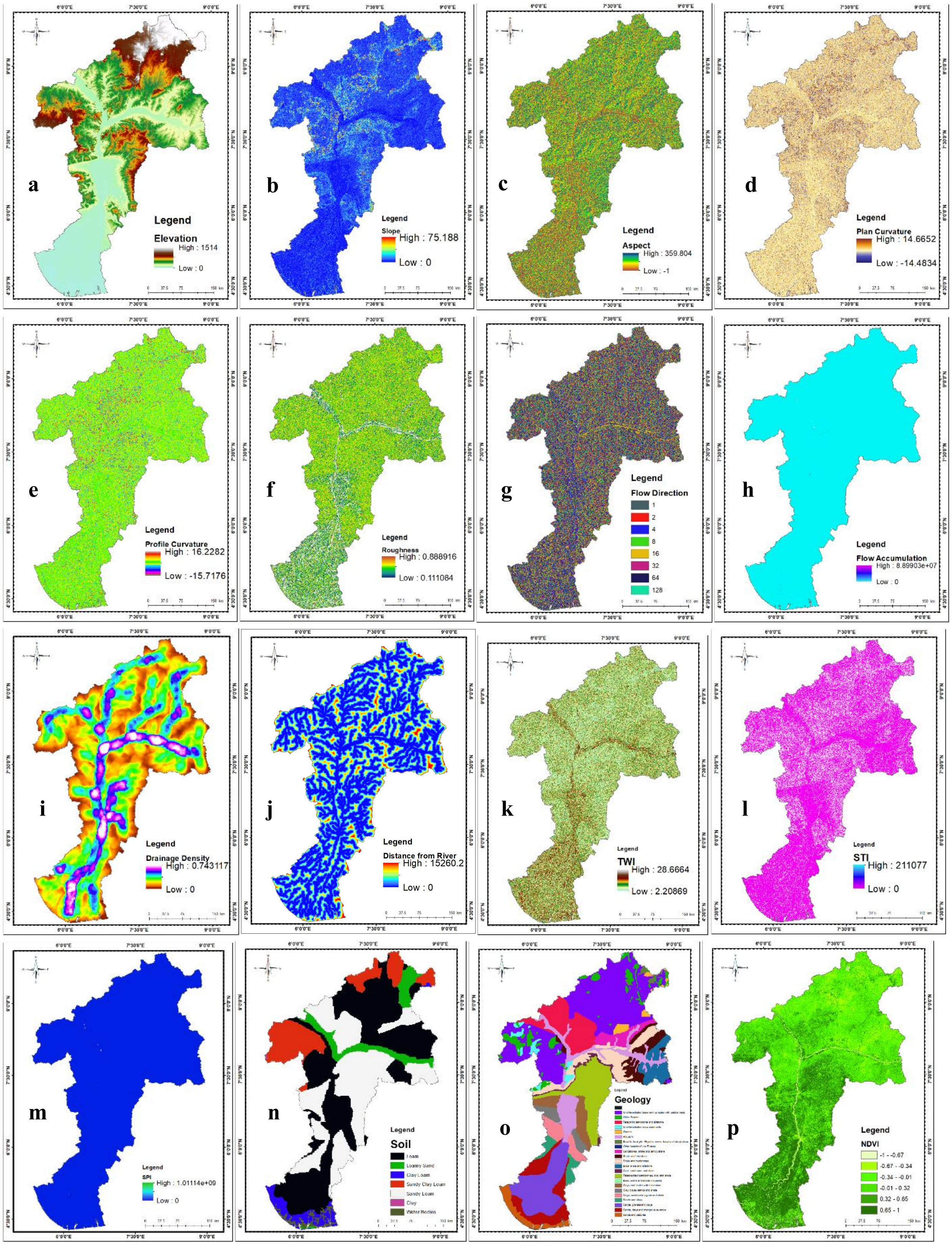
$$NDVI = \frac{NIR - R}{NIR + R}$$

Where NIR is the Near Infrared band of Sentinel-2, and R is the Red band of Sentinel-2.

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR}$$

Where, NIR is the Near Infrared band of Sentinel-2, and SWIR is the Short-Wave Infrared band of Sentinel-2.

Additionally, the soil and geology of the study area were selected for the flood susceptibility mapping. The type of soil of an area affects the drainage process of the area based on the soil characteristics. Soil types with low water permeability porosity or fine texture are known to be highly prone to floods. Similarly, the permeability of rock determines the rate of infiltration of water into the sub-zones.


a
Legend
Elevation
High : 1514
Low : 0
b
Legend
Slope
High : 75.188
Low : 0
c
Legend
Aspect
High : 359.804
Low : -1
d
Legend
Plan Curvature
High : 14.6652
Low : -14.4834
e
Legend
Profile Curvature
High : 16.2282
Low : -15.7176
f
Legend
Roughness
High : 0.888916
Low : 0.111084
g
Legend
Flow Direction
1
2
4
8
16
32
64
128
h
Legend
Flow Accumulation
High : 8.89903e+07
Low : 0
i
Legend
Drainage Density
High : 0.743117
Low : 0
j
Legend
Distance from River
High : 15260.2
Low : 0
k
Legend
TWI
High : 28.6664
Low : 2.20869
l
Legend
STI
High : 211077
Low : 0
m
Legend
SPI
High : 1.01114e+09
Low : 0
n
Legend
Soil
Loam
Loamy Sand
Clay Loam
Sandy Clay Loam
Sandy Loam
Clay
Water Bodies
o
Legend
Geology
p
Legend
NDVI
-1 - -0.67
-0.67 - -0.34
-0.34 - -0.01
-0.01 - 0.32
0.32 - 0.65
0.65 - 1

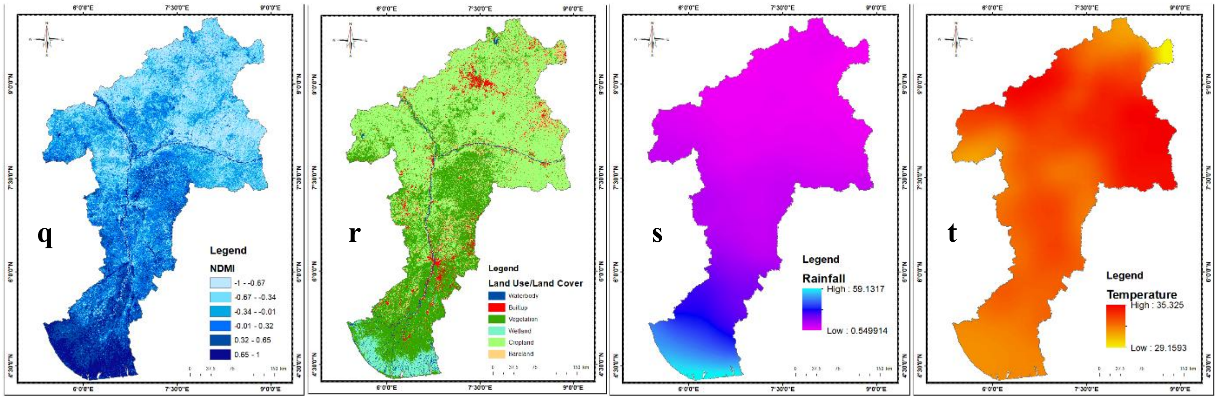


Figure 3: Flood influencing factors (a) Elevation, (b) Slope, (c) Aspect, (d) Plan Curvature, (e) Profile Curvature, (f) Roughness, (g) Flow Direction, (h) Flow Accumulation, (i) Drainage Density, (j) Distance from Rivers, (k) Topographic Wetness Index (TWI), (l) Sediment Transport Index (STI), (m) Stream Power Index (SPI), (n) Soil Types, (o) Geology, (p) Normalized Difference Vegetation Index (NDVI), (q) Normalized Difference Moisture Index (NDMI), (r) Land Use/Land Cover (LULC), (s) Annual Rainfall,and (t) Annual Temperature.

3.2.3 Extreme Gradient Boosting (XGBoost) Machine Learning Model

The ensemble machine-learning model, Extreme Gradient Boosting (XGBoost) was put out by Chen and Guestrin (2016). Through the use of gradient boosting, it applies machine learning methods to solve problems in parallel via tree-boosting. To prevent overfitting and enhance the model's capacity for generalization, regularization approaches wereemployed. The computational efficiency of the XGBoost algorithm is high. XGBoost can handle missing values, therefore missing values does notrequire any particular processing when applying the model (Ren *et. al.,* 2024). In this study, the XGBoost model was implemented in the prediction of the flood hazard zones of the study area. The 20 flood-influencing factors were considered the independent variables while the computed inventory data comprising the flooded and non-flooded dataset are considered the target variable.

3.2.4 Multicollinearity Investigation

Examining the link between variables is essential when making predictions in order to determine the level of correlation between the variables and ensure high prediction accuracy. When there is a high correlation between two or more independent variables, multicollinearity occurs which in turn causes errors in modeling and reduces the accuracy of the result, hence there is a need to remove one out of the two or two out of the three multicollinear variables leaving one out for better accuracy.The multicollinearity investigation in this study was conducted using Pearson's correlation coefficients technique. By pairing all the independent variables, Pearson's correlation coefficient revealed how strong the linear relationship is between each of them. The Pearson correlation coefficient is calculated following this formula (Ighile*et. al.,* 2022):

$$r = \frac{\sum(xi - x)(yi - y)}{\sqrt{(xi - x)^2 \sum(yi - y)^2}}$$

Where, r is the correlation coefficient, x_i and y_i are the values of variables x and y, and x and y are the mean values of each variable.

The values of the Pearson's correlation coefficient varied from -1 to 1. A perfect positive correlation is represented by a value of 1, a perfect negative correlation by a value of -1, and no correlation is represented by a value of 0. A table referred to as a correlation matrix, which displays the values of the linear associations between each of the independent variables, was used to convey the results of Pearson's correlation coefficient study. From the Pearson correlation coefficient matrix (Table 2), none of the factors shows to be multicollinear. They have no high multicollinearity and strong correlations, making them all necessary for the modeling.

Table 2: Correlation matrix of the flood influencing factors and the target variable with each causative factor IGR value

Factors: 1-Elevation; 2-Aspect; 3-Flow Direction; 4-Flow Accumulation; 5-Drainage Density; 6-Distance from the River; 7-Plan Curvature; 8-Profile Curvature; 9-Roughness; 10-Slope; 11-SPI; 12-STI; 13-TWI; 14-Soil; 15-Geology; 16-Precipitation; 17-Temperature; 18-NDMI; 19-NDVI; 20-LULC

Factor	Target	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0.405	1													
2	0.187	0.242	1												
3	0.038	0.085	0.091	1											
4	0.010	0.103	0.078	0.065	1										
5	0.509	0.496	0.162	0.072	0.054	1									
6	0.250	0.332	0.125	0.058	0.095	0.412	1								
7	-0.006	0.033	0.049	-0.004	0.055	0.030	-0.026	1							
8	0.048	0.044	0.163	0.060	0.088	0.033	0.077	-0.363	1						
9	-0.321	-0.329	-0.349	0.017	-0.038	-0.219	-0.143	-0.087	0.144	1					
10	0.261	0.373	0.249	0.055	0.115	0.252	0.190	0.034	0.152	-0.319	1				
11	0.029	0.196	0.115	0.131	0.335	0.094	0.173	0.100	0.152	0.011	0.271	1			
12	0.211	0.234	0.184	0.009	0.050	0.169	0.103	-0.031	0.089	-0.257	0.473	0.096	1		
13	0.274	0.294	0.333	0.026	0.117	0.201	0.138	0.280	0.005	-0.614	0.519	0.096	0.397	1	
14	-0.060	-0.064	-0.008	0.016	0.024	-0.087	-0.022	0.031	0.020	0.014	0.000	0.074	-0.022	0.011	1
15	0.301	0.248	0.113	0.032	0.065	0.276	0.154	0.077	0.010	-0.139	0.158	0.125	0.144	0.147	-0.060
16	0.056	0.375	0.073	0.002	-	-	0.003	0.017	0.011	-	0.144	-	0.112	0.154	0.417

					0.020	0.020				0.196		0.013			
17	-0.040	-0.102	-0.001	-0.048	-0.028	-0.126	-0.053	0.036	0.021	-0.052	-0.019	-0.010	-0.010	0.052	0.384
18	0.062	0.226	0.045	-0.026	0.031	-0.010	0.037	-0.011	-0.007	-0.158	0.069	0.008	0.065	0.102	0.216
19	0.026	-0.089	0.035	0.030	0.051	0.069	0.022	0.027	0.030	0.026	0.020	-0.013	-0.004	-0.001	0.083
20	0.020	0.127	0.033	-0.049	0.054	-0.052	-0.002	0.044	0.016	-0.138	0.005	0.017	0.058	0.100	0.126

3.2.5 Model Training and Evaluation

Training of a machine learning modelcan be described as the act of feeding the machine learning algorithm with labeled data to make it discover the underlying patterns and correlations of the dataset.To help the model understand the relationships and patterns in the dataset, the data must be exposed to the model. For the implementation of the XGBoost model, the computed dataset was split into a training dataset and a testing dataset in a 70:30. The training dataset was used to train the model,and the testing dataset was used to test the model.Analyzing the training dataset reveals a dataset's fitness while analyzing the testing dataset reveals the model's prediction ability (Bhattarai *et. al.,* 2024). Accuracy, f1-score, recall, precision, and ROC-AUC were used to evaluate and validate the model performance.

4.0 RESULTS

4.1 Flood Susceptibility Map

The mapping of flood susceptibility involves predicting the likelihood of flooding in different areas, while floodplain maps show where water will go based on past events.The generated flood susceptibility map of the study area is shown in Figure 4.Based on the natural breaks' classification method in ArcGIS, the flood susceptibility map was classified into five susceptibility zones: very low, low, moderate, high, and very high. Major cities and towns within the study area and States of the country were overlaid on the flood susceptibility layer.

It can be inferred from the resultant map that there is a non-uniform distribution of flood susceptibility across the study area. The study identified the central region states such asKogi State (beingthe confluence area of the Niger and Benue Rivers), the Nasarawa State, the Delta-Anambra States region, and some southern region; coastal statesof the study area to majorly exhibit higher susceptibility to flooding. In quantifying the area coverage of each susceptibility classes, 4.67% of the area is under the very high susceptible zone, followed by 4.98% under the high susceptible zone, 10.31% under the moderate susceptible zone, 11.13% under the low susceptible zone, and the very low susceptible zone taking the largest area of 68.91% of the study area (Table 3).

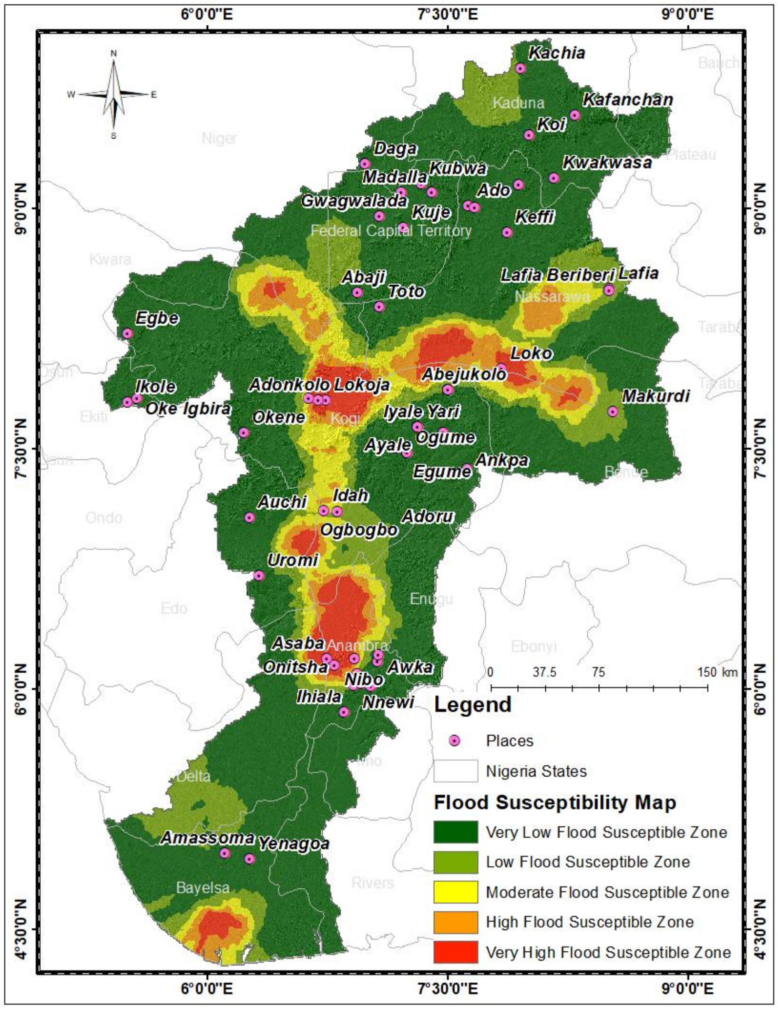


Figure 4: Flood Susceptibility Map

Table 3: Percent and Area extent of each susceptibility zone

Susceptibility Zone	**Area Extent (km^2)**	**Percent (%)**
Very low	82704.00	68.91
Low	13354.35	11.13
Moderate	12374.91	10.31
High	5981.12	4.98
Very high	5607.63	4.67

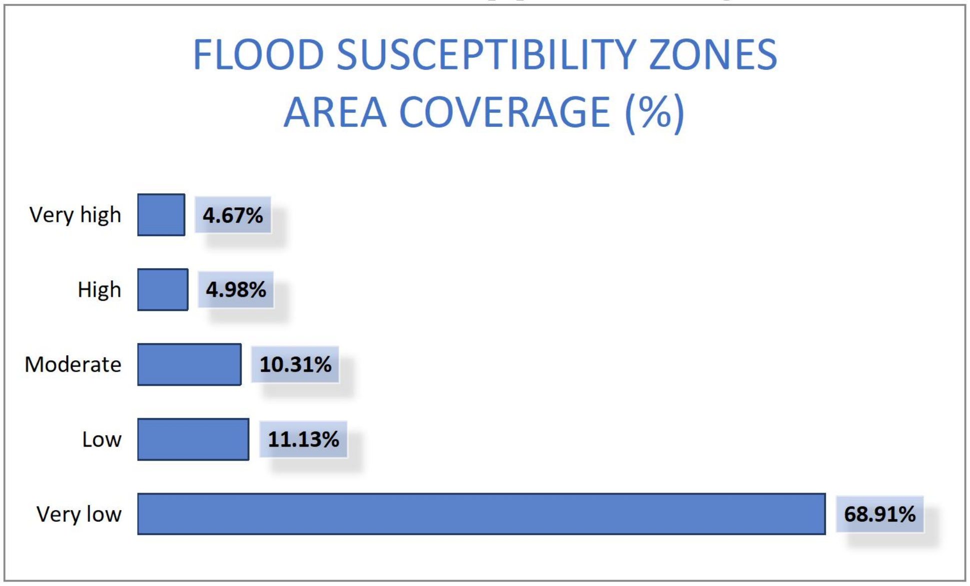


Figure 5:Chart showing the percentage of flood susceptible zones in the study area

4.2 Model Performance Evaluation and Validation

The result from the evaluation and validation of the model used for the flood susceptibility mapping shows a high performance of the model. The indicators used for the evaluation of the model's performance are precision, recall, f1-score, accuracy, and receiver operating characteristics-area under the curve (ROC-AUC). The model presented an overall accuracy of 90.5% while the precision, recall, and f1-score presented similar values of 0.94 and 0.81 for the non-flooded (0) and flooded points datasets respectively (Table 4). The receiver operating characteristics area under the curve (ROC-AUC) score is shown to be 0.88 (Figure 6).

Table 4: Performance parameters of the machine learning model

ML Model	**Flood Status**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
XGBoost	0	0.94	0.94	0.94	90.5%
	1	0.81	0.81	0.81	

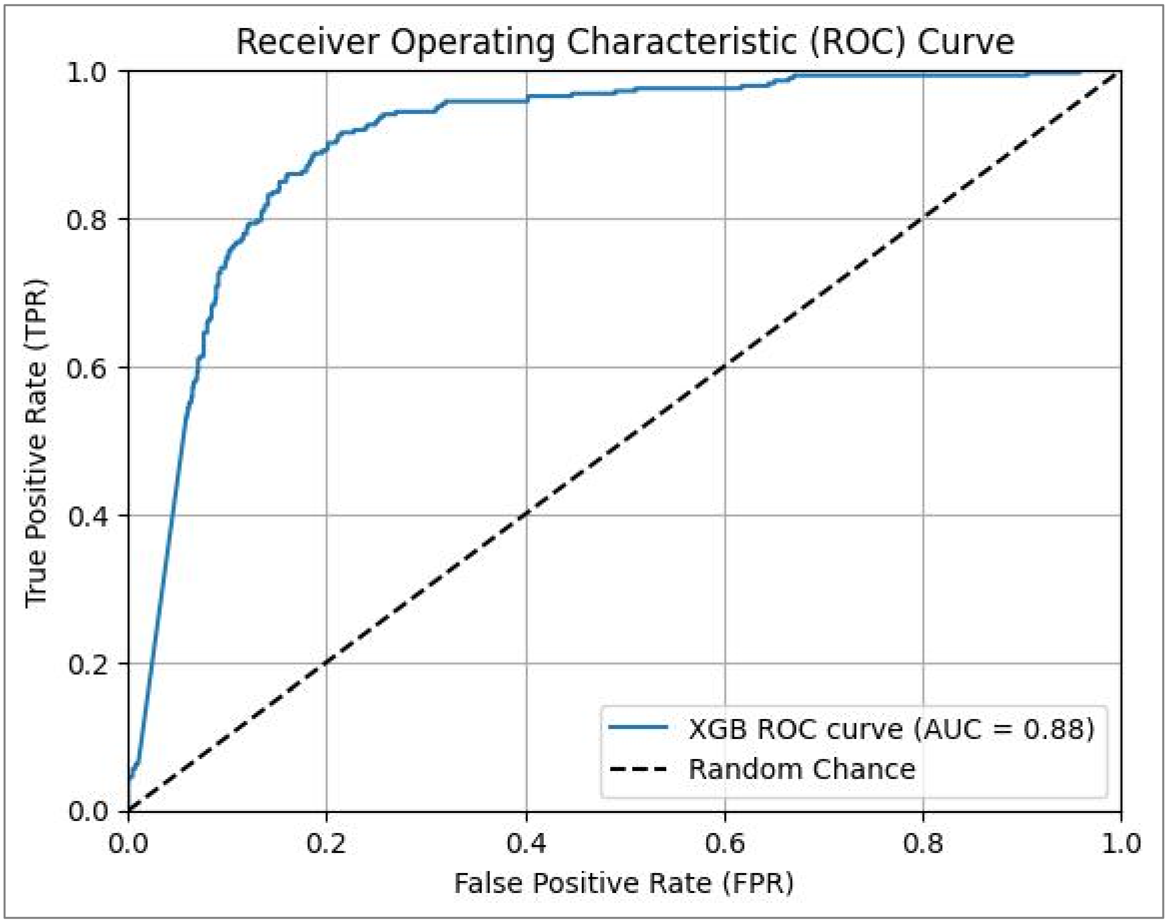


Figure 6: ROC-AUC plot of the machine learning model used for flood susceptibility mapping

5.0 DISCUSSION

The provision of the necessary spatial framework by GIS for data management, data analysis, and data visualization, and the strive by a machine learning algorithm to extract valuable insights from the geospatial and hydrological information aided a great deal in predicting and mapping flood-prone areas in the study area. Integration of GIS and machine learning techniques proved to be a powerful and valuable approach to enhancing flood hazard mapping. Given that the Niger River in Nigeria, Benue River being its main tributary, experiences flooding annually, hence mapping the degree of flood hazard in the river basin is important. Different studies have utilized machine learning techniques with GIS techniques to predict and map hazards such as flood (Ighile*et. al.,* 2022; Bhattarai *et. al.,* 2024), drought (Alawsi*et. al.,* 2022), and groundwater (Nourani and Mousavi, 2016).

In this study, the effectiveness of combining machine learning and GIS to generate enhanced results in flood hazard prediction was demonstrated. The extreme gradient boosting (XGBoost) model was applied to predict the flood hazard zones in the study area. Twenty factors influencing flood in the Niger River basin were used to produce the flood hazard map of the study area. Rainfall, flow accumulation, elevation, distance from the river, slope, and land use/land cover factors were demonstrated to be some of the most significant factors impacting floods in the area.

Rainfall as a dynamic factor has a significant impact on floods in the study area. The part of the Niger River in Nigeria which is downstream of the whole river suffers from flooding majorly during the rainfall season mostly when there is heavy rainfall at the upstream causing overflow downstream of the river. The accumulation of water flow from the upstream proves its effect mostly at the confluence area (i.e. Kogi State), the Delta-Anambra State, and the coastal area of the study area. Elevationand slope are also critical factors influencing floods in the study area. The speed and volume of runoff in the study area are relatively due to theelevation and steepness of the slope of the study area.The proximity to the low-lying river determines the level of risk to the flood hazard. In areas with dense land cover such as forests, the speed andamount of runoffand the resulting damage decrease (Avand *et. al.,* 2021). The increase in urbanization over time causing deforestation exposed the study area to flooding. The activities of humans such as settlements, farming, etc. increase the high susceptibility to flood in the areas. Due to the extensive activities of humans in Nigeria, flood probability may worsen over time (Ighile*et. al.,* 2022).

The XGBoost machine learning model evaluation proves the high performance of the model in this study, hence it's useful in mapping flood-prone areasin other studies. Incorporating additional data sources such as high-resolution imagery, and real-time hydrological data, could improve the model performance and provide more detailed flood predictions.

6.0 CONCLUSION

Leveraging the strengths of GIS and machine learning techniques, this study predicted and produced the flood susceptibility map of the study area. Twenty flood influencing factors which are elevation, slope, aspect, flow direction, flow accumulation, drainage density, distance from river, plan curvature, profile curvature, roughness, topographic wetness index (TWI), stream power index (SPI), sediment transport index (STI), normalized difference vegetation index (NDVI), normalized difference moisture index (NDMI), land use/land cover (LULC), soil, geology, temperature,and rainfall, were selected to successfully develop the flood susceptibility map.

A flood inventory map was initially created using historical flood occurrences data in Nigeria. The generated data were further divided into 70% training dataset and 30% testing dataset. Indicators such as precision, recall, f1-score, accuracy, and ROC-AUC were used to evaluate the performance of the XGBoost machine learning model. The model having a high-performance accuracy of 90.5% and a ROC-AUC score of 0.88. The flood susceptibility map developed in this study help highlight areas that are prone to flooding in the study area ranking the susceptibility between low to high levels.The successful integration of GIS with machine learning validates the potential to improve flood hazard prediction and mitigation efforts in the Niger River basin and other similar flooding environments in Nigeria.The findings from this study, proves the effectiveness of the integration as a relevant approach for academic and policymakers in comprehending flood occurrences. One of the limitations of this study is the lack of access to some areas for field survey. However, this study suggests that further similar research should focus on refining machine learning model by incorporating additional variable, and comparing different models for the flood hazard mapping.

REFERENCES

Ajin, R.S., R.R. Krishnamurthy, M. Jayaprakash and P.G. Vinod. (2013). Flood hazard assessment of Vamanapuram River Basin, Kerala, India: An approach using Remote Sensing & GIS techniques. Advances in Applied Science Research. 4. 263-274.

Alawsi, M.A.; Zubaidi, S.L.; Al-Bdairi, N.S.S.; Al-Ansari, N.; Hashim, K. Drought Forecasting: A Review and Assessment of the Hybrid Techniques and Data Pre-Processing. Hydrology 2022, 9, 115. https://doi.org/10.3390/hydrology9070115

Avand, M., Moradi, H. R., &Lasboyee, M. R. (2021). Spatial Prediction of Future Flood Risk: An Approach to the Effects of Climate Change. *Geosciences*, *11*(1), 25. https://doi.org/10.3390/geosciences11010025

Bhattarai, Y., Duwal, S., Sharma, S., &Talchabhadel, R. (2024). Leveraging machine learning and open-source spatial datasets to enhance flood susceptibility mapping in transboundary river basin. *International Journal of Digital Earth*, *17*(1).https://doi.org/10.1080/17538947.2024.2313857

Cabrera, N., & Lee, N. (2019). Flood-Prone Area Assessment Using GIS-Based Multi-Criteria Analysis: A Case Study in Davao Oriental, Philippines. *Water*, *11*(11), 2203. https://doi.org/10.3390/w11112203

Edamo, M. L., Ukumo, T. Y., Lohani, T. K., Ayana, M. T., Ayele, M. A., Mada, Z. M., & Abdi, D. M. (2022). A comparative assessment of multi-criteria decision-making analysis and machine learning methods for flood susceptibility mapping and socio-economic impacts on flood risk in Abela-Abaya floodplain of Ethiopia. *Environmental Challenges*, *9*, 100629. https://doi.org/10.1016/j.envc.2022.100629

Etuonovbe A. K. (2011). The devastating effect of flooding in Nigeria. In FIG Working Week. 2011, May. Accessed 10 March 2015; Available at: http://www.fig.net/pub/fig2011/papers/ts06j/ts06j_etuonovbe_5002.pdf.

Ghosh, A., Chatterjee, U., Pal, S. C., Islam, A. R. M. T., Alam, E., & Islam, M. K. (2023). Flood hazard mapping using GIS-based statistical model in vulnerable riparian regions of sub-tropical environment. *Geocarto International*, *38*(1). https://doi.org/10.1080/10106049.2023.2285355

Ibitoye, M. O., Komolafe, A. A., Adegboyega, A. a. S., Adebola, A. O., & Oladeji, O. D. (2019). Analysis of vulnerable urban properties within river Ala floodplain in Akure, Southwestern Nigeria. *Spatial Information Research*, *28*(4), 431–445. https://doi.org/10.1007/s41324-019-00298-6

Ighile, E.H.; Shirakawa, H.; Tanikawa, H. Application of GIS and Machine Learning to Predict Flood Areas in Nigeria. Sustainability 2022, 14, 5039. https://doi.org/10.3390/su14095039

Komolafe, A. A., Adegboyega, S. a. A., &Akinluyi, F. O. (2015). A Review of Flood Risk Analysis in Nigeria. *American Journal of Environmental Sciences*, *11*(3), 157–166. https://doi.org/10.3844/ajessp.2015.157.166

Maharjan, M., Timilsina, S., Ayer, S., Singh, B., Manandhar, B., &Sedhain, A. (2024). Flood susceptibility assessment using machine learning approach in the Mohana-Khutiya River of Nepal. *Natural Hazards Research*. https://doi.org/10.1016/j.nhres.2024.01.001

Mahdizadeh Gharakhanlou, N.; Perez, L. Spatial Prediction of Current and Future Flood Susceptibility: Examining the Implications of Changing Climates on Flood Susceptibility Using Machine Learning Models. Entropy 2022, 24, 1630.https://doi.org/10.3390/e24111630

Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N., & Ghazali, A. H. B. (2017). Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Geomatics Natural Hazards and Risk*, *8*(2), 1080–1102. https://doi.org/10.1080/19475705.2017.1294113

Morjani, Z. E. a. E., Ennasr, M. S., Elmouden, A., Idbraim, S., Bouaakaz, B., & Saad, A. (2016). Flood Hazard Mapping and Modeling Using GIS Applied to the Souss River Watershed. In *The handbook of environmental chemistry* (pp. 57–93). https://doi.org/10.1007/698_2016_69

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

Nourani, V.; Mousavi, S. Spatiotemporal groundwater level modeling using hybrid artificial intelligence-meshless method. J. Hydrol. 2016, 536, 10–25. http://doi.org/10.1016/j.jhydrol.2016.02.030

Park, S. J., & Lee, D. K. (2020). Prediction of coastal flooding risk under climate change impacts in South Korea using machine learning algorithms. *Environmental Research Letters*, *15*(9), 094052. https://doi.org/10.1088/1748-9326/aba5b3

Ren, H.; Pang, B.; Bai, P.; Zhao, G.; Liu, S.; Liu, Y.; Li, M. Flood Susceptibility Assessment with Random Sampling Strategy in Ensemble Learning (RF and XGBoost). Remote Sens. 2024, 16, 320. https://doi.org/10.3390/rs16020320

Tokunbo, A., &Ezigbo, B. (2012). Floods Claims 363 Lives, Displace 2.1 People Says NEMA