Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatial Heterogeneity of Flood–Salinity WASH Vulnerability: An Integrated FS-WVI and CatBoost–SHAP Assessment across Eight Unions of Subarnachar Upazila, Noakhali, Bangladesh

Md. Iqbal Hossain

hossain.iqbalrahi@gmail.com

Pabna University of Science and Technology https://orcid.org/0009-0003-2693-4178

Research Article

Keywords: Water, Sanitation and Hygiene (WASH), Compound Flood–Salinity Hazard, Explainable Machine Learning, Coastal Vulnerability, Remote Sensing, Drinking-Water Salinity, Coastal Bangladesh, Climate Adaptation

Posted Date: August 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10662024/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
The authors declare potential competing interests as follows: The authors declare that they have no financial or personal conflicts of interest that may have influenced the findings or conclusions of this study.

Statement on Participant Consent: This study was approved by the Pabna University of Science and Technology Ethics Committee on 31th March,2025. Prior to participation, all respondents were informed about the purpose, procedures, and voluntary nature of the study. Verbal consent was obtained from each participant, followed by written consent prior to the household survey. No minors, patients, or legally incapacitated individuals were involved in data collection.

Spatial Heterogeneity of Flood–Salinity WASH Vulnerability: An Integrated FS-WVI and CatBoost–SHAP Assessment across Eight Unions of Subarnachar Upazila, Noakhali, Bangladesh.

Abstract

Coastal Bangladesh is increasingly affected by the combined impacts of flooding and salinity, creating serious challenges for water, sanitation, and hygiene (WASH), particularly in low-lying coastal communities. The unexpected flash flood of August 2024 further demonstrated the need for integrated assessments that consider both environmental hazards and household vulnerability to support effective adaptation planning. This study developed a Flood–Salinity WASH Vulnerability Index (FS-WVI) by integrating remote sensing observations with household survey data to evaluate multidimensional WASH vulnerability in Subarnachar Upazila, Noakhali. The findings revealed pronounced spatial variation in vulnerability among the eight unions. Charaman Ullah 0.652, Mohammadpur 0.651, and Char Jubilee 0.647 exhibited the highest FS-WVI values, whereas Char Clark 0.297 showed the lowest vulnerability level. The results further demonstrated that WASH vulnerability was not determined by hazard exposure alone. Instead, different combinations of exposure, household sensitivity, and adaptive capacity deficit shaped the final vulnerability pattern across the study area. Exposure was the dominant vulnerability component in Char Jubilee, Charaman Ullah, and Mohammadpur, while Char Clark, Char Wapda, East Char Bata, and Char Bata were primarily influenced by household sensitivity. Char Jabbar was mainly affected by adaptive capacity deficit. The predictive analysis achieved 67.5% classification accuracy with a ROC–AUC of 0.721, while feature ranking consistently identified tubewell depth and the Compound Hazard Index as the strongest factors associated with household drinking-water salinity. The explainability analysis further confirmed that the influence of compound hazards varied according to groundwater access and household conditions, reinforcing that WASH vulnerability results from the interaction of environmental and socio-economic factors rather than hazard intensity alone. The proposed framework provides a practical and transferable approach for evidence-based WASH planning, climate adaptation, and targeted intervention in coastal Bangladesh.

Keywords: Water, Sanitation and Hygiene (WASH); Compound Flood–Salinity Hazard; Explainable Machine Learning; Coastal Vulnerability; Remote Sensing; Drinking-Water Salinity; Coastal Bangladesh; Climate Adaptation.

1. Introduction

Coastal communities face growing climate and water threats like floods, cyclones, saltwater intrusion, unsafe drinking water, poor sanitation, and disease(TIME, 2026). These problems are severe in Bangladesh, a low-lying, densely populated delta. Coastal areas often suffer from cyclones and saltwater in freshwater sources(ADAB, 2025; Mahmuduzzaman et al., 2014) and sudden inland floods can add to the crisis when communities are unprepared(Rabbani et al., 2018). In August 2024, eastern Bangladesh saw severe flooding in Feni, Noakhali, and Cumilla, caused by heavy rain and upstream water flows(The daily star, 2024). In fact, Indian authority opened the dam at the Feni border, and the Porshuram river channel opened to allow homogenous water flow, which caused flash flooding across the entire area(The Business Standard, 2025). The event was a sudden transboundary flood with poor warning and coordination(Reuters, 2024) and such flooding can mix with existing salinity, affecting water, sanitation, and hygiene (WASH). Recent research now uses broader frameworks to assess vulnerability, rather than looking at single hazards alone(Ahona et al., 2025). But many studies still treat flooding, salinity, health, and socio-economic issues separately, and misses the combined ways households face WASH insecurity(Rahman et al., 2026). Studies also show salinity heavily impacts water access, health, and livelihoods in coastal Bangladesh.

This problem is especially clear in Noakhali's coastal region. Estuaries, low-lying land, remote settlements, and unequal access to services create strong local differences in vulnerability(Mahmood et al., 2020). Subarnachar Upazila is a good case study because its unions differ in physical exposure, household conditions, water access, and the ability to respond to disruptions. The 2024 eastern flood showed that Noakhali is not safe from serious damage during major flood emergencies(Khatun et al., 2024). However, its long-term coastal risks are more often discussed in terms of cyclones, tidal processes, erosion, and salinity. The United Nations reported that Noakhali was one of the worst affected districts during that emergency(UN, 2024). At the same time, broader evidence from the coast shows that salinity pressure is not evenly spread that it cannot be understood from administrative statistics alone(Bhuyan et al., 2023). Remote sensing and GIS are useful tools for mapping flood extent and surface salinity over large areas. Household surveys can record things that satellites cannot observe directly. These include water shortages, disease outbreaks, sanitation failures, treatment costs, coping practices, housing conditions, and institutional support. Therefore, it is important to combine geospatial data with household evidence. This helps us understand how an environmental hazard becomes a real lived problem for water, sanitation, and hygiene. The present study adopts this combined approach. It links satellite-derived flood and salinity indicators with survey data from 384 households across eight unions.

Research on coastal vulnerability is growing, but important gaps remain. First, flood and salinity are usually treated as separate hazards with little attention paid to their joint effects on household WASH vulnerability(Bhuyan et al., 2023). Existing geospatial studies have demonstrated the benefits of including multiple coastal hazards but are generally based on broad physical threat rather than household drinking water, sanitation, health sensitivity and coping capacity(Deb et al., 2024; Iii et al., n.d.). Second, traditional vulnerability indices can identify high- and low-vulnerability areas(Himel et al., 2025), but often provide limited explanation of the nonlinear

relationships between environmental hazards, groundwater access, household resources, and WASH outcomes(Raisa et al., 2024). Machine-learning studies can also enhance prediction, but black-box outputs are difficult to translate to local planning when the direction and relative contribution of predictors are not explained(Hosan et al., 2026). The present research was motivated by these gaps. The main conceptual contribution is an integrated framework linking a remotely sensed Compound Hazard Index (CHI) with a household-level Flood–Salinity WASH Vulnerability Index (FS-WVI), an explainable machine-learning model, and union-level spatial interpretation. The novelty lies not in using GIS, vulnerability indices, or machine learning separately, but in connecting them within one analytical pathway to examine where vulnerability is concentrated, which factors influence drinking-water salinity, and why the dominant structure of vulnerability differs among nearby coastal unions.

To bridge these gaps, the study integrates spatial hazard assessment, multidimensional vulnerability analysis, and explainable machine learning. Compound environmental pressure is the product of flood extent, Soil Salinity Index-1 (SI1) and Normalised Difference Salinity Index (NDSI). Household indicators are grouped into exposure, sensitivity and adaptive capacity. The household survey collects local information on flooding, water salinity, health, income, housing, tubewell depth, coping strategies and institutional support, whereas remote sensing and GIS provide spatial evidence. We apply CatBoost to study drinking-water salinity, which can capture nonlinearities in heterogeneous environmental and household data, and has benefits over traditional Random Forest models due to ordered boosting and less bias in predictions(Nishat et al., 2025). SHAP is then used to explain the magnitude and direction of the effects of predictors. Together, these methods link regional hazard patterns to household vulnerability and drinking-water outcomes.

The household view is important because the 2024 flood was an unusual, fast-moving shock. Humanitarian reports noted serious flooding, road damage, isolation, and major disruption across the region(UNICEF, 2024). Feni was especially hard hit by flash floods from heavy rain and upstream flows(OCHA, 2024). These events raise a key question that a flood map cannot answer: why do households in the same event face different levels of WASH vulnerability? One household may be in a lower-risk area but still be highly vulnerable due to poor water sources, low income, weak housing, sanitation failure, health issues, or limited support. Another household may face higher physical risk but reduce its vulnerability through better water storage, alternative sources, stronger housing, or help from income, social ties, or institutions. This reasoning guides the study's three questions: (i) how flood and salinity pressures vary across space; (ii) what factors affect household drinking-water salinity; and (iii) how exposure, sensitivity, and adaptive capacity combine across unions. Thus, the survey does not just confirm spatial data, it provides the social evidence to explain why outcomes differ within the same coastal upazila.

The significance of this study is in the development of a locally grounded but potentially transferable framework for regional WASH and climate-risk planning. The analysis uses eight unions of Subarnachar and connects environmental hazard mapping with household conditions and explainable prediction. First, it develops an FS-WVI for identifying spatial differences of exposure, sensitivity, adaptive capacity and overall vulnerability. It then uses CatBoost and SHAP

to analyse the impact of environmental and socio-economic factors on the salinity of household drinking-water. Finally, it combines the index components with explainable machine-learning evidence to identify the dominant vulnerability structure of each union. This scope can help planners identify areas where priority should be given to hazard reduction, safe-water infrastructure, sanitation and health protection, or stronger adaptive resources. The framework is also of broader relevance to coastal districts where climate change, salinity, flash flooding and unequal household capacity interact. The study links regional earth observation data with local household experience to inform more spatially targeted WASH interventions and a better understanding of multidimensional vulnerability in coastal Bangladesh.

2. Materials and Methods

2.1 Study Area

Subarnachar Upazila is located in the southern part of Noakhali District, Bangladesh (***Figure.1***), covering approximately 576 km²(Wikipedia, 2026). It consists of eight unions with a total population of 375,910 according to the Bangladesh Population and Housing Census 2022. The upazila is characterized by low-lying floodplains, numerous tidal channels, and an agriculture-dependent economy(Sarwer et al., 2016). Rice cultivation, fishing, and livestock rearing are the primary livelihood sources. Due to its coastal setting, the area frequently experiences river flooding, tidal inundation, and increasing soil and water salinity.

Subarnachar was selected because it represents one of the most flood- and salinity-prone coastal upazilas in Bangladesh. The southern unions are directly exposed to tidal flooding and prolonged waterlogging during the monsoon season. Many households have limited economic resources and live in geographically isolated settlements with inadequate WASH infrastructure. These conditions make Subarnachar an appropriate case study for assessing compound flood–salinity impacts on household WASH vulnerability.

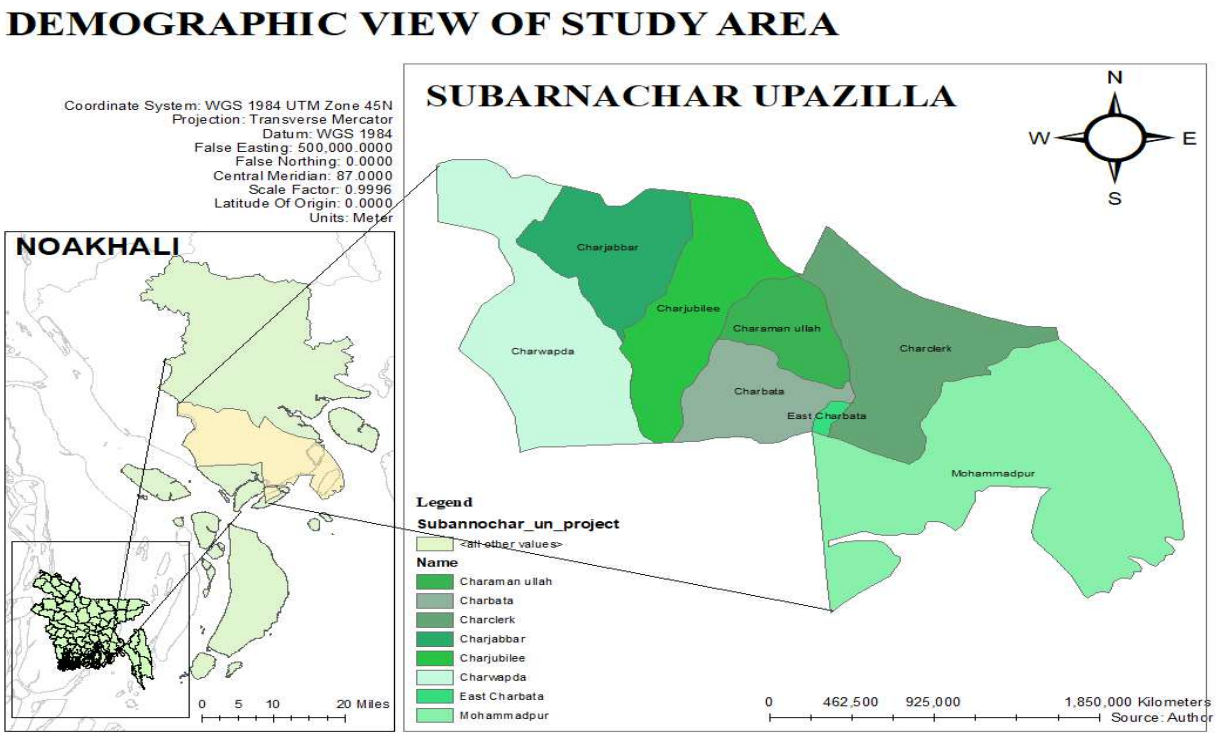


Figure 1. Geographic location and administrative boundaries of the eight unions in Subarnachar Upazila, Noakhali District, Bangladesh.

2.2 Conceptual Framework

This study employed an integrated methodological framework that combines remote sensing, household survey data, geographic information systems (GIS), and explainable machine learning to assess flood–salinity-induced WASH vulnerability in Subarnachar Upazila. The framework was designed to evaluate household vulnerability from both environmental and socio-economic perspectives. The overall process consisted of four sequential stages: data collection, data preprocessing, vulnerability assessment, and machine learning-based interpretation. Household survey data provided information on flood impacts, drinking water, sanitation, health, and socio-economic conditions, while satellite-derived flood and salinity datasets represented the spatial distribution of environmental hazards.

In the first analytical stage, flood extent, Soil Salinity Index-1 (SI1), and the Normalized Difference Salinity Index (NDSI) were processed from remote sensing data to develop the Compound Hazard Index (CHI), representing the combined effects of flood and salinity exposure. The CHI was integrated with household survey indicators to construct the Exposure Index (EI), Sensitivity Index (SI), and Adaptive Capacity Index (ACI). Before index development, all indicators were standardized using min–max normalization to ensure comparability among variables measured on different scales. These three component indices were subsequently combined to calculate the Flood–Salinity WASH Vulnerability Index (FS-WVI), which was mapped to identify spatial variations in vulnerability among the eight unions.

The second analytical stage focused on understanding the factors influencing household drinking-water salinity conditions using the CatBoost machine learning algorithm. Environmental and socio-economic variables were used as predictors, while SHAP (SHapley Additive exPlanations) was applied to quantify the contribution of each variable to model predictions and improve result interpretability. Finally, the spatial vulnerability patterns obtained from the FS-WVI were integrated with the dominant determinants identified by the CatBoost–SHAP model. This integrated interpretation enabled the identification of the principal drivers of vulnerability in each union and provided a scientific basis for location-specific WASH adaptation and management strategies under compound flood–salinity conditions.

2.3 Data Collection

2.3.1 Household Survey

Primary data were collected through a structured household questionnaire administered across the eight unions of Subarnachar Upazila between August and September 2024. The minimum sample size was determined using Cochran's formula for a large population(Ahmed, 2024). A 95% confidence level, 5% margin of error, and maximum response variability of 0.50 were assumed to ensure adequate representation of household conditions(SpssServices, 2025).

$$n_0 = \frac{Z^2 p(1-p)}{e^2} \tag{1}$$

where $n_0$is the required sample size, $Z$is the standard normal value at the selected confidence level, $p$is the estimated population proportion, and $e$is the acceptable margin of error. Using $Z = 1.96$, $p = 0.50$, and $e = 0.05$, the required sample size was:

$$n_0 = \frac{(1.96)^2(0.50)(1-0.50)}{(0.05)^2} = 384.16 \approx 384 \tag{2}$$

Thus, a target sample of 384 households was adopted. The total sample was then distributed proportionally among the eight unions according to their population shares(Arnab, 2017). The required sample for each union was calculated as:

$$Sample_i = \left(\frac{Population_i}{Total\ Population}\right) \times 384 \tag{3}$$

where $Sample_i$is the required number of households in the i-th union, $Population_i$is the population of the i-th union, and $Total\ Population$is the total population of Subarnachar Upazila. This proportional allocation ensured that unions with larger populations received larger target samples.

During field implementation, 384 completed questionnaires were successfully collected, including 80 households from Char Clark, 79 from Char Jabbar, 70 from Char Jubilee, 44 from East Char Bata, 39 from Char Bata, 32 from Char Amanullah, 28 from Char Wapda, and 12 from Mohammadpur. Minor differences between the calculated allocation and completed interviews occurred because of household availability and field accessibility during data collection. The final distribution is presented in ***Table 1***.

Table 1. Population distribution, proportional sample allocation, and completed household surveys across the eight unions of Subarnachar Upazila.

Union	Population	Sampling Weight (%)	Target Sample	Completed Survey
Char Clark	80,480	21.4	82	80
Char Jabbar	74,923	19.9	77	79
Char Jubilee	68,766	18.3	70	70
East Char Bata	43,111	11.5	44	44
Char Bata	38,143	10.1	39	39
Char Amanullah	30,840	8.2	32	32
Char Wapda	27,728	7.4	28	28
Mohammadpur	11,919	3.2	12	12
Total	**375,910**	**100**	**384**	**384**

2.3.2 Remote Sensing Data

Environmental data were derived from satellite-based remote sensing products processed in Google Earth Engine (GEE). Flood extent from Sentinel-1, Soil Salinity Index-1 (SI1), and the Normalized Difference Salinity Index (NDSI) from Landset-8/9 were generated for the entire Noakhali District using imagery acquired during August 2024, corresponding to the peak monsoon period. These datasets were selected to represent the compound environmental hazards associated with flooding and salinity intrusion in coastal Bangladesh. The processed raster datasets were subsequently exported for GIS-based analysis and Compound Hazard Index (CHI) development.

2.3.3 Data Preprocessing

The household survey dataset was examined before analysis to improve data quality and consistency. Incomplete records with missing information were removed, and abnormal observations were identified through descriptive statistics and boxplot inspection. All categorical responses were converted into standardized numerical scores according to predefined indicator classes. These scored variables were subsequently normalized using min–max normalization before calculating the Exposure Index, Sensitivity Index, Adaptive Capacity Index, and the final FS-WVI.

The remote sensing datasets were preprocessed within ArcGIS Pro after export from Google Earth Engine. Flood, SI1, and NDSI raster layers were clipped using the official Subarnachar Upazila boundary to extract the study area. All raster datasets were transformed into a common projected coordinate system and aligned to ensure consistent spatial resolution and geographic extent. The processed raster layers were then used to calculate the Compound Hazard Index (CHI) and derive union-level environmental indicators for subsequent spatial and machine learning analyses.

2.3.4 Variable Selection and Indicator Classification

The study variables were selected based on the conceptual framework of flood–salinity-induced WASH vulnerability and the availability of reliable household survey and remote sensing data.

The selected indicators were classified into three vulnerability components: Exposure, Sensitivity, and Adaptive Capacity, while the environmental variables used in the CatBoost model were also identified according to their analytical roles. A complete description of the selected indicators, their data sources, measurement units, scoring criteria, and component classification is presented in **Table 2**.

Table 2. Classification of study variables, data sources, measurement units, scoring system, and analytical application.

Dimension	Indicator	Measurement / Coding	Vulnerability Direction	Analytical Role
Exposure	Compound Hazard Index (CHI)	Continuous index	Higher = higher vulnerability	EI; CatBoost predictor
	Days of home submergence	Days / ordinal class		
	Flood depth at home	Feet / ordinal class		
	Crop and livestock loss	Percentage / loss class		
Sensitivity	Water taste/saltiness	1–5 scale	Higher = higher vulnerability	SI; CatBoost target*
	Days without safe water after flood	Days / ordinal class	Higher = higher vulnerability	SI
	Diarrhea during last flood, Skin disease or itching, Toilet flooded during last event	No / Yes	Yes = higher vulnerability	
	Treatment cost	BDT / cost class	Higher = higher vulnerability	
	Toilet type	Ordinal sanitation class	Poorer type = higher vulnerability	
	Household members	Number / size class	Higher = higher vulnerability	SI; CatBoost predictor
Adaptive Capacity	Monthly income	BDT / income class	Higher = lower vulnerability	ACI; CatBoost predictor
	Housing type	1 = Kacha; 2 = Semi-pucca; 3 = Pucca	Higher = lower vulnerability	
	Tubewell depth	Ordinal depth class	According to capacity scoring	
	Coping water strategy	Derived strategy score	Higher = lower vulnerability	
	Govt/NGO WASH support	Support score	Higher = lower vulnerability	

2.4 Compound Hazard Index Development

The Compound Hazard Index (CHI) was developed to represent the spatial co-occurrence of flood exposure and salinity-related surface conditions within flooded areas(Akter et al., 2024). Before integration, SI1 and NDSI were normalized to a common scale to reduce differences in their original value ranges. The normalized salinity indicators were then averaged and multiplied by the binary flood extent layer. In this formulation, flooded pixels receive CHI values according to their combined salinity condition, while non-flooded pixels receive zero values. Therefore, CHI should be interpreted as an index of combined flood–salinity pressure within the mapped flood extent,

rather than as a continuous measure of general salinity hazard across the entire study area(Fidelis et al., 2025).

$$\mathrm{CHI} = \mathrm{FE} \times \left(\frac{\mathrm{SI1_{norm}} + \mathrm{NDSI_{norm}}}{2}\right) \quad (4)$$

Where, FE represents binary flood extent, with 1 indicating flooded pixels and 0 indicating non-flooded pixels. $\mathrm{SI1_{norm}}$and $\mathrm{NDSI_{norm}}$represent the normalized salinity indicators. The resulting CHI raster was aggregated by union using zonal statistics to support spatial comparison and subsequent vulnerability assessment.

2.5 Development of the Flood–Salinity WASH Vulnerability Index (FS-WVI)

2.5.1 Vulnerability Indicators

The Flood–Salinity WASH Vulnerability Index (FS-WVI) was developed using the vulnerability framework proposed by the Intergovernmental Panel on Climate Change (IPCC), which defines vulnerability as a function of exposure, sensitivity, and adaptive capacity. Indicators were selected from household survey data and the Compound Hazard Index (CHI) based on their relevance to flood–salinity-induced WASH conditions (***Table 2***).

The **Exposure Index (EI)** represents the degree to which households are physically exposed to compound flood and salinity hazards. It includes environmental and flood-related indicators such as the Compound Hazard Index (CHI), flood depth inside the house, days of home submergence, and crop/livestock loss, which directly reflect the intensity of hazard exposure.

The **Sensitivity Index (SI)** describes the susceptibility of households to adverse WASH impacts after flood events. Indicators such as household size, drinking-water salinity, water shortage duration, diarrheal disease, skin disease, and sanitation conditions were used to represent the potential health and social consequences of flood–salinity hazards. The **Adaptive Capacity Index (ACI)** measures the ability of households to prepare for, respond to, and recover from these hazards. Monthly income, housing condition, tube-well depth, coping strategies for safe water, and Government/NGO support were considered indicators of household adaptive capacity.

2.5.2 Indicator Scoring and Data Normalization

All survey indicators were converted into standardized numerical scores according to predefined vulnerability classes (***Table 2***). Higher scores represented greater exposure and sensitivity, whereas higher adaptive-capacity scores indicated stronger resilience to flood–salinity impacts.

Because the indicators were measured using different units and scales, min–max normalization was applied to transform all variables into a comparable range between 0 and 1(Akanbi et al., 2015). The normalized value of each indicator was calculated using Equation 5(Sprinkle, 2024):

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (5)$$

where X_{norm} is the normalized value, (X) is the original indicator value, and X_{min} and X_{max} represent the minimum and maximum values of each indicator, respectively.

2.5.3 Exposure Index (EI)

The Exposure Index (EI) was calculated as the arithmetic mean of all normalized exposure indicators(Hirai & Comim, 2022). The selected variables included the Compound Hazard Index (CHI), flood depth inside the house, days of home submergence, and crop/livestock loss. The EI was computed using Equation (6):

$$EI = \frac{\sum_{i=1}^{n} E_i}{n} \tag{6}$$

where (E_i) represents the normalized exposure indicators and (n) is the total number of exposure variables.

2.5.4 Sensitivity Index (SI)

The Sensitivity Index (SI) was calculated as the average of all normalized sensitivity indicators representing household health, water accessibility, and sanitation conditions(Hashemi, 2020). The SI was determined using Equation (7):

$$SI = \frac{\sum_{i=1}^{n} S_i}{n} \tag{7}$$

where (S_i) denotes the normalized sensitivity indicators and (n) is the number of sensitivity variables.

2.5.5 Adaptive Capacity Index (ACI)

The Adaptive Capacity Index (ACI) was estimated as the arithmetic mean of all normalized adaptive-capacity indicators, including income, housing condition, tube-well depth, coping strategy, and institutional support. The ACI was calculated using Equation (8)(Pulina et al., 2026):

$$ACI = \frac{\sum_{i=1}^{n} A_i}{n} \tag{8}$$

where (A_i) represents the normalized adaptive-capacity indicators and (n) is the number of adaptive-capacity variables.

2.5.6 Composite Flood–Salinity WASH Vulnerability Index (FS-WVI)

The final Flood–Salinity WASH Vulnerability Index (FS-WVI) was obtained by integrating the three vulnerability components(Panpeng & Ahmad, 2017). Since higher adaptive capacity reduces overall vulnerability, the adaptive-capacity component was inversed before aggregation. The composite index was calculated using Equation:

$$FS - WVI = \frac{EI + SI + (1 - ACI)}{3} \tag{9}$$

where **EI**, **SI**, and **ACI** denote the Exposure Index, Sensitivity Index, and Adaptive Capacity Index, respectively. The resulting FS-WVI values ranged from **0** (lowest vulnerability) to **1** (highest vulnerability). Equal weighting was adopted because all vulnerability indicators were considered equally important in representing flood–salinity-induced WASH vulnerability and no objective basis existed for assigning differential weights. Union-level FS-WVI values were subsequently classified into five vulnerability categories (very low, low, moderate, high, and very high) using

the Natural Breaks (Jenks) classification method in ArcGIS Pro for spatial visualization and interpretation.

2.5.7 Spatial Aggregation and Mapping

The normalized household-level indicator values were aggregated at the union level by calculating the arithmetic mean of all surveyed households within each union. Consequently, union-level Exposure Index (EI), Sensitivity Index (SI), Adaptive Capacity Index (ACI), and the composite Flood–Salinity WASH Vulnerability Index (FS-WVI) were obtained for the eight unions of Subarnachar Upazila. The aggregated indices were joined to the union boundary shapefile in ArcGIS Pro using the union name as the common attribute. Finally, thematic maps of EI, SI, ACI, and FS-WVI were produced using the Natural Breaks (Jenks) classification method to visualize the spatial distribution of household WASH vulnerability across the study area.

2.6 CatBoost Machine Learning Model

The CatBoost algorithm was employed to identify the environmental and socio-economic factors influencing household drinking-water salinity under compound flood–salinity conditions(Saber et al., 2021). CatBoost is a gradient boosting decision tree algorithm that efficiently handles numerical(Medium, 2026) and categorical variables while minimizing overfitting through ordered boosting(Kulkarni, 2022). Compared with the conventional Random Forest algorithm, CatBoost generally provides higher predictive accuracy, better representation of nonlinear relationships, and improved robustness for relatively small and heterogeneous environmental datasets(Guan et al., 2025; Nogueira et al., 2026). Furthermore, CatBoost was selected because it performs well with relatively small datasets, minimizes prediction bias through ordered boosting, and requires limited parameter tuning compared with other gradient boosting algorithms(Prokhorenkova et al., 2018). These advantages make it suitable for integrating household survey data with environmental hazard indicators.

The binary drinking-water salinity condition (fresh water = 1, saline water = 0) was used as the target variable, while the Compound Hazard Index (CHI), days of home submergence, flood depth inside the house, crop/livestock loss, monthly income, housing type, household members, tube-well depth, coping strategy, and Government/NGO support were used as predictor variables. The dataset was randomly divided into **80% training data** and **20% testing data** using stratified sampling to preserve the class distribution.

CatBoost constructs the final prediction by sequentially combining multiple decision trees, which can be expressed as(Prokhorenkova et al., 2018):

$$F_M(x) = \sum_{m=1}^{M} \eta \, h_m(x) \tag{10}$$

where $F_M(x)$ is the final model prediction, $h_m(x)$ denotes the decision tree generated at the m_{th} boosting iteration, (M) is the total number of trees, and η is the learning rate controlling the contribution of each tree. The model was implemented in Python using 500 boosting iterations, a tree depth of 6, a learning rate of 0.05, Logloss as the loss function, and **AUC** as the evaluation

metric(Méndez et al., 2026). Automatic balanced class weights were applied to improve model performance, while Accuracy, Precision, Recall, F1-score, ROC–AUC, and the confusion matrix were used for model evaluation.

2.7 SHAP Explainability Analysis

To improve the interpretability of the CatBoost model, SHapley Additive exPlanations (SHAP) were employed to quantify the contribution of each predictor variable to model predictions. SHAP is based on cooperative game theory and estimates the contribution of each feature by comparing the prediction obtained with and without that feature(John et al., 2026). Unlike conventional feature importance methods, SHAP provides both the magnitude and direction of feature influence, making the machine learning model more transparent and interpretable(Ozsahin et al., 2024).

The SHAP value for each feature was calculated as(Lundberg & Lee, 2017):

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \tag{11}$$

where (ϕ_i) represents the SHAP value of feature (i), (F) denotes the complete set of predictor variables, (S) is any subset of features excluding (i), and $f(S \cup \{i\}$ represents the prediction function of the trained CatBoost model. Positive SHAP values indicate that a variable increases the probability of predicting fresh-water conditions, whereas negative SHAP values indicate a greater likelihood of saline drinking-water conditions(Sahwell & Solo-Gabriele, 2025). SHAP was preferred because it provides both the magnitude and direction of feature contributions at the global and individual prediction levels(Tempel et al., 2025).

The TreeExplainer algorithm was adopted because it is specifically optimized for tree-based ensemble models and efficiently computes exact SHAP values(Kopanja et al., 2024). Three complementary visualizations were generated to interpret model behaviour(Mountasir et al., 2026). The SHAP summary plot illustrated the overall distribution and direction of feature effects, the SHAP bar plot ranked variables according to their mean absolute SHAP values (global feature importance)(Huang et al., 2020), and the SHAP dependence plot examined the relationship between the Compound Hazard Index (CHI) and model predictions while accounting for interactions with other influential variables. These analyses enabled the identification of the principal environmental and socio-economic determinants of household drinking-water salinity.

2.8 Integration of FS-WVI and Explainable Machine Learning for Spatial Interpretation

The final stage of the study integrated the Flood–Salinity WASH Vulnerability Index (FS-WVI) with the CatBoost–SHAP results to interpret the spatial heterogeneity of household WASH vulnerability across the eight unions of Subarnachar Upazila. This integration aimed to explain not only the level of vulnerability but also the dominant vulnerability-related conditions and key environmental and socio-economic factors associated with spatial differences among the unions.

First, the union-level vulnerability structure was evaluated using the Exposure Index (EI), Sensitivity Index (SI), and Adaptive Capacity Deficit (ACD). Since higher adaptive capacity reduces vulnerability, the ACI was transformed into a vulnerability-oriented measure as ACD = (1 – ACI). This transformation ensured a consistent direction among the three components, where higher EI, SI, and ACD values all represented greater vulnerability. The component with the highest normalized value was identified as the dominant vulnerability-related component within each union. This comparison determined whether vulnerability was primarily associated with exposure, household sensitivity, or limited adaptive capacity.

Subsequently, the dominant vulnerability-related components were interpreted together with the global feature importance and SHAP results obtained from the CatBoost model. The SHAP-derived ranking of predictor variables was used to explain the major factors associated with household drinking-water salinity under compound flood–salinity conditions. This integrated interpretation linked union-level vulnerability structures with household-level environmental and socio-economic drivers, supported comparison among the eight unions, and helped identify priority areas for targeted WASH intervention and flood–salinity adaptation planning.

2.9 Statistical Analysis

All spatial processing, data analysis, and machine learning modelling were conducted using specialized GIS and Python-based software packages. The primary software, libraries, and their analytical applications are summarized in ***Table 3***.

Table 3. Software and analytical tools used in this study.

Software/Library	Library	**Version**	**Primary application**
Google Earth Engine	-	Cloud Platform	Processing flood extent, SI1, and NDSI raster datasets
ArcGIS Pro	-	3.1.4	Raster processing, zonal statistics, spatial analysis, and thematic mapping
Microsoft Excel	-	2021	Household survey coding, scoring, and preliminary data organization
Python		3.13	Data preprocessing, statistical analysis, and machine learning implementation
	Pandas	-	Data cleaning, manipulation, and preprocessing
	NumPy	-	Numerical computation and array processing
	Scikit-learn	-	Data splitting and model performance evaluation
	CatBoost	-	Gradient boosting classification model
	SHAP	-	Explainable machine learning and feature contribution analysis
	Matplotlib	-	Visualization of SHAP outputs and statistical figures

2.10 Methodological Workflow

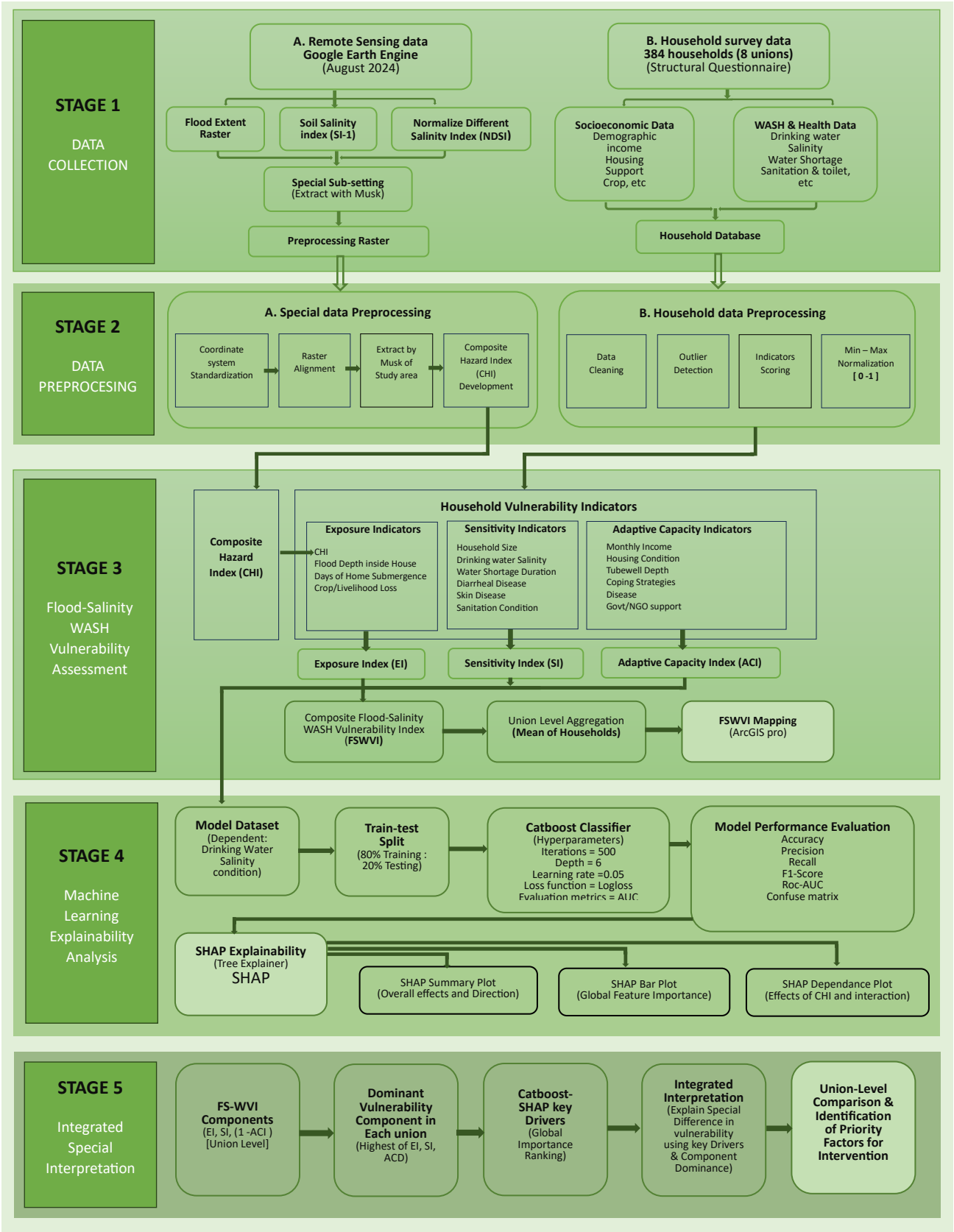
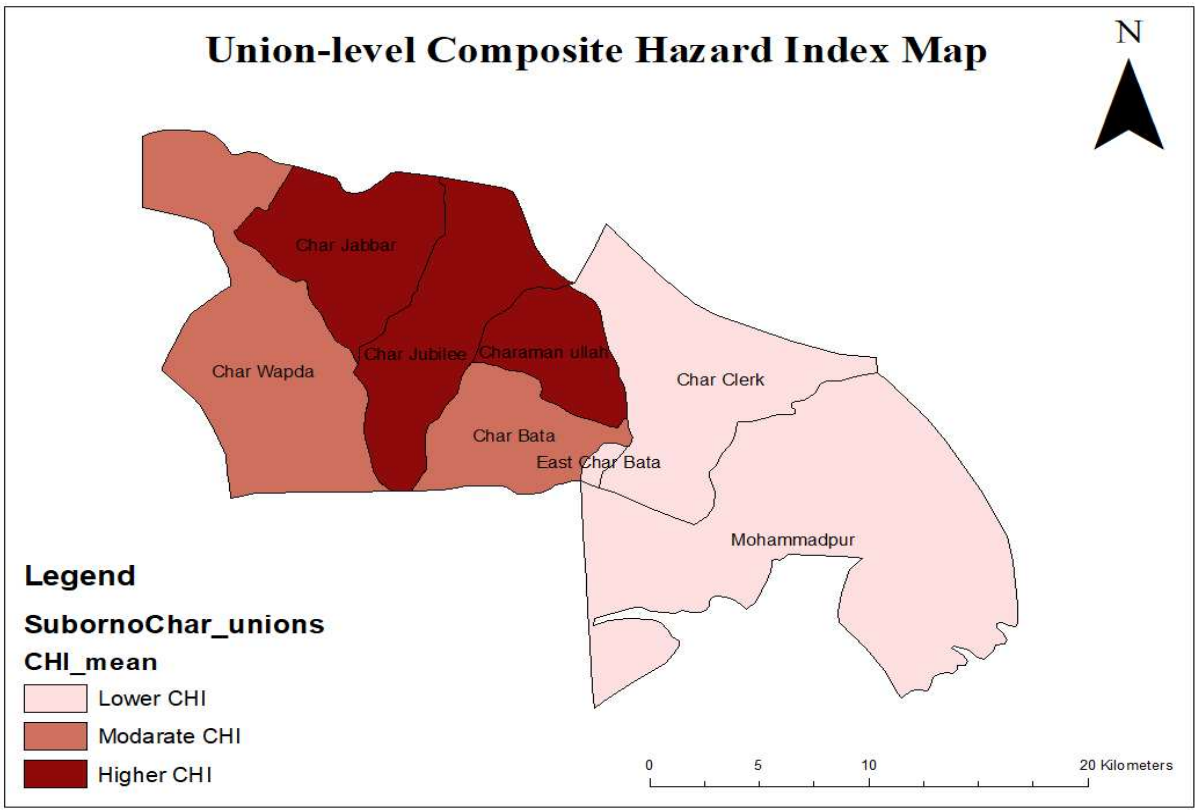


Figure 2. Detail Methodological Workflow

3.Results

3.1 Compound Hazard Index (CHI) Distribution

The spatial distribution of the Compound Hazard Index (CHI) is presented in ***Figure 3***. The CHI values varied considerably among the eight unions, ranging from *0.003769* to *0.149914*. Char Amanullah recorded the highest CHI value *0.149914*, followed by Char Jubilee *0.133644* and Char Jabbar *0.129376*. Moderate CHI values were observed in Char Wapda *0.080788* and Char Bata *0.074287*. In contrast, Char Clark *0.017543*, Mohammadpur *0.007818*, and East Char Bata *0.003769* exhibited comparatively lower CHI values. As illustrated in ***Figure 3***, higher CHI concentrations were primarily distributed in the western and central unions, whereas lower CHI concentrations were observed in the eastern and southeastern parts of Subarnachar Upazila.



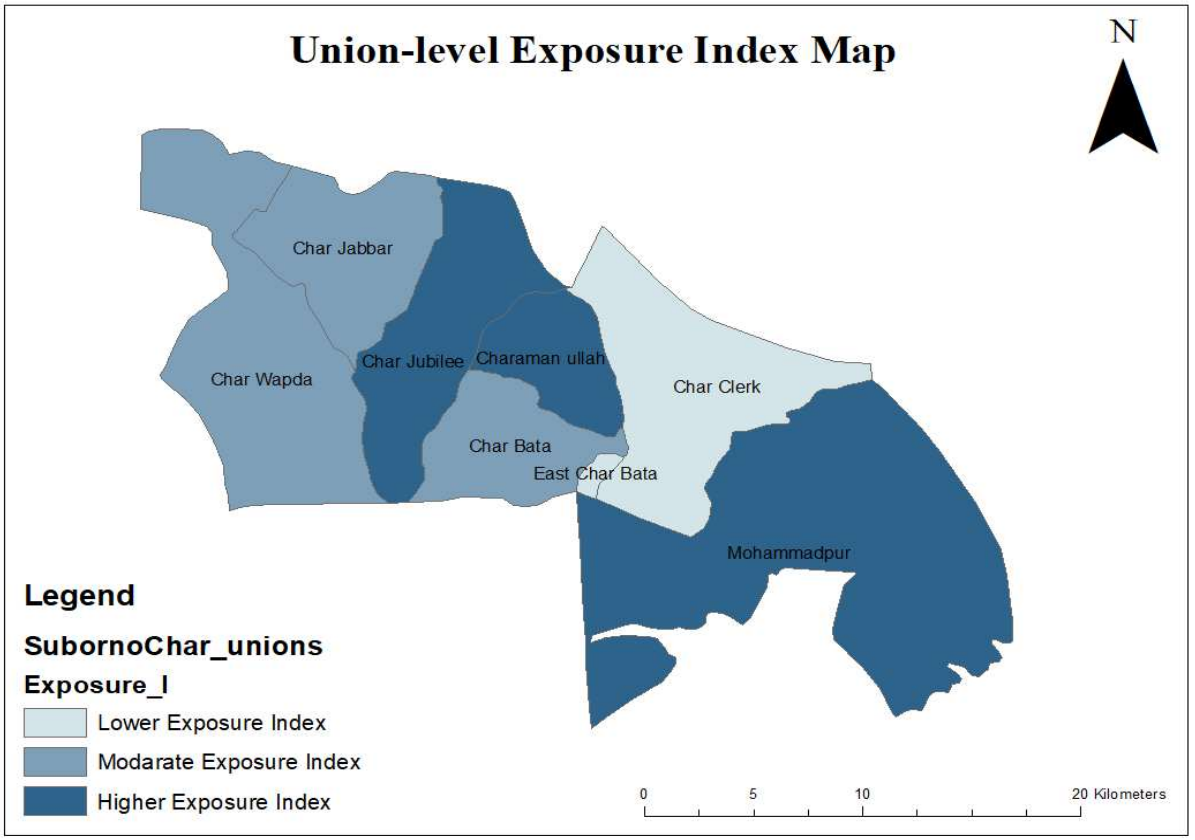
Union	*Char Wapda*	*Char Jabbar*	*Char Jubilee*	*Charaman Ullah*	*Char Bata*	*East Char Bata*	*Mohammadpur*	*Char Clark*
CHI_Mean	*0.080788*	*0.129376*	*0.133644*	*0.149914*	*0.074287*	*0.003769*	*0.007818*	*0.017543*

Figure 3. Spatial distribution of the Compound Hazard Index (CHI) across the eight unions of Subarnachar Upazila.

3.2 Flood–Salinity WASH Vulnerability Assessment.

3.2.1 Exposure Index (EI)

Using ***Eq.6*** The spatial distribution of the Exposure Index (EI) is presented in ***Figure 4***. The EI values ranged from *0.135278* to *0.789659* across the eight unions of Subarnachar Upazila. Charaman Ullah recorded the highest exposure value *0.789659*, followed by Char Jubilee *0.766650* and Mohammadpur *0.753586*. Moderate exposure levels were observed in Char Wapda *0.511120*, Char Bata *0.422040*, and Char Jabbar *0.391590*. Lower EI values were found in East Char Bata *0.277120* and Char Clark *0.135278*. As shown in map, higher exposure concentrations were mainly located in the central and eastern unions, whereas comparatively lower exposure concentrations were distributed in the western and southwestern parts of the study area.



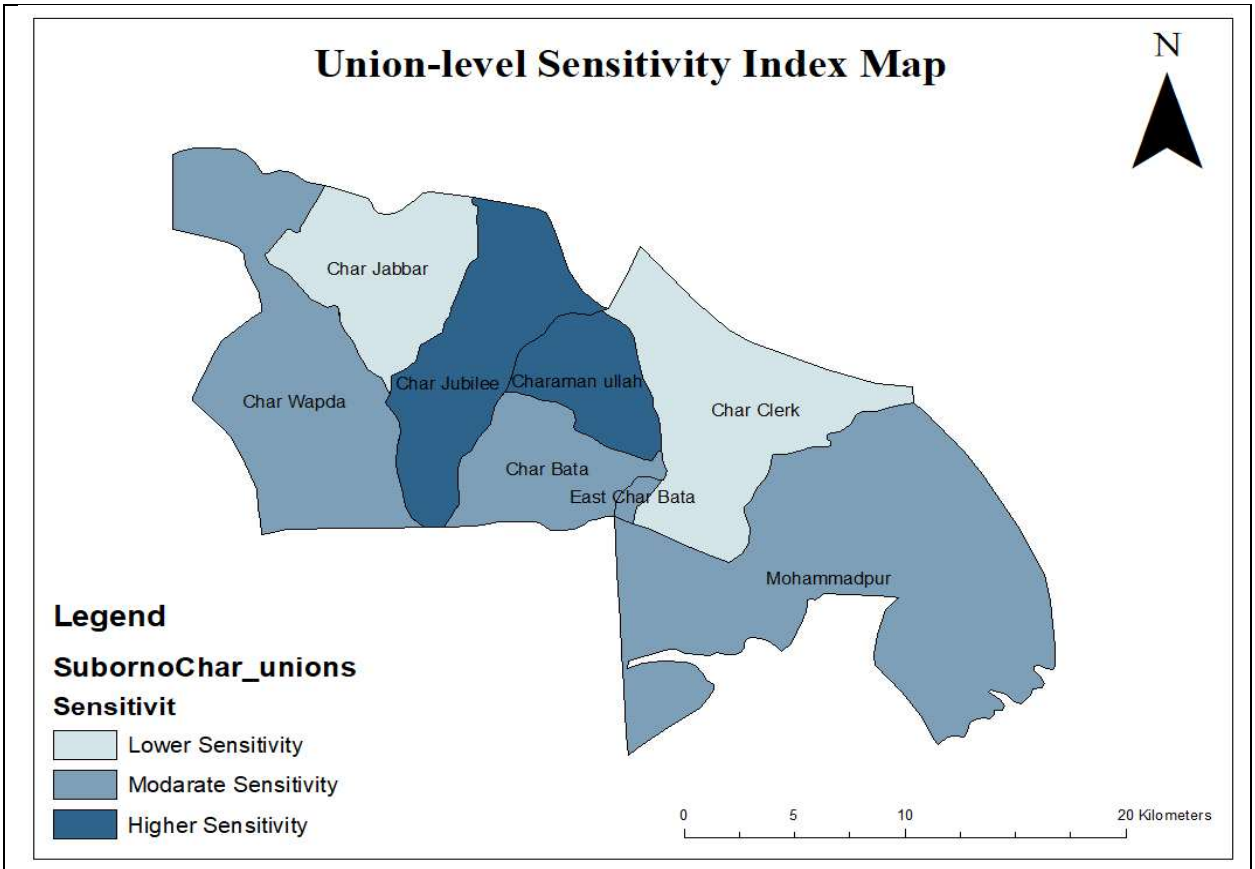
Union	*Char Wapda*	*Char Jabbar*	*Char Jubilee*	*Charaman Ullah*	*Char Bata*	*East Char Bata*	*Mohammadpur*	*Char Clark*
Exposure Index	0.51112	0.39159	0.76665	0.78965	0.42204	0.27712	0.753586	0.135278

Figure 4. Spatial distribution of the Exposure Index (EI) across the eight unions of Subarnachar Upazila.

3.2.2 Sensitivity Index (SI)

The spatial distribution of the Sensitivity Index (SI) is presented in ***Figure 5*** based on ***Eq7***. The SI values ranged from *0.431654* to *0.776152* among the eight unions of Subarnachar Upazila.

Charaman Ullah recorded the highest sensitivity value *0.776152*, followed by Char Jubilee *0.731311* and Char Wapda *0.609029*. Moderate sensitivity levels were observed in East Char Bata *0.599825*, Char Bata *0.586435*, and Mohammadpur *0.588889*. Lower SI values were found in Char Clark *0.494862* and Char Jabbar *0.431654*. According to visual map, Higher sensitivity concentrations were mainly distributed in the central unions, whereas lower sensitivity concentrations were observed in the northern and eastern parts of the study area.



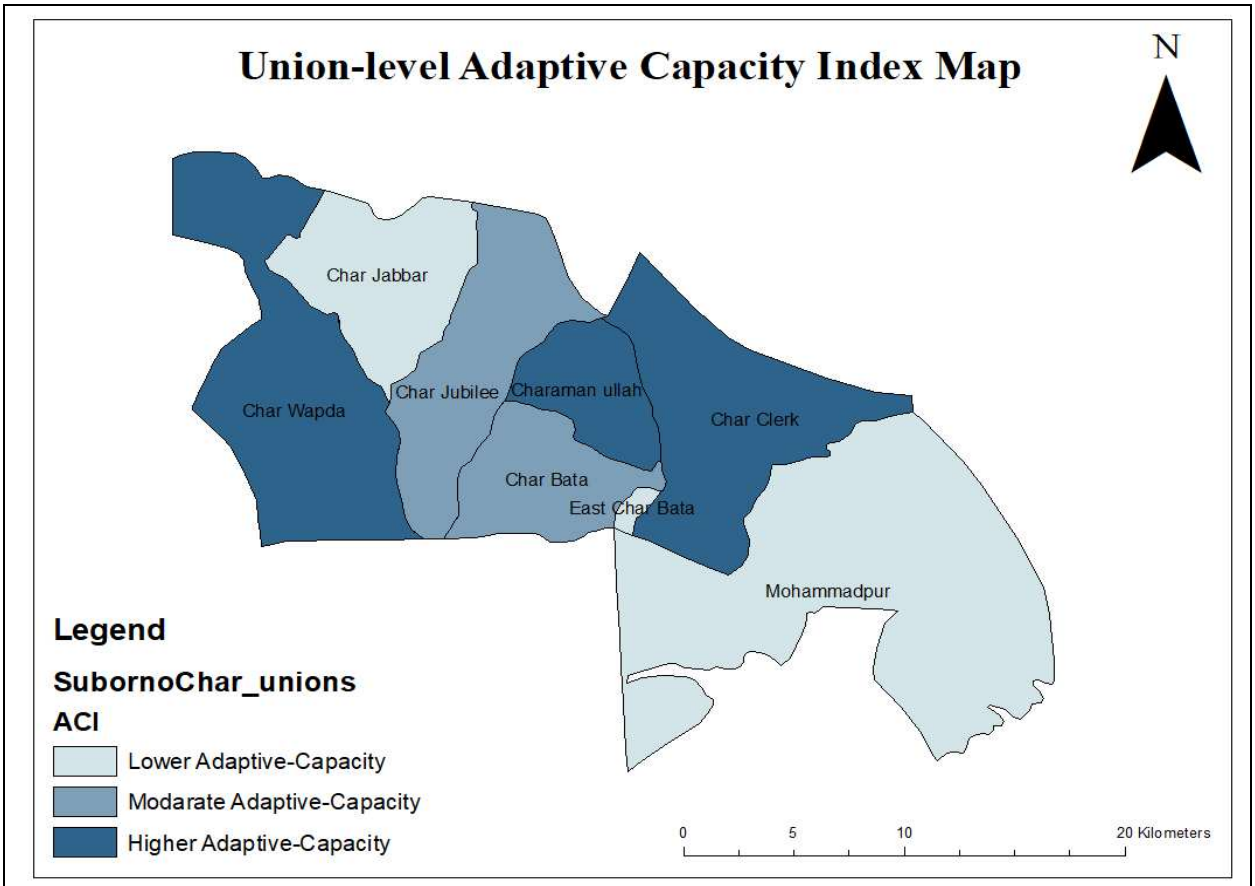
Union	*Char Wapda*	*Char Jabbar*	*Char Jubilee*	*Charaman Ullah*	*Char Bata*	*East Char Bata*	*Mohammadpur*	*Char Clark*
Sensitivity Index	0.6090	0.4316	0.7313	0.7761	0.5864	0.5998	0.5888	0.4948

Figure 5. Spatial distribution of the Sensitivity Index (SI) across the eight unions of Subarnachar Upazila.

3.2.3 Adaptive Capacity Index (ACI)

The spatial distribution of the Adaptive Capacity Index (ACI) is presented in ***Figure 6***. The ACI values ranged from *0.386769* to *0.738544* across the eight unions of Subarnachar Upazila, using ***Eq8***. Char Clark recorded the highest adaptive capacity *0.738544*, followed by Charaman Ullah *0.607887* and Char Wapda *0.602742*. Moderate ACI values were observed in Char Jubilee *0.556692*, Char Bata *0.506578*, and Char Jabbar *0.444860*. Lower adaptive capacity values were found in East Char Bata *0.421071* and Mohammadpur *0.386769*. In ***Figure 6***, higher adaptive

capacity was mainly concentrated in the western, central, and northeastern unions, whereas lower adaptive capacity was observed in the southeastern part of the study area.

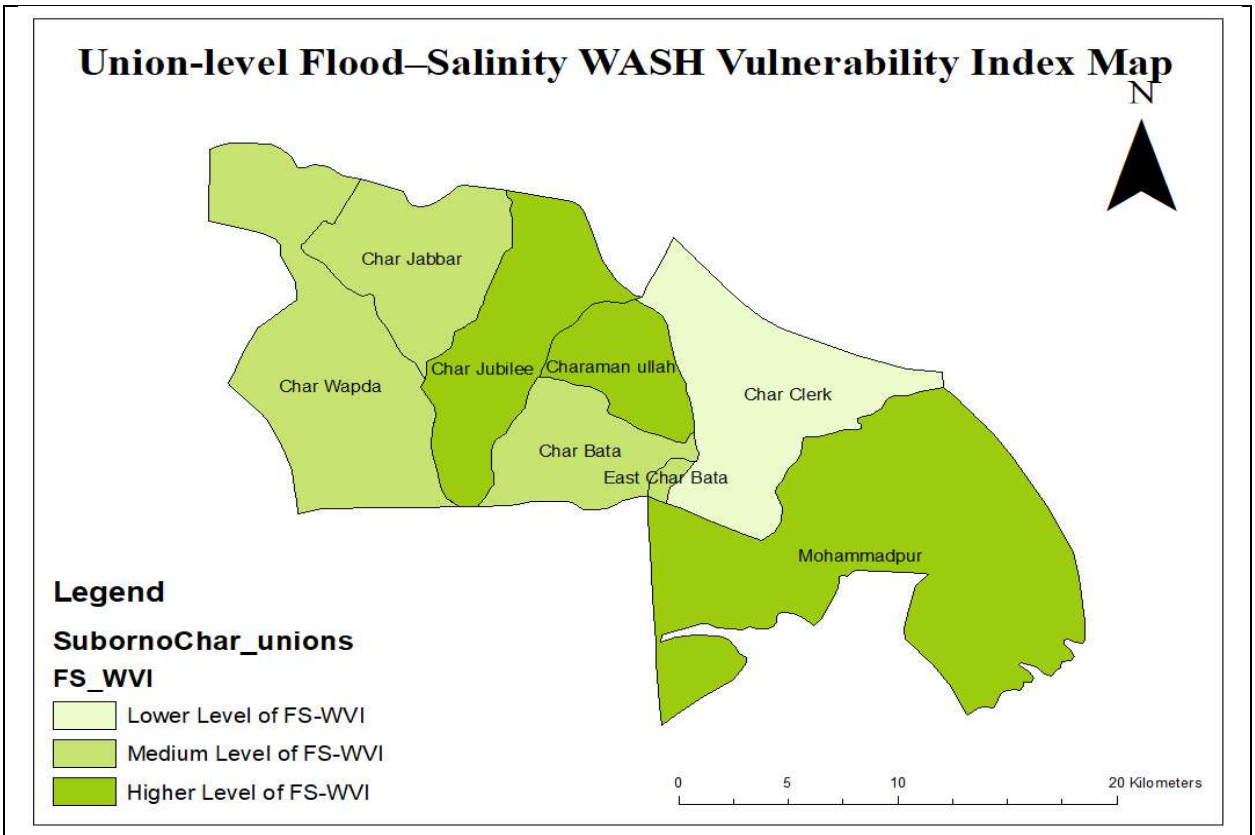


Union	*Char Wapda*	*Char Jabbar*	*Char Jubilee*	*Charaman Ullah*	*Char Bata*	*East Char Bata*	*Mohammadpur*	*Char Clark*
Exposure Index	0.6027	0.4448	0.5566	0.6078	0.5065	0.4210	0.38676929	0.7385

Figure 6. Spatial distribution of the Adaptive Capacity Index (ACI) across the eight unions of Subarnachar Upazila.

3.2.4 Composite Flood–Salinity WASH Vulnerability Index (FS-WVI)

The composite Flood–Salinity WASH Vulnerability Index (FS-WVI) for the eight unions is Calculated by ***Eq 9***. The FS-WVI values ranged from 0.297 to 0.652, indicating noticeable spatial variation across Subarnachar Upazila. Charaman Ullah recorded the highest FS-WVI value 0.652, followed closely by Mohammadpur 0.651 and Char Jubilee 0.647. Moderate vulnerability levels were observed in Char Wapda 0.505, Char Bata 0.500, East Char Bata 0.485, and Char Jabbar 0.459. Char Clark exhibited the lowest FS-WVI value 0.297. As illustrated in ***Figure 7***, higher vulnerability was mainly concentrated in the central and southeastern unions, whereas comparatively lower vulnerability levels were distributed in the western and northeastern parts of the study area.



Union	*Char Wapda*	*Char Jabbar*	*Char Jubilee*	*Charaman Ullah*	*Char Bata*	*East Char Bata*	*Mohammad pur*	*Char Clark*
Flood–Salinity WASH Vulnerability Index	0.505	0.459	0.647	0.652	0.500	0.485	0.651	0.297

Figure 7. Spatial distribution of the Flood–Salinity WASH Vulnerability Index (FS-WVI) across the eight unions of Subarnachar Upazila.

3.3 CatBoost Model Performance

The predictive performance of the CatBoost classifier for household drinking-water salinity is presented in ***Table 7***. The model was evaluated using an independent test dataset consisting of 77 households, including 37 saline-water households (Class 0) and 40 fresh-water households (Class 1). As shown in the Performance Metrics section of ***Table 7***, the CatBoost model achieved an overall accuracy of 0.675, indicating that approximately 67.5% of the test samples were correctly classified. The model also produced a precision of 0.642, recall of 0.850, F1-score of 0.731, and a ROC–AUC value of 0.721.

The Classification Report in ***Table 7*** further illustrates the predictive performance for each class. For saline drinking-water households (Class 0), the model achieved a precision of 0.75, recall of 0.49, and an F1-score of 0.59 based on 37 observations. For fresh drinking-water households (Class 1), the corresponding values were 0.64, 0.85, and 0.73, respectively, using 40 observations. These results indicate that the model correctly identified a larger proportion of fresh-water households than saline-water households.

The classification outcomes are further summarized in the Confusion Matrix of ***Table 7.*** The model correctly classified 18 saline-water households and 34 fresh-water households, while 19 saline-water households were predicted as fresh-water and 6 fresh-water households were predicted as saline-water. Overall, the CatBoost classifier demonstrated moderate predictive capability and provided a basis for subsequent exploratory feature importance and SHAP analyses for the subsequent feature importance analysis and SHAP-based model interpretation.

Table 7. Performance of the CatBoost model for drinking-water salinity classification.

Performance Metric			**Value**	
Training dataset			306 households (80%)	
Testing dataset			77 households (20%)	
Accuracy			**0.675**	
Precision			**0.642**	
Recall			**0.850**	
F1-score			**0.731**	
ROC–AUC			**0.721**	
Classification Report				
Class	**Precision**	**Recall**	**F1-score**	**Support**
0 (Saline water)	0.75	0.49	0.59	37
1 (Fresh water)	0.64	0.85	0.73	40
Confusion Matrix				
Actual / Predicted	**Saline (0)**		**Fresh (1)**	
Saline water (0)	18		19	
Fresh water (1)	6		34	

3.4 Feature Importance from CatBoost

The relative importance of the predictor variables estimated by the CatBoost model is presented in ***Table 8***. Among all variables, Tubewell Depth was identified as the most influential predictor, with a feature importance score of 29.00, substantially higher than the remaining variables. The Compound Hazard Index (CHI) ranked second with an importance score of 13.18, followed by Government/NGO Support 10.13 and Days of Submerged Home 9.57. Moderate importance scores were observed for Monthly Income 7.78, Flood Depth at Home 7.71, Crop Livestock Loss 7.58, and Housing Type 6.40. The lowest contribution was recorded for Household Members 5.18 and Coping Water Strategy 3.47. These rankings indicate the relative contribution of each predictor to the CatBoost model and provide the basis for the subsequent SHAP explainability analysis.

Table 8. CatBoost feature importance ranking for household drinking-water salinity classification.

Rank	Predictor Variable	Feature Importance
1	Tubewell Depth	**29.00**
2	Compound Hazard Index (CHI)	**13.18**
3	Government/NGO Support	**10.13**
4	Days of Submerged Home	**9.57**
5	Monthly Income	**7.78**
6	Flood Depth at Home	**7.71**
7	Crop Livestock Loss	**7.58**
8	Housing Type	**6.40**
9	Household Members	**5.18**
10	Coping Water Strategy	**3.47**

3.5 SHAP Explainability Analysis

3.5.1 Global Feature Importance

The global importance of the predictor variables derived from the SHAP analysis is presented in Figure 8 and ***Table 9***. The SHAP bar plot ranks each predictor according to its mean absolute SHAP value, which represents the average contribution of a variable to the CatBoost model predictions. As shown in Table 9, Tubewell Depth was the most influential variable with a mean SHAP value of 1.691, indicating the largest contribution to the model output. The Compound Hazard Index (CHI) ranked second 0.666, followed by Government/NGO Support 0.623 and Flood Depth at Home 0.509. Moderate contributions were observed for Housing Type 0.387, Monthly Income 0.373, and Days of Submerged Home 0.358. Lower SHAP values were recorded for Crop Livestock Loss 0.285, Household Members 0.206, and Coping Water Strategy 0.206. The ranking presented in ***Figure 8*** is generally consistent with the CatBoost feature importance analysis, confirming the dominant influence of environmental hazard indicators and household water-access characteristics on the model predictions.

Table 9. Mean absolute SHAP values and global feature importance ranking.

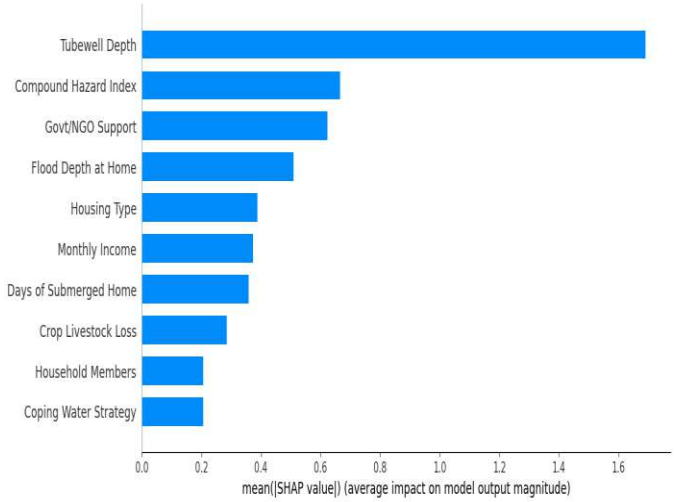


Figure 8. Global feature importance based on the mean absolute SHAP values for the CatBoost model predicting household drinking-water salinity.

Rank	Predictor Variable	Mean Absolute SHAP Value
1	Tubewell Depth	**1.691**
2	Compound Hazard Index (CHI)	**0.666**
3	Government/NGO Support	**0.623**
4	Flood Depth at Home	**0.509**
5	Housing Type	**0.387**
6	Monthly Income	**0.373**
7	Days of Submerged Home	**0.358**
8	Crop Livestock Loss	**0.285**
9	Household Members	**0.206**
10	Coping Water Strategy	**0.206**

3.5.2 SHAP Summary Plot

The contribution and distribution of individual predictor variables are presented in ***Figure 9***. The SHAP summary plot ranks the variables according to their influence on the CatBoost model, with Days of Submerged Home, Tubewell Depth, Compound Hazard Index (CHI), Flood Depth at Home, identified as the most influential predictors. Variables located near the bottom of the figure, including Government/NGO Support, Monthly Income, and Coping Water Strategy, exhibited comparatively lower contributions.

The horizontal position of each point represents its SHAP value, where positive values increase and negative values decrease the model prediction. The color gradient indicates the feature value, with red representing higher values and blue representing lower values. Variables such as Days of Submerged Home, and Flood Depth at Home showed wider SHAP distributions, indicating greater variability in their contributions across households, whereas Government/NGO Support, Monthly Income, and Coping Water Strategy displayed relatively smaller and more consistent effects.

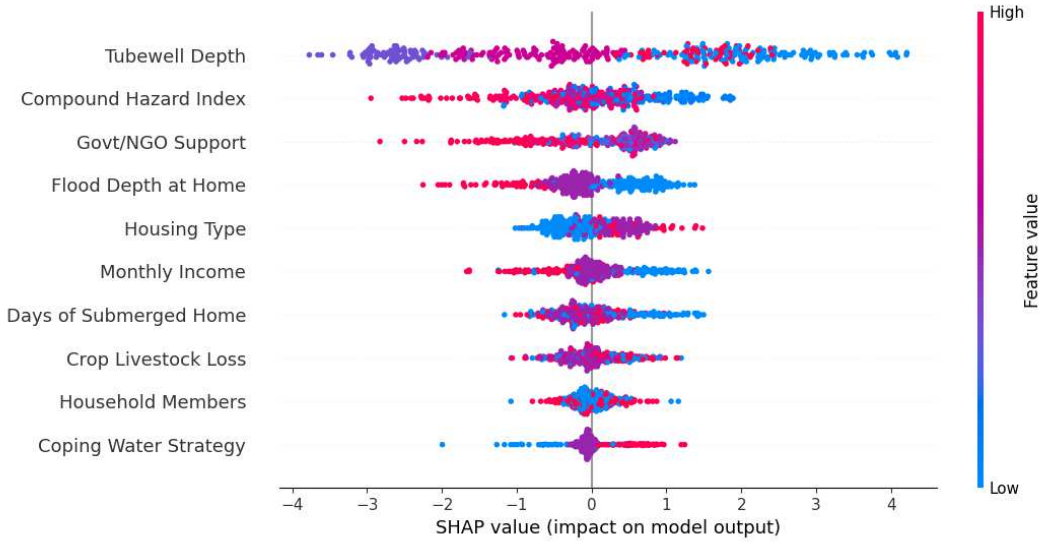


Figure 9. SHAP summary plot showing the direction, magnitude, and distribution of predictor variable contributions to the CatBoost model for household drinking-water salinity classification.

3.5.3 SHAP Dependence Plot

The interaction between the Compound Hazard Index (CHI) and its contribution to the CatBoost model prediction is illustrated in Figure 10. The SHAP dependence plot shows that the effect of CHI on the model output varies across different CHI levels. Lower CHI values roughly 0.00–0.08 are generally associated with positive SHAP values, indicating a greater contribution toward the prediction of fresh drinking water (Class 1). In contrast, higher CHI values around 0.13–0.15 are primarily associated with negative SHAP values, indicating an increased contribution toward the prediction of saline drinking water (Class 0). The color scale represents Tubewell Depth, revealing an interaction between CHI and tubewell depth in the prediction process. The dispersion of SHAP values at similar CHI levels further indicates that the influence of CHI varies among households depending on the combined effects of other predictor variables.

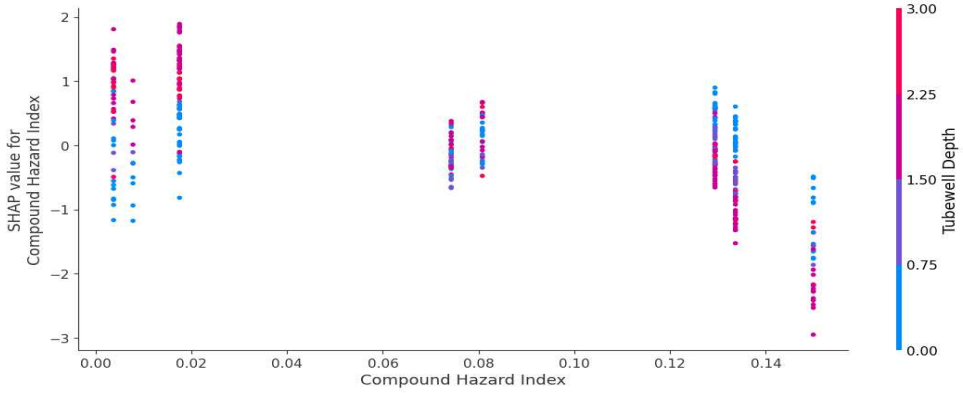


Figure 10. SHAP dependence plot illustrating the relationship between the Compound Hazard Index (CHI) and its contribution to the CatBoost model prediction, with point colors representing tubewell depth.

3.6 Integrated Spatial Interpretation

The integrated spatial interpretation revealed clear differences in the dominant sources of WASH vulnerability across the eight unions ***Table 10***. Exposure was the dominant component in Char Jubilee, Charaman Ullah, and Mohammadpur, which also showed comparatively high overall FS-WVI levels. In these unions, exposure exceeded both sensitivity and adaptive capacity deficit, indicating that flood-related and household exposure conditions made the largest contribution to the composite vulnerability structure. Sensitivity was dominant in Char Clark, Char Wapda, East Char Bata, and Char Bata. Among them, East Char Bata showed closely comparable sensitivity and adaptive capacity deficit values, although sensitivity remained slightly higher. In contrast, Char Jabbar was the only union dominated by adaptive capacity deficit, indicating that limited capacity represented its largest vulnerability-related component.

The spatial component patterns also showed broad links with the explainable machine-learning results. The CatBoost feature ranking in ***Table 8*** and the SHAP results in ***Figures 8–10*** identified tubewell depth, CHI, flood depth at home, and duration of home submergence as important influences on drinking-water salinity predictions. These variables correspond to environmental exposure, household sensitivity, and water-access capacity represented within the FS-WVI framework. Thus, the integrated analysis identified both the dominant vulnerability structure of each union and the household-level factors associated with drinking-water salinity outcomes.

Table 10. Dominant vulnerability-related component across the eight unions of Subarnachar Upazila.

Union	EI	SI	ACI	ACD (1 − ACI)	Dominant Vulnerability Component
Char Clark	0.135	0.495	0.739	0.261	Sensitivity
Char Jabbar	0.392	0.432	0.445	0.555	Adaptive Capacity Deficit
Char Jubilee	0.767	0.731	0.557	0.443	Exposure
Char Wapda	0.511	0.609	0.603	0.397	Sensitivity
East Char Bata	0.277	0.600	0.421	0.579	Sensitivity
Char Bata	0.422	0.586	0.507	0.493	Sensitivity
Charaman Ullah	0.790	0.776	0.608	0.392	Exposure
Mohammadpur	0.754	0.589	0.387	0.613	Exposure

4. Discussion

Although Noakhali District is not usually described as one of Bangladesh's major river-flood hotspots, its low-lying coastal and deltaic setting creates a different hazard context(Murshed et al., 2022). The district lies beside the Meghna estuary and Bay of Bengal, where tidal processes, coastal inundation, waterlogging, and salinity can affect land and water systems(ADAB, 2025; Khatun et al., 2025). Previous research has also documented spatial variation in soil salinity within Noakhali and water-quality concerns in Subarnachar, while broader studies show that salinity intrusion is a major pressure on freshwater security in coastal Bangladesh(Kawser et al., 2022). Against this background, the CHI pattern in **Figure 3** shows stronger compound hazard concentration in Charaman Ullah, Char Jubilee, and Char Jabbar, while East Char Bata and Mohammadpur have much lower CHI values. However, the final FS-WVI pattern in **Figure 7** is not identical. Charaman Ullah remains highly vulnerable, but Mohammadpur also reaches a very high vulnerability level despite its low mean CHI. Char Jubilee similarly combines high compound hazard with high overall vulnerability, whereas Char Clark remains the least vulnerable union. This comparison shows that the geography of compound hazard and the geography of WASH vulnerability overlap only partly. Such a pattern is reasonable because coastal WASH vulnerability is shaped not only by direct hazard intensity but also by water access, household sensitivity, infrastructure, livelihood conditions, and the capacity to cope with disruption.

The similarities and mismatches between CHI and FS-WVI are therefore among the most important findings of this study. Charaman Ullah and Char Jubilee show a clear similarity: both have high hazard conditions and high overall vulnerability. Mohammadpur presents the strongest mismatch, because its low CHI is combined with high FS-WVI. Char Jabbar shows the opposite pattern, with relatively high CHI but much lower overall vulnerability than the most vulnerable unions. This difference can be explained by the broader structure of FS-WVI. The index includes direct exposure indicators such as flood depth, duration of home submergence, and crop or livestock loss, together with sensitivity indicators related to water shortage, salinity, disease, sanitation, and household conditions. Therefore, a low union-mean CHI does not necessarily imply low household exposure or low WASH vulnerability. The role of adaptive capacity is equally important. Mohammadpur had the lowest ACI, whereas Char Clark had the highest ACI and the lowest final vulnerability. This pattern supports evidence from coastal Bangladesh that household infrastructure, economic capacity, social safety nets, and coping resources are important dimensions of resilience, while climate-sensitive livelihoods can increase vulnerability(Jalal et al., 2021; Muniruzzaman et al., 2026). In practical terms, stronger adaptive capacity can reduce the final effect of hazard and sensitivity, while limited income, weak water access, poor housing, and insufficient institutional support can maintain high vulnerability even where the mapped compound hazard is comparatively low.

The CatBoost results provide further evidence that household drinking-water salinity is controlled by interacting environmental and socio-economic conditions rather than by salinity exposure alone. As reported in **Table 7**, the model achieved moderate predictive performance, with an accuracy of 0.675, an F1-score of 0.731, and a ROC–AUC of 0.721; however, the lower recall for the saline-water class also shows that the model was less effective at identifying some saline

households. The feature ranking in **Table 8** is particularly important because tubewell depth was the strongest predictor, followed by CHI, Government/NGO support, and home-submergence duration. The dominant role of tubewell depth is plausible in a coastal aquifer setting because drinking-water salinity depends on groundwater conditions and the depth and location(Islam et al., 2013) from which water is abstracted; importantly, this should not be interpreted as a simple rule that "deeper is always fresher," because coastal aquifers can contain vertically variable fresh and saline layers(Pettigrew & Mccauley, 2026). Earlier research has shown that drinking-water sources in coastal Bangladesh are affected by salinity intrusion, and safe freshwater scarcity is a major regional problem(Abedin et al., 2014); a study specifically covering Subarnachar also reported concerns about drinking-water quality and sanitation(Kumar Roy & Munna, 2019). The importance of CHI further connects the remotely sensed hazard environment with household water outcomes, while the roles of support, income, housing, flood depth, and submergence duration indicate that access to safe water is also conditioned by resources, infrastructure, and post-flood disruption. This combination explains why the model did not reduce drinking-water salinity to one physical variable and why household-level conditions remain necessary for understanding coastal WASH risk.

The SHAP analysis strengthens this interpretation because it moves beyond a simple ranking and shows how predictor values change individual model outputs(Tempel et al., 2025). The global SHAP results in **Figure 8** identify tubewell depth and CHI as the two strongest contributors, which broadly supports the CatBoost ranking in **Table 8**. The summary plot in **Figure 9** further shows wide household-level variation in the effects of submergence duration, tubewell depth, CHI, and flood depth. This heterogeneity is expected in a WASH system where households with similar hazard exposure may use different water sources, have different economic resources, or receive different levels of support. SHAP is useful here because it can represent both the magnitude and direction of feature contributions(Li et al., 2025), making nonlinear tree-based predictions more transparent; explainable machine-learning approaches are increasingly used in water-quality assessment for this reason(Makumbura et al., 2024). Most importantly, the dependence pattern in **Figure 10** shows that lower CHI values generally contribute toward fresh-water predictions, while the highest CHI range contributes more toward saline-water predictions. The interaction coloring by tubewell depth suggests that the same CHI level does not produce an identical model response for every household. This finding also links back to FS-WVI: CHI contributes directly to the exposure component, but its household effect is modified by water-access and adaptive conditions. Thus, the index and SHAP results are complementary rather than competing. FS-WVI identifies where multidimensional vulnerability is concentrated, while SHAP helps explain which variables are associated with the household drinking-water outcome and how their effects vary.

The integrated interpretation in **Table 10** finally shows that vulnerability does not have the same structure across Subarnachar. Char Jubilee, Charaman Ullah, and Mohammadpur are exposure-dominated, and these unions also include the highest overall FS-WVI levels. This pattern indicates that direct flood-related and household exposure conditions make the largest contribution to their vulnerability structure. In contrast, Char Clark, Char Wapda, East Char Bata, and Char Bata are sensitivity-dominated, suggesting that WASH disruption, health conditions, household characteristics, and sanitation-related pressures are more prominent than exposure or adaptive

capacity deficit. Char Jabbar is the only union dominated by Adaptive Capacity Deficit (ACD), indicating that limited coping resources and weaker adaptive conditions make the largest contribution to its vulnerability structure. East Char Bata is also notable because its sensitivity and ACD values are very close, suggesting that household susceptibility and limited adaptive capacity jointly shape its vulnerability. This spatial variation is consistent with multidimensional coastal vulnerability research in Bangladesh, which shows that exposure, sensitivity, infrastructure, livelihoods, and coping resources vary among communities and should not be treated as a uniform coastal condition. At the regional level, the framework can help planners distinguish exposure-dominated areas from locations where health and WASH sensitivity or adaptive capacity deficits require greater attention, supporting more targeted interventions across unions. Future research should validate these relationships using seasonal groundwater salinity measurements, repeated household surveys, multi-year flood and salinity data, external CatBoost validation, and spatially explicit groundwater information. Such extensions could test the seasonal stability of the CHI–drinking-water relationship and assess whether the integrated FS-WVI framework can be transferred to other coastal upazilas of Bangladesh.

5. Conclusion

This study developed the FS-WVI by integrating compound flood–salinity conditions with household exposure, sensitivity, and adaptive capacity. The findings revealed clear spatial differences among the eight unions and showed that overall WASH vulnerability does not always follow the distribution of environmental hazards. Instead, household sensitivity, water-related conditions, and adaptive resources can strengthen or reduce vulnerability. The developed framework therefore provides a more comprehensive understanding of multidimensional coastal WASH vulnerability and its spatial variation.

The CatBoost and SHAP analyses addressed this limitation by identifying and explaining the environmental and socio-economic factors influencing household drinking-water salinity. The strong roles of tubewell depth, compound hazard conditions, and household-level factors showed that water outcomes emerge from interacting processes. This explainable machine-learning approach successfully connected prediction with transparent interpretation.

The integrated interpretation combined spatial FS-WVI patterns with explainable machine-learning evidence to identify distinct vulnerability structures across Subarnachar. Exposure dominated some unions, sensitivity shaped others, while adaptive capacity influenced how communities experienced and managed flood–salinity-related WASH risks. By linking the spatial vulnerability assessment with the major drivers of household drinking-water conditions, the study achieved its broader goal of explaining both where vulnerability is concentrated and which factors shape its local variation. The study is limited by its cross-sectional survey design, the temporal specificity of remote-sensing data, and the absence of seasonal groundwater salinity measurements. These limitations do not contradict the observed spatial and household patterns but indicate areas for further validation. Overall, the integrated framework offers a practical basis for targeted regional WASH planning and climate adaptation. Future researchers can extend this approach through multi-season observations, longitudinal household surveys, direct groundwater monitoring, climate scenarios, and comparative applications across other coastal regions.

Declaration of Generative AI and AI-assisted Technologies in the Manuscript Preparation Process

During the preparation of this work, the author used QuillBot and Grammarly to improve sentence structure and correct minor grammatical errors, as the author is a non-native English speaker. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

REFERENCES

Abedin, M. A., Habiba, U., & Shaw, R. (2014). Community Perception and Adaptation to Safe Drinking Water Scarcity: Salinity, Arsenic, and Drought Risks in Coastal Bangladesh. *International Journal of Disaster Risk Science*, *5*, 110–124. https://doi.org/https://doi.org/10.1007/s13753-014-0021-6

ADAB. (2025). *How the Climate Crisis Is Impacting Bangladesh*. Association of Development Agencies in Bangladesh. https://www.adab.org.bd/how-the-climate-crisis-is-impacting-bangladesh/

Ahmed, S. K. (2024). How to choose a sampling technique and determine sample size for research: A simplified guide for researchers. *Oral Oncology Reports*, *12*. https://doi.org/https://doi.org/10.1016/j.oor.2024.100662.

Ahona, N., Anik, A. H., & Tareq, S. M. (2025). A Systematic Review of Water, Sanitation and Hygiene of Bangladesh: Prospects and Challenges to Achieve Sustainable Development. *Water and Environment Journal*, *39*(4). https://doi.org/https://doi.org/10.1111/wej.70009

Akanbi, O. A., Amiri, I. S., & Fazeldehkordi, E. (2015). Chapter 4 - Feature Extraction, A Machine-Learning Approach to Phishing Detection and Defense. *Syngress*, 45–54. https://doi.org/https://doi.org/10.1016/B978-0-12-802927-5.00004-6.

Akter, T., Hoque, M. A.-A., Mukul, S. A., & Pradhan, B. (2024). Coastal Flood Induced Salinity Intrusion Risk Assessment Using a Spatial Multi-criteria Approach in the South-Western Bangladesh. *Earth Systems and Environment*. https://doi.org/10.1007/s41748-024-00399-9

Arnab, R. (2017). Chapter 7 - Stratified Sampling Survey Sampling Theory and Applications. *Academic Press*, 213–256. https://doi.org/https://doi.org/10.1016/B978-0-12-811848-1.00007-8.

Bhuyan, M. I., Mia, S., Supit, I., & Ludwig, F. (2023). Spatio-temporal variability in soil and water salinity in the south-central coast of Bangladesh. *CATENA*, *222*. https://doi.org/https://doi.org/10.1016/j.catena.2022.106786.

Deb, D., Uddin, M. M., Mahbub-E-Kibria, A. S. M., Das, M. K., & Hasan, M. (2024). Coastal vulnerability assessment to multi hazards in the exposed coast of Southeastern Coastal Region of Bangladesh. *Regional Studies in Marine Science*, *73*. https://doi.org/https://doi.org/10.1016/j.rsma.2024.103484.

Fidelis, G. D., Eziz, A., Ahmed, T., Azadi, H., Ozuzu, S. A., Ullah, S., & Kurban, A. (2025). , Geospatial analysis of Flood susceptibility through a combined bivariate statistical model and Sentinel-1 SAR data in Qarqan River basin, China,. *Heliyon*, *11*(16).

https://doi.org/https://doi.org/10.1016/j.heliyon.2025.e44074.

Guan, X., Xue, R., He, Z., Chen, S., & Chen, X. (2025). CatBoost-Optimized Hyperspectral Modeling for Accurate Prediction of Wood Dyeing Formulations. *Forests*, *16*(8), 1279. https://doi.org/https://doi.org/10.3390/f16081279

Hashemi, S. (2020). Sanitation Sustainability Index: A Pilot Approach to Develop a Community-Based Indicator for Evaluating Sustainability of Sanitation Systems. *Sustainability*, *12*. https://doi.org/10.3390/su12176937

Himel, T., Hossain, M., & Rahaman, K. (2025). Unraveling Climate Change Vulnerability and Adaptation in Flood-Affected Communities of Northern Bangladesh: A Multidimensional Poverty Perspective. *Environmental Development*, *54*. https://doi.org/10.1016/j.envdev.2025.101135

Hirai, T., & Comim, F. (2022). Measuring the sustainable development goals: A poset analysis (. *Ecological Indicators*, *145*. https://doi.org/https://doi.org/10.1016/j.ecolind.2022.109605.

Hosan, S., Haydar, M., Rafi, A. H., & Rokim, M. (2026). Natural spring-based water potential in Bangladesh using ensemble machine learning algorithms. *Journal of Water and Climate Change*, *17*(2). https://doi.org/https://doi.org/10.2166/wcc.2026.210

Huang, X., Kroening, D., Ruan, W., Sharp, J., Sun, Y., Thamo, E., Wu, M., & Yi, X. (2020). A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability (. *Computer Science Review*, *37*. https://doi.org/https://doi.org/10.1016/j.cosrev.2020.100270.

Iii, V., Exposure, M. H., Assessment, R., & Surge, S. (n.d.). *Volume III: Multi Hazard Exposure and Risk Assessment (Flood, Storm Surge, Drought, Landslide, Earthquake and Tsunami) | 1*. *III*.

Islam, M., Sakakibara, H., Karim, R., & Sekine, M. (2013). Potable water scarcity: Options and issues in the coastal areas of Bangladesh. *Journal of Water and Health*, *11*. https://doi.org/10.2166/wh.2013.215

Jalal, M. J. E., Khan, M. A., Hossain, M. E., Yedla, S., & Alam, G. M. M. (2021). Does climate change stimulate household vulnerability and income diversity? Evidence from southern coastal region of Bangladesh. *Heliyon*, *7*(9). https://doi.org/https://doi.org/10.1016/j.heliyon.2021.e07990.

John, M., Ebinezer, D., & Krishna, B. C. (2026). *Explainable artificial intelligence models using SHAP enhanced CatBoost , Bi-GRU with attention , and Tab Transformer*. 1–22.

Kawser, U., Nath, B., & Hoque, A. (2022). Observing the influences of climatic and environmental variability over soil salinity changes in the Noakhali Coastal Regions of Bangladesh using geospatial and statistical techniques,. *Environmental Challenges*, *6*. https://doi.org/https://doi.org/10.1016/j.envc.2021.100429.

Khatun, F., Kamal, M., Kabir, F. Al, & Huq, P. K. (2024). *Eastern Bangladesh Floods in 2024 Analysis of Immediate Damages and Recovery Efforts*. Earth Science Applied. https://cpd.org.bd/resources/2025/04/Eastern-Bangladesh-Floods-in-2024.pdf

Khatun, F., Saha, O. R., & Matin, N. (2025). Morphological changes in the offshore islands of Meghna estuary: Analysis of the erosion and accretion dynamics. *HydroResearch*, *8*, 294–306. https://doi.org/https://doi.org/10.1016/j.hydres.2024.12.006.

Kopanja, M., Hačko, S., Brdar, S., & Savić, M. (2024). Cost-sensitive tree SHAP for explaining cost-sensitive tree-based models. *Computational Intelligence*, *40*(3). https://doi.org/https://doi.org/10.1111/coin.12651

Kulkarni, C. (2022). Advancing Gradient Boosting: A Comprehensive Evaluation of the CatBoost Algorithm for Predictive Modeling. *Journal of Artificial Intelligence, Machine Learning and Data Science*, *1*. https://doi.org/10.51219/JAIMLD/chinmay-shripad-kulkarni/29

Kumar Roy, S., & Munna, M. (2019). Status of drinking water quality and sanitation facilities in Subarnachar and Maijdee in Noakhali, Bangladesh. *Asian Journal of Medical and Biological Research*, *5*. https://doi.org/10.3329/ajmbr.v5i1.41043

Li, W., Deng, M., Liu, C., & Cao, Q. (2025). Analysis of Key Influencing Factors of Water Quality in Tai Lake Basin Based on XGBoost-SHAP. *Water*, *17*(11), 1619. https://doi.org/https://doi.org/10.3390/w17111619

Lundberg, S. M., & Lee, S. (2017). *A Unified Approach to Interpreting Model Predictions*. *Section 2*, 1–10.

Mahmood, R., Ahmed, N., Zhang, L., & Li, G. (2020). Coastal vulnerability assessment of Meghna estuary of Bangladesh using integrated geospatial techniques,. *International Journal of Disaster Risk Reduction*, *42*. https://doi.org/https://doi.org/10.1016/j.ijdrr.2019.101374.

Mahmuduzzaman, M., Ahmed, Z. U., A. K. M. Nuruzzaman, F., & Ahmed, R. S. (2014). Causes of Salinity Intrusion in Coastal Belt of Bangladesh. *International Journal of Plant Research*, *4*(4A), 8–13. https://doi.org/10.5923/s.plant.201401.02.

Makumbura, R. K., Mampitiya, L., Rathnayake, N., Meddage, D. P. P., Henna, S., Dang, T. L., Hoshino, Y., & Rathnayake, U. (2024). Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature. *Results in Engineering*, *23*. https://doi.org/https://doi.org/10.1016/j.rineng.2024.102831.

Medium. (2026). *CatBoost*. https://medium.com/@mausamsingh025/catboost-b74ceb6e291c

Méndez, M., Díaz, G., Merayo, M. G., & Núñez, M. (2026). Combining a novel metric and categorical boosting to better estimate rare event likelihood: Use in traffic accidents. *Applied Soft Computing*, *198*. https://doi.org/https://doi.org/10.1016/j.asoc.2026.115261.

Mountasir, N., Lafabregue, B., Albert, B., & Lachiche, N. (2026). DEPlot: A visualization tool for comparative analysis of regression models. *Array*, *30*. https://doi.org/https://doi.org/10.1016/j.array.2026.100953.

Muniruzzaman, M., Islam, S., & Eniola, A. (2026). *Community Resilience to Climate Change in Coastal Bangladesh: Challenges and Adaptation Strategies*. https://www.researchgate.net/publication/401825705_Community_Resilience_to_Climate_

Change_in_Coastal_Bangladesh_Challenges_and_Adaptation_Strategies

Murshed, S., Griffin, A. L., Islam, M. A., Wang, X. H., & Paull, D. (2022). Assessing multi-climate-hazard threat in the coastal region of Bangladesh by combining influential environmental and anthropogenic factors. *Progress in Disaster Science*, *16*. https://doi.org/https://doi.org/10.1016/j.pdisas.2022.100261.

Nishat, M. H., Rahman, H., Khan, B., Ahmed, T., & Hossain, S. N. (2025). Comparative analysis of machine learning models for predicting water quality index in Dhaka ' s rivers of Bangladesh. *Environmental Sciences Europe*. https://doi.org/10.1186/s12302-025-01078-w

Nogueira, L. S. R., Carvalho, M. A. S. de, Santos, B. de O., Yonaba, R., Bamal, A., Uddin, M. G., Bodini, M., & Goliatt, L. (2026). A comparative study of ensemble and non-ensemble machine learning methods for predicting river pollution index. *Ecological Informatics*, *94*. https://doi.org/https://doi.org/10.1016/j.ecoinf.2026.103617

OCHA. (2024). *Bangladesh: Eastern Flash Floods 2024 Situation Report No. 02 (As of 30 August 2024)*. United Nations Office for the Coordination of Humanitarian Affairs. https://www.unocha.org/publications/report/bangladesh/bangladesh-eastern-flash-floods-2024-situation-report-no-02-30-august-2024

Ozsahin, D. U., Ameen, Z. S., Hassan, A. S., & Mubarak, A. S. (2024). Enhancing explainable SARS - CoV - 2 vaccine development leveraging bee colony optimised Bi - LSTM , Bi - GRU models and bioinformatic analysis. *Scientific Reports*, 1–20. https://doi.org/10.1038/s41598-024-55762-7

Panpeng, J., & Ahmad, M. M. (2017). Vulnerability of Fishing Communities from Sea-Level Change: A Study of Laemsing District in Chanthaburi Province, Thailand. *Sustainability*, *9*(8), 1388. https://doi.org/https://doi.org/10.3390/su9081388

Pettigrew, K. A., & Mccauley, D. (2026). Groundwater for Sustainable Development Coastal salinity challenges in Bangladesh : Impacts on health , agriculture , and livelihoods. *Groundwater for Sustainable Development*, *33*(March 2025), 101603. https://doi.org/10.1016/j.gsd.2026.101603

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). *CatBoost : unbiased boosting with categorical features*. *Section 4*, 1–11.

Pulina, A., Bellocchi, G., Iocola, I., Mula, L., Altea, L., Piras, F., Seddaiu, G., & Roggero, P. P. (2026). , Assessing the adaptive capacity to climate change using fuzzy logic-based approaches: the case of Sardinian agriculture (. *Journal of Environmental Management*, *404*. https://doi.org/https://doi.org/10.1016/j.jenvman.2026.129435.

Rabbani, G., Munira, S., & Saif, S. (2018). *Coastal Community Adaptation to Climate Change-Induced Salinity Intrusion in Bangladesh*. https://doi.org/10.5772/intechopen.80418

Rahman, A., Hagare, D., Siddiqui, Z., Ouarda, T. B. M. J., Muhitur, M., & Rahman. (2026). Salinity-Induced Water Insecurity in Coastal Bangladesh. *4th International Conference on Water and Environmental Engineering - Proceedings of ICWEE 2025*, 392–403. https://researchoutput.csu.edu.au/en/publications/salinity-induced-water-insecurity-in-coastal-bangladesh/

Raisa, S., Sarkar, S. K., & Sadiq, M. A. (2024). TY - JOUR PY - 2024/03/02 SP - T1 - Advancing groundwater vulnerability assessment in Bangladesh: A comprehensive machine learning approach. *Groundwater for Sustainable Development*, *25*. https://doi.org/10.1016/j.gsd.2024.101128

Reuters. (2024). *Floods, landslides in India's Tripura displace tens of thousands*. https://www.reuters.com/world/india/floods-landslides-indias-tripura-displace-tens-thousands-2024-08-23/

Saber, M., Boulmaiz, T., Guermoui, M., Abdrabo, K., Kantoush, S., Sumi, T., Hamouda, B., Nohara, D., & Mabrouk, E. (2021). Examining LightGBM and CatBoost Models for Wadi Flash Flood Susceptibility Prediction. *Geocarto International*, *37*. https://doi.org/10.1080/10106049.2021.1974959

Sahwell, P. J., & Solo-Gabriele, H. M. (2025). Control of freshwater flows can counteract the effects of global climate change on salinity within urban estuaries. *Science of The Total Environment*, *1001*. https://doi.org/https://doi.org/10.1016/j.scitotenv.2025.180544.

Sarwer, G., Ali, M., Bhowmik, S., Asadujjaman, M., & Sharmin, S. (2016). *Pond Farming and Livelihood Status of Fish Farmers in Subarnachar, Noakhali, Bangladesh*. https://www.researchgate.net/publication/307559577_Pond_Farming_and_Livelihood_Status_of_Fish_Farmers_in_Subarnachar_Noakhali_Bangladesh

Sprinkle. (2024). *What is Normalisation of Data: Its Types, Goals and Use in Data Analytics and Machine Learning*. Sprinkle Data. https://www.sprinkledata.ai/blogs/what-is-normalisation-of-data-its-types-goals-and-use-in-data-analytics-and-machine-learning/

SpssServices. (2025). *Cochran's Sample Size Calculator: A Comprehensive Guide*. https://spssservices.com/cochran-sample-size-calculator-guide/

Tempel, F., Ihlen, E. A. F., Adde, L., & Strümke, I. (2025). Explaining Human Activity Recognition with SHAP: Validating insights with perturbation and quantitative measures. *Computers in Biology and Medicine*, *188*. https://doi.org/https://doi.org/10.1016/j.compbiomed.2025.109838.

The Business Standard. (2025). *August 2024 floods in Bangladesh and Tripura: A nexus of erratic rainfall, vanishing waterways, and neglect*. https://www.tbsnews.net/features/panorama/august-2024-floods-bangladesh-and-tripura-nexus-erratic-rainfall-vanishing

The daily star. (2024). What really caused the floods? *The Daily Star*. https://www.thedailystar.net/opinion/views/news/what-really-caused-the-floods-3687066

TIME, B. (2026). *Bangladesh's Southern Coast Fights for Survival*. Sariatpur Printing Press 28/B. https://www.businesstimes-bd.com/bangladesh/8773

UN. (2024). *Bangladesh: Eastern Flash Floods*. https://bangladesh.un.org/sites/default/files/2024-08/Final_Sit Rep 01 HCTT_RCO_ 25 August 2024 FINAL_0_0.pdf

UNICEF. (2024). *Two million children at risk as worst floods in three decades lash through eastern Bangladesh - UNICEF*. Press Release. https://www.unicef.org/press-releases/two-

million-children-risk-worst-floods-three-decades-lash-through-eastern-bangladesh

Wikipedia. (2026). *Subarnachar Upazila*. Encyclopedia. https://en.wikipedia.org/wiki/Subarnachar_Upazila