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.

Identifying Associated Factors of Common Mental Health Disorders in Riverbank Erosion Areas of Bangladesh Using Machine Learning Algorithms

Md. Tahidur Rahman

tahidur.iu@gmail.com

Islamic University

Md. Lotifur Rohman

Islamic University

Shyam Sundar Sarkar

Islamic University

Md. Fahim

R.P. Shaha University

Md. Merajul Islam

Jatiya Kabi Kazi Nazrul Islam University

Md. Maniruzzaman

Khulna University

Research Article

Keywords: Depression, Anxiety, Stress, Riverbank erosion, Loss of house, Machine Learning, Bangladesh

Posted Date: May 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9455372/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Riverbank erosion is a recurrent environmental hazard in Bangladesh that leads to displacement, loss of livelihood, loss of land, and social instability. These disruptions place affected populations at heightened risk of common mental health disorders (CMHDs), including depression, anxiety, and stress. However, evidence on the relative importance of socioeconomic, environmental, and displacement-related determinants remains limited.

Objective

This study aimed to identify key determinants of common mental health disorders among individuals living in riverbank erosion areas of Bangladesh using machine learning algorithms.

Methods

We analyzed cross-sectional data collected from erosion-affected communities, where household heads were interviewed face-to-face between September 2021 and January 2022 to assess Common Mental Health Disorders (CMHDs) using the Depression, Anxiety, and Stress Scale (DASS-21) (Depression ≥ 10, Anxiety ≥ 8, or Stress ≥ 15). Descriptive statistics were conducted to determine the prevalence of CMHDs, while χ^2-test and logistic regression (LR) analysis were applied to identify statistically significant risk factors associated with CMHDs. Moreover, three optimized machine learning (ML) algorithms- Random Forest (RF), XGBoost, and LR were implemented for predicting CMHDs.

Results

A total of 711 households were included, and randomly allocated to training (70%; n = 497) and testing (30%; n = 214) datasets. Among 497 respondents, 398 (80.08%) were exposed to riverbank erosion. The prevalence of common mental health disorders was significantly higher among the exposed group than the non-exposed group (exposed: 84.17% versus non- exposed: 15.83%, $P < 0.001$). The Random Forest classifier was found to be the best predictive model, with the accuracy (0.776) and precision (0.807), while loss of house, displacement, and loss of land were found to be the most influential predictors of CMHDs risk.

Conclusion

This study identified and predicted key risk factors for common mental health disorders among exposed individuals in riverbank erosion areas of Bangladesh using ML algorithms, which may assist

policymakers in mitigating the burden of CMHDs, with particular attention to housing loss, displaced populations, and low-income individuals.

1. Introduction

According to the World Health Organization (WHO), depression and anxiety are the two most Common Mental Health Disorders globally, affecting approximately 300 million people worldwide [1]. These disorders can impair daily functioning, lower work or academic performance, strain relationships, and, in severe cases, lead to suicide. At the societal level, they contribute to a significant economic burden [2]. The Depression Anxiety Stress Scales (DASS) were systematically developed based on the tripartite model of depression and anxiety to assess symptoms of Common Mental Health Disorders (CMHDs) [3]. Globally, it is estimated that about 280 million people suffer from depression, while around 272.5 million are affected by anxiety disorders [4, 5].

Natural disasters affect millions of people each year, not only physically but also psychologically [6]. Among such disasters, riverbank erosion has been associated with increased rates of depression, anxiety, and stress in both developed and developing countries. Climate change has further intensified the frequency and severity of such extreme weather events, including riverbank erosion, which now occur beyond the range of normal natural variability, thereby disrupting social, economic, and ecological systems [7, 8]. In South Asia, Bangladesh reports a higher prevalence of mental illness compared to neighbouring countries [9]. Among Bangladeshi adults, 16.8% of males and 17% of females experience mental health issues, yet 32.3% of them do not seek medical care [10]. Around 20 of the 64 districts in Bangladesh are highly vulnerable to riverbank erosion [11]. In 2019 alone, approximately 5,081 households were displaced, and nearly 2,270 hectares of land were lost due to erosion [12, 13]. According to Khatun et al. [14], if natural disasters keep happening more often and become stronger, about 30,366,230 households in Bangladesh could be affected by 2030. These extreme weather events contribute not only to widespread displacement and forced migration but also to deteriorating conditions that influence both physical and mental well-being [15]. We conducted a comprehensive literature review and found no population-based study in Bangladesh that specifically investigates the prevalence and risk factors of CMHDs in river-prone areas. Globally, only a limited number of studies have explored the common mental health impacts in other populations, such as the prevalence and predictors of CMHDs among pregnant women in Malaysia and a rural community-based cohort of northern Vietnamese women [16, 17] and the association between Common Mental Health Disorders and body weight in Eastern Saudi Arabia [18]. To the best of our knowledge, no prior research in Bangladesh or elsewhere has identified the key risk factors of Common Mental Health Disorders (CMHDs), based on Depression, Anxiety, and Stress (DAS) scores, among individuals affected by riverbank erosion. Therefore, this study aims to investigate how river erosion influences Common Mental Health Disorders in Bangladesh and to identify the risk factors associated with these conditions among affected populations.

2. Methods

A. Study area and sampling procedure

We purposefully chose three districts of Kushtia, Tangail, and Rajbari due to the high risk of river erosion. We initially selected three erosion-prone Upazila Kumarkhali in the Kushtia district, Bhuyapur Upazila in the Tangail district, and Rajbari Sadar Upazila in the Rajbari district. From each of these subdistricts, one union was selected: Shiladah Union from Kumarkhali, Arjuna Union from Bhuyapur, and Mijanpur Union from Rajbari Sadar. We then randomly selected seven villages (Ramkrisnopur, Charjoukuri, Ramcondropur, Chilimpur, Dhunchi, Sonakandor and Charnarayanpur) from Shiladah union, two villages (Kalua, Kumarkandi) from Arjuna union, and four villages (Arjuna, Tarai, Dhubliya, and Kuthiboira) from Mijanpur union. A total of 13 villages were chosen from the three districts. The detail description of the selected household sample is presented in Fig. 1. We calculated the sample size using the formula for a two-stage cluster sampling design:

Step 1: Calculate initial sample size (n_0):

$$n_0 = \frac{Z^2 P(1-P)}{d^2}$$

Where:

- $Z = 1.96$ (for 95% confidence), $P = 0.5$ (assumed prevalence), $d = 0.05$ (margin of error) [17]

$$n_0 = \frac{(1.96)^2 \times\ 0.5 \times\ 0.5}{(0.05)^2} = \frac{3.8416 \times\ 0.25}{0.0025} = \frac{0.9604}{0.0025} = 384.16 \approx\ 384$$

So, the initial sample size $n_0 = 384$

Step 2: Adjust for design effect ($Deff$):

$$Deff = 1 + (m-1)\rho$$

Where:

- $m = 50$ (average cluster size), $\rho\ = 0.01$ (intra-cluster correlation)

$$Deff = 1 + (50-1) \times\ 0.01 = 1 + 49 \times\ 0.01 = 1 + 0.49 = 1.49 \approx\ 1.50$$

Step 3: Apply design effect to get adjusted sample size (n_{deff})

$$n_{\text{deff}} = n_0 \times\ Deff = 384 \times\ 1.50 = 576$$

So, the sample size after adjusting for the design effect is 576.

Step 4: Adjust for non-response rate (r)

$$n_{\text{final}} = n_{\text{deff}} \times \ (1 + r) = 576 \times \ (1 + 0.10) = 576 \times \ 1.10 = 633.6 \approx \ 634$$

Therefore, the minimum required sample size was 634, and a total of 711 households were finally selected for the study to ensure greater precision, comprising 100 from Shiladha Union (Kushtia District), 201 from Arjuna Union (Tangail District), and 410 from Mijanpur Union (Rajbari district).

B. Study variables

Outcome variables: The outcome variable in this study is Common Mental Health Disorders (CMHD), defined based on three mental health indicators: depression, anxiety, and stress. These were assessed using the Depression Anxiety Stress Scales–21 items (DASS-21), which comprises three subscales—each with 7 items—corresponding to depression, anxiety, and stress. Participants responded to each item on a 4-point Likert scale ranging from **0** ("Did not apply to me at all") to 3 ("Applied to me very much or most of the time"). The total score for each subscale ranges from 0 to 21 and **is** multiplied by two (×2) to align with the DASS scoring guidelines. The following cut-off scores were used to determine the presence of CMHD symptoms: Depression: ≥10, Anxiety: ≥8, Stress: ≥15 [16, 19]. We defined a binary variable, where "1" indicates the presence of Common Mental Health Disorders (CMHD), and "0" indicates their absence.

C. Measures

Depression, Anxiety and Stress Scales

Depression

Involves symptoms like sadness, loss of interest, sleep or appetite problems, guilt, trouble concentrating, and fatigue. It can be caused by genetic, biological, environmental, sociodemographic, and psychological factors [20].

Anxiety

Anxiety is a reaction to perceived threats, marked by troubling thoughts, tension, fast heartbeat, sweating, dizziness, and chest pain. It may result from natural disasters, poverty, family issues, personality traits, or brain chemistry [21].

Stress

arises when people feel overwhelmed by demands. Symptoms include mood changes, flashbacks, avoidance, and heightened alertness. Causes include trauma, loss, serious illness, or accidents [22].

Predictor variables: The predictor variables included socioeconomic, demographic, and riverbank erosion-related factors, as well as the household's exposure status. Riverbank erosion-related variables included the distance of the homestead from the river (in miles), time since last displacement, amount of agricultural land (yes/no), loss of agricultural land (yes/no), cattle loss (yes/no), death of a relative

(yes/no), home loss (yes/no), social detachment (yes/no), substance misuse (yes/no), social encouragement for rehabilitation (yes/no), and hope for land recovery (yes/no). Demographic characteristics comprised sex (male/female), age group (in years), and number of children. Socioeconomic factors included the respondent's occupation, household income, and educational status. Displacement, a direct consequence of riverbank erosion, was categorized into three groups based on the time since the last displacement: non-displaced, displaced within the past three years, and displaced more than three years ago. Age was categorized into three groups: 22–37 years, 38–45 years, and 46–80 years. Similarly, the number of children was grouped into three categories: 1–2, 3–4, and more than 4.

D. Feature Selection

Feature selection was performed to identify the most relevant predictors for the machine learning models. In the first stage, a chi-square test was applied to identify variables significantly associated with the outcome variable, and only statistically significant variables held. In the second stage, logistic regression analysis was conducted on those chi-square significant variables to further assess their independent effect. A p-value < 0.05 was used to determine the statistical significance of these associations. These selected features were subsequently used for model development and evaluation on the testing dataset.. All statistical analyses were conducted using Stata version 15 (StataCorp, College Station, TX, USA) and Python (Jupyter Notebook; ipykernel 6.25.1).

E. Machine Learning Framework and Algorithms

The overview of the proposed machine-learning (ML) framework for predicting common mental health disorders (CMHD) is illustrated in Fig. 2. First, the complete dataset (including participants with and without CMHD) is randomly divided into two subsets: 70% training data and 30% testing data. Using the training subset, a univariate feature screening step is performed in which each candidate predictor is evaluated against the CMHD outcome using the Chi-square test of independence, and p-values are computed for all features. Features meeting the predefined significance threshold ($p < 0.05$) are retained as the optimal feature set [23].

Next, the reduced training data (containing only the optimal features) is used to build an ML-based prediction system by training three classifiers: Logistic Regression (LR), Random Forest (RF), and XGBoost

i. Logistic Regression (LR): Utilized as a linear baseline to compare against ensemble methods [24].
ii. Random Forest (RF): An ensemble bagging technique that builds several decision trees to manage feature interactions and increase classification accuracy [25].
iii. XGBoost: A gradient boosting framework that builds sequential trees to minimize residual errors through an iterative optimization [26].

F. Hyperparameter Tuning and Selection Criteria

Grid search [27] with fivefold cross-validation was used for systematic hyperparameter optimization to ensure the robustness and generalizability of the prediction models. By dividing the training data into five separate subsets and iteratively training the models on four folds while validating on the remaining one, this method made it possible to thoroughly evaluate different parameter combinations. This method guarantees that the chosen parameters function consistently across several data segments and reduces the possibility of overfitting [28]. The hyperparameters to be used in optimization are given in Table 1.

Table 1
Hyperparameter optimization of algorithms using the grid search method

Algorithms	Search range of each parameter
LR	'C': [0.001, 0.01, 0.1, 1, 10, 100], 'penalty': ['10', '12','13', '14'], 'solver': ['lbfgs', 'liblinear']
RF	'n_estimators': [100, 200, 300, 400, 450, 500, 550, 600], 'max_depth': [None, 10, 20], 'min_samples_split': [2, 5, 8, 10], 'max_features': ['sqrt', 'log2']
XGBoost	'n_estimators': [100, 200, 300, 400], 'max_depth': [3, 5, 7], 'learning_rate': [0.01, 0.1, 0.2], 'subsample': [0.8, 1.0], 'colsample_bytree': [0.8, 0.9, 1.0, 1.2]

The model parameters were estimated through maximum likelihood estimation (MLE) using the training dataset. The fitted model was subsequently employed to predict class membership (CMHD vs. healthy controls) in the test dataset. Furthermore, class probabilities for the response variable were computed to quantify the likelihood of each predicted outcome.

G. Model Evaluation and Feature Importance

A wide range of evaluation metrics were used to evaluate the optimized models' performance:

i. Accuracy: The model's overall accuracy [29].
ii. The model's capacity to distinguish between CMHD and non-CMHD situations is measured by the Area under the Receiver Operating Characteristic Curve (AUC) [30].
iii. F1-Score: The harmonic mean of recall and precision, guaranteeing a fair assessment for classification [31].
iv. Recall: Particularly given top priority in order to identify as many susceptible people as possible [32].

3. Results

Data were collected from 711 households to assess common mental health conditions in river erosion–affected areas of Bangladesh. The dataset was randomly divided into a training set (70%; n = 497) and a testing set (30%; n = 214). All subsequent analyses were conducted using the training sample (n = 497).

A. Bivariate analysis of CMHD and associated factors

The largest proportion of participants was from the Rajbari district (n = 281; 56.54%), followed by Tangail (n = 141; 28.37%) and Kushtia (n = 75; 15.09%). Approximately 47.48% of respondents were uneducated,

and 78.27% did not own any agricultural land. Of the total respondents, 397 (80.08%) were exposed to riverbank erosion, while 99 (19.92%) were not. Among the total participants, 6.64% reported the loss of cattle, 10.87% lost a relative, 44.27% lost their homestead, and 77.06% lost agricultural land due to erosion (Table 2). According to regional distribution, Rajbari had the highest percentage of respondents (56.54%), followed by Tangail (28.37%) and Kushtia (15.09%) (Table 2).

Figure 3 illustrates that among the exposed respondents, the highest proportion (84.17%) exhibited abnormal mental health conditions, indicating a strong association between erosion exposure and mental disorders.

The bivariate distribution of Common Mental Health Disorders (CMHD) across different predictors is presented in Table 2. We observed that the proportion of CMHD was higher in Rajbari (80.43%) and Tangail (78.01%) than in Kushtia (57.33%), a difference that is statistically significant ($p < 0.001$). The prevalence of CMHD increased with age, peaking at 84.66% in the oldest age group (46–80 years; $p = 0.007$). Compared to their counterparts, respondents who were uneducated (80.83%), had more than four children (90.20%), did not own agricultural land (80.46%), and had lower monthly earnings (> = 6000$ taka; 82.93%) showed the highest rates of CMHD. These differences were statistically significant ($p < 0.05$).By profession, the highest rates of CMHD were found among day laborers (87.23%) and drivers (83.33%). Among respondents who were exposed and relocated within the last three years, the prevalence of CMHD was significantly higher (97.56%) compared to those who did not migrate (59.01%; $p < 0.001$). We found that the prevalence of CMHD was statistically higher among those who lost their cattle (93.94% vs. 75.00%, $p = 0.023$), houses (94.55% vs. 61.73%, $p < 0.001$), and agricultural land (84.86% vs. 47.37%, $p < 0.001$) than among those who did not experience such losses. Furthermore, respondents who required social encouragement (88.56% vs. 67.91%, $p < 0.001$), experienced social detachment (91.82% vs. 68.93%, $p < 0.001$), and had hope for land return (87.25% vs. 73.42%, $p < 0.001$) also exhibited significantly higher rates of CMHD.

Table 2
Prevalence of common mental health disorders with different predictors.

Characteristics	Overall N = 497(100%) n (%)	Common mental health disorders		χ^2-value	p-value
		Normal N = 118 (23.74%)	Abnormal N = 379 (76.26%)		
n (%)		n (%)			
Socio-demographic and economic factors					
Exposed					
Yes	398 (80.08)	63 (15.83)	335 (84.17)	66.929	< 0.001
No	99 (19.92)	55 (55.56)	44 (44.44)		
Area					
Kushtia	75 (15.09)	32 (42.67)	43 (57.33)	17.773	< 0.001
Tangail	141 (28.37)	31 (21.99)	110 (78.01)		
Rajbari	281 (56.54)	55 (19.57)	226 (80.43)		
Age group(year)					
22–37	170 (34.21)	45 (26.47)	125 (73.53)	9.825	0.007
38–45	164 (33.00)	48 (29.27)	116 (70.73)		
46–80	163 (32.8)	25 (15.34)	138 (84.66)		
Educational status					
Uneducated	236 (47.48)	45 (19.07)	191 (80.83)	4.943	0.026
Educated	261 (52.52)	73 (27.97)	188 (72.03)		
Profession					
Housewife	230 (46.28)	59 (25.65)	171 (74.35)	9.709	0.137
Farmer	62 (12.47)	13 (20.97)	49 (79.03)		
Business	55 (11.07)	19 (34.55)	36 (65.45)		
Day labor	47 (9.46)	6 (12.77)	41 (87.23)		
Service	23 (4.63)	7 (30.43)	16 (69.57)		
Driver	30 (6.04)	5 (16.67)	25 (83.33)		

Characteristics	Overall N = 497(100%) n (%)	Common mental health disorders		χ^2-value	p-value
		Normal N = 118 (23.74%)	Abnormal N = 379 (76.26%)		
n (%)		n (%)			
Others	50 (10.06)	9 (18.00)	41 (82.00)		
No of children					
1–2	196 (39.44)	52 (26.53)	144 (73.47)	6.374	0.041
3–4	250 (50.30)	61 (24.40)	189 (75.60)		
> 4	51 (10.26)	5 (9.80)	46 (90.20)		
Monthly earnings (taka)					
≤ 6000	41 (8.25)	7 (17.07)	34 (82.93)	13.752	0.003
6001–10000	157 (31.59)	26 (16.56)	131 (75.25)		
10001–15000	202 (40.64)	50 (24.75)	152 (83.44)		
≥ 15000	97 (19.52)	35 (36.08)	62 (63.92)		
Agricultural land					
No	389 (78.27)	76 (19.54)	313 (80.46)	16.432	< 0.001
Yes	108 (21.73)	42 (38.89)	66 (61.11)		
Amount of agricultural land (decimal)					
No agricultural land	389 (78.27)	76 (19.54)	313 (80.46)	18.43	< 0.001
1–66	77 (15.49)	28 (36.36)	49 (63.64)		
67–99	11 (2.21)	5 (45.45)	6 (54.55)		
≥ 100	20 (4.02)	9 (45.00)	11 (55.0)		
Factors contributing to riverbank erosion					
Exposure category					
Exposed	398 (80.08)	63 (15.83)	335 (84.17)	66.929	< 0.001
Non-exposed	99 (19.92)	55 (55.56)	44 (44.44)		
Duration of internal displacement (Years)					

Characteristics	Overall N = 497(100%) n (%)	Common mental health disorders		χ^2-value	p-value
		Normal N = 118 (23.74%)	Abnormal N = 379 (76.26%)		
n (%)		n (%)			
Not migrate	222 (44.67)	91 (40.99)	131 (59.01)	78.148	< 0.001
≤ 3	164 (33)	4 (2.44)	160 (97.56)		
> 3	111 (22.33)	23 (20.72)	88 (79.28)		
Distance of homestead from the river (in miles)					
≤ 0.2					0.978
> 0.2					
Loss of cattle					
No	464 (93.36)	116 (25.00)	384 (75)	5.103	0.023
Yes	33 (6.64)	2 (6.06)	31 (93.94)		
Death of relative					
No	443 (89.13)	333 (75.17)	110 (24.83)		0.143
Yes	54 (10.87)	46 (85.19)	8 (14.81)		
Loss of agricultural land					
No	144 (22.94)	60 (52.63)	54 (47.37)	66.136	< 0.001
Yes	383 (77.06)	58 (15.14)	325 (84.86)		
Substance misuse					
No	481 (96.78)	117 (24.32)	364 (75.68)	1.885	0.1698
Yes	16 (3.22)	1 (6.25)	15 (93.75)		
Social encouragement					
No	296 (59.56)	95 (32.09)	201 (67.91)	27.07	< 0.001
Yes	201 (40.44)	23 (11.44)	178 (88.56)		
Loss of home					
No	277 (55.73)	106 (38.27)	171 (61.73)	71.114	< 0.001
Yes	220 (44.27)	12 (5.45)	208 (94.55)		

Characteristics	Overall N = 497(100%) n (%)	Common mental health disorders		χ^2-value	p-value
		Normal N = 118 (23.74%)	Abnormal N = 379 (76.26%)		
n (%)		n (%)			
Social detachment					
No	338 (68.01)	105 (31.07)	233 (68.93)	30.038	< 0.001
Yes	159 (31.99)	13 (8.18)	146 (91.82)		
Hope for land return					
No	395 (79.48)	105 (26.58)	290 (73.42)	7.826	< 0.001
Yes	102 (20.52)	13 (12.75)	89 (87.25)		

P-values for each variable were computed using the χ^2 test statistic.

A Parcats Plot (Fig. 4) illustrates a comprehensive snapshot of the study and visualizes the interaction between the most significant determinants identified by the chi-square tests: Displacement, Exposure, Loss land, and Loss house. The concentration of CMHD was significantly higher among those who had frequent displacements (> 3 times), even if the percentage of people with lower displacement frequency (< = 3 times) was higher. Participants who lost their homes showed the lowest levels of psychological resilience (green routes), while those who did not lose any assets showed the highest levels. Critical differences in stressors at the community, household, and environmental levels were further highlighted by the much greater trajectories toward CMHD of those who experienced the junction of high exposure and dual land-house loss compared to those who did not.

B. Binary Logistic Regression Analysis of CMHD and Associated Factors

The logistic regression analysis's findings are shown in Table 3. Higher probabilities of aberrant outcomes were highly correlated with exposure (OR = 6.65, 95% CI: 4.12–10.73; p < 0.001). The probabilities of aberrant outcomes were considerably greater for participants who lost agricultural land (OR = 6.23, 95% CI: 3.92–9.88; p < 0.001), livestock (OR = 5.17, 95% CI: 1.22–21.92; p = 0.026), and home (OR = 10.74, 95% CI: 5.72–20.18; p < 0.001). Conversely, having formal education (OR = 0.61, 95% CI: 0.40–0.93; p = 0.020), having a larger monthly income (≥ 15000 taka: OR = 0.36, 95% CI: 0.15–0.91; p = 0.030), owning agricultural land (OR = 0.38, 95% CI: 0.24–0.61; p < 0.001). In addition to psychosocial characteristics including social encouragement (OR = 3.66, 95% CI: 2.22–6.02; p < 0.001) and social detachment (OR = 5.06, 95% CI: 2.74–9.34; p < 0.001), respondents with more than four children had greater odds (OR = 3.32, 95% CI: 1.25–8.82; p = 0.016). Geographically, respondents from Rajbari (OR = 1.90, 95% CI: 1.11–3.25; p = 0.019) and Tangail (OR = 2.48, 95% CI: 1.33–4.62; p = 0.004) had greater

odds than those from Kushtia. Increased risks of aberrant outcomes were also substantially correlated with older age (46–80 years: OR = 1.99, 95% CI: 1.15–3.43; p = 0.014) and internal displacement (≤ 3 years: OR = 27.79, 95% CI: 9.94–77.64; p < 0.001; >3 years: OR = 2.66, 95% CI: 1.56–4.52; p < 0.001).

Table 3
Risk Factors Associated with Common Mental Health Disorders: Binary Logistic Regression Analysis.

Characteristic	OR (95% CI)	p-value
Agricultural land		
No (Ref.)	1	
Yes	0.38 (0.24–0.61)	< 0.001
Exposed		
No (Ref.)	1	
Yes	6.65 (4.12–10.73)	< 0.001
Loss of agricultural land		
No (Ref.)	1	
Yes	6.23 (3.92–9.88)	< 0.001
Loss of cattle		
No (Ref.)	1	
Yes	5.17 (1.22–21.92)	0.026
Monthly earnings (taka)		
≤ 6000	1	
6001–10000	1.04 (0.42–2.59)	0.937
10001–15000	0.63 (0.26–1.50)	0.293
≥ 15000	0.36 (0.15–0.91)	0.03
Loss of home		
No (Ref.)	1	
Yes	10.74 (5.72–20.18)	< 0.001
Hope for land return		
No (Ref.)	1	
Yes	79504926.22 (0.00 - inf)	0.998
No of children		
1–2 (Ref.)	1	
3–4	1.12 (0.73–1.72)	0.608

Characteristic	OR (95% CI)	p-value
Agricultural land		
> 4	3.32 (1.25–8.82)	0.016
Social encouragement		
No (Ref.)	1	
Yes	3.66 (2.22–6.02)	< 0.001
Social detachment		
No (Ref.)	1	
Yes	5.06 (2.74–9.34)	< 0.001
Area		
Kushtia (Ref.)	1	
Tangail	2.48 (1.33–4.62)	0.004
Rajbari	1.90 (1.11–3.25)	0.019
Amount of agricultural land (decimal)		
No agricultural land (Ref.)	1	
1–66	0.42 (0.25–0.72)	0.001
67–99	0.29 (0.09–0.98)	0.046
> 100	0.30 (0.12–0.74)	0.009
Educational status		
Uneducated (Ref.)	1	
Educated	0.61 (0.40–0.93)	0.02
Age group (year)		
22–37 (Ref.)	1	
38–45	0.87 (0.54–1.40)	0.569
46–80	1.99 (1.15–3.43)	0.014
Duration of internal displacement (Years)		
Not migrate	1	
≤ 3	27.79 (9.94–77.64)	< 0.001
> 3	2.66 (1.56–4.52)	< 0.001

Ref: Reference.

C. Hyperparameter Optimization of ML Algorithms and Comparative Analysis

To complement the conventional statistical analysis, three supervised machine-learning algorithms—Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)—were trained to predict the risk of common mental health disorders (CMHD) in the study population. The final optimized hyperparameter configurations for each model are presented in Table 4.

Table 4
Optimized value of the Algorithms

Algorithms	Optimized values of Hyperparameters
LR	'C': 0.001, 'penalty': 'l2', 'solver': 'lbfgs
RF	'max_depth': 10, 'max_features': 'log2', 'min_samples_split': 10, 'n_estimators': 500
XGBoost	'colsample_bytree': 0.8, 'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 200, 'subsample': 1.0

Model performance was evaluated using Accuracy, Area Under the Receiver Operating Characteristic Curve (AUC), and F1-score, with Precision and Recall additionally reported to characterize classification trade-offs. The detailed comparative results of the optimized models are summarized in Table 5.

Table 5
Comparative performance metrics of optimized machine learning algorithms.

Model	Accuracy	AUC Score	F1-Score	Precision	Recall
Logistic Regression	0.762	0.746	0.865	0.762	1
Random Forest	0.776	0.757	0.863	0.807	0.926
XGBoost	0.771	0.763	0.864	0.788	0.957

Overall, the three optimized models showed similar F1-scores (0.863–0.865), indicating a comparable balance between precision and recall in identifying CMHD in this dataset. The Random Forest model achieved the highest accuracy (0.776) and the highest precision (0.807), suggesting fewer false-positive classifications relative to the other approaches. In contrast, XGBoost yielded the highest AUC (0.763), indicating the strongest overall discriminative ability across decision thresholds. Logistic Regression demonstrated perfect recall (1.000), meaning all CMHD cases were identified by the model, but this came with lower precision (0.762), consistent with a larger number of false positives. Taken together, the findings indicate that tree-based ensemble models (RF and XGBoost) provided slightly stronger discrimination and/or precision, whereas Logistic Regression maximized sensitivity in this sample. Although XGBoost showed the highest AUC, Random Forest demonstrated the best overall balance across accuracy, precision, recall, and F1-score. Therefore, Random Forest was selected as the most appropriate algorithm for predicting CMHD, as shown in Figs. 5 and 6.

D. Variable Significance and Feature Contribution using Random Forest

Since the Random Forest model showed the strongest overall predictive performance, we examined which predictors contributed most to its classification of CMHD risk. Feature contribution was quantified using the model's built-in tree-based feature. To improve interpretability, importance values were aggregated at the variable level as presented in Fig. 7.

As shown in Fig. 7, house loss was by far the most influential predictor (importance = 0.126), indicating that housing disruption is the dominant signal used by Random Forest when distinguishing individuals with CMHD from those without. The next most influential predictors were displacement (0.114) and land loss (0.101), highlighting the importance of erosion-related asset loss and mobility as major stress pathways. Exposure to erosion also contributed meaningfully (0.102), suggesting that direct environmental risk remains informative even after accounting for downstream losses. Beyond these primary shock-related factors, several household and psychosocial conditions contributed additional predictive information, including household income (0.099), number of child (0.058), and cope with that situation (cope_withh) (0.066). Overall, Fig. 7 shows that Random Forest predictive power is driven mainly by displacement-related shocks—especially house loss—followed by land loss and displacement history, with exposure, number of child, coping, and income providing incremental contributions to CMHD risk classification.

4. Discussion

This is a population based cross-sectional study, offer a representative estimate of the prevalence of CMHD and explores its association with individuals living in river erosion-prone areas. A comprehensive review of the literature found that no similar study has been conducted in the country. This study was designed to fill that gap by measuring the prevalence of CMHD and identifying the key risk factors contributing to these mental health issues in populations displaced or affected by riverbank erosion. Understanding these factors is crucial for developing targeted interventions and support programs to improve mental health outcomes in these vulnerable communities. However, our study found that people exposed to river erosion had significantly higher rates of CMHDs based on the DAS score compared to those not exposed (**Figure. 3**). These results align with earlier studies [16, 33–35]. Thomas et al. [35] and Arobi et al. [36] found that people exposed to river erosion had a higher risk of developing DASS disorders compared to those not exposed.

The reason is that exposed individuals lost various assets such as houses, property, and cattle, and some turned to drugs to cope with stress, anxiety, and frustration. Additionally, river erosion exposure may have worsened existing psychological distress, sleep problems, physical complaints, and psychosocial and behavioral issues [6]. We also found that the rates of abnormal Common Mental Health Disorders differed significantly among the three districts studied. Based on DAS scores, people in Rajbari and Tangail districts had a much higher risk of abnormal Common Mental Health Disorders compared to those in Kushtia district. One possible explanation is that during the dry season, people

affected by erosion in Tangail and Rajbari districts had access to riverbeds along the Padma and Brahmaputra rivers due to their previous land boundaries. We also found that females were more likely to have DAS disorders than males, which coincided with earlier studies [33, 37]. According to DAS scores, Common Mental Health Disorders were significantly more common among illiterate people than those with education, consistent with findings from Altaf et al. and Weyerer et al. [33, 38]. Older age groups, especially those aged 38–45 and above 45, had higher rates of abnormal Common Mental Health Disorders compared to younger people (≤ 37 years), a pattern supported by previous research [33, 39]. In the study population, we found that the prevalence of abnormal CMHD increased as the number of children in a family grew. This may be because having more children adds to the economic burden caused by river erosion. Respondents with larger families had higher chances of developing CMHD compared to those with fewer children, as managing a big family under financial stress was especially difficult for affected individuals. This finding aligns with previous research [33]. Additionally, the odds of having abnormal CMHD were higher among landless participants living within 0.2 miles of the riverbank compared to those who owned land and lived farther than 0.2 miles from the river.

Further, we use advanced machine learning (ML) techniques to greatly extend traditional statistical methods by capturing high-dimensional, non-linear interactions between socio-environmental stresses. With a peak accuracy of 0.776 and a precision of 0.807, the Random Forest model proved to be the best classifier, surpassing the baseline Logistic Regression (AUC = 0.746). This discrepancy in performance suggests that complicated convergences, which are frequently oversimplified by traditional linear models, impact CMHD risk in populations affected by disasters.

The feature importance analysis revealed "Loss of House" to be the main factor influencing CMHD classification, with the highest importance scores in the algorithmic framework, in line with the statistical evaluation. This finding indicates that the primary cause of psychological discomfort is the sudden loss of shelter. Algorithmic paths also demonstrated that secondary factors increase this risk; as the Lollipop Chart (Fig. 7) illustrates, significant contributions from Land Ownership and Displacement together boost the model's predictive power. From the standpoint of public health, the Random Forest model's high Accuracy (0.776).

These results show that ML-driven predictive modelling is an effective tool for policymakers, despite the drawbacks of cross-sectional data and the "black-box" character of ensemble models. After erosion incidents, it makes it easier to quickly identify high-risk homes, enabling more focused mental health interventions.

Study limitations

This study had several limitations that should be considered when interpreting the findings. First, it was a cross-sectional study done in only three river erosion-prone districts (Kushtia, Rajbari, and Tangail) with a relatively small sample size, which means the findings may not apply to the whole country. Another limitation is that the study focused only on three major mental health conditions—depression, anxiety, and stress—as indicators of Common Mental Health Disorders (CMHD). Other important mental health

issues, like neurodevelopmental disorders, sleep disorders, schizophrenia, bipolar disorder, obsessive-compulsive disorder, panic disorder, mixed anxiety and depression, and social phobia, were not covered in this study.

5. Conclusions

This study demonstrates that people exposed to riverbank erosion had worse mental health and were more likely to develop common mental health disorders (CMHD) than those who weren't exposed. Using advanced machine learning modeling, risk factors were identified, with the sudden "Loss of House", "Loss of land", and subsequent "Displacement" emerging as the most critical predictors of psychosocial distress. Additionally, our study highlights that certain demographic groups are more vulnerable to mental health disparities, such as women, the elderly, people without agricultural land, and those with bigger families (more than four children). Property loss and economic instability combine to create high-risk pathways for CMHD, as demonstrated by the Random Forest model's superior performance (Accuracy = 0.776), which highlights the complex and non-linear nature of these socio-environmental stressors. Special attention should also be given to vulnerable groups, including women, the elderly, landless individuals, and those with larger families, to reduce mental health inequalities. Introducing low-interest loans in erosion-prone areas could further help recovery and improve people's ability to adapt.

Declarations

Ethical approval

The study's procedures were conducted in accordance with the protocol approved by the Ethical Review Committee of Islamic University, Kushtia-7003, Bangladesh (Reference No. FS/IU/2025/17). All participants were informed prior to data collection about the study's objectives. The questionnaire was translated into the native language (Bangla) and read out to the respondents. When it was decided to conduct an interview, the educated respondents gave their written consent, and the uneducated participants signed the consent form with their fingerprints. Before taking part in this study, each participant gave their informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

The authors did not receive any specific funding for this study.

Author Contribution

Conceptualization: Md. Tahidur Rahman and Md Lotifur Rohman; Data curation and statistical analysis: Md. Tahidur Rahman, Md Lotifur Rohman, Md. Maniruzzaman and Md. Merajul Islam; Drafting the manuscript: Md. Tahidur Rahman, Md Lotifur Rohman, Shyam Sundar Sarkar and Md. Fahim; Review and editing: Md. Tahidur Rahman, Md Lotifur Rohman, Md. Maniruzzaman and Md. Merajul Islam; Contribution to discussion, interpretation, and finalization: Md. Tahidur Rahman, Md Lotifur Rohman, Shyam Sundar Sarkar and Md. Fahim; Supervision: Md. Tahidur Rahman, Md. Maniruzzaman and Md. Merajul Islam; Validation. All authors have read and approved the final version of the manuscript.

Acknowledgement

The authors would like to acknowledge to Dr. Md. Sazzed Hossain Zahid, Associate professor, Department of English, Islamic University Kushtia, Bangladesh, for his constructive feedback and assistance in improving the language of the manuscript.

Data Availability

The datasets used in the current study are available from the corresponding author upon reasonable request.

References

1. Lee EH, Moon SH, Cho MS, Park ES, Kim SY, Han JS, Cheio JH. The 21-Item and 12-Item Versions of the Depression Anxiety Stress Scales: Psychometric Evaluation in a Korean Population. Asian Nurs Res (Korean Soc Nurs Sci). 2019;13. 10.1016/j.anr.2018.11.006.
2. Rapaport MH, Clary C, Fayyad R, Endicott J. Quality-of-life impairment in depressive and anxiety disorders. Am J Psychiatry. 2005;162. 10.1176/appi.ajp.162.6.1171.
3. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. 1995.
4. WHO. Depressive disorder (depression). 2023.
5. Baxter AJ, Vos T, Scott KM, Norman RE, Flaxman AD, Blore J, Whiteford HA. The regional distribution of anxiety disorders: Implications for the Global Burden of Disease Study, 2010. Int J Methods Psychiatr Res. 2014;23. 10.1002/mpr.1444.
6. Norris FH, Friedman MJ, Watson PJ, Byrne CM, Diaz E, Kaniasty K. 60,000 Disaster victims speak: Part I. An empirical review of the empirical literature, 1981–2001. Psychiatry. 2002;65. 10.1521/psyc.65.3.207.20173.
7. Costello A, Abbas M, Allen A, et al. Managing the health effects of climate change. Lancet and University College London Institute for Global Health Commission. Lancet. 2009;373. 10.1016/S0140-6736(09)60935-1.

8. Watts N, Amann M, Ayeb-Karlsson S, et al. The Lancet Countdown on health and climate change: from 25 years of inaction to a global transformation for public health. Lancet. 2018;391. 10.1016/S0140-6736(17)32464-9.
9. Ogbo FA, Mathsyaraja S, Koti RK, Perz J, Page A. The burden of depressive disorders in South Asia, 1990–2016: findings from the global burden of disease study. BMC Psychiatry. 2018;18. 10.1186/s12888-018-1918-1.
10. WHO: National Mental Health Survey of Bangladesh, 2018–2019: provisional fact sheet. Dhaka. 2019. Published Online First: 2019.
11. Masum JH. Climatic Hazards in Bangladesh-A Literature Review Jahangir Hasan Masum. Published Online First. 2020. 10.13140/RG.2.2.10239.94882.
12. Munna TI. River erosion: A silent disaster in Bangladesh. 2019.
13. Kamal M. Riverbank Erosion and Migration: A Study on Displaced People from Governance and Cultural Perspective. 2019, XIII:23–34.
14. Khatun F, Bhattacharya D, Rahman M, Moazzem KG, Khan TI, Sabbih MA, Saadat S. Four Years of SDGs in Bangladesh: Measuring progress and charting the path forward. 2020.
15. IPCC: Intergovernmental Panel on Climate Change working group II. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects, Polar regions. New York: Cambridge Univ; 2014.
16. Nagandla K, Nalliah S, Yin L, et al. Prevalence and associated risk factors of depression, anxiety and stress in pregnancy. Int J Reprod Contracept Obstet Gynecol Published Online First. 201610.18203/2320 – 1770.ijrcog20162132.
17. Meltzer H, Bebbington P, Brugha T, Farrell M, Jenkins R. The relationship between personal debt and specific common mental disorders. Eur J Public Health. 2013;23:108–13. 10.1093/eurpub/cks021.
18. Darwish MA, Al Turki ZA, Sabra AA. Relation between Common Mental Disorders and Body Weight using Arabic DASS 21 scale (Depression, Anxiety and Stress Scale) among Adult Saudi Females Attending Primary Care. East Saudi Arabia. 2014.
19. Sekoni O, Mall S, Christofides N. The relationship between protective factors and common mental disorders among female urban slum dwellers in Ibadan, Nigeria. PLoS ONE. 2022;17:1–16. 10.1371/journal.pone.0263703.
20. NIH. Depression: Transforming the understanding and treatment of mental illnesses. 2024.
21. Park SG, Min KB, Chang SJ, Kim HC, Min JY. Job stress and depressive symptoms among Korean employees: The effects of culture on work. Int Arch Occup Environ Health. 2009;82. 10.1007/s00420-008-0347-8.
22. Eske J. What is acute stress disorder? MedicalNewsToday. 2019.
23. Pearson K. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. London, Edinburgh, Dublin Philos Mag J Sci. 1900, 50:157–75.

24. Lavalley MP. Statistical Primer for Cardiovascular Research Logistic Regression. 2008, 2395–9. 10.1161/CIRCULATIONAHA.106.682658
25. Desarrollo U, Desarrollo U, Desarrollo U. Integration of RNA Editing with Multiomics Data Improves Machine Learning Models for Predicting Drug Responses in Breast Cancer Patients. 2024.
26. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proc ACM SIGKDD Int Conf Knowl Discov Data Min. 2016, 13-17-Augu:785–94. 10.1145/2939672.2939785
27. Liashchynskyi P. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. arXiv Prepr ArXiv191206059. 2019.
28. Yang L, Shami A. On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice. 1–69.
29. Swets JA, Swets JA. Measuring the Accuracy of Diagnostic Systems Linked references are available on JSTOR for this article: Measuring the Accuracy of Diagnostic Systems. 1988, 240:1285–93.
30. Lobo JM, Jiménez-valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. 2007, 1–7. 10.1111/j.1466-8238.2007.00358.x
31. Sokolova M. Beyond Accuracy, F-score and ROC : a Family of Discriminant Measures for Performance Evaluation.
32. Buckland M, Gey F. The Relationship between Recall and Precision. Published Online First: 1994. 10.1002/(sici)1097–4571(199401)45.
33. Hossain A, Alam MJ, Haque MR. Effects of riverbank erosion on mental health of the affected people in Bangladesh. PLoS ONE. 2021;16. 10.1371/journal.pone.0254782.
34. Munro A, Kovats RS, Rubin GJ, et al. Effect of evacuation and displacement on the association between flooding and mental health outcomes: a cross-sectional analysis of UK survey data. Lancet Planet Heal. 2017;1. 10.1016/S2542-5196(17)30047-5.
35. Jom Thomas J, Prakash B, Kulkarni P, Narayana Murthy MR. Prevalence and severity of depression among people residing in flood affected areas of Kerala. Int J Community Med Public Heal. 2019;6. 10.18203/2394–6040.ijcmph20190600.
36. SarzamArobi, Naher J, Soron TR. Impact of River Bank Erosion on Mental Health and Coping Capacity in Bangladesh. Glob Psychiatry. 2019;2. 10.2478/gp-2019-0011.
37. Yeshaw Y, Mossie A. Depression, anxiety, stress, and their associated factors among Jimma university staff, Jimma, southwest Ethiopia, 2016: A cross-sectional study. Neuropsychiatr Dis Treat. 2017;13. 10.2147/NDT.S150444.
38. Weyerer S, Eifflaender-Gorfer S, Köhler L, et al. Prevalence and risk factors for depression in non-demented primary care attenders aged 75 years and older. J Affect Disord. 2008;111. 10.1016/j.jad.2008.02.008.
39. Pinheiro KAT, Horta BL, Pinheiro RT, Horta LL, Terres NG, Da Silva RA. Common mental disorders in adolescents: A population based cross-sectional study. Rev Bras Psiquiatr. 2007;29. 10.1590/s1516-44462006005000040.

Figures

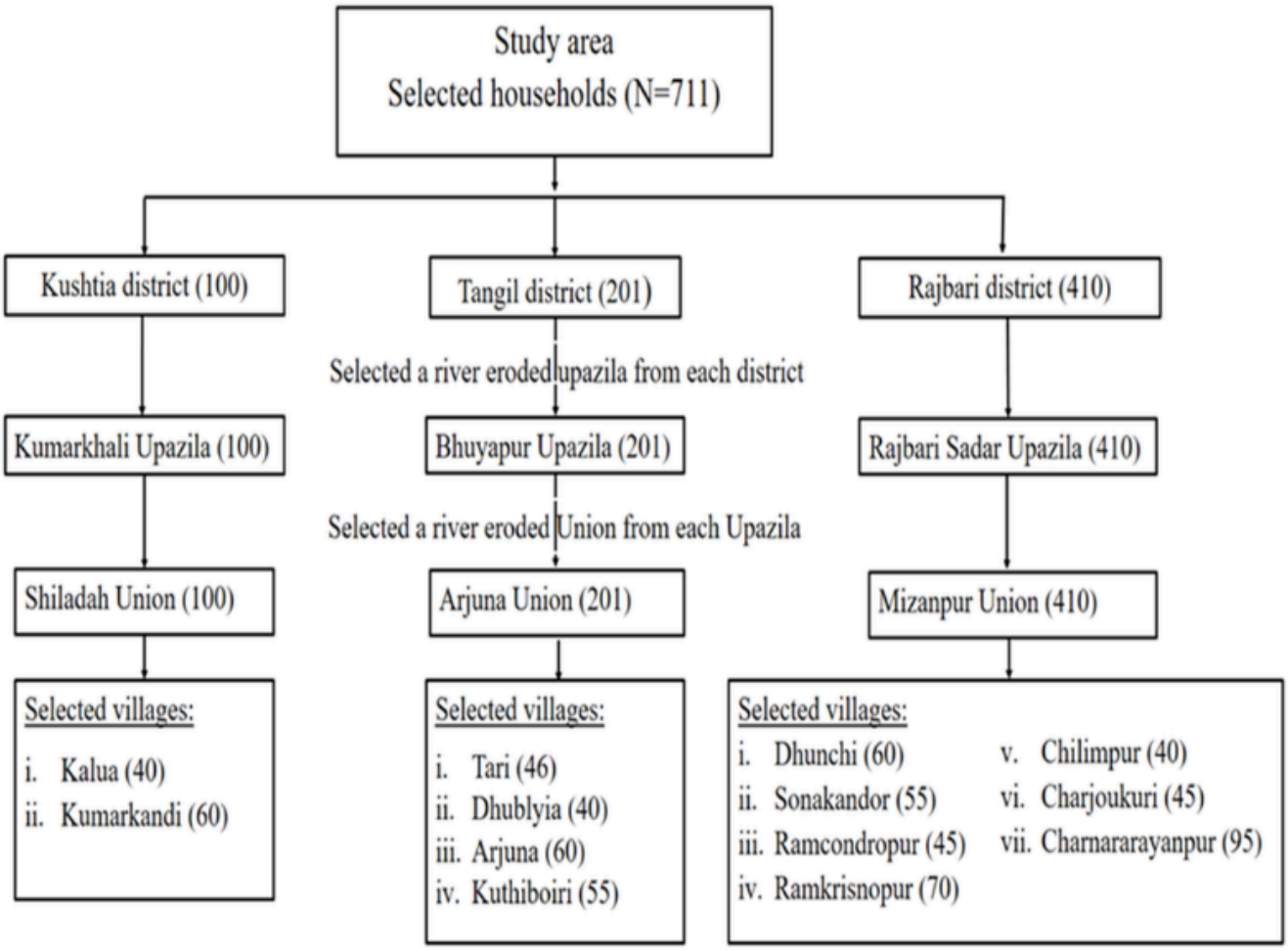


Figure 1

Distribution of the sampled households.

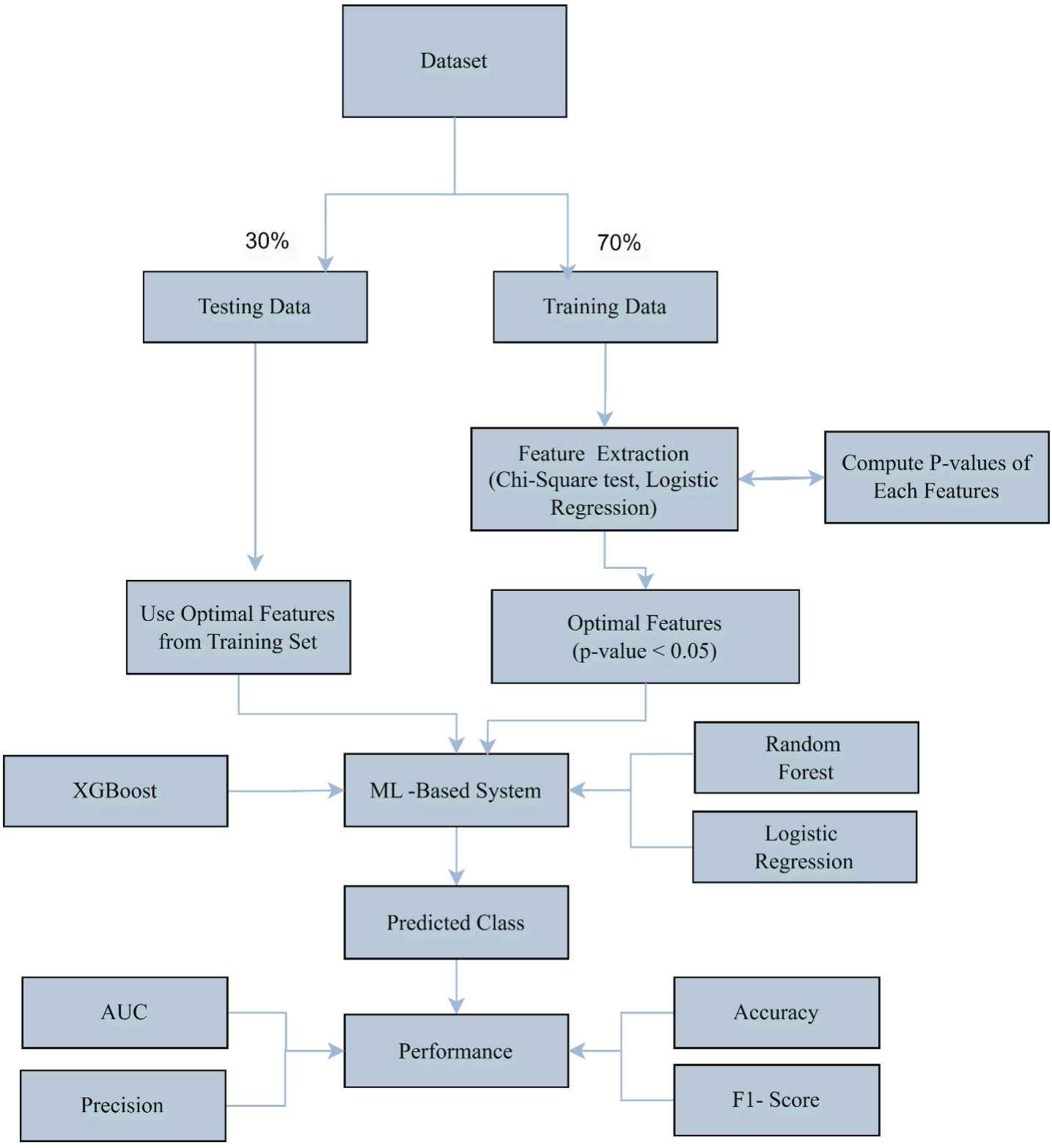


Figure 2

A proposed ML-based framework for the prediction of CMHD.

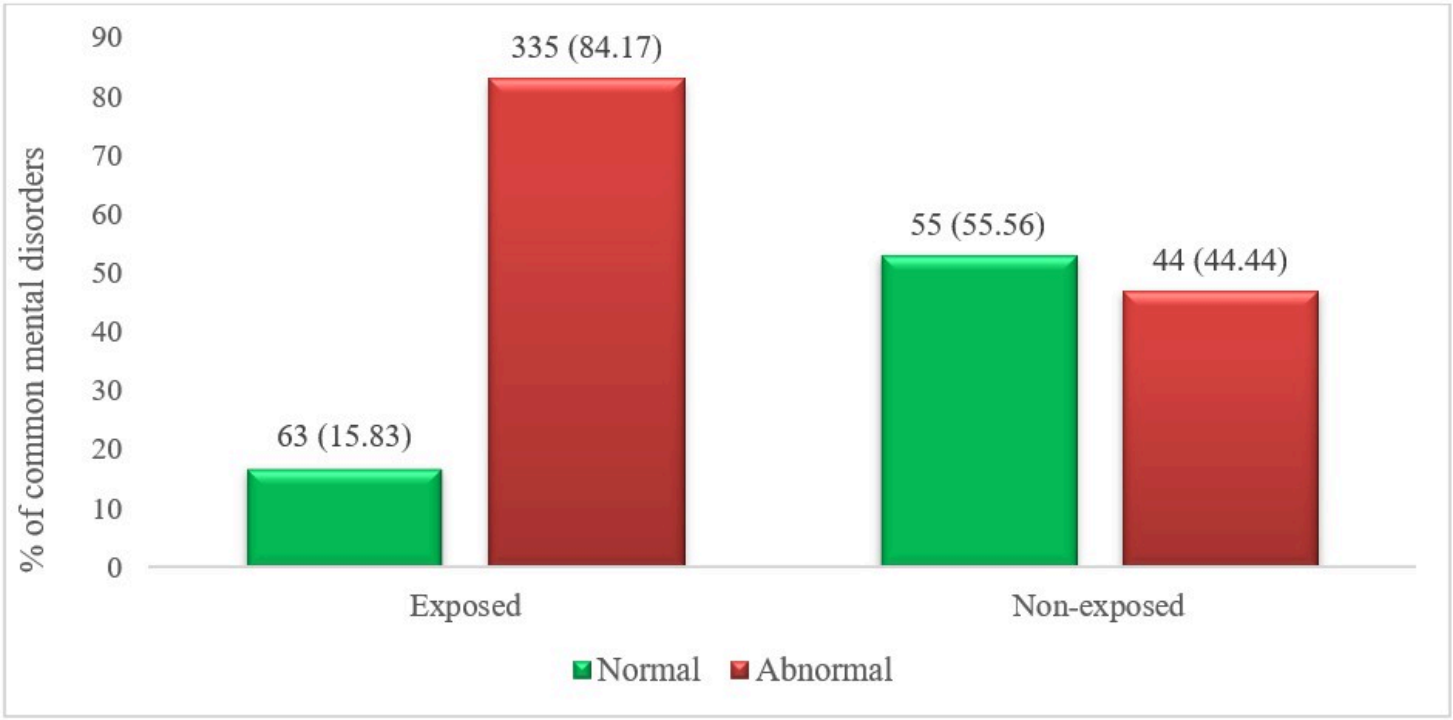


Figure 3

Proportion of respondents with Common Mental Health Disorders in exposed and non-exposed groups: N =497.

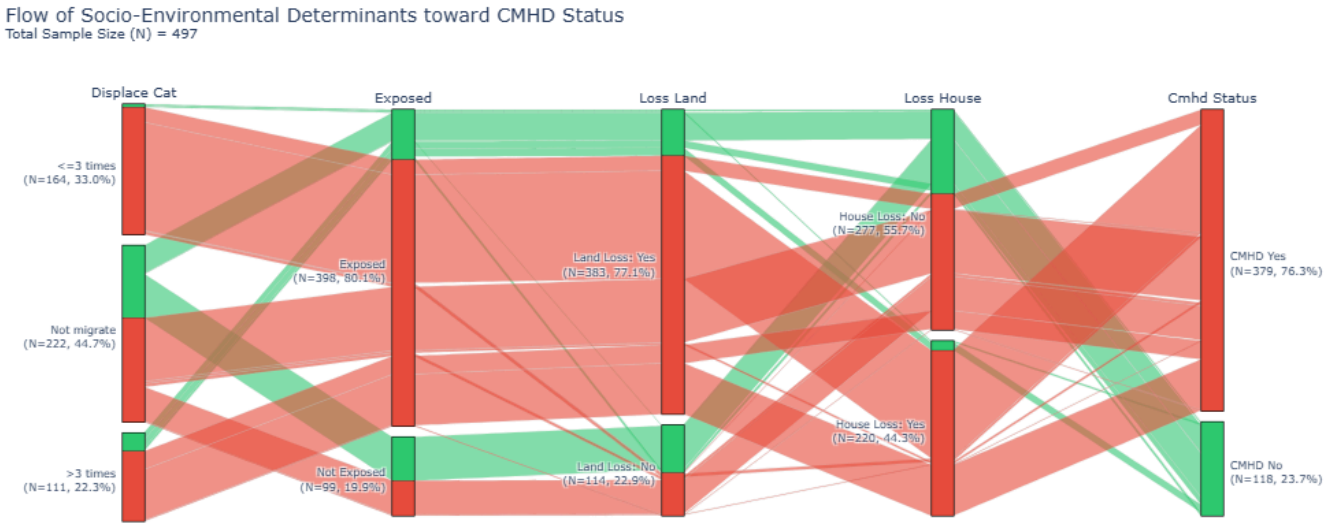


Figure 4

Flow of socio- environmental factors toward CMHD Status

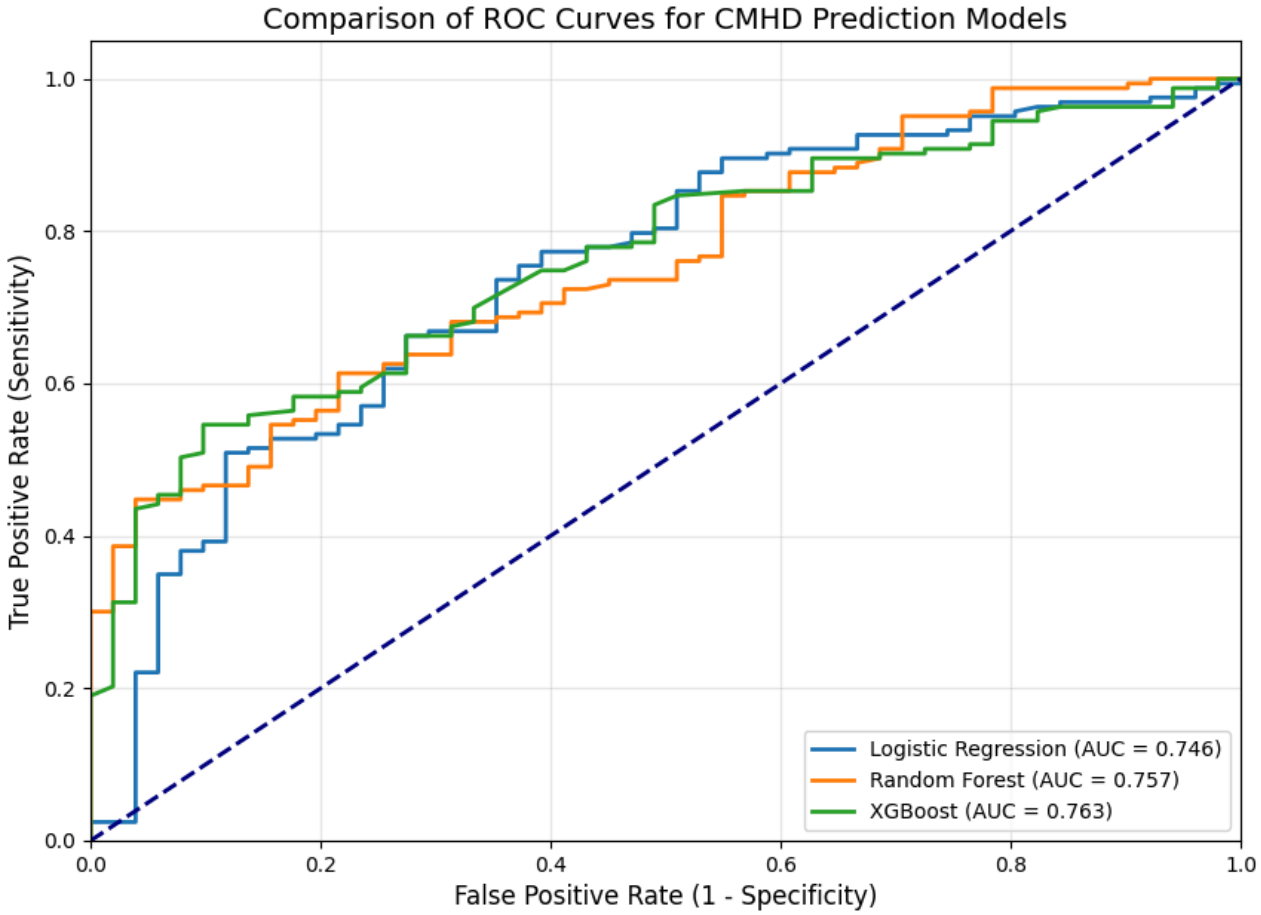


Figure 5

Comparison of Receiver Operating Characteristic (ROC) curves for three machine learning models in predicting CMHD

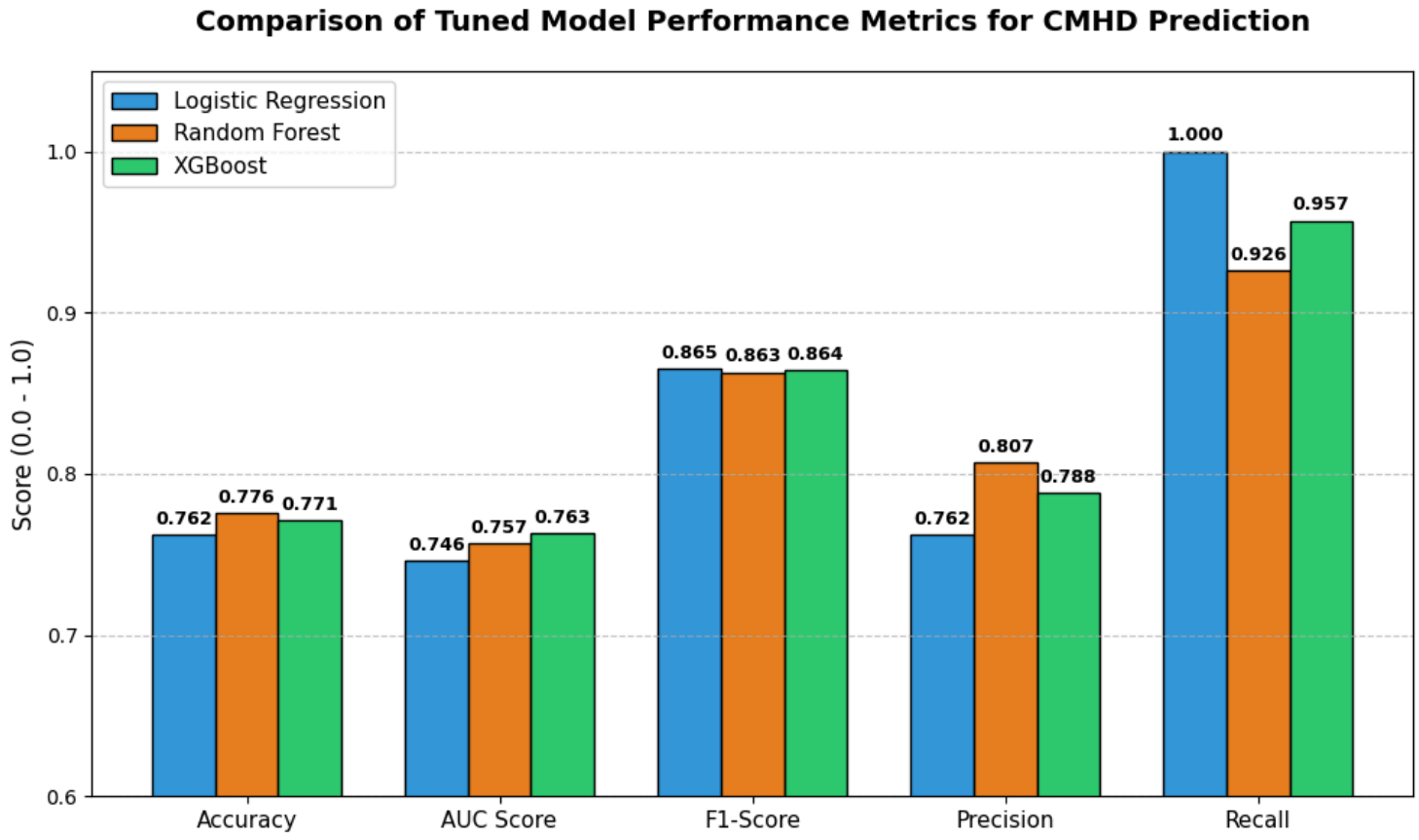


Figure 6

Comparison of Tuned Model Performance Metrics for CMHD Prediction

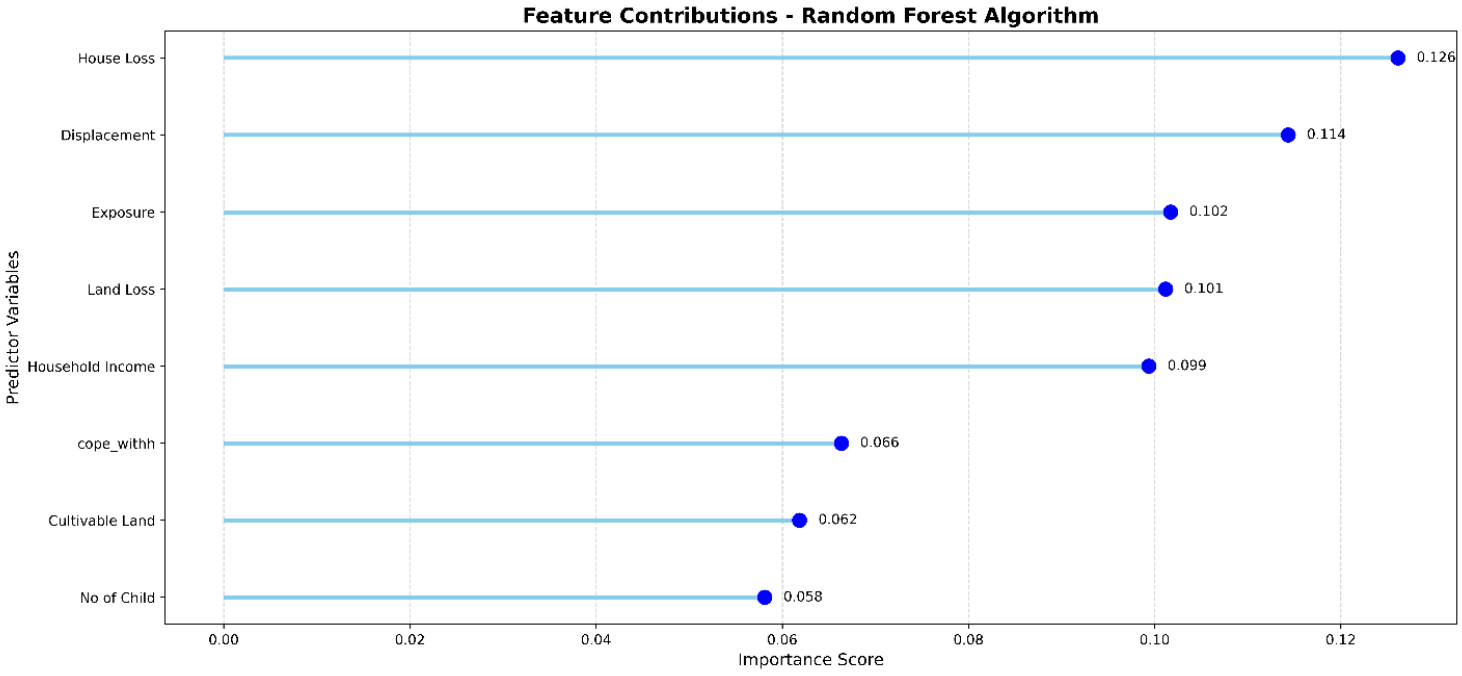


Figure 7

CMHD Features Contribution (Random Forest).