International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 25(2), November, 2025, pp 101-113
DOI: https://doi.org/10.3329/ijss.v25i2.85772

Data-Driven Dengue Prevention Strategies in Bangladesh using Explainable Artificial Intelligence and Causal Inference

Md. Siddikur Rahman[1]* and Md. Abu Bokkor Shiddik[2]

[1]Department of Statistics, Begum Rokeya University, Rangpur
Email: siddikur@brur.ac.bd
[2]Department of Statistics, Begum Rokeya University, Rangpur
Email: abubokkor.brur60@gmail.com
*Correspondence should be addressed to Md. Siddikur Rahman
(Email: siddikur@brur.ac.bd)

[Received July 28, 2025; Accepted November 13, 2025]

Abstract

Dengue is one of the major public health problems in Bangladesh. This country experiences periodic dengue outbreaks and it is particularly vulnerable to climate-related factors. The increasing dengue burden in the country necessitates reliable prediction-based instruments to inform public health strategies. In this study, we employed high-performance machine learning (XGBoost), combined with explainable artificial intelligence (XAI). This evaluates the predictive relevance of climate, socioeconomic, healthcare, and land-use variables for dengue incidence in Bangladesh from January 2010 to December 2024. We also applied causal inference (CAI) techniques by using the *DoWhy* framework. It enables robust interpretation of the variables' impact and their directional influence on dengue incidence. The model performance was assessed using RMSE, MAE, and MAPE matrices. Dengue cases increased suddenly in Bangladesh during 2019, 2023, and 2024, with each year exceeding 100,000 total cases and monthly peaks of 50,000 cases. The XGBoost model was the top-performing model (RMSE: 11,365.2; MAE: 7,014.68; MAPE: 764.36). It offers the highest predictive accuracy of dengue risk. Climate indicators were the strongest contributor to dengue prediction (41.87%), followed by sociodemographic (35.03%), healthcare (19.16%), and landscape factors (3.95%). Our findings show that machine learning, specifically XGBoost, can effectively predict dengue cases and help identify key ecological and structural drivers. It highlights key ecological and structural drivers, supporting the integration of real-time, multisectoral data into public health planning and decision-making. It also improved digital literacy, health infrastructure, and early warning systems, which are essential for mitigating dengue outbreaks in Bangladesh and similar endemic countries.

Keywords: Dengue, XGBoost, SHAP, Public health, Bangladesh.

AMS Classification: 62P10, 68T09.

1. Introduction

In Bangladesh, Dengue is one of the most serious public health issue with increasingly frequent and prolonged outbreaks[1]. In the past years, it was seasonal disease, but now it poses a persistent threat throughout the year. In 2023, Bangladesh reported the highest number of dengue cases (over 300,000) and deaths (more than 1,500)[2,3]. Urban centsres (such as Dhaka, Chattogram, and Khulna) are particularly vulnerable because of rapid urbanization, shifting unusual climatic patterns, and unequal access to healthcare infrastructure[4].

Some studies often relied on traditional statistical models, including regression analyses and time-series forecasts, to explore the relationship between climatic variables, population density, and dengue incidence[5]. While these methods are informative, they assume linear relationships and are unable to detect non-intuitive interactions within the complex datasets. Their explanatory power remains limited in addressing the underlying drivers of disease emergence [6]. Some recent studies have incorporated that, the machine learning (ML) models improved the prediction accuracy[7–9]. For example, extreme gradient boosting (XGBoost) models proved to be quite effective in handling high-dimensional, non-linear data and producing accurate predictions[10]. However, many of these models operate as opaque systems—often referred to as "black boxes" with the foundation for predictions hidden [11]. This lack of interpretability restricts the use of machine learning in policymaking because, in policymaking, transparency and intelligibility are essential for successful implementation[11].

Explainable artificial intelligence (XAI) techniques are becoming increasingly relevant in addressing these limitations in model interpretability. Among all the XAI techniques, SHAP (Shapley Additive Explanations) values help clarify how individual predictors influence model outcomes as well as improve their transparency[12]. As prediction is valuable, public health planning also requires insight into the mechanisms behind disease transmission. Causal Inference (CAI) methods can help uncover these mechanisms. The CAI approaches estimate the directional influence of variables and it explores the counterfactual scenarios, such as assessing the impact of increased healthcare access or modified urban development on dengue outcomes[13]. Despite their usefulness, the integration between CAI and XAI-supported machine learning models remains limited in dengue-focused research, leaving room for methodological enhancement.

In this study, we address that gap by combining XGBoost modelling with SHAP-based interpretation and selected CAI techniques for dengue prediction. We used the data from 2010 to 2023. The dataset included climatic trends, healthcare infrastructure, land use, and demographic indicators. We also identified relevant predictors of dengue incidence by using these methods. We then estimate potential causal effects to provide insight into factors that may influence dengue outbreak dynamics. Rather than proposing a definitive solution, this framework offers a methodological contribution that improves both interpretability and policy relevance in dengue predictions. By linking prediction with explanation, it supports more informed analysis and may help guide targeted interventions. The goal is to advance understanding and decision-support strategies within Bangladesh’s broader dengue mitigation efforts.

2. Materials and methods

2.1 Study area and data collection

This study focuses on Bangladesh which is a South Asian nation with a tropical monsoon climate and a wide range of environmental and socio-demographic parameters that influence dengue epidemics (Figure 1). We collected monthly dengue case data from the DGHS and IEDCR. The data were used for national analysis from 2010 to 2024. Supplementary variables from the World Bank and WHO included GDP growth, unemployment rate, literacy rate, population density, and healthcare access indicators such as hospital beds and physician density. Climatic data (temperature, humidity, rainfall, pressure, and wind speed) collected from NASA, while Landscape metrics data (fractions of forest, agricultural, and arable land) collected from World Bank databases. Variables were selected for their relevance to dengue risk and temporal resolution, yielding comprehensive insights into spatial and seasonal trends described in previous studies [14,15] and Table 1.

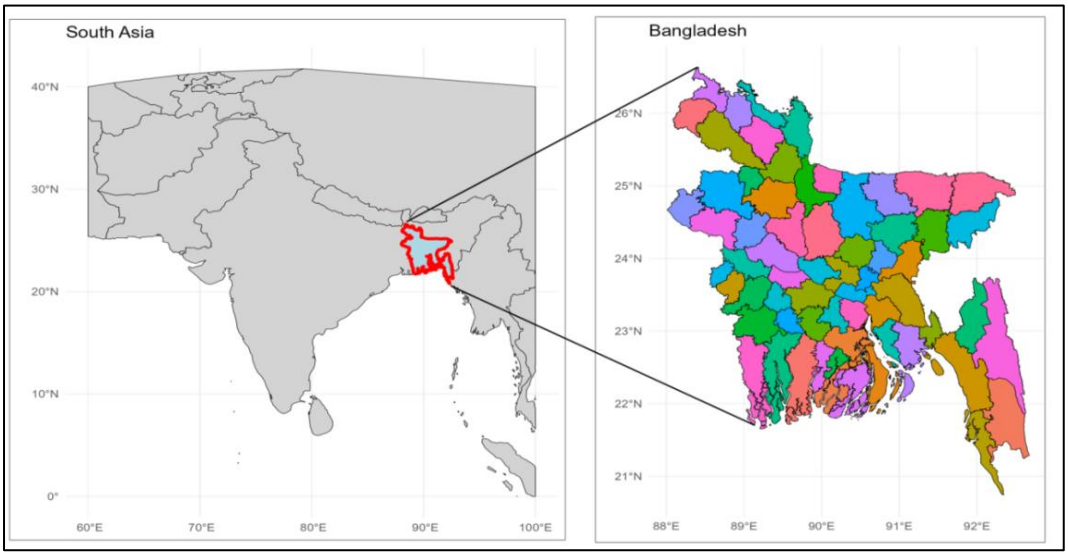

Figure 1: Study area. The map was made in R (version 4.5.1, https://cran.rstudio.com) by using the rnaturalearth package, which provides natural earth map data for visualization. We used the *sf* package to handle spatial data and make sure the geographic boundaries were represented accurately.

Table 1: Description of the variables and data sources used in this study

Category	Code	Variable Description	Data Source	Temporal Resolution
Time Variable	Year	Calendar Year (2010-2024)		Yearly
	Months	Calendar Months (Jan-Dec)		
Dengue Data	Cases	Number of reported dengue cases	DGHS & IEDCR	Monthly
Socio-demographic	Y1	GDP growth (annual %)	https://data.worldbank.org/indicator	Yearly
	Y2	Unemployment rate (% of total labor force)	https://data.worldbank.org/indicator/	Yearly
	Y3	Adult literacy rate (% of population aged ≥15 years)	https://data.worldbank.org/indicator/SE.ADT.LITR.ZS	Yearly
	Y4	Population density	https://data.worldbank.org/indicator	Yearly

Healthcare & Accessibility	Y5	Number of hospital beds per 10,000 population	https://data.worldbank.org/indicator	Yearly
	Y6	Number of physicians per 10,000 population	https://data.who.int/indicators	Yearly
	Y7	Domestic general government health expenditure (%)	https://data.who.int/indicators	Yearly
	Y8	Community health workers (per 1,000 people)	https://data.worldbank.org/indicator/SH.MED.CMHW.P3	Yearly
Climate Variables	Y9	Mean monthly temperature (°C)	https://www.nasa.gov/	Monthly
	Y10	Relative humidity (%)	https://www.nasa.gov/	Monthly
	Y11	Monthly total rainfall (mm)	https://www.nasa.gov/	Monthly
	Y12	Surface atmospheric pressure (kPa)	https://www.nasa.gov/	Monthly
	Y13	Maximum wind speed at 50 meters (m/s)	https://www.nasa.gov/	Monthly
Landscape Variables	Y14	Forest area (% of total land area)	https://data.who.int/indicators	Yearly
	Y15	Agricultural land (% of total land area)	https://data.who.int/indicators	Yearly
	Y16	Arable land (% of total land area)	https://data.who.int/indicators	Yearly

Note: *yearly values were held constant across all months within the corresponding year.*

2.2 Data pre-processing and statistical analysis

The dataset is spanning January 2010 to December 2024 and undergoing systematic harmonization to support machine learning and causal inference modelling. Missing values in continuous variables were imputed using linear interpolation, defined as following equation [14]:

$$y_t = y_{t-1} + \left(\frac{y_{t+1} - y_{t-1}}{2}\right) \tag{1}$$

To standardize variable scales and improve model convergence, Z-score normalization was applied by using the following equation[15]:

$$z = \frac{y-\mu}{\sigma} \tag{2}$$

Multicollinearity was evaluated using the Variance Inflation Factor (VIF) by using the following equation[16]:

$$VIF_i = \frac{1}{1-R_i^2} \tag{3}$$

Where R_i^2 denotes the coefficient of determination for predictor ii; predictors with VIF values exceeding 5 were excluded to ensure model stability and interpretability. Outliers were detected using the *Mahalanobis Distance* by using the following equation[17]:

$$D^2 = (y-\mu)^T \Sigma^{-1} (y-\mu) \tag{4}$$

The Mahalanobis Distance values which are higher than the critical chi-square cut-off ($p < 0.001$), we reduced their impact by winsorizing the extreme values at the 1st and 99th percentiles. This way, the data kept its original structure while minimizing the influence of outliers on the results.

All the pre-processing and statistical analyses were conducted by using the RStudio v4.5.1[18] software. with the libraries *dplyr, car, and MVN. Additionally,* and Python Kernel 3.12was utilized within JupyterLab v4.2.5[19] with the libraries *pandas*, *numpy*, *scikit-learn*, *statsmodels*, and *dowhy*. This integrated workflow ensured analytical rigor, transparency, and consistency across the modeling pipeline.

2.3 Machine learning model selection and evaluation

We split the dataset into a training set (January 2010 to December 2021) and a testing set (January 2022 to December 2024) to develop a predictive framework for monthly dengue cases in Bangladesh. Then we employed four machine learning model: Generalized linear model with elastic net, Decision Tree, Random Forest, and eXtreme Gradient Boosting (XGBoost) to predict dengue cases. Each model was chosen for its methodological diversity. These offers complementary capabilities in capturing linear trends, nonlinear interactions, and variable importance.

2.3.1 Generalized linear model (GLM)

The GLM regression with elastic net extension incorporates both L1 (Lasso) and L2 (Ridge) regularization penalties. It enables accurate feature selection while mitigating multicollinearity. The optimization function[20] is defined as the following equation:

$$\min_{\beta}\{\frac{1}{2n}\sum_{i=1}^{n}(x_i - y_i\beta)^2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2\} \tag{5}$$

Where, x_i is the response variable, y_i is the predictor matrix, and λ_1, λ_2 are the response variable, predictor matrix , and the control regularization strength respectively. Elastic Net is interpretable and statistically robust, although it has limitations for modelling nonlinear relationships.

2.3.2 Decision tree (DT)

The DT algorithms recursively split data based on the minimization of mean squared error (MSE). It defined as the following equation[21]:

$$MSE(t) = \frac{1}{n_t}\sum_{i=1}^{n_t}(x_t - \bar{x}_t)^2 \tag{6}$$

here, n_t is the number of samples in node t and $\bar{x}_t$ is the node's mean target value. While decision trees provide transparency and handle mixed data types, they are susceptible to overfitting and variance.

2.3.3 Random forest (RF)

The RF model builds an ensemble of decision trees through bootstrap sampling. It aggregates predictions via averaging the outputs of individual trees in regression tasks. The formula[9] is denoted by the flowing equation:

$$\hat{f}_{RF}(y) = \frac{1}{B}\sum_{b=1}^{B} T_b(y) \tag{7}$$

here, B denotes the number of trees and $T_b(y)$ represents the prediction from tree$b$. This model enhances generalizability and resilience to noise, albeit at the expense of reduced interpretability due to ensemble complexity.

2.3.4 XGBoost

The XGBoost utilizes gradient-based boosting to optimize prediction accuracy. It incorporate regularization to prevent overfitting sequentially. Its objective function[9] is

$$L = \sum_{i=1}^{n} l\left(x_i, \hat{x}_i^{(t)}\right) + \sum_{i=1}^{T} \Omega(f_t) \tag{8}$$

here, l is the loss function, $\hat{x}_i^{(t)}$ is the predicted value at iteration t, and $\Omega(f_t)$ penalizes complexity. XGBoost offers superior predictive accuracy and scalability but lacks intrinsic interpretability without auxiliary tools such as SHAP.

2.3.5 Model performance matrices

Model performance was assessed using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), calculated respectively with the flowing equations:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2} \tag{9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}_i| \tag{10}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{x_i - \hat{x}_i}{x_i}\right| \tag{11}$$

These metrics provided a consistent basis for comparing predictive accuracy across models and informed subsequent selection for explain ability and causal inference integration[9]. All models were implemented by using R version 4.5.1[18], ensuring computational consistency and reproducibility.

2.4 Cross-validation and hyperparameter tuning

We applied five-fold cross-validation across all machine learning models by using the training set (2010-21) to improve the model's generalizability and minimize overfitting. This step involved partitioning the dataset into five subsets, where each fold served as a validation set once, enabling robust internal performance evaluation. Hyper-parameter tuning was model-specific: we used the *cv.glmnet* package for elastic net regularization. It select optimal penalty strength (λ) while fixing the mixing ratio (α = 0.5). We tuned DT for complexity control by using *rpart*. It adjust tree depth and minimum splits. We also tuned RF focusing on *mtry*, which implemented through the caret interface with 500 trees. Also XGBoost underwent grid search using caret and *xgboost*, optimizing *nrounds*, *eta*, *max_depth*, and other learning parameters. All procedures were executed in R version 4.5.1[18] software, which ensures the reproducibility and model stability across diverse algorithmic frameworks.

2.5 Explainable artificial intelligence (XAI)

XAI techniques enhances interpretability by revealing how input features influence in the model predictions. This is crucial for translating machine learning outputs into actionable epidemiological insights. We integrated SHAP contributions with the best-performing ML model in this study by using the following equation[9]:

$$\phi_i = \sum_{S \subseteq N\{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!}\left[f_{S\cup\{i\}}(y) - f_s(y)\right] \tag{12}$$

Here, the SHAP value for feature i is ϕ_i, S is any subset of the predictor set N excluding i, and is the model's output functionis f. This approach enabled localized and consistent decomposition of predictions, which helps identify key predictors that contributing to monthly dengue predictions.

2.6 Causal inference

To estimate directional effects between selected predictors and dengue cases, we conducted causal inference by using the *DoWhy* framework in Python[22]. By using this method, we identified the top five contributors via SHAP analysis were designated as treatments in separate causal models, with remaining variables treated as potential confounders. The causal effect of each treatment on monthly dengue cases was then estimated by using the backdoor criterion with linear regression, providing interpretable estimates of directional associations. The causal effect of each treatment on monthly dengue cases was computed by using the backdoor criterion with linear regression. The average treatment effect (ATE)[23] was estimated by using the flowing equation:

$$ATE = \frac{d}{dT} E[x|T, Z] \tag{13}$$

Where the treatment is T, Z is the confounders, and $E[x|T, Z]$is the conditional expectation of dengue cases. This strategy provided interpretable estimates of causal impact beyond correlation, helping prioritize predictors most influential to outbreak dynamics.

3. Results

3.1 Temporal trends and predictor associations in dengue case burden in Bangladesh from 2010 to 2024

Between 2010 and 2024, reported dengue cases in Bangladesh exhibited substantial year-to-year variation, with pronounced peaks in 2019 (101,324 cases), 2023 (100,194 cases), and 2024 (101,211 cases) (Figure 2A and Table 2). The most extreme monthly dengue cases occurred in August 2019 (52,636 cases), which is consistent with post-monsoon seasonality. In contrast, in the earlier years such as 2010 and 2014 recorded minimal transmission, with average monthly cases of 34.08 and 31.25, respectively. Rising means and standard deviations in recent years reflect both increased magnitude and volatility in the case distribution (Figure 2A and Table 2).

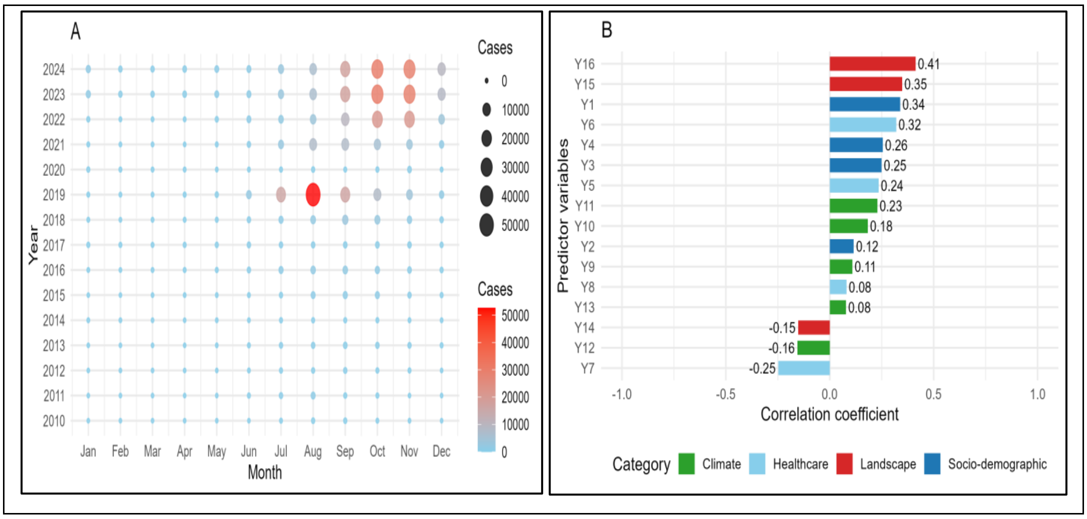

Figure 2: A) Temporal distribution of dengue cases and B) correlation analysis of predictive variables with dengue incidence in Bangladesh (2010–2024). Y1: GDP growth (annual %); Y2: Unemployment rate (% of total labor force); Y3: Adult literacy rate (% of population aged ≥15 years); Y4: Population density; Y5: Number of hospital beds per 10,000 population; Y6: Number of physicians per 10,000 population; Y7: Domestic general government health expenditure (%); Y8: Community health workers (per 1,000 people); Y9: Mean monthly temperature (°C); Y10: Relative humidity (%); Y11: Monthly total rainfall (mm); Y12: Surface atmospheric pressure (kPa); Y13: Maximum wind speed at 50 meters (m/s); Y14: Forest area (% of total land area); Y15: Agricultural land (% of total land area); Y16: Arable land (% of total land area).

Correlation analysis demonstrated strong positive associations between dengue cases and arable land proportion(r = 0.4129), agricultural land use(r = 0.3478), and GDP growth rate(r = 0.3384) (Figure 1B). Moderate relations were observed for adult literacy rate (r = 0.2491) and hospital bed availability(r = 0.2352). Notably, domestic health expenditure showed an negative relationship (r = –0.2477) on dengue cases, which suggests its potential mitigating influence on dengue cases (Figure 2B).

Table 2: Summary statistics (yearly) of the reported dengue cases in Bangladesh (2010–24).

Year	**Min**	**Q_1**	**Median**	**Q_3**	**Mean ± SD**	Total cases
2010	0	0	0	24.5	34.08 ± 66.38	409
2011	0	0	0	88.25	113.25 ± 233.49	1359
2012	0	0	8	94.5	55.92 ± 81.71	671
2013	0	0	51	253.25	145.75 ± 177.14	1749
2014	0	7.75	13	66.25	31.25 ± 33.27	375
2015	0	5	51.5	394.5	263.5 ± 375.09	3162
2016	3	32.75	199.5	963.75	505 ± 588.63	6060
2017	36	87.25	200.5	361.75	230.75 ± 165.27	2769
2018	7	28.25	294	1343	845.67 ± 1064.1	10148
2019	17	53	1565.5	10148	8443.67 ± 15226.6	101324
2020	10	24.5	46	109.5	99.42 ± 148.17	1193
2021	3	27.25	739.5	4039.75	2369.08 ± 3054.99	28429
2022	20	100.25	1154	5276	5257.42 ± 8157.92	63089
2023	311	609	1862	11068	8349.5 ± 11489.47	100194
2024	311	609	1860.5	11833	8434.25 ± 11496.45	101211

Min: Minimum; Q1: 1st quartile; Q3: 3rd quartile; SD: standard deviation;

3.2 Selection of optimal machine learning model

Figure 3: A) Comparative evaluation metrics of four machine learning models based on training and test datasets. RMSE: Root Mean Square Error; MAE: Mean Absolute Error; MAPE: Mean Absolute Percentage Error; GLM: Generalized Linear Model; RF: Random Forest; XGB: Extreme Gradient Boosting; DT: Decision Tree, B) Category-wise SHAP value contributions from the XGBoost model, showing aggregated importance of socio-demographic, healthcare, climate, and landscape predictors. C) Top five predictors in predicting dengue cases ranked by mean SHAP contributions, with percentages indicating the proportion of total SHAP contribution for these predictors. D) Estimated causal relationships of the five most influential variables, with numerical values representing the Average Treatment Effect (ATE), indicating the direction and magnitude of their effect on dengue incidence. Y1: GDP growth (annual %); Y10: Relative humidity (%); Y7: Domestic general government health expenditure (%); Y11: Monthly total rainfall (mm); Y3: Adult literacy rate (% of population aged ≥15 years).

XGBoost was identified as the most effective model for predicting dengue cases over the 2010–24 period (Figure 3A). It achieved the lowest mean absolute error (7014.68), mean absolute percentage error (764.36), and root mean squared error (11365.2), indicating high forecasting precision and minimal proportional deviation. While Random Forest delivered a moderately good performance (RMSE: 11,986.07; MAPE: 1,470.34), it did not surpass XGBoost in overall accuracy. In our overall findings, GLM-ElasticNet performed the least effectively across all the metrics (Figure 3A).

3.3 Key drivers and causal effects on dengue cases

We find that climate related indicators contributed the most significantly to dengue prediction (SHAP: 41.87%), followed by socio-demographic (SHAP: 35.03%), healthcare (19.16%), and landscape factors (SHAP: 3.95%) (Figure 3B). GDP growth (SHAP: 26.75%; ATE: 4333.69) had the highest predictive contribution and exerted the strongest positive causal effect, suggesting economic fluctuations play a critical role in dengue dynamics. Relative humidity (SHAP: 17.75%; 73.94) and monthly total rainfall (SHAP: 11.85%; ATE: 55.10) also had positive causal relationships with dengue case counts. In contrast, domestic general government health expenditure (SHAP: 14.45%; ATE: −1788.98) and adult literacy rate (SHAP: 6.21%; ATE: −262.18) showed negative causal effects on the prediction (Figure 3 C, D). These findings underscore that, the combined relevance of explainable AI and causal inference in quantifying the directional impact of predictors on dengue incidence in Bangladesh (Figure 3).

4. Discussion

In this study, we build a new framework that combines ML, XAI, and CAI to predict dengue cases in Bangladesh. We found that climate-related variables (humidity, rainfall, and temperature) play a crucial role in dengue prevalence. These results align with some earlier studies that the *Aedes* mosquito is the primary vector for dengue, which thrives in warm, humid tropical environments [24]. We also find in this study that climate patterns, socioeconomic, and infrastructural factors contribute to the spread of dengue. Economic growth (GDP growth) has an indirect role. The rapid expansion often leads to crowded living areas and poor drainage, allowing water to accumulate and mosquitoes to breed [25]. On the other hand, healthcare expenditure and adult literacy appeared to offer adverse effects on dengue. Institutional preparedness and community awareness play crucial roles in reducing risk [26,27]. These trends suggest that dengue is no longer merely a seasonal ecological phenomenon, but a result of broader systemic transitions [28]. The interplay between climatic and socioeconomic structures suggests that the rising dengue cases are as much a result of infrastructural lag and public health disparities as they are of environmental shifts[29].

Bangladesh’s urban centers often lack adequate drainage systems. It creates habitats conducive to mosquito breeding more and more[30]. Additionally, limited public awareness and uneven access to healthcare intensify the risk in marginal communities [31]. From a policy standpoint, the integration of XAI and CAI offers actionable insights into dengue prevention. These tools not only build trust among decision-makers but also identify leverage points for intervention. Real-time climate surveillance, adaptive public health infrastructure, and education campaigns focused on community-level preventive measures are urgent priorities now [15,32]. Establishing centralized AI-enabled decision-support systems could help allocate resources more efficiently by anticipating high-risk zones before outbreaks occur in our country[15].

4.1 Limitations

Our study has some limitations. We used retrospective data, which may limit our ability to capture evolving transmission dynamics and emerging socioeconomic shifts. The variable selection, guided by SHAP and causal inference, may overlook latent confounders that are not represented in the available datasets in this study. Geographical data were limited in our study, which restricted district-level validation and spatial resolution. The causal effects are robustly estimated, but it can rely on assumptions that may not fully account for unobserved heterogeneity. Additionally, the exclusion of behavioral and vector-specific data limits biological realism.

4.2 Conclusion

The main objective of our study is to show the strength of integrating ML with XAI and CAI to predict dengue incidence as well as to find the key drivers related to dengue in Bangladesh. Our findings provide a comprehensive view of the multifaceted nature of dengue transmission by identifying climatic factors, socioeconomic indicators, and healthcare investments as major contributors. XGBoost confirms its suitability for modelling complex disease patterns, especially in contexts characterized by diverse and interrelated variables. Future studies should be focused on real-time surveillance and dynamic modelling approaches to capture evolving trends related to dengue. In this way, dengue prediction could become a practical tool for preventing outbreaks rather than just reacting to them. Adding information on human behavior, mosquito populations, and data at district-level resolution could make predictions more precise. Linking these insights to policymaking could help establish early warning systems, focus interventions where they are needed most, and enhance coordination across different sectors. This way, dengue prediction could become a practical tool for preventing outbreaks rather than just reacting to them, which was very useful for this study.

Acknowledgements: We thank the Ministry of Health and Family Welfare, Government of the People's Republic of Bangladesh, the Bangladesh Bureau of Statistics (BBS), the World Bank, WHO, and NASA for facilitating our access to essential data. The author very grateful to the reviewer for recognizing the potential of this study and offering meaningful suggestions for improvement.

References

[1] Hossain, M. S., Noman, A. A., Mamun, S.A.A., and Mosabbir, A. A. (2023). Twenty-two years of dengue outbreaks in Bangladesh: epidemiology, clinical spectrum, serotypes, and future disease risks. Trop Med Health. 2023;51: 37. doi:10.1186/s41182-023-00528-6

[2] The 2023 fatal dengue outbreak in Bangladesh highlights a paradigm shift of geographical distribution of cases | Epidemiology & Infection | Cambridge Core. [cited 27 July 2025]. Available: https://www.cambridge.org/core/journals/epidemiology-and-infection/article/2023-fatal-dengue-outbreak-in-bangladesh-highlights-a-paradigm-shift-of-geographical-distribution-of-cases/E244119B1E995CB573953777F2FCEE4E

[3] Dengue - Bangladesh. [cited 27 July 2025]. Available: https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON481

[4] Frontiers | Unplanned urbanization and health risks of Dhaka City in Bangladesh: uncovering the associations between urban environment and public health. [cited 27 July 2025]. Available: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2023.1269362/full

[5] Islam, S., Haque, C. E., Hossain, S., and Hanesiak, J. (2021). Climate Variability, Dengue Vector Abundance and Dengue Fever Cases in Dhaka, Bangladesh: A Time-Series Study. Atmosphere. 2021;12: 905. doi:10.3390/atmos12070905

[6] Nonlinearity in the Epidemiology of Complex Health and Disease Processes | Theoretical

Medicine and Bioethics. [cited 27 July 2025]. Available: https://link.springer.com/article/10.1023/A:1009979306346

[7] Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: A machine learning approach | PLOS One. [cited 27 July 2025]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0270933

[8] Machine learning prediction in cardiovascular diseases: a meta-analysis | Scientific Reports. [cited 27 July 2025]. Available: https://www.nature.com/articles/s41598-020-72685-1

[9] Dengue Early Warning System and Outbreak Prediction Tool in Bangladesh Using Interpretable Tree-Based Machine Learning Model - Rahman - 2025 - Health Science Reports - Wiley Online Library. [cited 27 July 2025]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/hsr2.70726

[10] Khan, M. I., and Abbas, Y. M. (2023). Robust extreme gradient boosting regression model for compressive strength prediction of blast furnace slag and fly ash concrete. Mater Today Commun. 2023;35: 105793. doi:10.1016/j.mtcomm.2023.105793

[11] Tursunalieva, A., Alexander, D.L.J., Dunne, R., Li J, Riera, L., and Zhao, Y. (2024). Making Sense of Machine Learning: A Review of Interpretation Techniques and Their Applications. Appl Sci. 2024;14: 496. doi:10.3390/app14020496

[12] Sahlaoui, H., Alaoui, E.A.A., Nayyar, A., Agoujil, S., Jaber, M. M. (2021). Predicting and Interpreting Student Performance Using Ensemble Models and Shapley Additive Explanations. IEEE Access. 2021;9: 152688–152703. doi:10.1109/ACCESS.2021.3124270

[13] Causal inference concepts can guide research into the effects of climate on infectious diseases | Nature Ecology & Evolution. [cited 27 July 2025]. Available: https://www.nature.com/articles/s41559-024-02594-3

[14] Moritz, S., Sardá, A., Bartz-Beielstein, T., Zaefferer, M., AND Stork, J. (2015). Comparison of different Methods for Univariate Time Series Imputation in R. arXiv; 2015. doi:10.48550/arXiv.1510.03924

[15] Utilizing artificial intelligence to predict and analyze socioeconomic, environmental, and healthcare factors driving tuberculosis globally | Scientific Reports. [cited 27 July 2025]. Available: https://www.nature.com/articles/s41598-025-96973-w

[16] Akinwande, M. O., Dikko, H. G., Samson, A. (2015). Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. Open J Stat. 2015;05: 754. doi:10.4236/ojs.2015.57075

[17] Kumari, P., and Gupta, S. (2024). Comparative analysis between Euclidean distance metric andMahalanobis Distance Metric. Int J Innov Res Technol Sci. 2024;12: 271–277.

[18] Download R-4.5.1 for Windows. The R-project for statistical computing. [cited 27 July 2025]. Available: https://cran.r-project.org/bin/windows/base/

[19] Project Jupyter. [cited 27 July 2025]. Available: https://jupyter.org

[20] Variable selection for uncertain regression models based on elastic net method: Communications in Statistics - Simulation and Computation: Vol 0, No 0. [cited 27 July 2025]. Available: https://www.tandfonline.com/doi/abs/10.1080/03610918.2024.2410392

[21] Nagpal, Y., Gupta, A., Satija, T., Maheshwari, G., and Aggarwal, S. (2025). Prediction of Performance Parameters of Journal Bearing Using Machine Learning Algorithms. J Inst Eng India Ser C. 2025 [cited 27 July 2025]. doi:10.1007/s40032-025-01226-4

[22] Dawoud, A., and Talupula, S. (2025). ProRCA: A Causal Python Package for Actionable Root Cause Analysis in Real-world Business Scenarios. arXiv; 2025. doi:10.48550/arXiv.2503.01475

[23] Wang, R., Zhao, Y-Q., Dukes, O, and Zhang, B. (2025). Nested Instrumental Variables Design: Switcher Average Treatment Effect, Identification, Efficient Estimation and Generalizability. arXiv; 2025. doi:10.48550/arXiv.2405.07102

[24] Samsudin, N. A., Othman, H., Siau, C. S., Zaini, Z-'Izzat I. (2024). Exploring community needs in combating aedes mosquitoes and dengue fever: a study with urban community in the recurrent hotspot area. BMC Public Health. 2024;24: 1651. doi:10.1186/s12889-024-18965-1

[25] Franklinos, L.H.V., Jones, K.E., Redding, D.W., and Abubakar, I. (2019). The effect of global change on mosquito-borne disease. Lancet Infect Dis. 2019;19: e302–e312. doi:10.1016/S1473-3099(19)30161-6

[26] Upsurge of dengue outbreaks in several WHO regions: Public awareness, vector control activities, and international collaborations are key to prevent spread - Sarker - 2024 - Health Science Reports - Wiley Online Library. [cited 27 July 2025]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/hsr2.2034

[27] Health seeking behaviours, dengue prevention behaviours and community capacity for sustainable dengue prevention in a highly dengue endemic area, Sri Lanka | BMC Public Health. [cited 27 July 2025]. Available: https://link.springer.com/article/10.1186/s12889-023-15404-5

[28] Riad, M. H., Cohnstaedt, L. W., Scoglio, C. M. (2021). Risk Assessment of Dengue Transmission in Bangladesh Using a Spatiotemporal Network Model and Climate Data. Am J Trop Med Hyg. 2021;104: 1444. doi:10.4269/ajtmh.20-0444

[29] Gibb, R., Colón-González, F. J., Lan, P. T., Huong, P. T., Nam, V. S., Duoc, V. T., et al. (2023). Interactions between climate change, urban infrastructure and mobility are driving dengue emergence in Vietnam. Nat Commun. 2023;14: 8179. doi:10.1038/s41467-023-43954-0

[30] Dhar-Chowdhury, P., Haque, C. E., Lindsay, R., Hossain, S. (2025). Socioeconomic and Ecological Factors Influencing Aedes aegypti Prevalence, Abundance, and Distribution in Dhaka, Bangladesh. 2016 [cited 27 July 2025]. doi:10.4269/ajtmh.15-0639

[31] Dengue: knowledge gaps, unmet needs, and research priorities - The Lancet Infectious Diseases. [cited 27 July 2025]. Available: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(16)30473-X/fulltext

[32] Artificial Intelligence-Based Waste Management: A Review of Classification, Techniques, Issues, and Challenges - Yevle - 2025 - WIREs Data Mining and Knowledge Discovery - Wiley Online Library. [cited 27 July 2025]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.70025