Title: Explainable and physics-informed machine learning for seasonal water quality prediction in the monsoon-driven Padma River Basin, Bangladesh

Author names: Abu Reza Md Towfiqul Islam , Md Abdullah-Al Mamun , Md Nashir Uddin , Sheikh Fahim Faysal Sowrav , M Nur E Alam , Shahidur R Khan , Mohaiminul Haider Chowdhury , Tasrina Rabia Choudhury

Abstract: River water quality in monsoon-driven subtropical basins exhibits strong seasonal variability driven by hydroclimatic forcing and increasing anthropogenic pressure, posing challenges for reliable assessment and management. Despite advances in water quality modeling, most Water Quality Index (WQI) prediction frameworks require extensive sampling and lack interpretability, limiting rapid baseline assessment during critical periods. This study develops the first integrated Explainable Artificial Intelligence (XAI) framework combining Machine Learning (ML), Deep Learning (DL), and Physics-Informed Neural Networks (PINNs) to predict, interpret, and spatially characterize seasonal water quality dynamics in the Padma River Basin, Bangladesh. Forty-four surface water samples collected during winter and monsoon seasons were evaluated using WQI assessment, explainable modeling, probabilistic uncertainty analysis, and spatial regionalization. Results show that seasonal variability dominates over spatial variability ($p < 0.0001$), with winter low-flow conditions promoting solute concentration and localized degradation, while monsoon discharge drives basin-wide dilution and recovery. Model performance is strongly region-dependent: Deep Neural Networks achieve the highest accuracy in winter ($R^2 = 0.98$; RMSE = 1.16), whereas Ridge Regression and Voting Ensemble models perform more robustly during the monsoon ($R^2 \approx 0.97$; RMSE ≈ 1.01). Explainable AI analysis identifies NO_3^- emerged as the dominant contaminant (24.0 ± 36.3 mg/L winter, 47.5 ± 68.7 mg/L monsoon, with isolated samples exceeding WHO limits), whereas pH and DO exhibit dual seasonal influences. PINN-based data augmentation improves model generalization under limited sampling while preserving hydrochemical consistency. Monte Carlo simulations quantify prediction uncertainty and reveal seasonal shifts in WQI probability distributions, while spatial autocorrelation analysis identifies localized winter degradation hotspots and widespread monsoon improvement. The proposed physics-informed and explainable AI framework enhances predictive reliability, interpretability, and decision relevance, offering a transferable approach for uncertainty-aware water quality assessment and adaptive management in monsoon-affected, data-limited river basins.

Keywords: Artificial intelligence; Bangladesh; DNN; Monsoon River Basin; Padma River; Water Quality Index.

- DOI: 10.1007/s10653-026-03031-z