Hybrid Machine Learning and Swarm Intelligence Techniques for Enhanced Cyanobacterial Bloom Prediction

Dendani Nadjette *
Computer Science Department, Badji Mokhtar University Annaba, Algeria

Bounekhla Oumaima
Computer Science Department, Badji Mokhtar University Annaba, Algeria

Hadiby Rayenne
Public Health, Paris Saclay University, France

Abstract

Cyanobacterial harmful algal blooms (CyanoHABs) pose increasing threats to freshwater ecosystems, potable water safety, and public health, particularly in regions affected by eutrophication and climate change. Accurate prediction of these events is crucial for effective environmental monitoring and mitigation. This study introduces a robust machine learning framework for CyanoHAB prediction, comparing advanced ensemble models with and without Particle Swarm Optimization (PSO) hyperparameter tuning.

The approach was applied to two datasets: (1) a data-limited set from Algeria's Mexa Dam, augmented using Conditional Generative Adversarial Networks (CGANs), and (2) a comprehensive dataset from South Korean rivers (Nakdong, Han, and Geum) with frequent bloom events. Evaluated algorithms included Random Forest (RF), XGBoost, LightGBM, CatBoost, and Histogram-based Gradient Boosting (HistGB), along with hybrid and stacked models. PSO optimization significantly improved performance, with RF+PSO+CV achieving an R^2 of 0.992 on the Mexa dataset, and RF+PSO reaching an R^2 of 0.994 with a MAE of 5.388 on the South Korean dataset. Hybrid models demonstrated strong generalization across diverse conditions.

Data preprocessing involved K-Nearest Neighbors imputation and domain-specific feature engineering. CGAN-based augmentation enhanced model robustness in data-scarce scenarios.

This study offers three key contributions: (1) a high-performance, PSO-optimized machine learning pipeline for CyanoHAB forecasting, (2) a comparative analysis of modeling strategies across datasets, and (3) insights into generative augmentation techniques for ecological applications. The results highlight AI-driven approaches' potential for proactive water resource management.

Keywords

Cyanobacterial Harmful Blooms, Machine Learning, Particle Swarm Optimization, deep learning Water Quality.