

# ABSTRACT

Microbial health assessment and water quality (WQ) parameter prediction in inland waters are challenging, primarily due to the complex interplay of dynamic drivers, such as hydrologic, meteorological, chemical, and biological processes that co-occur across multiple spatio-temporal scales. This complexity is further exacerbated by the fact that field observations, while foundational, are constrained by irregularity and sparsity of sampling, as well as inherent latencies in laboratory analysis across space and time. Remote sensing (RS) has transformed many areas of Earth observation (EO), particularly aquatic monitoring by providing high-frequency spectral retrievals along with radiometric and pigment-specific signatures. Additionally, incorporating RS-based environmental covariates such as hydro-meteorological, land-cover, and lake-morphometric features proves invaluable in elucidating water quality dynamics. Recently, artificial intelligence (AI)-based models utilizing RS-based data sources have been increasingly applied to environmental monitoring tasks. These data sources create new opportunities; however, they also introduce a central, data-driven challenge: learning from limited labeled observations of highly heterogeneous environmental features, and simultaneously preventing models from becoming opaque, site-specific, or reliant on spurious correlations within the training data. Therefore, for models that guide public health decisions, as in WQ and microbial applications, predictive performance alone is insufficient. Models must be interpretable and follow mechanistic reasoning to produce trustworthy, generalized predictions.

This dissertation addresses these challenges by developing science-guided trustworthy AI through mechanism-aligned architectures for WQ and microbial prediction. The dissertation is organized into four manuscript-based studies that incorporate environmental dynamics into domain-guided feature representation, mechanism-aligned model structures, and self-supervised learning. Chapter 2 formulates a motivational baseline for interpretable WQ prediction and elucidates the limitations of sparse sampling combined with delayed biogeochemical responses. The study utilizes heterogeneous data sources (i.e., satellite-derived multispectral imagery, meteorological forcings, and spatio-temporal variables) to predict eight WQ parameters at an average 3-day lead time, where the methodological framework couples a hybrid feature selection strategy with time-aware AI architectures that explicitly learn lagged temporal dynamics and employ localized interpretability techniques to identify robust environmental drivers. It demonstrates that although multisource machine learning can learn complex patterns from irregular data, the sole reliance on post-hoc explainability is insuffi-

cient. This is because these methods only explain statistical patterns learned during training, rather than forcing the model to follow mechanistic constraints as it learns.

Chapter 3 addresses this limitation by developing a mechanism-aware Shared Mechanistic Network (SMN). Because bloom states often persist from one week to the next, models can achieve high overall accuracy through temporal autocorrelation, rather than actively learning the biogeochemical mechanisms driving the system. This chapter, therefore, presents a novel persistence-aware learning in an architecture organized around six ecologically structured feature groups (five dynamic concept encoders together with static lake context). A concept-bottleneck branch and a mixture-of-experts branch share a mechanism-aligned backbone and are jointly trained for interpretable mechanism-aligned representations for a nonlinear prediction.

Chapter 4 addresses the challenges posed by the scarcity of labeled bloom observations. It develops a SimCLR-enabled wide-and-deep learning framework that uses NASA PACE hyperspectral imagery via self-supervised learning and combines these representations with domain-guided features for one-week-ahead cyanobacterial bloom prediction. The framework is trained and validated on Lake Erie and evaluated zero-shot on Lake Okeechobee as a cross-lake generalization test under ecological and optical domain shift.

Chapter 5 extends this self-supervised learning in a physics-informed design via the Spectral Targeted Masking (SpecTM) framework for hyperspectral foundation-model pretraining. In this work, a representative example of cyanotoxin (microcystin concentration) prediction is presented as a proof of concept, where phycocyanin, chlorophyll-a, and red-edge-sensitive bands are masked to elucidate spectral covariance with bloom dynamics.

This dissertation provides a sequence of methods for improved model interpretability and science-guided frameworks in environmental AI. It shows that trustworthy AI for water-quality and microbial prediction requires a design where domain science must guide the path from feature development to mechanism-aligned representation, even under limited ground-truth data. The dissertation explores diverse RS data sources, explainable machine learning, concept bottleneck modeling, contrastive learning, and physics-informed spectral masking to advance trustworthy AI in EO domains for improved environmental modeling.