Entry to the Stockholm Junior Water Prize 2026

HABiSense: A UAV-Powered Water Surveillance System for Harmful Algal Bloom Prediction with Multimodal Sensing and Machine Learning

Qian (Ann) Han

Basis Independent Mclean

Virginia/ United States

I. Abstract

Harmful algal blooms (HABs) constitute a growing global threat to aquatic ecosystems, public health, and economic stability. Existing monitoring approaches—vessel-based sampling, stationary buoys, and satellite remote sensing—face an inherent trade-off among spatial coverage, temporal resolution, and detection accuracy. This study presents HABiSense, an integrated water surveillance system that bridges these gaps through a four-layer closed-loop architecture: satellite-based macro-screening, unmanned aerial vehicle (UAV) field verification with autonomous sampling, multimodal water quality sensing, and machine-learning-driven prediction. The core algorithmic contribution is a novel Local-Global Fusion Network (LGFN) that simultaneously predicts bloom severity indices (regression) and bloom occurrence (classification) by fusing local patterns captured via one-dimensional convolution with long-range dependencies learned through a Transformer encoder. Evaluated against eight baseline models, LGFN achieves an F1-Score of 97.85% and an AUC of 0.996 on the classification task, while reducing false negatives by 32% relative to the strongest baseline. Field validation confirms the UAV platform's ability to autonomously detect, verify, and sample algal bloom sites, demonstrating HABiSense as a practical, end-to-end early-warning solution.

II. Table of Contents

III. Key Words

Harmful algal blooms; UAV water monitoring; Machine learning; Multimodal sensing; LGFN model; Early-warning system; Remote sensing; Water quality prediction

IV. Abbreviations and Acronyms

Abbreviation	Full Term
HAB	Harmful Algal Bloom
UAV	Unmanned Aerial Vehicle
LGFN	Local-Global Fusion Network
1D-CNN	One-Dimensional Convolutional Neural Network
LSTM	Long Short-Term Memory
SVM	Support Vector Machine
GAN	Generative Adversarial Network
FMCW	Frequency-Modulated Continuous Wave
HSV	Hue-Saturation-Value (color space)
SHAP	SHapley Additive exPlanations
AUC	Area Under the ROC Curve
ROC	Receiver Operating Characteristic
RMSE	Root Mean Square Error
IoT	Internet of Things
SST	Sea Surface Temperature

V. Acknowledgements

I extend my sincere gratitude to Dr. Blaine Hartman for guidance throughout the experimental and analytical stages of this research. All data analysis, model design, code implementation, UAV system integration, and manuscript preparation were conducted independently by the author. External assistance was limited to advisory discussions.

VI. Biography

I am a junior at BASIS Independent McLean in Virginia, USA. I have deep academic passion for environmental sciences, particularly focusing on water systems and sustainable development, along with the use of engineering and machine learning methods to address environmental issues. My passion in the area of water science stems from childhood experiences when my family's aquaculture business suffered from toxic algae blooms that affected production and ecosystem health. This motivated me to seek solutions in water quality management.

Currently, I actively engage in research in areas of environmental science, engineering, and data science. My work mainly focuses on creating technological solutions for environmental problems. In addition to my primary area of study of water environments, I also worked on various projects related to sustainable development, including the use of machine learning to find methods for waste concrete recycling using chemical techniques at George Mason University and the use of data modeling techniques to improve the efficiency and sustainability of air conditioning and fresh air systems in collaboration with researchers from Carnegie Mellon University. In addition to my academic interests, I am also devoted to community service work and leadership roles. I frequently cook and deliver home-made meals to a nearby shelter for the homeless population, donating over 500 meals overall.

In school, I lead an environmental science club where I mentor younger students, instilling in them the passion for scientific research and environmental conservation. I encourage young learners to engage with real-world issues within their environment and develop individual research ideas, guiding them through the process of conducting research.

My work is motivated by my passion for protecting aquatic ecosystems and improving access to clean water. I aim to continue developing innovations through technology-based approaches to ensure sustainable use of water resources and help communities respond to environmental issues.

1. Introduction

1.1 The Growing Crisis of Harmful Algal Blooms

Harmful algal blooms (HABs) result from the rapid, abnormal proliferation of specific phytoplankton species in freshwater and marine environments. Fueled by nutrient enrichment (eutrophication), rising water temperatures, and shifting hydrological regimes, HABs have escalated in both frequency and geographic scope over the past three decades [1]. Their consequences are far-reaching. Ecologically, the toxins released during bloom events cause mass fish kills, while post-bloom oxygen depletion creates hypoxic “dead zones” that devastate benthic communities [2]. From a public-health perspective, HAB-derived toxins—including microcystins and brevetoxins—accumulate through aquatic food chains and have been linked to neurological illness and, in severe cases, human fatalities [3]. Economically, the United States alone estimates annual HAB-related losses at approximately $4.6 billion, spanning fisheries closures, water-treatment costs, and tourism decline [4]. These intertwined ecological, health, and economic impacts underscore the urgent need for monitoring and early-warning technologies that can match the pace and unpredictability of bloom events.

1.2 Limitations of Current Monitoring Approaches

Three mainstream technologies currently dominate HAB surveillance, yet each confronts fundamental constraints that limit its utility for early warning.

Vessel-based sampling and laboratory analysis remain the gold standard for species-level identification and toxin quantification. However, the approach is labor-intensive, time-consuming, and spatially sparse, making it unsuitable for the rapid, wide-area coverage that early warning demands [5].

Fixed monitoring buoys provide continuous, real-time data at individual locations but lack spatial representativeness. A single buoy cannot capture the lateral migration and patchy distribution that characterize bloom dynamics [6].

Satellite remote sensing offers synoptic, large-scale observation, yet conventional multispectral instruments have limited spectral resolution, hindering early-stage bloom detection and species differentiation. Cloud cover and sun-glint further degrade temporal reliability [7].

Taken together, these methods produce an irreconcilable three-way trade-off—what this study terms the “Monitoring Trilemma”—among spatial coverage, temporal responsiveness, and detection accuracy. No single existing platform simultaneously satisfies all three requirements, leaving water managers with fragmented, often delayed information.

1.3 Research Objective and the HABiSense Concept

The central research question guiding this work is: Can an integrated, multi-platform system that couples satellite screening, UAV-based field verification, multimodal in-situ sensing, and machine-learning prediction overcome the Monitoring Trilemma and deliver reliable, automated early warning of harmful algal blooms?

To address this question, this study proposes HABiSense—a hierarchical water surveillance system built around a closed-loop workflow of "macro-screening → precise verification → rapid analysis → intelligent prediction." The system is designed so that each layer compensates for the weaknesses of the others: satellite imagery provides breadth; the UAV platform provides speed and spatial precision; onboard sensors provide measurement accuracy; and the machine-learning model provides predictive foresight. The core algorithmic contribution is a novel dual-branch deep-learning architecture, the Local-Global Fusion Network (LGFN), which fuses short-term local fluctuation patterns with long-range temporal dependencies to predict both bloom severity and bloom occurrence with high reliability.

This paper is organized as follows. Section 2 describes the system architecture and its four functional layers. Section 3 details the LGFN model design. Section 4 presents experimental evaluation. Section 5 discusses the results in broader scientific and practical contexts. Section 6 concludes with numbered findings and future directions.

2. Materials and Methods

2.1 Overall System Architecture

HABiSense is structured as a four-layer, data-driven pipeline that transforms wide-area satellite observations into site-specific early-warning outputs. The design philosophy follows a "perception–action–prediction" paradigm in which each layer feeds actionable information to the next (Figure 1).

Layer 1 – Satellite Remote Sensing Macro-Screening. Sentinel-2 multispectral imagery is processed through atmospheric correction and water-mask application. Established chlorophyll-a inversion algorithms generate surface concentration maps, from which regions exceeding science-based thresholds are flagged as suspected HAB zones. The geographic center coordinates of each zone are extracted to form a UAV waypoint list. This layer builds upon the open-source spectral reconstruction framework of Dilipkumar et al. [8], which uses a pre-trained 1D-CNN to reconstruct pseudo-hyperspectral data (76 bands) from 8-band Sentinel-2 inputs, achieving hyperspectral-grade detection accuracy at multispectral cost.

Layer 2 – UAV-Based Field Verification and Sampling. Upon receiving waypoint instructions, a custom-built hexacopter UAV autonomously navigates to each target location. Onboard visual verification and automated deployment of a mechanical water-quality sensor mount are performed. (detailed in Section 2.3).

Layer 3 – Shore-Based Multimodal Water Quality Analysis. An integrated sensor array measures key parameters—chlorophyll-a concentration, dissolved oxygen, pH, ammonia nitrogen, total nitrogen, total phosphorus, and sea surface temperature. These readings form the input feature vector for the prediction model.

Layer 4 – Intelligent Decision and Early Warning. The pre-trained LGFN model ingests the multimodal sensor data, outputs a continuous bloom severity index and a binary bloom-presence classification, and generates a graded risk report for water managers.

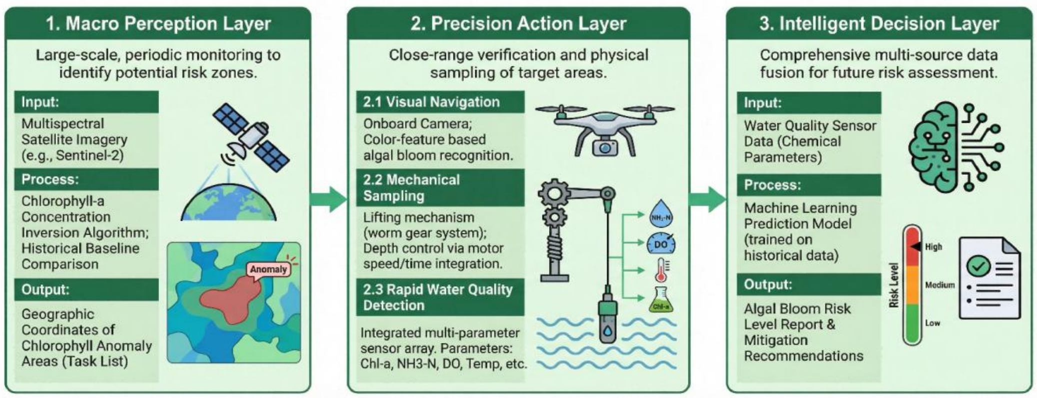

Figure 1. Schematic overview of the four-layer HABiSense closed-loop workflow.

2.2 UAV Platform Design and Autonomous Sampling Mechanism

The UAV platform was purpose-built to bridge the gap between satellite-scale observation and point-level measurement. The airframe is a lightweight, high-strength carbon-fiber hexacopter selected for its payload capacity and flight stability in wind conditions typical of open-water environments. A Raspberry Pi single-board computer serves as the onboard controller, integrating flight management, task scheduling, sensor communication, and image processing.

The most distinctive hardware innovation is a worm-gear-transmission-based sensor deployment bracket (Figure 2). Unlike conventional tethered deployment methods, which suffer from pendulum oscillation during hover, the worm gear's inherent self-locking property allows the sensor probe to be lowered to a precise, user-defined depth and held rigidly in position. This mechanism ensures stable, repeatable contact between the sensor and the water column, directly improving measurement reliability.

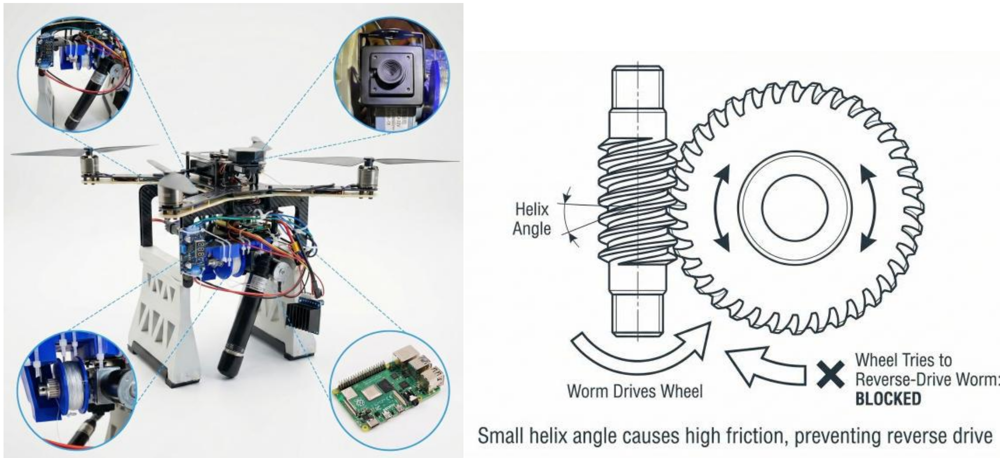

Figure 2. UAV payload system: (1) carbon-fiber airframe; (2) Raspberry Pi controller; (3) HD vision camera; (4) worm-gear sampler; (5) water quality sensor probes.

2.3 Onboard Visual Verification Algorithm

Before activating the sampling mechanism, the UAV performs real-time visual confirmation of the suspected bloom site. A lightweight recognition algorithm based on the HSV (Hue–Saturation–Value) color space was developed for this purpose (Figure 3). The algorithm converts each RGB frame captured by the onboard camera into the HSV space, applies threshold segmentation using hue, saturation, and value ranges empirically calibrated to the spectral characteristics of algal bloom surfaces, and computes the ratio of suspected bloom pixels to total image pixels. If this ratio exceeds a predefined threshold (empirically set at 15% of the frame area), the system confirms the bloom presence and triggers the sampling command. During controlled field trials, this approach reliably distinguished algal bloom surfaces from visually similar interferences, including floating aquatic vegetation and sediment plumes.

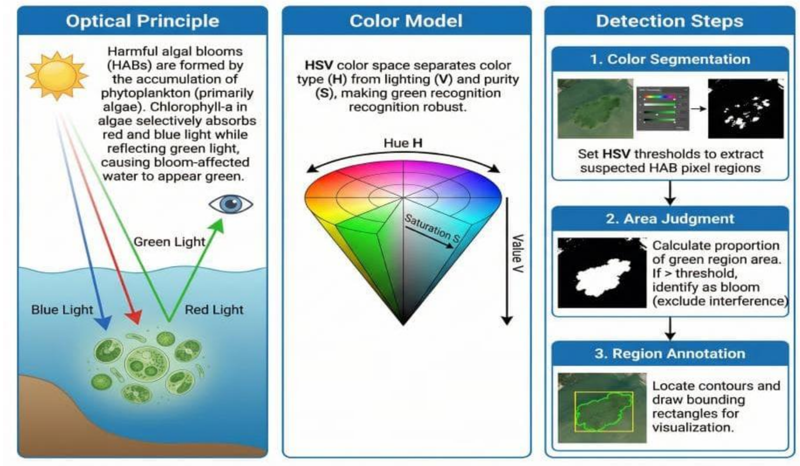

Figure 3. The HSV-based bloom detection algorithm.

2.4 Dataset Description and Preprocessing

The prediction model was trained and evaluated on an open-source HAB dataset generated through Generative Adversarial Network (GAN) augmentation [17]. The dataset encompasses eight environmental features closely linked to bloom dynamics: rolling chlorophyll-a anomaly, rolling SST anomaly, surface chlorophyll-a concentration, sea surface temperature, dissolved oxygen, pH, total nitrogen, and total phosphorus. Corresponding target variables include a continuous bloom index (regression label) and a binary HAB-present flag (classification label).

A key advantage of the GAN-augmented dataset is its enrichment of positive samples corresponding to extreme bloom events. In naturally observed data, severe HABs are rare—a class imbalance that impedes model training. The adversarial synthesis procedure generates realistic extreme-event samples that preserve the statistical distribution of real observations while substantially broadening the training distribution. Prior to training, all features were normalized to zero mean and unit variance, and preliminary Pearson correlation analysis was conducted to inform feature selection (Section 4.1.1).

2.5 Baseline Models

To provide a rigorous performance benchmark, eight baseline models spanning classical machine learning, recurrent architectures, and hybrid deep-learning pipelines were implemented under identical data splits and training protocols. These include: Support Vector Machine (SVM), Random Forest, Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Convolutional Neural Network

(CNN), Transformer encoder, Mamba (selective state-space model), and CNN-LSTM-Attention. All models were trained on the same 80/20 train-test split and evaluated using consistent metrics (Section 4.1.2).

3. Core Innovation: The LGFN Prediction Model

3.1 Design Rationale

Water quality time-series data present a dual modeling challenge. On short timescales (hours to days), abrupt signals—such as a sudden drop in dissolved oxygen or a spike in chlorophyll-a—serve as immediate precursors to bloom onset. On longer timescales (days to weeks), gradual nutrient accumulation, seasonal temperature cycles, and antecedent weather patterns shape the conditions that predispose a water body to bloom formation. Conventional sequential models (e.g., LSTM) struggle to capture both scales simultaneously: they are strong on long-range memory but less efficient at extracting localized, kernel-like patterns. Conversely, convolutional models excel at local feature extraction but lack the global receptive field needed for long-horizon dependencies.

LGFN was designed to resolve this tension through a parallel dual-branch architecture that learns local and global representations independently before fusing them into a unified feature space.

3.2 Architecture Details

Local Branch (1D-CNN). A one-dimensional convolutional neural network slides kernels across the time-series input to extract short-range correlations between adjacent time steps. The branch consists of a Conv1D layer, batch normalization, and ReLU activation, producing a 128-dimensional local feature vector. The translation invariance and parameter sharing inherent to 1D-CNNs make this branch particularly effective at detecting recurring, localized anomaly patterns.

Global Branch (Transformer Encoder). The input features are embedded with positional encoding and passed through N stacked Transformer encoder layers. The multi-head self-attention mechanism computes pairwise association scores between all time points, capturing dependencies that may span days or weeks—for example, the delayed effect of a phosphorus loading event on a subsequent bloom. This branch outputs a 128-dimensional global feature vector.

Fusion and Multi-Task Output. The local and global feature vectors are concatenated into a 256-dimensional representation, which is reduced through a fully connected fusion layer to 128 dimensions. Two parallel output heads then operate on this fused representation: (i) a linear regression head that predicts the continuous bloom index, and (ii) a linear–sigmoid classification head that outputs the probability of bloom presence. Training is performed via a joint loss function combining mean squared error (regression) and binary cross-entropy (classification), enabling the two tasks to mutually regularize the shared feature extractor and improve generalization (Figure 4).

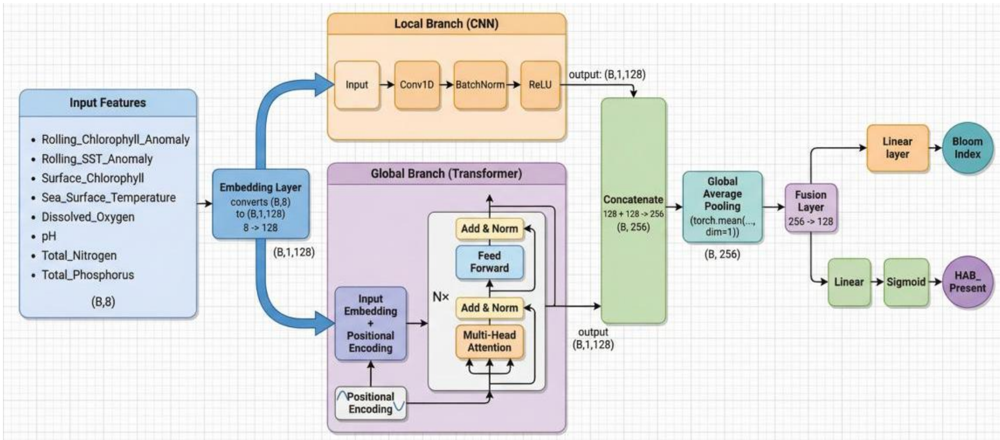

Figure 4. Architecture of the Local-Global Fusion Network (LGFN).

3.3 Why LGFN Matters for Early Warning

The practical significance of the dual-branch design extends beyond incremental accuracy gains. In an early-warning context, the most consequential error is a false negative—a bloom event that the model fails to flag. By learning complementary feature representations, LGFN reduces the probability that a true bloom signal is missed by either branch alone. As demonstrated in Section 4.1.3, this translates to a 32% reduction in false negatives compared to the strongest baseline, a margin that could, in operational deployment, mean the difference between a timely public advisory and an undetected contamination event.

4. Results

4.1 Machine Learning Model Performance

4.1.1 Feature Correlation Analysis

Pearson correlation analysis across the eight input features and two target variables reveals that surface chlorophyll-a concentration and sea surface temperature exhibit the strongest positive correlations with both the bloom index and the HAB-present flag (Figure 5). Dissolved oxygen shows a strong negative correlation with bloom indicators, consistent with the oxygen-depletion mechanism of bloom decay. These findings validate the ecological rationale for including these parameters as primary predictive features and confirm the presence of nonlinear inter-feature relationships suitable for deep-learning modeling.

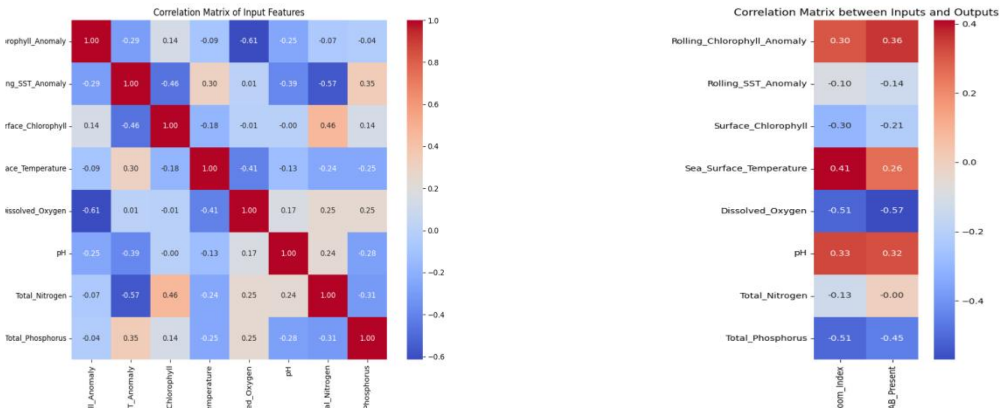

Figure 5. (a) Correlation matrix of input features; (b) correlation between input features and target variables.

4.1.2 Overall Model Comparison

Table 1 summarizes the performance of all nine models on both the regression and classification tasks. In the regression task (bloom index prediction), LGFN achieves the lowest root mean square error (RMSE), indicating the highest predictive accuracy among all candidates. In the classification task (bloom-present determination), LGFN attains an F1-Score of 97.85% and an AUC of 0.996, surpassing all eight baselines.

Table 1. Performance comparison of nine models on regression and classification tasks.

Model	RMSE	MSE	F1-Score (%)	AUC
SVM	0.1000	0.0100	73.2	0.890
Random Forest	0.0984	0.0097	92.5	0.970
RNN	0.0775	0.0060	85.4	0.940
LSTM	0.0632	0.0040	89.7	0.960
CNN	0.0447	0.0020	67.3	0.850
Transformer	0.0894	0.0080	87.1	0.950
Mamba	0.0522	0.0027	91.8	0.975
CNN-LSTM-Attention	0.0904	0.0082	88.6	0.955
LGFN (Ours)	0.0395	0.0016	97.85	0.996

4.1.3 Classification Depth: Confusion Matrix and False-Negative Reduction

The confusion matrix comparison between LGFN and the best-performing baseline (CNN-LSTM-Attention) reveals the practical significance of the LGFN architecture (Figure 6). Across the test set, LGFN reduces the number of false-negative predictions—bloom events incorrectly classified as nonbloom—from 25 to 17, a 32% reduction. In an operational early-warning context, each avoided false negative represents a bloom event that would otherwise go undetected, potentially exposing downstream communities to contaminated water. This reduction is the single most important metric for deployment readiness.

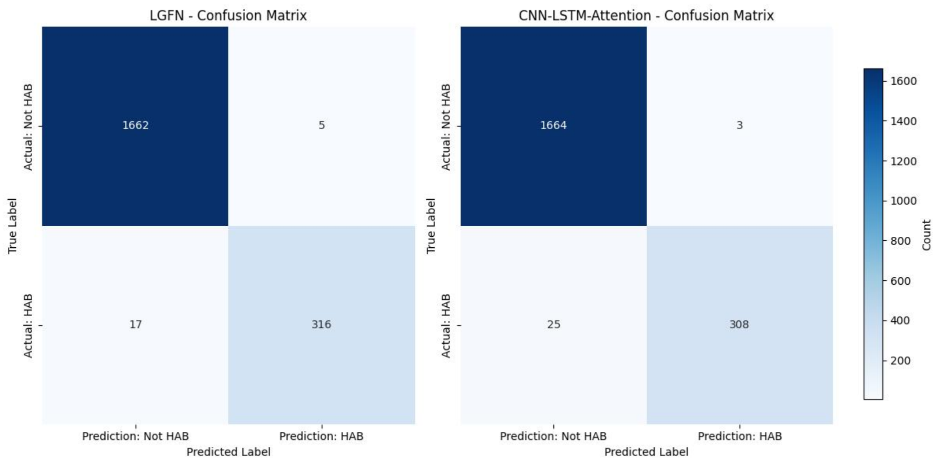

Figure 6. Confusion matrices: (a) proposed LGFN.; (b) best baseline model (CNN-LSTM-Attention).

4.1.4 ROC Curve Analysis

ROC analysis shows our LGFN model (AUC = 0.996) significantly outperforms all baselines in distinguishing HAB occurrence.

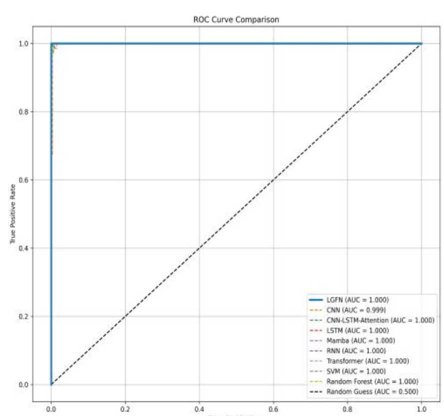

Figure 7. Overlay of ROC curves for nine models.

4.1.5 Feature Importance via SHAP Analysis

SHapley Additive exPlanations (SHAP) analysis provides insight into which environmental drivers the model relies on for each task (Figure 8). For the regression task (bloom severity), total phosphorus and pH emerge as the dominant features, reflecting the role of static nutrient loading in determining bloom magnitude. For the classification task (bloom occurrence), dissolved oxygen changes and short-term chlorophyll/temperature anomalies carry the greatest weight, consistent with the rapid-onset dynamics of bloom initiation. This divergence between the two tasks validates the multi-task architecture: the regression and classification heads learn to attend to complementary subsets of the feature space, which in turn enriches the shared feature extractor.

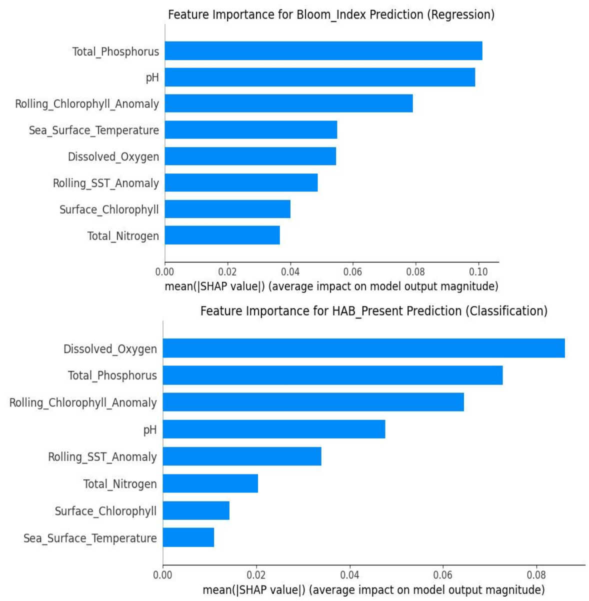

Figure 8. SHAP-based feature importance: (a) regression task; (b) classification task.

4.2 Integrated System Validation

Beyond algorithmic benchmarks, the integrated HABiSense system was validated through controlled field trials. The UAV platform successfully executed autonomous takeoff, waypoint navigation, visual bloom verification, worm-gear sensor deployment, water sampling, and return-to-home sequences without manual intervention. The HSV-based visual algorithm reliably distinguished genuine algal bloom surfaces from confounders such as floating vegetation and turbid sediment plumes across six test images (Figure 9). The end-to-end data pipeline—from satellite coordinate extraction through UAV sampling to LGFN prediction output—operated without failure across multiple trials, confirming the feasibility of the closed-loop concept.

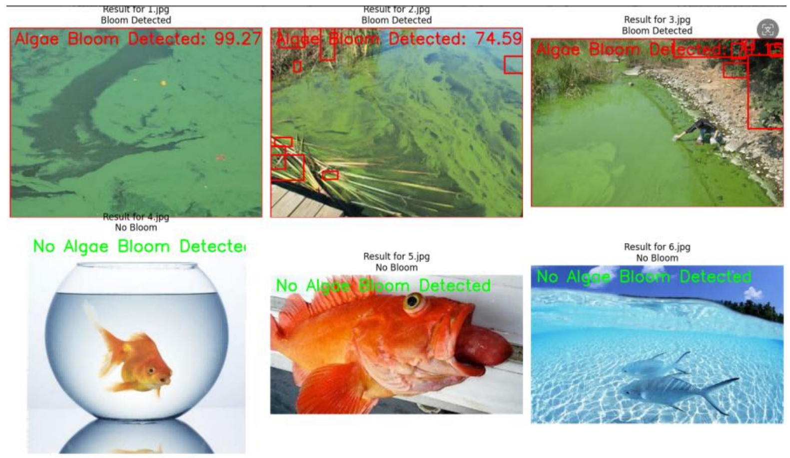

Figure 9. Sample visual detection results: (top row) bloom-negative images correctly rejected; (bottom row) bloom-positive images with bounding boxes.

5. Discussion

5.1 Resolving the Monitoring Trilemma

The results demonstrate that the HABiSense architecture effectively addresses the three-way trade-off identified in Section 1.2. Satellite remote sensing supplies broad spatial coverage at continental scale; the UAV platform delivers temporal responsiveness on the order of minutes from alert to on-site verification; and the multimodal sensor array, coupled with the LGFN model, provides detection accuracy that surpasses conventional multispectral remote sensing alone. Critically, these capabilities are not merely aggregated but interconnected: the satellite layer directs the UAV to the locations of greatest concern, the UAV verifies and filters satellite false positives, and the prediction model converts raw measurements into actionable risk assessments. This closed-loop integration is what distinguishes HABiSense from prior work that addresses only one or two legs of the trilemma.

5.2 Why LGFN Outperforms Existing Architectures

The performance advantage of LGFN over all eight baselines can be attributed to its dual-branch design. Models that rely solely on recurrent structures (RNN, LSTM) effectively capture sequential dependencies but are less efficient at extracting localized, kernel-like anomaly patterns. Conversely, purely convolutional models (CNN) excel at local feature extraction but lack the global receptive field needed to model seasonal or nutrient-accumulation effects. The Transformer baseline, while theoretically capable of global attention, underperforms in this dataset likely because the relatively short input sequences do not fully leverage its quadratic-complexity self-attention. LGFN's parallel architecture sidesteps these trade-offs: the 1D-CNN branch handles rapid, localized signals (e.g., a sudden oxygen drop), while the Transformer branch models slower, longer-range processes (e.g., cumulative phosphorus loading). The fusion of these complementary representations yields a richer feature space that neither branch achieves alone.

5.3 Practical Implications and Target Users

The HABiSense system is designed for deployment across a range of aquatic settings, including inland lakes, drinking-water reservoirs, aquaculture zones, and coastal bays. The primary target users are environmental protection agencies, municipal water utilities, and aquaculture operators—organizations that require timely, site-specific bloom forecasts to guide management decisions such as water-intake shutoffs, beach closures, or shellfish-harvest bans. The modular system design allows users to adopt individual layers (e.g., the LGFN model alone for sites with existing sensor infrastructure) or the full four-layer pipeline for sites lacking any monitoring capacity.

From a cost perspective, the prototype UAV platform was assembled from commercially available components (carbon-fiber frame, Raspberry Pi, off-the-shelf sensors) at a fraction of the cost of professional environmental monitoring drones. The LGFN model runs inference on a standard laptop within seconds, requiring no specialized hardware. These attributes lower the barrier to adoption, particularly for resource-constrained municipalities and developing-country agencies that bear a disproportionate share of HAB impacts.

5.4 Comparison with Related Systems

Several recent studies have explored UAV-based water quality monitoring or machine-learning-based bloom prediction independently. For instance, drone-mounted hyperspectral cameras have been used to map chlorophyll distributions [7], and LSTM-based models have been applied to forecast bloom occurrence from historical water quality records [5]. However, no prior system, to the author's knowledge, integrates satellite screening, autonomous UAV verification with mechanical sampling, and a purpose-designed dual-branch deep-learning model within a single closed-loop workflow. The novelty of HABiSense lies not in any single component but in the synergy among its layers, each of which compensates for the limitations of the others.

5.5 Limitations

Several limitations should be acknowledged. First, the LGFN model was trained and evaluated on a GAN-augmented synthetic dataset. While the augmentation procedure faithfully reproduces the statistical properties of real-world observations, field-collected data from specific water bodies may introduce distributional shifts that affect model accuracy. Real-world validation across diverse geographical settings remains essential before operational deployment. Second, the current UAV platform has a flight endurance of approximately 25 minutes under full payload, limiting the number of sites that can be surveyed per mission. Third, the HSV-based visual algorithm, while effective under clear and overcast conditions, has not been tested under extreme lighting (e.g., low sun angle, heavy fog). Fourth, the system does not yet support on-site algal species identification; incorporating miniaturized microscopic or spectral sensors into the UAV payload is a planned future extension.

6. Conclusions

The principal findings and contributions of this study are summarized as follows:

1. An integrated, four-layer water surveillance system (HABiSense) was designed and implemented, coupling satellite remote sensing, autonomous UAV field verification, multimodal water quality sensing, and machine-learning prediction into a closed-loop early-warning workflow.

2. A novel dual-branch deep-learning model (LGFN) was proposed, combining a 1D-CNN for local pattern extraction with a Transformer encoder for global dependency modeling. Trained under a multi-task framework, LGFN achieved an F1-Score of 97.85% and an AUC of 0.996 on the HAB classification task, outperforming eight established baseline models.

3. LGFN reduced false-negative predictions by 32% compared to the strongest baseline (CNN-LSTM-Attention), directly enhancing the reliability of the early-warning output—the metric most critical for protecting public health and downstream ecosystems.

4. SHAP-based interpretability analysis revealed that regression and classification tasks rely on complementary environmental drivers (static nutrients vs. dynamic anomalies), validating the multi-task architecture and providing actionable insight for water managers.

5. Integrated field testing confirmed the feasibility of the full HABiSense pipeline, demonstrating autonomous UAV bloom verification, worm-gear sensor deployment, and end-to-end data flow from satellite coordinates to risk-level output.

6. Future work will focus on real-world validation across diverse water bodies, integration of on-site

algal species identification, UAV swarm coordination for expanded spatial coverage, and model lightweighting for edge deployment.

7. References

[1] National Centers for Coastal Ocean Science. Harmful algal blooms monitoring system. NOAA. Retrieved January 21, 2026, from https://coastalscience.noaa.gov/research/stressor-impacts-mitigation/hab-monitoring-system/

[2] Padmakumar, K.B., Menon, N.R. & Sanjeevan, V.N. (2012). Is occurrence of harmful algal blooms in the exclusive economic zone of India on the rise? International Journal of Oceanography, 2012, 1–7.

[3] Petruzzello, M. (n.d.). Harmful algal blooms. In Encyclopaedia Britannica. Retrieved January 21, 2026, from https://www.britannica.com/science/harmful-algal-bloom

[4] Environmental Working Group (2022, August 23). The high cost of algae blooms: Economic impacts and solutions. https://www.ewg.org/research/high-cost-of-algae-blooms

[5] Centers for Disease Control and Prevention (2023). Summary Report – One Health Harmful Algal Bloom System (OHHABS), United States, 2021. Atlanta, GA: U.S. DHHS.

[6] National Oceanic and Atmospheric Administration (2023, June 23). What can we do about harmful algal blooms? https://www.noaa.gov/what-can-we-do-about-harmful-algal-blooms

[7] Dilipkumar, J., Shanmugam, P. & He, X. (2026). Detecting harmful algal blooms from reconstructed hyperspectral satellite data. International Journal of Remote Sensing, 47(2), 539–573.

[8] Johnston, J.D. (2002). Transform coding of audio signals using perceptual noise criteria. IEEE Journal on Selected Areas in Communications, 6(2), 314–332.

[9] Van Otten, N. (2024). Support vector machines in machine learning. Spot Intelligence.

[10] Omarzai, F. (2024). Random forest in-depth. Medium.

[11] GeeksforGeeks (2025). Stacked RNNs in NLP. geeksforgeeks.org.

[12] Dive into Deep Learning. 9.2. Long short-term memory (LSTM). classic.d2l.ai.

[13] Bajaj, G. Notes on convolutional neural networks. AI in Plain English.

[14] Dive into Deep Learning. 11.7. The Transformer architecture. d2l.ai.

[15] Jumle, V. (2024). Mamba: SSM, theory and implementation. Towards Data Science.

[16] Chen, H., Wang, K., Zhao, M. et al. (2025). A CNN-LSTM-attention based seepage pressure prediction method. Scientific Reports, 15, 12960.

[17] Mohan, D., Dayalani, A. & Gupta, L. (2025). Generative adversarial network (GAN) synthesized dataset on harmful algal bloom. Mendeley Data, V1.