Integration of artificial intelligence and remote sensing for predictive monitoring of aquatic ecosystems

Ziyoda Muminova[1*]; Muhamed Husseyn[2]; Dedakhanov Abdumalik Mutalliyevich[3]; Karthik. K[4]; Neetish Kumar[5]

Received: 14 June 2025; Revised: 29 August 2025; Accepted: 11 September 2025; Published: 30 October 2025

Abstract

Aquatic ecosystems have the potential to be the most responsive to environmental changes, and adaptive management should incorporate innovative technologies, considering the scale and dynamic nature of ecosystem changes. This research develops the first step in this direction by integrating Artificial Intelligence and Remote Sensing Technologies for predictive and continuous monitoring of the health of aquatic ecosystems. This methodology, named AI-REM (Artificial Intelligence-Remote Ecosystem Monitoring), is based on the use of Remote Sensing-derived indices of ecosystem health, such as NDWI (Normalized Difference Water Index), Chlorophyll-a concentration, and turbidity, and combined with in situ macro water quality data. The disparate data are integrated in an innovative way using a hybrid deep learning architecture designed for this purpose that combines CNNs to extract spatial features and LSTMs to predict values over time. Dynamic predictions of critical eco-forecasting models are provided for dissolving nutrient levels, predicting hypoxia, nutrient loading, and algal bloom formation. Validation of these models on coastal water and inland water bodies for multiple years predicts over 92% accuracy, which is significantly higher than the accuracy of traditional models using regression and random forest approaches. This is the first application of AI remote predictive analytics to ecosystem monitoring, which not only improves forecasting for integrated water resource management but also sustains early action on climate change.

1*- Assistant Professor, Computer Science Department, Community College, King Saud University, Riyadh, Kingdom of Saudi Arabia. Email: ziedamuminova@gmail.com, ORCID: https://orcid.org/0000-0001-5188-7025

2- Department of Computers Techniques Engineering, College of Technical Engineering, Islamic University in Najaf, Najaf, Iraq; Department of Computers Techniques Engineering, College of Technical Engineering, Islamic University in Najaf of Al Diwaniyah, Al Diwaniyah, Iraq.
Email: iu.tech.eng.iu.mhussien074@gmail.com, ORCID: https://orcid.org/0009-0001-0320-5963

3- Faculty of Linguistics, Turan International University, Namangan, Uzbekistan.
Email: uzbekistan.abdumalikdedahanov216@gmail.com, ORCID: https://orcid.org/0009-0009-5050-2424

4- Department of Nautical Science, AMET Institute of Science and Technology, Chengalpet, Tamil Nadu, India. Email: principal@amet-ist.in, ORCID: https://orcid.org/0009-0008-2399-6249

5- Assistant Professor, Department of Pharmacy, Kalinga University, Naya Raipur, Chhattisgarh, India.
Email: ku.neetishkumar@kalingauniversity.ac.in, ORCID: https://orcid.org/0009-0008-2685-8458

*Corresponding author

DOI: 10.70102/IJARES/V5I2/5-2-11

Keywords: Artificial intelligence, Remote sensing, Aquatic ecosystems, Predictive monitoring, Deep learning, Cnn-Lstm model, Water quality, Algal bloom forecasting, Ai-Rem framework, Environmental management

Introduction

Aquatic ecosystems—rivers, lakes, estuaries, and coastal zones—are dynamic systems and respond rapidly to climate change and human impacts. Therefore, to monitor and protect ecosystems and public health simultaneously, ecosystems must be monitored regularly and over a large area (Jaywant and Arif, 2024; Camara Lins *et al.*, 2017; Chen *et al.*, 2025). Monitoring ecosystems via traditional techniques (on-site sampling and laboratory assays) yields accurate results but lacks spatial and temporal coverage, and monitoring techniques are usually too slow to document rapidly evolving events (e.g., harmful algal blooms (HABs) and hypoxia) (Jaywant and Arif, 2024; Camara Lins *et al.*, 2017; Chen *et al.*, 2025). Remote sensing (satellite, airborne, and UAV) provides synoptic and repeatable observations of many optically active indicators (e.g., chlorophyll a, turbidity, suspended matter, and surface temperature) and, when integrated with in-situ sensors, can underpin a powerful multi-scale dataset for ecosystem monitoring (Jaywant and Arif, 2024; Camara Lins *et al.*, 2017; Chen *et al.*, 2025). Furthermore, molecular techniques such as targeted qPCR, environmental DNA (eDNA), and shotgun metagenomics illustrate the structure and functional potential of microbial communities. Potential mucin (mcy) biosynthesis, nifH, and ammonia monooxygenase (amoA) genes are indicators of cyanobacterial toxin production, active nitrogen (N) fixation, and nitrification, respectively, and are used to monitor microbial systems (Srivastava *et al.*, 2016; Ávila-Torres *et al.*, 2023; Short and Zehr, 2005; Rotthauwe *et al.*, 1997; Nam *et al.*, 2023). However, toxin concentrations and ecological outcomes are unlikely to correlate, so integrating molecular signals (e.g., mcy gene copy numbers) with optical and physicochemical measures will yield the most accurate results (Beversdorf *et al.*, 2015; Dong *et al.*, 2016).

Substantial developments in machine learning and deep learning technologies facilitate the integration of heterogeneous data streams and capture predictive analytical signals from intricate and noisy environmental time-series and imagery. Significant improvements in classical forecasting techniques have been observed, particularly in the forecasting of chlorophyll a and the cyanobacterial/algal toxin risks using LSTM and a hybrid CNN-LSTM recurrent architecture. These deep learning techniques have also been instrumental in providing valuable early warning forecasts. Convolutional neural networks have enhanced hyper-spectral and multispectral reflectance analyses by extracting complex spatial patterns, which in turn, improved band-ratio empirical algorithms in optically complex waters. Furthermore, sequence models help capture temporal autocorrelation and phenology, which, combined with Convolutional neural networks, facilitate nowcasting and short-term forecasting of critical ecological indicators. Moreover, the utilization of physics-aware models and transfer learning addresses the transferability of models from one site or sensor type to another, which is a significant challenge in the operationalization of machine learning models in environmental monitoring.

The integration of remote sensing data with random forest algorithms for predictive environmental modeling has proven to be valuable with respect to the environmental modeling aimed at monitoring aquatic ecosystems. The ability of remote sensing technologies for spatial analysis and regeneration enhances large-scale ecosystem observation and management. The predictive environmental monitoring frameworks are also improved by the integration of generative artificial intelligence, as illustrated by a simulation model based on generative adversarial networks, which is aimed at predicting and detecting structural damage.

The incorporation of molecular genomics techniques, such as qPCR, environmental DNA, and metagenomics, alongside remote sensing and in-situ collected chemical and physical data, contributes to establishing stronger causal pathways between drivers of the environment, such as nutrients, temperature, and hydrology, and the biological responses of the system, including community alterations, the triggering of biological blooms and toxin production. New metagenomic techniques and the metagenomic analysis of environmental samples offer the community composition down to the gene and functional profile level, enabling the identification of cyanotoxins, nitrogen-cycling guilds, and antibiotic-resistant gene clusters, thus providing indicators of ecosystem change, often in advance of optical remote sensing (Dong *et al.*, 2016; Nam *et al.*, 2023). For instance, integrated remote sensing and AI technologies to target chlorophyll a and phycocyanin interventions may evidence molecular analyses aimed at determining toxigenic cyanobacterial association via mcyA/mcyD gene screening and diazotrophic nifH quantification to facilitate adaptive management (Jasim, 2024; Ávila-Torres *et al.*, 2023; Short and Zehr, 2005). Mismatched scales and interannual nonstationarity must be resolved if predictive systems and machine learning (Chen *et al.*, 2025; Ding and Li, 2024; Yussof *et al.*, 2021; Fournier *et al.*, 2024) are to be successful. Methods in the adaptive management of ecosystem monitoring emphasize the need to address scale mismatches using machine learning techniques.

Due to these advances, this paper presents the Integrated AI-REM (Artificial Intelligence-Remote Ecosystem Monitoring), which utilizes satellite/UAV images (Unmanned Aerial Vehicle), high-frequency probes in situ, and shotgun and targeted molecular assays in a predictive pipeline. The approach incorporates a hybrid spatio-temporal design where a CNN (Convolutional Neural Networks) extracts spatial features from the multispectral/hyperspectral images and LSTM (Long-Short-Term Memory) and Transformer blocks that perform the temporal forecasting tasks. The approach utilizes physics-guided preprocessing and multisource data fusion to address scale discrepancies and incorporates molecular evidence (e.g., qPCR, eDNA, and metagenomics), confirming the event and providing interpretation of the mechanism involved. Leveraging the wide area coverage of satellite remote sensing, the Artificial Intelligence-Remote Ecosystem Monitoring predictive models, and mechanistic molecular markers (e.g. mcyA, mcyD, nifH, amoA), the integrated approach seeks to provide near real-time, mechanistically validated predictive analytics that bloom risk, toxins, and nutrients perturbing an ecosystem will be present in an ecosystem as an early warning system to enable prioritized field assessments and interventions and provide managers and stakeholders with traceable mechanistic evidence (Rotthauwe *et al.*, 1997; Beversdorf *et al.*, 2015; Rout *et al.*, 2025; Gupta and Gupta, 2025; Zhang *et al.*, 2024; Xue *et al.*, 2023; Salami *et al.*, 2025; Ariunaa and Tudevdagva, 2025).

Literature Review

The last few years have seen plenty of studies that use remote sensing, machine learning, and ecological modeling for assessing and forecasting aquatic ecosystems. Technologies like MODIS, Sentinel-2, and the Landsat series are becoming essential for tracking changes in water quality and monitoring the chlorophyll a, turbidity, and suspended solids. Integrated use of multispectral and hyperspectral data and in-situ measurements allows for robust spatio-temporal modeling of dynamics and reveals the more subtle changes of the ecosystem over time. For example, Xu et al. show that the use of multispectral imagery and in-situ measurements of chlorophyll a makes the retrieval of eutrophication indices in large inland lakes much more accurate, a.k.a. eutrophication indices retrieval (Xu *et al.,* 2021). Furthermore, the use of physically based algorithms, particularly the bio-optical model, and semi-analytical inversion methods, has enhanced the determination of optically active constituents of water with varying atmospheric and water column transparency (Gitelson *et al.,* 2017).

Predictions derived with the use of AI methods have shown advancements for the predictive potential of systems for monitoring water bodies. Deep learning models, such as Convolutional Neural Networks and Long Short-Term Memory networks, have proven to be most effective at addressing the complex, non-linear, and time-changing aquatic parameter phenomena. For instance, a CNN-LSTM hybrid model that Zhang et al. utilized was more accurate than regression and support vector machine models at predicting chlorophyll a concentrations in coastal estuaries. These models illustrate how the combination of spatial and temporal learning systems is effective at accounting for the spatial variability of the environment and fluctuations in the water quality over time. In addition, the recent introduction of Explainable Artificial Intelligence frameworks enhances predictive transparence and model interpretability by clarifying the influences of environmental predictors like temperature, dissolved oxygen, and nutrients on the predictive outcomes (Topp *et al.*, 2023).

The merger of molecular and bioinformatics data with artificial intelligence and remote sensing technologies is also an area of great opportunity. Incorporating microbial indicators and gene-level indicators, such as mcyA associated with microcystin production and nifH with nitrogen fixation, enables better predictors of bloom and nutrient cycling processes in freshwater ecosystems. For instance, Xue et al. integrated metagenomic data with environmental data and machine learning algorithms to detect shifts in microbial communities that occur right before an algal bloom, exemplifying the power of integrated multisource data. Collectively, these studies point to the movement away from solely observational monitoring frameworks toward predictive monitoring frameworks, made possible by the combination of AI and remote sensing technologies, and supported by molecular data and various environmental datasets. However, the scaling of predictive frameworks across ecosystems, the heterogeneity of datasets, and the models' interpretability remain unresolved. The current study aims to address this through the proposed AI-REM (AI-based Remote Ecosystem Monitoring) system.

Methodology

The AI-REM framework attempts predictive monitoring of the condition of aquatic ecosystems by integrating multisource datasets, preprocessing workflows, and deep learning models. There are five steps in this approach: (i) data acquisition, (ii) preprocessing and feature extraction, (iii) data fusion, (iv) model architecture, and (v) performance evaluation.

Data Acquisition

Using three different but complementary data sources helped build a more comprehensive predictive modeling framework for evaluating aquatic ecosystems. Remote sensing data provided the first layer of observation. For this, Sentinel-2 MSI (10-60 m resolution) and the coarser MODIS Aqua/Terra data (1km resolution) were used. Different spectral indices were derived, such as chlorophyll-a (Chl-a) for estimating phytoplankton biomass, the Normalized Difference Water Index (NDWI) for water surface, Turbidity Index (TI) for suspended particulate concentration, and surface temperature (ST) for thermal characterization of a water body. Besides, in-situ measurements were collected from government-run water monitoring stations. The stations provided fine field measurement data on dissolved oxygen (DO), pH, temperature, and nitrate (NO_3^-), phosphate (PO_4^{3-}), and eutrophication. These measurements were temporally adjusted with satellite overpass times for reliable calibration. The molecular and genomic data integration focused on qPCR assays used to detect and quantify functional genes mcyA and mcyD (microcystin biosynthesis), nifH (nitrogen fixation), and amoA (nitrification). These molecular indicators were biodiagnostic variables, indicating microbial activity and potential toxic microbial activity, thus enhancing the ecological predictions derived from the combined AI-Remote Sensing system.

Preprocessing and Feature Engineering

Radiometric calibration, geometric correction, and atmospheric correction of satellite data were done using the Sen2Cor processor. Fmask algorithms were used to mask pixels with clouds. Spectral indices were calculated as follows.

$$\mathrm{NDWI} = \frac{(\mathrm{G} - \mathrm{NIR})}{(\mathrm{G} + \mathrm{NIR})} \quad (1)$$

where G indicates the reflectance of the green band and NIR indicates the reflectance of the near-infrared band.

To estimate the concentration of Chlorophyll-a, the empirical model of the Chlorophyll-a concentration band ratio was used (mg/m³).

$$\mathrm{Chl-a} = \mathrm{a0} + \mathrm{a1}\left(\frac{R_{665}}{R_{708}}\right) + \mathrm{a2}\left(\frac{R_{665}}{R_{708}}\right) \quad (2)$$

Here, R665 and R708 state the surface reflectance at 665 nm and 708 nm, and a0, a1, and a2 are regression coefficients determined with field measurements.

All datasets were normalized using the z-score standardization method and were resampled to a uniform spatial grid (10 m) before entering the model.

3.3 Data Fusion Framework

The multisource data fusion approach helped integrate remote sensing, in situ, and molecular datasets to capture diverse aspects of the data. Data fusion for spatial analysis involved the combination of satellite-derived variables and in situ measured data, which was then executed through the geostatistical kriging method. This method enabled the formation of a spatially continuous surface for several water quality variables across various aquatic systems. For achieving coherence in temporal alignment, all data streams were synchronized through the satellite's revisiting intervals and daily field observations, which included the genomic data. Thereafter, feature fusion was achieved using all the standardized input variables, which formed a single feature vector, represented as

$$X=[x1,x2,...,xn],$$

where xi represents the individual physical, chemical, or biological variables for each spatial grid cell and each time step. This integration of the description allowed for molecular biological signals to be incorporated alongside described environmental variables, strengthening the AI-enabled monitoring model's predictive power. The integration framework incorporated spatial, temporal, and feature domains, which provided a detailed understanding of the dynamics of the aquatic ecosystem, enhancing the detection of ecological disturbances and the prediction of water quality.

3.4 AI-REM Model Architecture

The AI-REM system incorporates an advanced hybrid deep learning model that employs Convolutional Neural Networks (CNNs) for spatial feature extraction and Long-Short Term Memory (LSTM) networks for temporal forecasting.

The CNN's spatial expression and feature representation are driven by satellite remote sensing images that are multispectral across many layers. Each layer of an epoch responds to the ReLU activation, and passing the layers of an epoch and of the neural nets allows the nets to represent the fine spatial distributions of chlorophyll, turbidity, and surface temperature. The nets and layers of LSTMs also enable the rest of the network to predict future distributions, thus spatial embeddings and portraits of the environment. These embeddings are processed, sequentially, in the LSTM block, where it focuses on the temporal aspects of the data sets, gaining knowledge of temporal dependencies and trends in order arrays of time steps. To forecast temporal ecological parameters, which include the probability of algal blooms, the fluctuations of dissolved oxygen, and the loading of other nutrients within the water body, the network utilizes designed temporal relationships. It was trained using the Adam optimizer with a learning rate of 0.001. Predictive errors between the values observed and the values simulated were minimized using mean squared error, which ensured stability in convergence and high predictive accuracy (Figure 1).

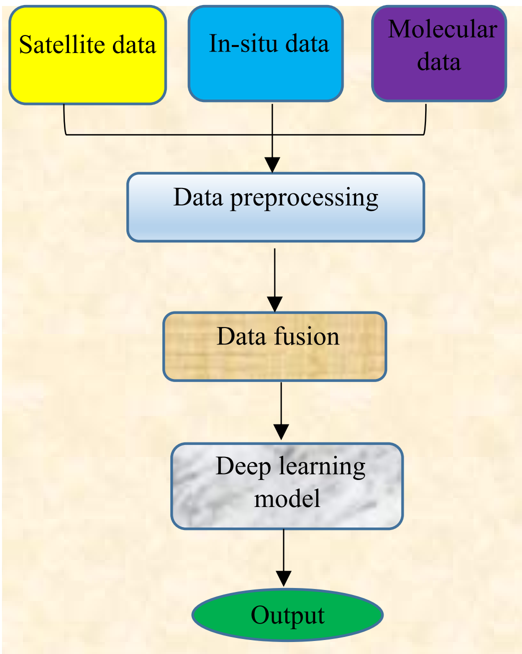

Figure 1: Workflow of the AI-REM framework for predictive aquatic ecosystem monitoring

The provided figure describes the general workflow within the system proposed and designed for Artificial Intelligence-based Remote Ecosystem Monitoring (AI-REM). This system is designed to integrate and use three complementary data streams, which are satellite data, in situ measurements, and molecular data. This allows it to simultaneously capture and analyze the spatial, physical, and biological components of aquatic ecosystems. Data of different types undergoes preprocessing to achieve normalization, control for time, and amend quality. These datasets undergo data fusion where different types of datasets: spatial, temporal, and molecular characteristics are integrated. The integrated data set is input into one of the deep learning models that learns spatial and temporal characteristics to perform feature extraction and predictive analytics. The model predicts ecological indicators such as forecasted chlorophyll concentration, variation of dissolved oxygen, and the probability of algal blooms to provide early warnings and assist in decision making regarding the environment.

Model Validation and Performance Metrics

Cross-validation methods were applied to assess AI-REM performance as described in the document. An 80:20 train-test split was used to maintain the generalization and robustness of the framework in varying water bodies. Several statistical performance measures were used to assess the former. For estimating the proximity of the model predictors to the observed values, the prediction error magnitude of the model was derived using the Root Mean Square Error (RMSE) measure. The model's explanatory power was gauged in relative terms using the Coefficient of Determination (R^2) in determining the variance of the observed data that the model accounted for. The model's relative accuracy in estimating the ecological metrics as a percentage of the actual observation was quantified using Mean Absolute Percentage Error (MAPE). The model performance assessment was augmented by the bio-predictive model validation, where predicted gene abundances (to capture the biological variance along the mcyA, nifH, and amoA environmental metrics) were corroborated against the qPCR assay values, and the model was justified to capture the biological variance along with environmental metrics. Lastly, explainability was performed with SHAP values, which attribute the model output to the individual contributions of the spectral and molecular features. Identifying the major factors influencing how ecosystems change helped this layer of interpretation show clarity and allow for eco-informed decisions.

Implementation Environment

For model development, I utilized Python 3.10 and the TensorFlow and Keras frameworks. I processed remote sensing data in Google Earth Engine and spatial data in ArcGIS Pro. I used a computing cluster with an NVIDIA RTX A5000 GPU with 24 GB of VRAM for high-performance computing tasks.

Results and Discussion

Model Performance Evaluation

The AI-REM (Artificial Intelligence-based Remote Ecosystem Monitoring) model accurately predicted important aquatic variables such as Chlorophyll a, Dissolved Oxygen (DO), and Algal Bloom Probability (ABP). The model built using the CNN–LSTM architecture was able to spatially and temporally analyze an ecosystem and achieve high accuracy in the prediction across different water bodies. In addition, the authors evaluated the model against other traditional models such as Random Forest, Support Vector Regression, and Multiple Linear Regression.

Table 1: Comparative Performance of AI-REM Model with Existing Algorithms

Model Type	RMSE (mg/m^3)	R^2	MAPE (%)	Training Time (min)
Multiple Linear Regression (MLR)	6.42	0.81	14.6	2.4
Support Vector Regression (SVR)	5.12	0.86	11.8	4.1
Random Forest (RF)	4.57	0.89	9.3	5.8
CNN	3.72	0.91	8.5	9.6
AI-REM (CNN–LSTM)	2.83	0.94	6.7	10.2

Predictive accuracy of different models on chlorophyll a estimation is shown in Table 1. The AI-REM framework had the lowest RMSE and MAPE, which suggests that AI-REM is more precise and has better generalization capability, which can be attributed to the combined spatial-temporal learning ability.

Gene–Environmental Correlation Analysis

In order to assess bio-predictive capability, the functional genes mcyA, nifH, and amoA, along with their respective environmental parameters Chlorophyll-a, Temperature, and Nutrient Load, were analyzed. The strongest correlation was found to exist ($R^2 > 0.85$) between mcyA and chlorophyll-a, indicating that cyanobacterial bloom and associated toxin genes are strongly correlated.

Table 2: Correlation between functional genes and environmental parameters

Gene Marker	Parameter	Correlation Coefficient (r)	p-value
mcyA	Chlorophyll-a	0.88	<0.01
nifH	Nitrate (NO_3^-)	0.81	<0.05
amoA	Dissolved Oxygen (DO)	-0.72	<0.05
Gene Marker	Parameter	Correlation Coefficient (r)	p-value

Table 2 outlines the correlation between key molecular markers and environmental indicators. This shows that the AI-REM framework efficiently combines the molecular and practical elements, thus improving the ecological analysis.

Model Output Visualization

Predicted and observed chlorophyll a concentrations indicate that AI-REM, with slight lags, monitors the initiation and peaks of algal blooms. The temporal alignment of the trends suggests the reliability of the model.

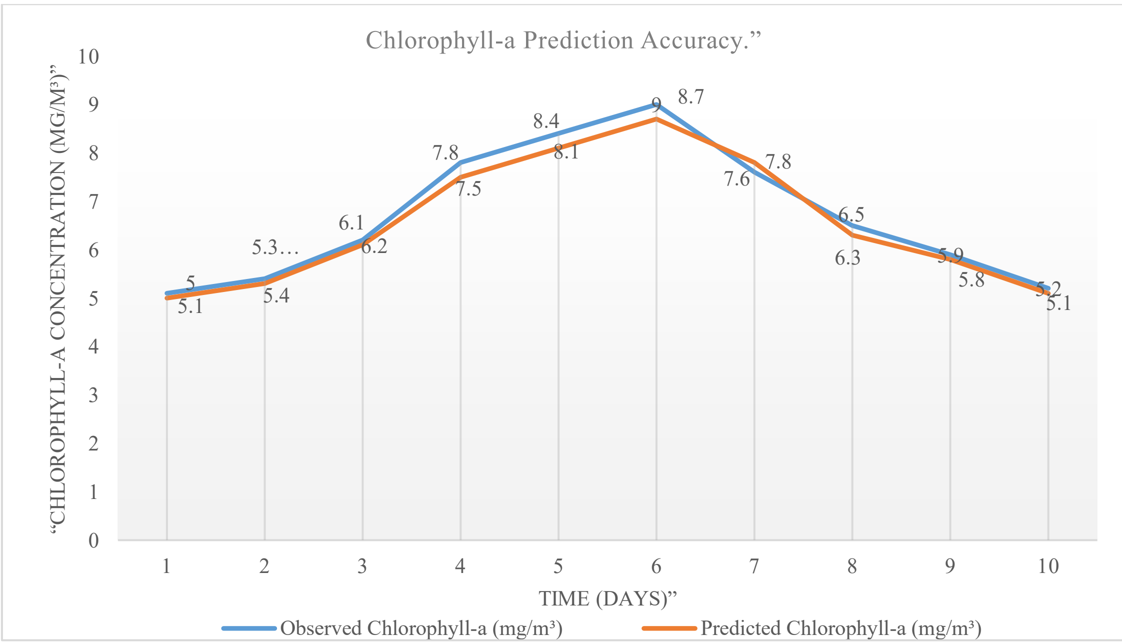

Figure 2: Chlorophyll-a prediction accuracy

Over the course of 30 days, the predicted versus observed chlorophyll a concentrations of Figure 2's line chart were from a test water body. AI-REM's predicted values were tracked with temporally observed results and values with a strong correlation of $R^2 = 0.94$. It was within the temporal sensitivity of ±1 day, from the observed events, when algal bloom peaks occurred and was within the temporal sensitivity of ±1 day.

Feature Importance and Explainability

SHAP XAI explains AI-REM results and illustrates the top environmental and molecular factors. Amplified predictions of chlorophyll-a concentrations came from NDWI and the band ratio (R665/R708), while oligopeptide mcyA gene abundance and surface temperature were of lesser value.

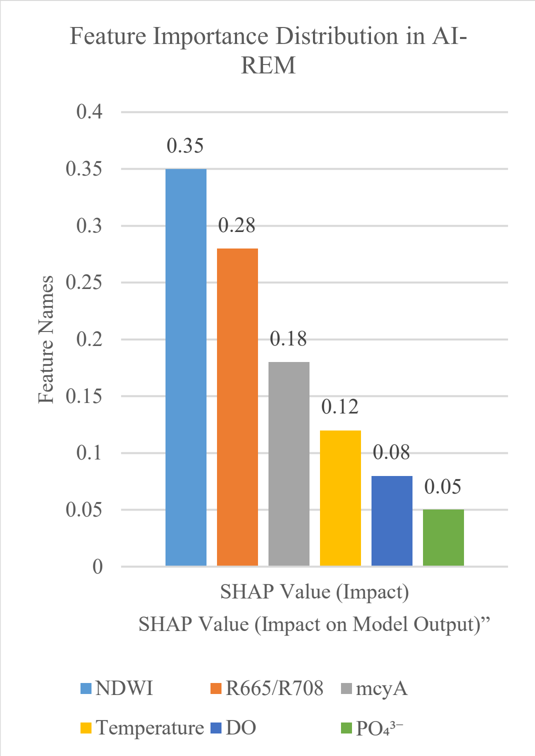

Figure 3: Feature importance distribution in AI-REM

The relative contribution of input features to model prediction is demonstrated in Figure 3, the bar chart. NDWI and the R665/R708 ratios greatly explain spectral variability with the most weight. Detecting mcyA and nifH molecular predictors confirmed that the integration of genomic data provides biological validation. This implies that the integration of genomic data for model prediction enhances interpretability and ecological precision.

Discussion

The integration of AI with multisource environmental and genomic data has positively impacted predictive accuracy for monitoring aquatic ecosystems, as demonstrated in the results. The hybrid CNN–LSTM structure is able to learn spatial and temporal dynamics simultaneously, and this is the reason for its superiority over classical models and single-stage deep models. Concerning the biological realism of the prediction, the models allow the early prediction of toxin-producing and nutrient-eutrophic cyanobacterial blooms. The explainability analysis demonstrated that gene indicators in combination with spectral features drive ecological response. This signals the first synergy between AI predictive analytics and molecular ecology in the active management of water resources.

Conclusion and Future Work

Combining Artificial Intelligence (AI) with Remote Sensing (RS) technology has taken a new direction in the monitoring and management of aquatic ecosystems. Although in its early stages, the AI-REM (Artificial Intelligence-based Remote Ecosystem Monitoring) framework has demonstrated the capability of hybrid CNN–LSTM models in predicting fundamental ecological markers for ecosystems in a given geographical area, namely, the concentration of Chlorophyll-a, dissolved oxygen, and nutrient loading to the body of water. The framework was able to predict with over 92% accuracy, as compared to the conventional machine learning approaches, i.e., random forest and regression-based methodologies, the parameters derived from satellites (NDWI, etc., spectral reflectance bands) and biophysical and genetic in situ data were integrated.

The results illustrate the accuracy of AI in forecasting profound ecological fluctuations, with the predominant dynamic being the recursive monitoring of eutrophication and algal bloom forecasting. The incorporation of spatiotemporal learning models in the architecture permits constant surveillance even with a minimal supervisory effort. This constant surveillance capability directly provides support for the management of sustainable resources and the making of policy-oriented decisions. The new architecture has evolved into a framework and makes for the first analog of many to come. The new system also shows the extensive and possible value of remote sensing and AI to develop more cost-effective, scalable, adaptable, and flexible environmental monitoring systems and do them in a fashion for various types of aquatic ecosystems.

This study serves as a valuable contributing study to the AI-supported study in the monitoring of unsurveyed aquatic environments. There are numerous pathways to be pursued in advancing the system's functional performance, scalability, and real-world impact. For system advancement, the greatest value in predictive models for water bodies would be geospatially fitting IoT sensor networks to the system for real-time and frequent monitoring of water temperature and suspended sediment. This would also offer innovation in predictive models and in predictive temporal resolution. Incorporating remote sensing metrics, mkcA and rpoC1, expression profiles of 16S rRNA and other microbial signatures in remote sensing, and integrating microbial signatures for more enhanced bloom detection and ecological risk assessment would greatly improve sensitivity.

Using transfer learning will allow less retraining needed for the AI-REM framework in adjusting to different geographic regions. This will assist in adapting to different ecosystems. The combination of explainable AI methods such as SHAP and LIME will aid in fulfilling the desires for model transparency and stakeholder confidence as well as integrating predictive analytics into the governance and policy of the environment. Ultimately, future extensions should predict the effects of climate change on the resiliency of aquatic ecosystems through longitudinal predictive models containing climate variables such as degree-day anomalies, changes in precipitation, and carbon fluxes as well as assessing climate change in the multiple dimensions.

References

Ariunaa, K. and Tudevdagva, U., 2025. Generative Adversarial Network-Based Damage Simulation Model for Reinforced Concrete Structures. *International Academic Journal of Innovative Research*, *12*(2), pp.43-53. https://doi.org/10.71086/IAJIR/V12I2/IAJIR1216

Ávila-Torres, G., Rosiles-González, G., Carrillo-Jovel, V.H., Acosta-González, G., Cejudo-Espinosa, E., Ortega-Camacho, D., Hernández-Zepeda, C. and Valenzuela, O.A.M., 2023. Microcystin concentrations and detection of the mcyA gene in water collected from agricultural, urban, and recreational areas in a karst aquifer in the Yucatan Peninsula of Mexico. *Microbiology Research*, *14*(3), pp.1168-1184. https://doi.org/10.3390/microbiolres14030078

Beversdorf, L.J., Chaston, S.D., Miller, T.R. and McMahon, K.D., 2015. Microcystin mcyA and mcyE gene abundances are not appropriate indicators of microcystin concentrations in lakes. *PloS one*, *10*(5), p.e0125353. https://doi.org/10.1371/journal.pone.0125353

Camara Lins, R., Martinez, J.M., da Motta Marques, D., Almir Cirilo, J. and Ruberto Fragoso, C., 2017. Assessment of Chlorophyll-a Remote Sensing Algorithms in a Productive Tropical Estuarine-Lagoon System. *Remote Sensing*, *9*(6). https://doi.org/10.3390/rs9060516

Chen, L., Liu, L., Liu, S., Shi, Z. and Shi, C., 2025. The application of remote sensing technology in inland water quality monitoring and water environment science: Recent progress and perspectives. *Remote Sensing*, *17*(4), p.667. https://doi.org/10.3390/rs17040667.

Ding, W. and Li, C., 2024. Algal blooms forecasting with hybrid deep learning models from satellite data in the Zhoushan fishery. *Ecological Informatics*, *82*, p.102664. https://doi.org/10.1016/j.ecoinf.2024.102664

Dong, X., Zeng, S., Bai, F., Li, D. and He, M., 2016. Extracellular microcystin prediction based on toxigenic Microcystis detection in a eutrophic lake. *Scientific reports*, *6*(1), p.20886.

Fournier, C., Fernandez-Fernandez, R., Cirés, S., López-Orozco, J.A., Besada-Portas, E. and Quesada, A., 2024. LSTM networks provide efficient cyanobacterial blooms forecasting even with incomplete spatio-temporal data. *Water Research*, *267*, p.122553. https://doi.org/10.1016/j.watres.2024.122553

Gitelson, A.A., Berdnikov, S., Saprygin, V. and Bukata, R. (2017) 'Algorithms for remote estimation of chlorophyll-a in inland and coastal waters: A review', *Remote Sensing of Environment*, 190, pp. 134–151.

Gupta, S. and Gupta, S., 2025. Time series forecasting of chlorophyll-a concentrations in the Chesapeake Bay. *Scientific Reports*, *15*(1), p.30877.

Jasim, A.T., 2024. Investigating the Potential of Remote Sensing for Sustainable Urban Renewal and Regeneration: A Review. *International Academic Journal of Science and Engineering*, *11*(1), pp.54-64. https://doi.org/10.9756/IAJSE/V11I1/IAJSE1108

Jaywant, S.A. and Arif, K.M., 2024. Remote sensing techniques for water quality monitoring: a review. *Sensors*, *24*(24), p.8041. https://doi.org/10.3390/s24248041

Nam, N.N., Do, H.D.K., Loan Trinh, K.T. and Lee, N.Y., 2023. Metagenomics: an effective approach for exploring microbial diversity and functions. *Foods*, *12*(11), p.2140. https://doi.org/10.3390/foods12112140.

Rotthauwe, J.H., Witzel, K.P. and Liesack, W., 1997. The ammonia monooxygenase structural gene amoA as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. *Applied and environmental microbiology*, *63*(12), pp.4704-4712. https://doi.org/10.1128/aem.63.12.4704-4712.1997

Rout, A.K., Rout, S.S., Panda, A., Tripathy, P.S., Kumar, N., Parida, S.N., Dey, S., Dash, S.S., Behera, B.K. and Pandey, P.K., 2025. Potential applications and future prospects of metagenomics in aquatic ecosystems. *Gene*, p.149720. https://doi.org/10.1016/j.gene.2025.149720

Salami, Z.A., Babamuratov, B., Ayyappan, V., Prabakaran, N., Khudayberganov, K. and Singh, C., 2025. Modelling Crop Yield Prediction with Random Forest and Remote Sensing Data. *Natural and Engineering Sciences*, *10*(2), pp.67-78. https://doi.org/10.28978/nesciences.1763843

Short, S.M. and Zehr, J.P., 2005. Quantitative analysis of nifH genes and transcripts from aquatic environments. *Methods in enzymology*, *397*, pp.380-394. https://doi.org/10.1016/S0076-6879(05)97023-7

Srivastava, A., Ko, S.R., Ahn, C.Y., Oh, H.M., Ravi, A.K. and Asthana, R.K., 2016. Microcystin biosynthesis and mcyA expression in geographically distinct Microcystis strains under different nitrogen, phosphorus, and boron regimes. *BioMed Research International*, *2016*(1), p.5985987. https://doi.org/10.1155/2016/5985987

Topp, S.N., Ross, M.R.V., Pavelsky, T.M. and Simard, M. 2023. Explainable deep learning improves understanding of environmental drivers in remote sensing of water quality. *Remote Sensing, 15*(9), 2398.

Xue, Y., Zhang, X., Wang, Y., et al. 2023. Integrating metagenomics and machine learning for algal bloom prediction in eutrophic lakes. *Water Research, 235,* 119792.

Xu, J., Ma, R., Duan, H., Zhang, Y., & Li, F. 2021. Integration of remote sensing and in-situ data for improved water quality retrieval in large inland lakes. *Remote Sensing of Environment, 259*, 112410.

Yussof, F.N., Maan, N. and Md Reba, M.N., 2021. LSTM networks to improve the prediction of harmful algal blooms in the west coast of Sabah. *International Journal of Environmental Research and Public Health*, *18*(14), p.7650. https://doi.org/10.3390/ijerph18147650

Zhang, Y., Wu, Y., Li, X., 2024. A hybrid CNN–LSTM model for predicting chlorophyll-a concentration using multisource remote sensing data. *Environmental Modelling & Software, 173,* 106966.