Journal of Climate Change and Space Science (JCCaSS) 2(2) (2026): 4
https://doi.org/10.17576/jccass.0202.2026.04

Machine Learning and Remote Sensing Technologies for Algae Monitoring and Prediction in Freshwater

Muhammad Sarfraz Ahmad Mangat[a], Teh Sabariah binti Abd Manan[a,b*], Sharifah Sakinah Syed Abd Mutalib[c*], Hiroki Matsuda[d*], Yasuhiro Tawara[d], Rana Muhammad Mubeen Muhsin[a], Zarimah Hanafiah[e,f], Nur Liyana Mohd Kamal[g], Salmia Beddu[g], Daud Mohamad[g], Nur Azwa Muhamad Bashar[h], Wan Hanna Melini Wan Mohtar[f,i], Muhammad Raza Ul Mustafa[j], Mohamed Hasnain Isa[k], Hamidi Abdul Aziz[b], Mohd Suffian Yusoff[b]

[a]*Institute of Tropical Biodiversity and Sustainable Development, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu Darul Iman, Malaysia.*
[b]*School of Civil Engineering, Engineering Campus, Universiti Sains Malaysia, Nibong Tebal 14300, Malaysia*
[c]*Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu Darul Iman, Malaysia.*
[d]*Geosphere Environmental Technology Corp. (GETC), 101-0063 2-1 Awajicho, Chiyoda-ku, Tokyo, Japan*
[e]*Chemical Engineering Department, Faculty of Engineering, Universiti Malaya, 50603 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur.*
[f]*Civil Engineering Department, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor Darul Ehsan, Malaysia.*
[g]*Department of Civil Engineering, Universiti Tenaga Nasional, Jalan Ikram-Uniten, 43000 Kajang, Selangor, Malaysia.*
[h]*Civil Engineering Studies, College of Engineering, Universiti Teknologi MARA, Pulau Pinang Branch, Permatang Pauh Campus, 13500, Pulau Pinang, Malaysia.*
[i]*Environmental Management Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 Selangor Darul Ehsan, Malaysia.*
[j]*Civil and Environmental Engineering Department, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia*
[k]*Civil Engineering Programme, Faculty of Engineering, Universiti Teknologi Brunei, Tungku Highway, Gadong, BE1410, Brunei Darussalam.*

**Corresponding author: tehsabariah@umt.edu.my; s.sakinah@umt.edu.my; h.matsuda@getc.co.jp*

Received: 25 November 2025; Accepted: 22 February 2026; Published: 31 July 2026

ABSTRACT

Harmful algal blooms (HABs) in freshwater systems are increasingly driven by nutrient enrichment and climate change, posing serious ecological, health, and economic risks. Traditional monitoring methods, including field sampling and chlorophyll-a analysis, are labor-intensive, time-consuming, and spatially limited, making real-time assessment challenging. This review synthesizes recent advances in integrating remote sensing and machine learning (ML) for HAB detection, monitoring, and prediction. This review followed a structured review approach. Peer-reviewed studies were retrieved from major scholarly databases using keyword combinations related to HABs, remote sensing, and machine and deep learning, then screened using predefined inclusion criteria and synthesized by categorizing papers by sensor/platform, model type, target variables, and application. We highlight studies using satellites such as MODIS, Sentinel-2, Landsat-8, and Sentinel-3, and ML models including Random Forest, Support Vector Machine, Artificial Neural Networks, Convolutional Neural Networks, and Long Short-Term Memory networks. These approaches allow accurate mapping of chlorophyll-a, phycocyanin, and cyanobacterial toxins, with R^2 values from 0.67 to 0.94 and predictive lead times up to one month. Explainable artificial intelligence (XAI) enhances model transparency and interpretability, supporting regional transferability and informed management decisions. We also review the ecological and human health impacts of HAB toxins, including microcystins, anatoxin-a, cylindrospermopsin, saxitoxins, and brevetoxins. Challenges such as data scarcity, ecological variability, and computational demands are discussed, alongside future directions focusing on hybrid ML models, autonomous sensing, and standardized datasets. Integrating remote sensing and ML offers a scalable, real-time, and interpretable framework for HAB management, contributing to public health protection and sustainable freshwater ecosystem management.

Keywords: Harmful algal blooms; Machine learning; Remote sensing; Explainable Artificial Intelligence; water resource management

1. INTRODUCTION

Algae provide numerous benefits owing to their capacity to produce both oxygen and biomass (Chia et al., 2022) (Shao et al., 2021). Various forms of renewable energy such as raw material and biofuel can be derived from this biomass. Algae can also convert wastewater contaminants into less harmful compounds while simultaneously generating oxygen through photosynthesis, contributing to natural water purification processes (Abd Manan et al., 2022). However, rising water contamination from high levels of nitrogen and phosphorus discharge leads to significant algal growth (Akinnawo, 2023) (Bali & Gueddari, 2019). Seasonal fluctuations, particularly monsoon-driven runoff, significantly elevate organic matter inputs into aquatic systems, increasing total organic carbon (TOC) levels (Mangat, 2024) and further intensifying bloom-favourable conditions. This enrichment of nutrients plays a major role in the formation of HABs (EPA, 2025) that threaten drinking water supplies, biodiversity and ecological sustainability (Amorim & Moura, 2021). Pollution from multiple anthropogenic sources continues to degrade freshwater systems, intensifying the need for enhanced monitoring and mitigation frameworks (Muhsin et al., 2025). Due to HABs, developed countries such as China, Australia, and European Commission have dealt with major challenges in ensuring water quality (J. Wang et al., 2022) (Herath, 1997). In 2014, Lake Erie producing algal bloom implement a three-day drinking water ban in Toledo, Ohio (Ho & Michalak, 2015) Lake Taihu, the largest freshwater lake in China has experienced decades of growing algal blooms driven by eutrophication (X. Yan et al., 2024).

Due to eutrophication, HABs can quickly emerge as thick blooms that obstruct sunlight, deplete oxygen and trigger hypoxic or black water condition that disturb aquatic ecosystem. Cyanobacteria and other algae produce harmful toxins such as hepatotoxins and neurotoxins that may accumulate through the food web. This could lead to serious illnesses in humans and animals such as liver disease, neurological, skin disease and digestive problems (Saha et al., 2022) (*Algal Blooms*, 2025). Climate change contributes to the rise in HAB incidents by affecting water temperature, hydrodynamic mixing and rainfall patterns. These changes alter nutrient fluxes, stimulate phytoplankton growth and are accompanied by more frequent droughts and intense rainfall events (EPA, 2025) (Dai et al., 2023). These complex interactions make it critical to develop timely and efficient algae monitoring systems. Traditional field sampling and lab-based techniques such as microscopic identification of algae and chlorophyll-a analysis are time consuming, labour-intensive and unsuitable for large or real time monitoring.

Modern remote sensing enables high resolution, broad-area assessment of water quality indicators provides critical input for machine learning models to improve forecasting. Satellite-based platforms like Google Earth Engine allow real-time spatial analysis, making them invaluable for dynamic HAB monitoring. Two major modelling frameworks exist: process-based (PB) models, which simulate bloom mechanisms through nutrient cycling and physical interaction using differential equations (W. Wang et al., 2025). In contrast, data-driven (DD) machine learning models utilized environmental variables to learn bloom patterns and generate prediction (Lin et al., 2023). PB models are useful for understanding bloom mechanism but depend heavily on empirical data and calibration. Machine learning (ML) models provide efficient and accurate predictions of bloom dynamics when trained on large, high-quality datasets (Gupta et al., 2023).

ML models minimize dependence on experimental inputs, enhance prediction and support scalable applications. However, model accuracy can vary depending on input data quality and system specific characteristics (N. Gupta et al., 2021). Accurate modelling depends on appropriate input variables and model structure particularly since bloom related factors vary significantly among aquatic systems. Recent study highlights the use of explainable artificial intelligence techniques to improve the interpretability of black box ML models, thereby increasing transparency and trust in environmental monitoring. Recent developments in ML include image classification techniques of identifying algal presence and categorizing bloom intensity, supporting early warning efforts. When paired with remote sensing, ML enables real-time forecasting and alerting of blooms, much like modern weather system to protect public health and aquatic ecosystems (Caballero et al., 2025). This combined approach has become essential for managing the growing risks of HABs under continued nutrient loading and climate change.

Freshwater harmful algal blooms (HABs) are increasing due to nutrient enrichment and climate variability, yet conventional monitoring based on field sampling and laboratory analysis remains time-consuming, costly, and spatially limited, which constrains timely detection and early warning. In this context, this study asks how recent advances in remote sensing and machine learning can be combined to improve HAB detection, monitoring, and short-term forecasting in freshwater systems, and what key limitations must be addressed to enable reliable and transferable operational use. Accordingly, the objectives are to synthesize recent remote-sensing platforms and spectral indicators used for HAB-related water-quality retrieval; compare commonly used machine learning and deep learning models for HAB detection and prediction, including typical inputs and reported performance; examine how explainable AI approaches are being adopted to enhance interpretability and management relevance; and identify current challenges, knowledge gaps, and future research directions toward scalable, real-time HAB monitoring and decision support. The novelty of this paper lies in providing an integrated, management-oriented synthesis that connects sensor/index choices to model development and forecasting use-cases while emphasizing interpretability and transferability which are the elements that are often treated separately in previous reviews.

2. METHODOLOGY

This study adopts a structured narrative review design to synthesize recent advances in the use of remote sensing and machine learning for freshwater harmful algal bloom (HAB) detection, monitoring, and forecasting. The review focuses on peer-reviewed studies that apply satellite-based or airborne remote sensing data in combination with machine learning or deep learning models. An analytical synthesis approach was used to categorize studies based on data sources, modeling techniques, target variables, and application objectives (detection, monitoring, or forecasting), enabling comparative assessment of methodological trends and research gaps.

Relevant literature was collected through searches of major scientific databases, including Web of Science, Scopus, and Google Scholar. The search was conducted using combinations of keywords such as harmful algal blooms, cyanobacteria, chlorophyll-a, remote sensing, satellite imagery, machine learning, deep learning, and forecasting. Searches were limited to English-language journal articles published primarily within the last decade to capture recent methodological developments, while a small number of earlier foundational studies were included where necessary for context.

Studies were included if they (i) focused on freshwater or inland water bodies, (ii) employed remote sensing data as a primary or supporting input, (iii) applied machine learning or deep learning techniques for HAB-related detection, classification, or forecasting, and (iv) reported sufficient methodological detail to allow comparison of inputs, models, and outputs. Studies focusing exclusively on marine systems, purely laboratory-based analyses, or non–data-driven modeling approaches were excluded. Following title and abstract screening, full-text assessment was conducted to confirm eligibility. The final set of selected studies is summarized in Table 1, which provides a comparative overview of data sources, modeling techniques, target variables, and key outcomes used in analytical synthesis.

Table 1 summarises recent advances demonstrating significant progress in harmful algal bloom prediction and monitoring. Remote sensing now provides essential spectral information for aquatic systems. Machine learning models use these multisensory data with high efficiency. Field observations further enhance reliability across varied environments. In Lake Erie, Random Forest delivered strong performance for chlorophyll a. Chlorophyll a and microcystin showed an R^2 of 0.83. Key predictors included Bands 04, 03, 11, 16, and 21. These wavelengths captured bloom signals with high sensitivity. In the Geum River Basin, Faster R CNN produced reliable detection. Precision reached 0 point 60, and recall reached 0 point 69. Radar backscatter improved detection during cloudy weather. This capability addressed a common optical limitation. In the Persian Gulf, an artificial neural network performed strongly. Accuracy reached 88 points 7% with an area under the curve of 0 point 90. Sea surface temperature and salinity variability dominated bloom behavior. The model successfully predicted blooms one month ahead. In Biscayne Bay, Support Vector Machine provided strong results. The model achieved an R squared of 0 point 74. Precipitation and lagged discharge increased bloom likelihood. Wind speed reduced bloom development. In Grahamstown Dam, Self Organizing Maps identified clear optical patterns. Landsat 8 and Sentinel 2 performed the best. Curvature Around Red offered strong cyanobacteria sensitivity. Near Infrared minus Red also performed reliably. Across New York lakes, Extra Trees Regression reached high accuracy. It achieved an R squared of 0 point 72. Root mean squared error was 8 point 19. Land cover and lake morphology improved predictions. In Lake Erie forecasting, ensemble models performed strongly. 10 days predictions reached R squared near 0 point 68. Legacy phosphorus and water level influenced bloom intensity. In the Galician Rias Baixas, the DoME model performed best. It achieved an R squared of 0 point 77. Forecast skill extended 3 days before decline. Deep learning systems such as HABNet showed strong robustness across conditions. These findings confirm the value of integrated remote sensing and machine learning frameworks.

Table 1: Summary of peer-reviewed studies integrating remote sensing and machine learning for freshwater HAB detection, monitoring, and forecasting.

Sr.no	Study Area	Data Source	Model	Input Variables	Findings	Reference
1	Western Lake Erie Basin (North America).	Sentinel-3 OLCI Level-2 satellite data (2016-2021); Ground-based datasets from NOAA GLERL and Ohio Sea Grant Stone Lab.	Random Forest	16 Sentinel-3 spectral bands as predictors; Four water quality proxies Chlorophyll-a (Chl-a), Phycocyanin (PC), Microcystin (MC), and Secchi Depth (SD).	RF model performed best for Chlorophyll-a. Strong linear correlation between Chl-a and MC (R^2=0.83). Bands 04, 03, 11, 16, and 21 were most influential.	(Joshi et al., 2024)
2	Geum River Basin, South Korea.	Sentinel-1 SAR C-band (2020–2022); Sentinel-2 MSI; Environmental data from NIER.	Faster R-CNN (Deep Learning).	VV and VH backscatter (σ^0) values; NDCI index; water temperature, chlorophyll-a, dissolved oxygen.	Model detected algal blooms with 0.60 precision, 0.69 recall, 0.64 F1; detected blooms mainly in summers 2021–2022; SAR proved effective under cloudy conditions.	(Phetanan et al., 2025)
3	Persian Gulf & Gulf of Oman.	MODIS (2004–2023), ERA5 reanalysis, HYCOM salinity.	Artificial Neural Network.	Chl-a, SST, salinity, wind speed, temporal & spatial SDs of SST/salinity, SST gradient.	ANN achieved 88.7% accuracy and AUC 0.90; SST and salinity variability strongly influenced blooms; model predicts next-month HABs using climatological factors.	(Shahmiri et al., 2025)
4	North Biscayne Bay, Miami-Dade County, South Florida.	DERM water quality data (1998–2020), DBHYDRO discharge data, gridMET climate data, NLCD landcover, NEX-GDDP-CMIP6 climate projections.	Support Vector Machine, Random Forest, Multi-Layer Perceptron.	Climate variables (temperature, humidity, wind, radiation, precipitation), discharge, nutrients, turbidity, pH, DO, water temperature, land use.	SVM performed best (R^2=0.74 test). Precipitation & 1-month lag discharge are top positive drivers; current wind speed is strongest negative driver. Wet season shows higher concentrations than dry season. Long-term HAB risk increases.	(Z. Yan & Alamdari, 2025)
5	Grahamstown Dam, New South Wales, Australia.	Satellite imagery (Landsat 8, Sentinel-2, Planetscope); field cyanobacterial cell density data (2015–2020).	Self-Organizing Map (SOM).	Spectral reflectance bands (Blue, Green, Red, NIR, SWIR); cyanobacterial cell densities.	Landsat 8 and Sentinel-2 performed best for detecting harmful algal blooms (HABs); spatial resolution had little impact; best empirical algorithms were “Curvature Around Red (CAR)” and “NIR minus Red (NmR)”; Planetscope data less reliable due to calibration issues.	(Liu et al., 2022)

6	Lakes across New York state, USA.	Landsat-8 (2013–2023), Landsat-9 (2021–2023), Sentinel-2 (2019–2023), LAGOS-NE dataset, USGS, CSLAP, Jefferson Project buoy data.	Extra Trees Regression, Random Forest, Gradient Boosted Regression, and Support Vector Regression.	Spectral bands (visible, NIR, SWIR) from Landsat and Sentinel; lake morphology (surface area); watershed land cover.	ETR achieved $R^2 = 0.72$, RMSE = 8.19 µg/L, outperforming all other models. Combining Landsat & Sentinel imagery with non-optical variables improved prediction accuracy. The model accurately mapped temporal-spatial patterns of algal bloom (Chl-a) and captured peak bloom periods.	(Greene et al., 2025)
7	Lake Erie (Western Basin), North America.	Remote sensing data (MERIS and Sentinel-3 satellites); Hydrometeorological and nutrient data from Maumee, Sandusky, Raisin, and Cuyahoga Rivers; NOAA and EPA datasets.	LASSO (Least Absolute Shrinkage and Selection Operator); Random Forest, Artificial Neural Network.	Cyanobacterial Index (CI), Wind speed, Air temperature, Water level, Solar radiation, Streamflow, Secchi depth, Total Phosphorus (TP), Total Kjeldahl Nitrogen (TKN), Legacy nutrients (TP discharged over past 10 years), Observation time step (10-day intervals).	RF and Ensemble models outperformed LASSO and ANN; best forecast accuracy for 10-day lead time ($R^2 \approx 0.68$); legacy nutrients and water level were important predictors; solar radiation and temperature had weaker influence; uncertainty mainly due to satellite CI errors and limited in-situ data.	(A. Gupta et al., 2023)
8	Galician Rías Baixas, Northwestern Spain.	Biological data (Dinophysis acuminata cell counts) from INTECMAR; Chlorophyll-a from Copernicus satellite data Physical variables (temperature, salinity, velocity, BV frequency, etc.) from CROCO 3D ocean hydrodynamic model; Upwelling Index (NCEP).	Random Forest, Multi-Layer Perceptron, Support Vector Regressor, k-Nearest Neighbors, Hoeffding Tree Regressor, Hoeffding Adaptive Tree Regressor.	Biological: *D. acuminata* cell counts, Chlorophyll-a; Physical: temperature, salinity, velocity components, BV frequency, Upwelling Index, depth, month (sin transformed); Derived: mean and std. dev. across layers and sections.	DoME (Batch Learning) achieved best performance with $R^2 = 0.77$, MAE ≈ 87.78, RMSE ≈ 187.6 for 3-day-ahead predictions; RF followed with $R^2 = 0.67$; Stream Learning models (HATR, HTR) achieved $R^2 \approx 0.4$; DoME provided explainable, equation-based predictions; Upwelling events limited forecast horizon (~3 days).	(Molares-Ulloa et al., 2025)

9	Lake Erie	MODIS satellite imagery.	HABNet (Deep Learning CNN).	Remote-sensing reflectance values from MODIS ocean color bands.	HABNet accurately detects HABs; outperformed statistical and threshold-based methods; shows high robustness to spatial–temporal variability.	(Paul R. Hill et al., 2020)
10	Aqaba Gulf, Red Sea.	Sentinel-2A and Sentinel-3 satellite imagery, remote sensing indices (FAI, RI, NDVI, NDWI, etc.), USGS Earth Explorer, ESRI LU/LC datasets.	Multilayer Perceptron, Neural Network, Random Forest, Partial Least Squares Regression.	Remote sensing indices: FAI, MPH, NDVI, NDWI, NDSII, SWM, SABI, NDCI, VNRI, RI; spectral bands (B2–B11).	MLP and PLSR effectively predicted petroleum pollution; MLP achieved $R^2 = 0.941$, outperforming PLSR. RF model improved calibration of FAI using RI ($R^2 = 0.95$ training, 0.91 testing). FAI identified as best algal bloom indicator ($R^2 = 0.998$). Integration of RS and ML provides accurate estimation of algal blooms and oil contamination.	(Abd El-Hamid et al., 2022)

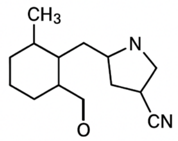
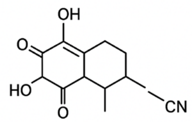
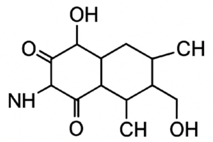
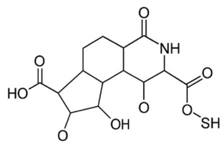
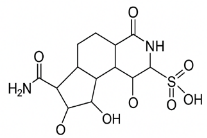
3. DRIVERS AND IMPACTS OF HARMFUL ALGAL BLOOMS (HABs)

HABs in freshwater ecosystems are increasingly widespread, posing serious threats to water quality, public health and economic stability. The excessive presence of nitrogen and phosphorus from anthropogenic sources such as agriculture, domestic wastewater and industry are the primary drivers (Goel & Motlagh, 2013). The addition of these nutrients accelerates the eutrophication process, promoting a spike primary production that favors algal bloom development. Climate change is increasingly seen as a compounding stressor that accelerates frequency and intensity of HABs. Global warming supports heat tolerant cyanobacteria species and shift in rainfall patterns drive greater nutrient loading and thermal stratification collectively intensifying bloom formation (Wells et al., 2020). Although algae are fundamental to aquatic productivity and oxygenation, overgrowth leads to ecological imbalance and functional degradation (Ahmad et al., 2024). HABs pose a significant threat due to their ability to deplete dissolved oxygen, resulting in hypoxic or anoxic condition. The excessive growth and breakdown of algae enhance microbial respiration which consumes available oxygen and create threaten sensitive aquatic species like fish and benthic invertebrates (Lan et al., 2024).

Severe hypoxia turns to black water events which are distinguished by large scale fish mortality and the death of vulnerable species (Small et al., 2014). The risk of ecological degradation is heightened in shallow, nutrient-rich lakes due to their sensitivity to internal phosphorus loading and sediment resuspension (Du et al., 2022) (Y. Wang et al., 2021). Equally concerning is the production of toxins by certain cyanobacterial species, such as Microcystis, Anabaena and Planktothrix. These organisms are capable of synthesizing hepatotoxins (e.g., microcystins), neurotoxins (e.g., anatoxin-a) and cytotoxins that can pose serious health threats to both humans and wildlife (David Kuo & Xagoraraki, 2016). Human health impacts highlight the dangers of exposure in the United States, Microcystis blooms in the Ohio and Potomac Rivers caused gastrointestinal illness in 5,000–8,000 individuals consuming contaminated water (Chorus et al., 2000). In 1979, a Cylindrospermopsis raciborskii bloom in Palm Island, Australia, led to severe illness in over 100 Aboriginal community members (Griffiths & Saker, 2003). In 1993 correlation between cyanobacteria-contaminated drinking water and liver cancer in China (YU, 1995). These cases underline the significant human health risks posed by cyanotoxins.

HABs impose significant economic cost on industries such as aquaculture, tourism and public water treatment with annual damages in U.S alone estimated at $49 million (A National Assessment of Harmful Algal Blooms in US Waters, 2000). Countries such as China and Australia also face significant financial pressures, as widespread eutrophication has driven the need for extensive environmental management efforts. In table 2 impact of HABs explained that produce toxins that pose serious health risks. Anatoxin a causes rapid respiratory paralysis along with drowsiness and muscle pain. This toxin also leads to speech disturbances and excessive salivation. Brevetoxin exposure produces asthma like symptoms with coughing and chest tightness. Many individuals also experience shortness of breath after brevetoxin exposure. Microcystins cause nausea, vomiting, and abdominal discomfort in exposed populations. Severe exposure can result in significant liver inflammation. Cylindrospermopsin produces gastrointestinal illness with fever and headache in many cases. It also causes hepatotoxic effects and can damage kidney tissues. Saxitoxins are associated with paralytic shellfish poisoning. Affected individuals experience tingling, ataxia, and respiratory muscle paralysis. These toxins represent major public health concerns in bloom affected waters.

Table 2: Impact of HABs on human health.

Toxins	Structure	Effect	Reference
Anatoxin-a		Respiratory paralysis, drowsiness, speech disturbances, excessive salivation, and muscle pain	(Backer et al., 2025)
Brevetoxin		Asthma-like symptoms, coughing, shortness of breath, and chest tightness.	(Morris, 2014)
Microcystins		Nausea, vomiting, abdominal pain, diarrhea, severe acute exposure can cause liver inflammation	(Brad Kelechava, 2018)
Cylindrospermopsin (CYN)		Gastrointestinal illness (nausea, vomiting, diarrhea), fever, headache, hepatotoxicity and potential kidney damage	(EPA, 2015)
Saxitoxins		Paralytic Shellfish Poisoning; Tingling, ataxia, respiratory muscle paralysis	(Gad, 2014)

4. TRADITIONAL METHODS FOR ALGAE MONITORING AND THEIR LIMITATIONS

Conventional monitoring of HABs in freshwater system depends heavily on in-situ sampling and laboratory-based analysis, which are time consuming, labour-intensive and require complex sample preparation and pre-enrichment steps (Chin Chwan Chuong et al., 2022). These techniques include microscopic identification of phytoplankton species, quantification of algal pigments. Microscopy is highly accurate for species identification but involve lengthy processing and requires specialized taxonomic expertise (McQuatters-Gollop et al., 2017). Standard sampling protocols generally include collecting surface and subsurface water at regular intervals, followed by taxonomic identification in laboratory using light or fluorescence microscopy. Although necessary for identifying toxic versus non-toxic cyanobacteria, the technique lacks the speed required for real time forecasting or early warning application.

Analytical measurement of chlorophyll-a, the primary pigment responsible for photosynthesis in algae is widely employed as an indirect indicator of algal biomass (García–Nieto et al., 2024). Pigment concentration can be accurately measured using spectrophotometric, fluorometric and high-performance liquid chromatography (HPLC) techniques (Pinckney et al., 1994). These techniques however require physical sampling, preservation and extraction of pigments which add more time and cost to efforts in monitoring. Furthermore, chlorophyll-a act as common type of algae without labelling as toxin species. To complement these methods, mechanistic models have been developed to simulate bloom dynamics based on physical, chemical, and biological variables. These process-based (PB) models use differential equations to represent interactions among hydrodynamics, nutrient cycling, and phytoplankton growth (W. Wang et al., 2025). While PB models can offer deep insights into system behaviour, they require intensive data input, precise calibration, and site-specific parameterization (Pethybridge et al., 2019) (Sungmin et al., 2020).

Traditional method's performance is often limited when applied to ecosystems with high spatial or temporal variability. Another major limitation of them is the infrequent and spatially sparse sampling as most water quality programs rely on weekly or monthly intervals that can miss rapid blooms development (Rome et al., 2021). Moreover, due to logistical barriers, monitoring efforts are often limited to accessible regions, overlooking vast areas within lakes or reservoirs. This leads to spatial bias and potentially underrepresents bloom extent or severity. Traditional monitoring methods are further limited by operational constraints, rendering them inadequate for addressing the serious challenges posed by climate change. As climate-driven shifts make bloom dynamics more unpredictable the demand for frequent and spatially comprehensive monitoring becomes increasingly critical. These emerging challenges are exposing the limitation of conventional monitoring methods.

5. REMOTE SENSING TECHNOLOGIES FOR HAB DETECTION

The increasing spatial and temporal variability of HABs in freshwater systems has highlighted the limitations of ground-based monitoring and preferred the adoption of remote sensing technologies. Satellite-based remote sensing offers a cost-effective solution for large-scale, high-frequency monitoring of algal bloom indicators such as chlorophyll-a, turbidity, surface reflectance and water temperature (Janatian et al., 2025). Remote sensing complements ground-based approaches effectively and is now central to real-time detection and (S. Wang & Qin, 2025) Bs events (S. Wang & Qin, 2025).These systems detect the unique spectral patterns of phytoplankton and other light-interacting substances in surface water. Algal blooms often chang water colour because of pigments like chlorophyll-a and phycocyanin which absorb strongly in the blue and red spectra while reflecting green and near-infrared wavelengths (Zhao et al., 2020). Environmental assessments further emphasize that satellite-derived observations provide robust insights into surface-level ecological conditions, strengthening the reliability of remote sensing for large-scale environmental monitoring (Naqvi, 2025). Changes in spectral reflectance facilitates the use of specialized indices, including NDCI, FAI, and MCI to monitor and detect algal blooms in freshwater environments.

Satellite systems including MODIS, Sentinel-2 and Landsat-8 deliver consistent and scalable coverage of freshwater ecosystems. Each satellite offers unique spatial, spectral and temporal capabilities enabling targeted approaches to water quality monitoring strategies. For example, the landsat-8 OLI (30m) and sentinel-2 MSI (10-20 m) sensors provide sufficient spatial resolution to monitor localized bloom events in smaller lakes, which are often beyond the mapping capacity of traditional monitoring (Mamun et al., 2024). Remote sensing offers a major advantage by enabling the analysis of bloom behaviour over spatial and temporal across multiple water systems. These datasets can be incorporated into monitoring frameworks to support time series analysis, anomaly detection and change detection. Remote sensing also enables quantitative assessment by facilitating pigment retrieval and algal biomass estimation. Estimating chlorophyll-a and phycocyanin from reflectance data is enabled by a suite of developed empirical and semi-analytical algorithms (Pyo et al., 2017) (Lundberg & Lee, 2017). Despite its many strengths, remote sensing is not without limitations. Image quality can be compromised by atmospheric interference, cloud cover and sunlight while optical sensors are limited to surface level observation due to their shallow penetration.

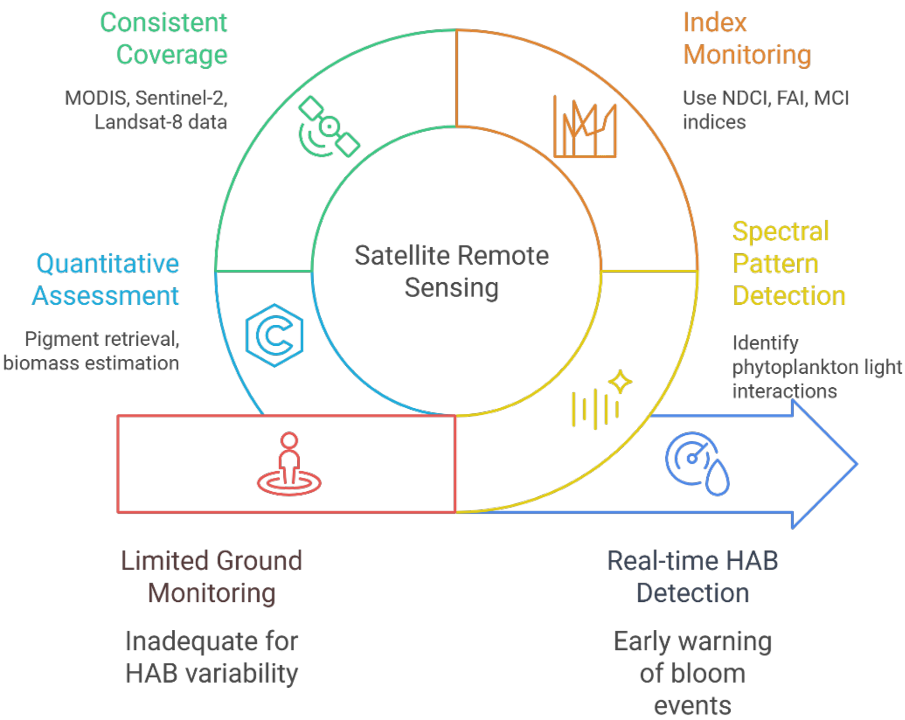


Figure 1. Remote Sensing for HABs monitoring.

6. MACHINE LEARNING FOR HAB PREDICTION AND MONITORING

Machine learning has recently gained prominence as a powerful tool for detecting, classifying and predicting HABs in freshwater systems. Traditional modelling methods simulate basic system response, but they are limited in capturing the complex, non-linear process that influence bloom development. In contrast, ML algorithms can learn directly from data without relying on predefined ecological assumptions, making it particularly suitable for HABs forecasting (Cho et al., 2014). Machine learning algorithms can be broadly categorized into supervised learning. Forecasting HABs, researcher commonly employ supervised learning algorithms such as ANN, decision trees, SVM and various ensemble techniques.

ANN: $z = \sum_{i=1}^{n} w_i + x_i + b$ (Lee, 2025) (1)

Where,

x_i = input feature.
w_i = weight.

b = bias.
z = weighted sum (linear transformation).
σ = activation function (nonlinear transformation).
a = output.

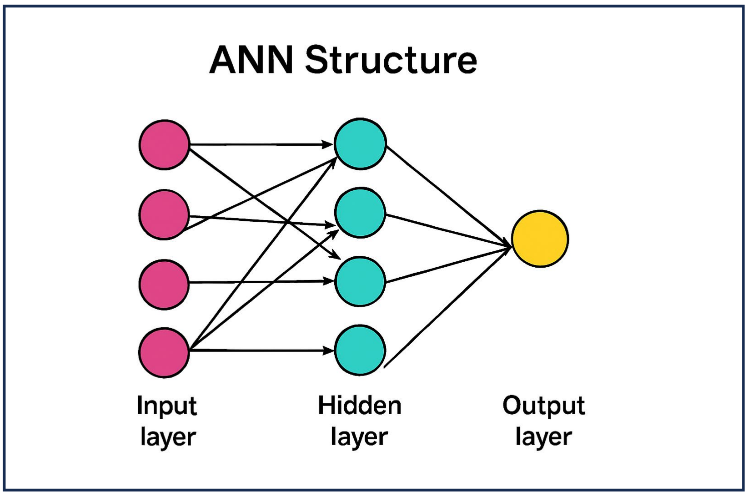


Figure 2. ANN structural diagram.

Decision tree: $f(x) = \sum_{m=1}^{M} c_m \, I\{X \in R_m\}$ (Strobl et al., 2008) (2)

Where,

M =represents the total number of leaf (terminal) nodes in the decision tree.
R_m=denotes the region or subset of the feature space associated with the m-th leaf node.
C_m=is the predicted output for all data points that fall within region R_m.
$I(.)$ = is an indicator function that returns 1 if $x \in R_m$, and 0 otherwise.

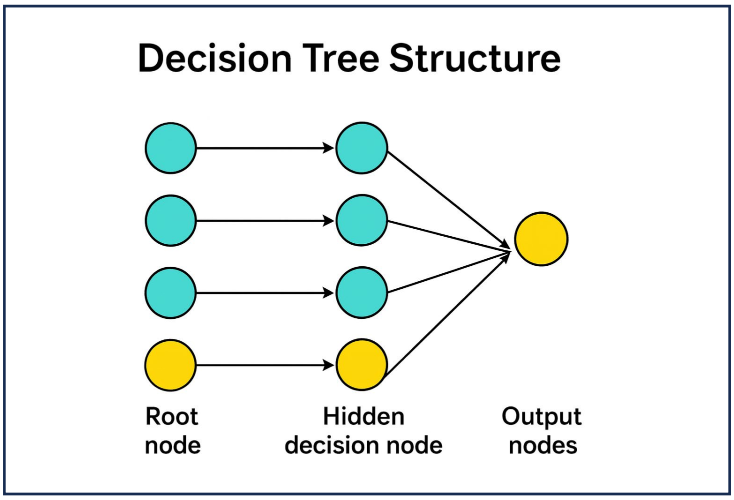


Figure 3. Decision Tree structural diagram.

SVM: $\mathrm{wx} + \mathrm{b} = 0$ (GeeksforGeeks, 2025) (3)

Where,
w = is the weight vector, representing the normal to the separating hyperplane.
x = is the input feature vector.
b = is the bias term (or intercept).

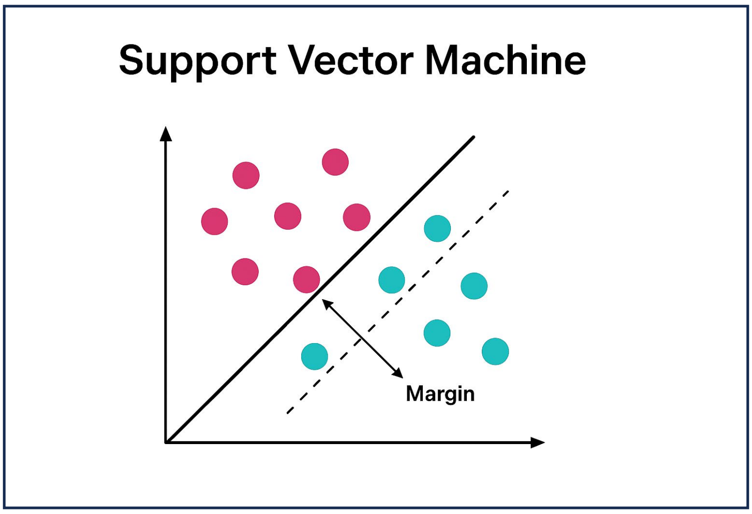


Figure 4. Support Vector Machine (SVM) structural diagram.

These models learn relationship between input variable and target outputs from historical data (Shu & Ye, 2023). When adequately trained, they deliver accurate predictions under a range of environmental conditions. A major advantage of ML based forecasting is its capacity to incorporate diverse data sources such as satellite imagery, meteorological data and hydrological records to proves a holistic view of bloom dynamics. Deep learning models convolutional neural network (CNN) and long short-term memory (LSTM) networks are increasingly being applied for HAB prediction and classification tasks.

CNN: $Z_k^{(l)} = \sum_{d=1}^{D_l-1} \sum_{m=0}^{h_l-1} \sum_{n-0}^{w_l-1} W_{k,d}^{(l)}(m,n) X_d^{(l-1)}(i+m, j+n) + b_k^{(l)} A_k^{(l)}(i,j) = \emptyset\,(Z_k^{(l)}(i,j))$ (LeCun et al., 1998) (4)

Where,

$X^{(l-1)}$= input feature map.
$W_{k,d}^{(l)}$= convolution filter for output channel $k$and input channel d.
$b_k^{(l)}$= bias for output channel k.
$Z_k^{(l)}$= pre-activation feature map.
$A_k^{(l)}$= output after activation.
$\emptyset$ = activation function (e.g., ReLU).

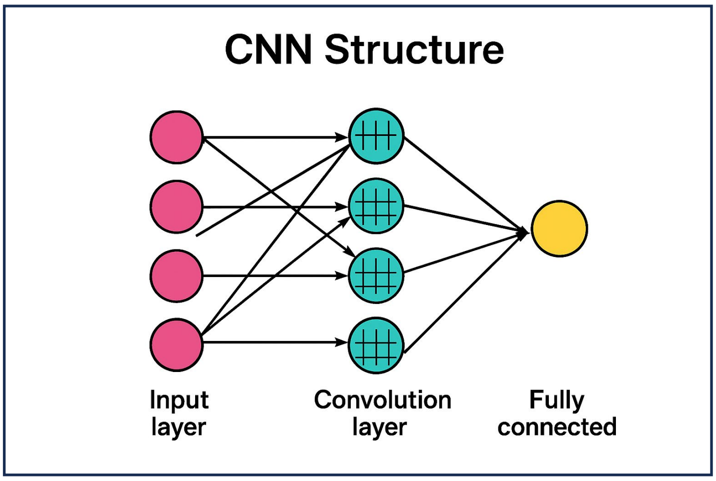


Figure 5. CNN structural diagram.

LSTM: $$i_t = \sigma\,(w_i[h_{t-1}, x_t] + b_i) \quad \text{(Thakur, 2018)} \qquad (5)$$

$$f_t = \sigma\left(w_f[h_{t-1}, x_t] + b_i\right)$$

$$o_t = \sigma\,(w_o[h_{t-1}, x_t] + b_o)$$

where,

i_t= input gate.
f_t= forget gate.
o_t= output gate.
σ = sigmoid activation function.
w_x= weight matrix for gate x.
h_{t-1}= hidden state from the previous LSTM block (at time $t-1$).
x_t= input at the current timestep.
b_x= bias term for gate x.

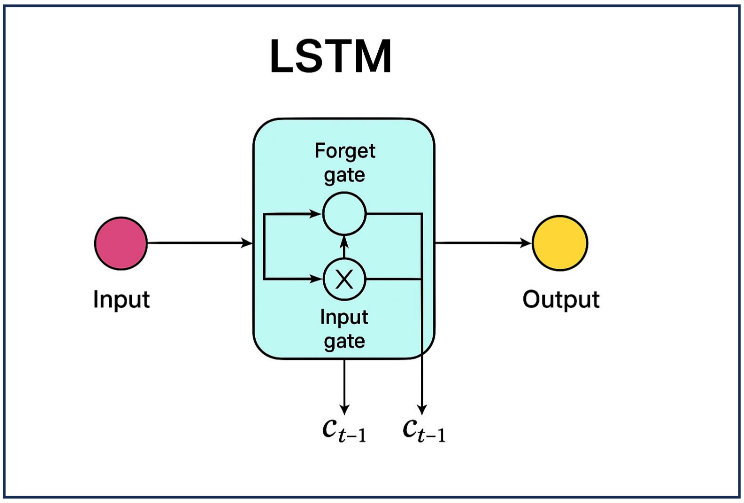


Figure 6. LSTM structural diagram.

Thes models offer advantages in feature extraction and temporal pattern recognition but require large volumes of high-quality labelled data and significant computational resources. ML performance is strongly influenced by data integrity including completeness, quality and preprocessing steps. Inadequate handling of missing data, imbalanced classes may unable predictive accuracy and transferability (Maharana et al., 2022). Feature selection plays a critical role in model optimization; the inclusion of irrelevant or collinear variables can degrade performance and increase computational cost. Despite existing limitations, ML has shown significant promise in enhancing HAB monitoring through early warning system, improved sampling strategies and informed adaptive management. As freshwater systems face rising challenges from climate and human induced change, ML driven models are emerging as critical tools for effective HAB prediction and adaptive management.

7. EXPLAINABLE AI AND MODEL INTERPRETABILITY IN HAB RESEARCH

Despite high predictive accuracy of ML in HAB modelling, their black-box nature hinders their integration into real world environmental management practices. Many high-performing models, especially deep learning frameworks such as artificial neural network (ANN) and convolutional neural network (CNN) function as "Black-box" delivering accurate prediction without revealing how individual input variable influence outcomes (Vilone & Longo, 2020). The absence of interpretability undermines trust, restricts regulatory approval and impedes scientific insight especially in ecological sensitive settings where transparent and accountable decision making is essential. To overcome these limitations, the integration of XAI has gained significant traction. Through various techniques, XAI sheds lights on the predictive behaviour of complex ML models, clarifying the influence of input variables on outcomes (Sadeghi et al., 2024). Such approaches are especially important in HAB monitoring where uncovering casual links between environmental factors and bloom occurrence is critical for effective mitigation and management.

To enhance model interpretability, XAI techniques like SHAP and Lime are used to determine the relative importance of environmental predictors. These insights not only enhance trust in model prediction but also enable more precise and effective management interventions. XAI also aid model transferability (Ali et al., 2023). Comparing dominant predictive features across regions allows researcher to evaluate the transferability of model between different ecosystems. This enhances model generalizability while informing region specific adjustment. Explainable model outputs also help close the communication gap between technical experts and environmental decision-makers. Understanding how and why a model predicts a bloom boosts decision-maker confidence and action. In comparison, unexplained outcomes maybe questioned or dismissed. However, challenges remain. Not all XAI tools work equally well across model types and poorly designed explanations can mislead users. Further research is required to standardized interpretability techniques and assess their applicability in ecological modelling.

8. DISCUSSION

The findings synthesized in this review indicate a clear shift in HAB research from traditional empirical and threshold-based approaches toward data-driven frameworks that integrate remote sensing and machine learning for detection, monitoring, and forecasting. From a conceptual perspective, this trend reflects a broader transition in environmental monitoring toward predictive and decision-support paradigms, where spatiotemporal generalization and early warning are prioritized over site-specific diagnostics. Compared with earlier reviews that focused primarily on spectral index development or algorithm performance in isolation, the present synthesis highlights how model effectiveness is strongly conditioned by sensor characteristics, feature selection strategies, and the temporal structure of input data. Consistent with previous studies, machine learning and deep learning models generally outperform conventional statistical approaches in capturing non-linear HAB dynamics; however, our analysis further reveals that issues of interpretability, transferability, and data imbalance remain under-addressed. By explicitly examining these limitations across studies summarized in Table 1, this review advances prior work by framing performance not only in terms of accuracy, but also in terms of operational reliability and management relevance.

ML has shown increasing success in predicting HABs posing significant challenges that impact its performance, scalability and reliability. A major challenge lies in availability and consistency of monitoring data. In many freshwater systems lack adequate toxin specific data and exhibit irregular sampling in both spatial and temporal, undermining model training and predictive accuracy (Zhu et al., 2022). Another key challenge lies in ecological heterogeneity across freshwater systems. Key drivers of bloom development differ substantially across systems, depending on parameters like lake depth, stratification patterns and sediment behaviour (Ding & Qin, 2024).

These variations reduce the generalizability of models and necessitate site-specific input feature and parameter tuning. Although deep neural networks and other advanced ML models provide high predictive accuracy but require large, labelled datasets and offer limited interpretability. XAI technique like SHAP offer valuable interpretability by highlighting the most influential feature driving model predictions (Demiray et al., 2025); however, there remains a need for standardized interpretability tools that are specified tailored to ecological and bloom forecasting.

The implications of these findings are significant for both future research and water resource management. For researchers, the results underscore the need to move beyond model-centric optimization toward integrated frameworks that combine multi-sensor remote sensing, temporal modeling, and explainable AI to improve robustness and stakeholder trust. Future studies should prioritize standardized benchmarking datasets, cross-region validation, and the incorporation of domain knowledge to enhance model transferability across freshwater systems. From a policy and management perspective, the growing evidence supporting remote sensing and machine learning integration suggests strong potential for real-time HAB early warning systems that can inform proactive interventions, such as nutrient management, reservoir operation, and public health advisories. However, realizing this potential will require closer collaboration between scientists, water authorities, and policymakers to ensure that model outputs are interpretable, timely, and aligned with regulatory decision-making processes. Future efforts can also focus on creating hybrid models that combine ecological theory with machine learning and expanding autonomous sensing tools for real time water quality monitoring. These advancements will be crucial for building flexible, transparent and regionally adaptable HAB prediction tools.

9. Conclusion

This review synthesizes evidence that integrating remote sensing with machine learning has become a practical pathway for freshwater HAB detection and short-term forecasting, particularly when satellite-derived indicators (e.g., chlorophyll-a, phycocyanin and surface reflectance) are combined with hydrometeorological and catchment variables. Across the studies reviewed, supervised machine learning and deep learning approaches (e.g., Random Forest, SVM, ANN/CNN and LSTM) consistently support almost real-time bloom mapping and forecasting, with reported predictive performance commonly in the range of R^2 approximately between 0.67 and 0.94 and forecast horizons extending to approximately one month, depending on data availability and system dynamics. Importantly, the literature indicates that performance is not driven by algorithms alone but by data quality, temporal structure, feature selection, and transferability across water bodies; in parallel, interpretability remains a bottleneck for many deep models, limiting scientific insight and operational trust. Practically, these findings support the development of operational early warning systems for water agencies by enabling routine risk screening, targeted field sampling, faster public health advisories, and more proactive interventions such as nutrient-control prioritization and reservoir management during high-risk periods. To strengthen reproducibility and real-world deployment, future research should move toward (i) standardized benchmark datasets and reporting protocols (including clear validation across regions and seasons), (ii) hybrid modeling that combines process knowledge with data-driven learning, (iii) uncertainty quantification and explainable AI to communicate drivers and confidence to decision-makers, and (iv) scalable monitoring pipelines that fuse multi-sensor satellite products with in-situ/IoT observations for continuous updating. Overall, remote sensing and machine learning integration is positioned as a deployable, management-oriented tool to reduce HAB impacts on ecosystems, water security, and public health under continuing nutrient enrichment and climate pressure.

10. REFERENCES

A National Assessment of Harmful Algal Blooms in US Waters. (2000). Https://Www.Iatp.Org/Sites/Default/Files/National_Assessment_of_Harmful_Algal_Blooms_in.Htm.

Abd El-Hamid, H. T., Reda, R., Zarzoura, F., Hafiz, M. A., & El-Alfy, M. A. (2022). Integrated approach of remote sensing and machine learning to simulate and predict petroleum pollution and algal blooms along Aqaba Gulf. *Biocatalysis and Agricultural Biotechnology*, 46, 102528. https://doi.org/10.1016/J.BCAB.2022.102528

Abd Manan, T. S. B., Khan, T., Wan Mohtar, W. H. M., Machmudah, A., Dutykh, D., Qazi, S., Ahmad, A., & Wan Rasdi, N. (2022). Bioremediation of wastewater using algae for potential renewable bioenergy cogeneration. *Algal Biotechnology: Integrated Algal Engineering for Bioenergy, Bioremediation, and Biomedical Applications*, 47–62. Elsevier. https://doi.org/10.1016/B978-0-323-90476-6.00019-4

Ahmad, N. B., Jaafaru, M. S., Isa, Z., Abdulhamid, Y., Kakudi, R. A., Ugya, A. Y., & Meguellati, K. (2024). High pollution loads engineer oxygen dynamics, ecological niches, and pathogenicity shifts in freshwater environments. *Journal of Hazardous Materials Advances*, 14, 100425, https://doi.org/10.1016/j.hazadv.2024.100425.

Akinnawo, S. O. (2023). Eutrophication: Causes, consequences, physical, chemical and biological techniques for mitigation strategies. *Environmental Challenges*, 12, 100733, https://doi.org/10.1016/j.envc.2023.100733.

Algal Blooms. (2025). Https://Www.Niehs.Nih.Gov/Health/Topics/Agents/Algal-Blooms.

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99, 101805, https://doi.org/10.1016/j.inffus.2023.101805

Amorim, C. A., & Moura, A. do N. (2021). Ecological impacts of freshwater algal blooms on water quality, plankton biodiversity, structure, and ecosystem functioning. *Science of the Total Environment*, 758, 143605, https://doi.org/10.1016/j.scitotenv.2020.143605

Backer, L. C., McNeel, S. V., Barber, T., Kirkpatrick, B., Williams, C., Irvin, M., Zhou, Y., Johnson, T. B., Nierenberg, K., Aubel, M., LePrell, R., Chapman, A., Foss, A., Corum, S., Hill, V. R., Kieszak, S. M., & Cheng, Y. S. (2025). Recreational exposure to microcystins during algal blooms in two California lakes. *Toxicon*, 55(5), 909–921. https://doi.org/10.1016/j.toxicon.2009.07.006

Bali, M., & Gueddari, M. (2019). Removal of phosphorus from secondary effluents using infiltration–percolation process. *Applied Water Science*, 9, 54. https://doi.org/10.1007/s13201-019-0945-5

Brad Kelechava. (2018). *Microcystins: A Toxic Threat to Drinking Water*. The ANSI Blog. https://blog.ansi.org/ansi/microcystins-cyanobacteria-drinking-water/

Caballero, C. B., Martins, V. S., Paulino, R. S., Butler, E., Sparks, E., Lima, T. M., & Novo, E. M. L. M. (2025). The need for advancing algal bloom forecasting using remote sensing and modeling: Progress and future directions. *Ecological Indicators*, 172, 113244. https://doi.org/10.1016/j.ecolind.2025.113244

Chia, S. R., Nomanbhay, S. B. H. M., Chew, K. W., Munawaroh, H. S. H., Shamsuddin, A. H., & Show, P. L. (2022). Algae as potential feedstock for various bioenergy production. *Chemosphere*, 287(1), 131944. https://doi.org/10.1016/j.chemosphere.2021.131944

Chin Chwan Chuong, J. J., Rahman, M., Ibrahim, N., Heng, L. Y., Tan, L. L., & Ahmad, A. (2022). Harmful Microalgae Detection: Biosensors versus Some Conventional Methods. *Sensors*, 22(9), 3144. https://doi.org/10.3390/s22093144

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Proceedings of the 2104 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–1734, Doha, Qatar. https://doi.org/10.3115/v1/d14-1179

Dai, Y., Yang, S., Zhao, D., Hu, C., Xu, W., Anderson, D. M., Li, Y., Song, X. P., Boyce, D. G., Gibson, L., Zheng, C., & Feng, L. (2023). Coastal phytoplankton blooms expand and intensify in the 21st century. *Nature*, 615, 280–284. https://doi.org/10.1038/s41586-023-05760-y

David Kuo, H. W., & Xagoraraki, I. (2016). Contaminants Associated with Drinking Water. In S.R. Quah (Ed), International encyclopedia of public health (2nd ed, Vol.2, pp. 148 – 158. Elsevier. https://doi.org/10.1016/B978-0-12-803678-5.00489-6

Demiray, B. Z., Mermer, O., Baydaroğlu, Ö., & Demir, I. (2025). Predicting Harmful Algal Blooms Using Explainable Deep Learning Models: A Comparative Study. *Water*, 17(5), 676. https://doi.org/10.31223/X58D9F

Ding, K., & Qin, B. (2024). Global Drivers of Algal Bloom Occurrence: The Role of Climate, Hydrology, and Morphometry. *Water,* 16(24), 3624. https://doi.org/10.3390/w16243624

Du, Y. X., An, S. L., He, H., Wen, S. L., Xing, P., & Duan, H. T. (2022). Production and transformation of organic matter driven by algal blooms in a shallow lake: Role of sediments. *Water Research*, 219, 118560. https://doi.org/10.1016/j.watres.2022.118560

EPA. (2015). *Health Effects Support Document for the Cyanobacterial Toxin Cylindrospermopsin*, U.S. Environmental Protection Agency.

EPA, U. (2025). *Climate Change and Freshwater Harmful Algal Blooms.* Https://Www.Epa.Gov/Habs/Climate-Change-and-Freshwater-Harmful-Algal-Blooms.

Gad, S. E. (2014). Saxitoxin. In *Encyclopedia of Toxicology* (3rd ed, pp. 218-220). Elsevier. https://doi.org/10.1016/B978-0-12-386454-3.00925-8

García–Nieto, P. J., García–Gonzalo, E., Alonso Fernández, J. R., & Díaz Muñiz, C. (2024). Forecast of chlorophyll-a concentration as an indicator of phytoplankton biomass in El Val reservoir by utilizing various machine learning techniques: A case study in Ebro river basin, Spain. *Journal of Hydrology*, 639, 131639. https://doi.org/10.1016/j.jhydrol.2024.131639

GeeksforGeeks. (2025). *Support Vector Machine (SVM) Algorithm.* https://www.geeksforgeeks.org/machine-learning/support-vector-machine-algorithm/

Goel, R. K., & Motlagh, A. M. (2013). Biological Phosphorus Removal. In *Comprehensive Water Quality and Purification* (Vol. 3, pp. 150–162). Elsevier. https://doi.org/10.1016/B978-0-12-382182-9.00051-7

Greene, J. A., Metlitsky, L., Levine, A., Foley, E., Henry, M., Azarderakhsh, M., Blake, R. A., & Norouzi, H. (2025). A new perspective on estimating Chlorophyll-a concentrations using machine learning and remote sensing: a case study of New York state lakes. *Ecological Indicators*, 180, 114316. https://doi.org/10.1016/J.ECOLIND.2025.114316

Gupta, A., Hantush, M. M., & Govindaraju, R. S. (2023). Sub-monthly time scale forecasting of harmful algal blooms intensity in Lake Erie using remote sensing and machine learning. *Science of the Total Environment*, 900, 165781. https://doi.org/10.1016/j.scitotenv.2023.165781

Gupta, N., Patel, H., Afzal, S., Panwar, N., Mittal, R. S., Guttula, S., Jain, A., Nagalapatti, L., Mehta, S., Hans, S., Lohia, P., Aggarwal, A., & Saha, D. (2021). Data Quality Toolkit: Automatic assessment of data quality and remediation for machine learning datasets. arXiv preprint arVXiv:2018.05935.

Herath, G. (1997). Freshwater algal blooms and their control: Comparison of the European and Australian experience. *Journal of Environmental Management*, 51 (2), 217–227. https://doi.org/10.1006/jema.1997.0138

Ho, J. C., & Michalak, A. M. (2015). Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie. *Journal of Great Lakes Research*, 41 (2), 317–32. https://doi.org/10.1016/j.jglr.2015.01.001

Janatian, N., Raudsepp, U., Broomandi, P., Fickas, K., Olli, K., Heimovaara, T., Mannik, A., Uiboupin, R., & Pahlevan, N. (2025). A review on remote-sensing-based harmful cyanobacterial bloom monitoring services. In *Remote Sensing Applications: Society and Environment,* 37, 101488. https://doi.org/10.1016/j.rsase.2025.101488

Joshi, N., Park, J., Zhao, K., Londo, A., & Khanal, S. (2024). Monitoring Harmful Algal Blooms and Water Quality Using Sentinel-3 OLCI Satellite Imagery with Machine Learning. *Remote Sensing*, 16(13). https://doi.org/10.3390/rs16132444

Lan, J., Liu, P., Hu, X., & Zhu, S. (2024). Harmful Algal Blooms in Eutrophic Marine Environments: Causes, Monitoring, and Treatment. *Water (Switzerland)*, 16 (17). https://doi.org/10.3390/w16172525

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2323. https://doi.org/10.1109/5.726791
Lee, F. (2025). *What Is a Neural Network*? IBM. https://www.ibm.com/think/topics/neural-networks
Lin, S., Pierson, D. C., & Mesman, J. P. (2023). Prediction of algal blooms via data-driven machine learning models: an evaluation using data from a well-monitored mesotrophic lake. *Geoscientific Model Development*, 16 (1), 35–46. https://doi.org/10.5194/gmd-16-35-2023
Liu, S., Glamore, W., Tamburic, B., Morrow, A., & Johnson, F. (2022). Remote sensing to detect harmful algal blooms in inland waterbodies. *Science of The Total Environment*, 851, 158096. https://doi.org/10.1016/J.SCITOTENV.2022.158096
Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4766 – 4775).
Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3 (1), 91–99. https://doi.org/10.1016/j.gltp.2022.04.020
Mamun, M., Hasan, M., & An, K. G. (2024). Advancing reservoirs water quality parameters estimation using Sentinel-2 and Landsat-8 satellite data with machine learning approaches. *Ecological Informatics*, 81, 102608. https://doi.org/10.1016/j.ecoinf.2024.102608
Mangat, M. S. A. (2024). Spatial-Temporal Pattern Of Total Organic Carbons (Toc) In Largest Southeast Asia Man-Made Lake. ResearchGate. https://www.researchgate.net/publication/395103934_SPATIAL-TEMPORAL_PATTERN_OF_TOTAL_ORGANIC_CARBONS_TOC_IN_LARGEST_SOUTHEAST_ASIA_MAN-MADE_LAKE
McQuatters-Gollop, A., Johns, D. G., Bresnan, E., Skinner, J., Rombouts, I., Stern, R., Aubert, A., Johansen, M., Bedford, J., & Knights, A. (2017). From microscope to management: The critical value of plankton taxonomy to marine policy and biodiversity conservation. *Marine Policy*, 83, 1–10. https://doi.org/10.1016/j.marpol.2017.05.022
Molares-Ulloa, A., Rocruz, E., Rivero, D., Padin, X. A., Nolasco, R., Dubert, J., & Fernandez-Blanco, E. (2025). Towards improved harmful algal bloom forecasts: A comparison of symbolic regression with DoME and stream learning performance. *Computers and Electronics in Agriculture*, 233, 110112. https://doi.org/10.1016/J.COMPAG.2025.110112
Morris, J. G., Jr. (2014). Human illness associated with harmful algal blooms. In J. E. Bennett, R. Dolin, & M. J. Blaser (Eds.), Mandell, Douglas, and Bennett's principles and practice of infectious diseases (Vol. 2, pp. 3192–3195). Elsevier Saunders. https://doi.org/10.1016/B978-1-4557-4801-3.00286-1
Muhsin, R. M. M., Abd Manan, T. S. B., Bidai, J., Mangat, M. S. A., Mohd Hanafiah, Z., Gohari, A., Ahmad, N., Ahmad, F., Beddu, S., Mohd Kamal, N. L., Mohamad, D., Aldala'in, S. A. H., Mustafa, M. R. U., Mohtar, W. H. M. W., Hasnain Isa, M., Yusoff, M. S., & Abdul Aziz, H. (2025). Polycyclic aromatic hydrocarbons (PAHs) occurrences in water bodies, extraction techniques, detection methods, and standardized guidelines for PAHs in aqueous solutions. *Science of The Total Environment*, 972, 179123. https://doi.org/10.1016/J.SCITOTENV.2025.179123
Naqvi, S. F. N. (2025). View of Spatiotemporal Analysis of Air Quality Index (AQI) in Pakistan: Implications for Public Health and Policy Interventions. *International Journal of Biosciences Review*. https://www.ijbr.com.pk/IJBR/article/view/2477/2269
Paul R. Hill, Anurag Kumar, M. T., & David R. Bul. (2020). HABNet: Machine Learning, Remote Sensing-Based Detection of Harmful Algal Blooms*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 3229 – 3239.
Pethybridge, H. R., Weijerman, M., Perrymann, H., Audzijonyte, A., Porobic, J., McGregor, V., Girardin, R., Bulman, C., Ortega-Cisneros, K., Sinerchia, M., Hutton, T., Lozano-Montes, H., Mori, M., Novaglio, C., Fay, G., Gorton, R., & Fulton, E. (2019). Calibrating process-based marine ecosystem models: An example case using Atlantis. *Ecological Modelling*, 412, 108822, https://doi.org/10.1016/j.ecolmodel.2019.108822
Phetanan, K., Kwon, D. H., Lee, J., Jeong, H., Nam, G., Hwang, E., Pyo, J. C., & Cho, K. H. (2025). SAR remote sensing for monitoring harmful algal blooms using deep learning models. *GIScience and Remote Sensing*, 62(1). https://doi.org/10.1080/15481603.2025.2524202
Pinckney, J., Papa, R., & Zingmark, R. (1994). Comparison of high-performance liquid chromatographic, spectrophotometric, and fluorometric methods for determining chlorophyll a concentrations in estaurine sediments. *Journal of Microbiological Methods*, 19 (1), 59–66. https://doi.org/10.1016/0167-7012(94)90026-4
Pyo, J. C., Pachepsky, Y., Baek, S. S., Kwon, Y. S., Kim, M. J., Lee, H., Park, S., Cha, Y. K., Ha, R., Nam, G., Park, Y., & Cho, K. H. (2017). Optimizing semi-analytical algorithms for estimating chlorophyll-a and phycocyanin concentrations in inland waters in Korea*. Remote Sensing*, 9 (6). https://doi.org/10.3390/rs9060542
Rome, M., Beighley, R. E., & Faber, T. (2021). Sensor-based detection of algal blooms for public health advisories and long-term monitoring. *Science of the Total Environment*, 767, 144984. https://doi.org/10.1016/j.scitotenv.2021.144984

Sadeghi, Z., Alizadehsani, R., CIFCI, M. A., Kausar, S., Rehman, R., Mahanta, P., Bora, P. K., Almasri, A., Alkhawaldeh, R. S., Hussain, S., Alatas, B., Shoeibi, A., Moosaei, H., Hladík, M., Nahavandi, S., & Pardalos, P. M. (2024). A review of Explainable Artificial Intelligence in healthcare. *Computers and Electrical Engineering*, 118 (Part A), 109370. https://doi.org/10.1016/j.compeleceng.2024.109370
Saha, P., Bose, D., Stebliankin, V., Cickovski, T., Seth, R. K., Porter, D. E., Brooks, B. W., Mathee, K., Narasimhan, G., Colwell, R., Scott, G. I., & Chatterjee, S. (2022). Prior exposure to microcystin alters host gut resistome and is associated with dysregulated immune homeostasis in translatable mouse models. *Scientific Reports*, 12, 115116. https://doi.org/10.1038/s41598-022-15708-3
Shahmiri, A., Seyed-Djawadi, M. H., & Siadatmousavi, S. M. (2025). AI-driven forecasting of harmful algal blooms in Persian Gulf and Gulf of Oman using remote sensing. *Environmental Modelling & Software*, 185, 106311. https://doi.org/10.1016/J.ENVSOFT.2024.106311
Shao, H., Kiyomoto, S., Kawauchi, Y., Kadota, T., Nakagawa, M., Yoshimura, T., Yamada, H., Acker, T., & Moore, B. (2021). Classification of various algae canopy, algae turf, and barren seafloor types using a scientific echosounder and machine learning analysis. *Estuarine, Coastal and Shelf Science*, 255, 107362. https://doi.org/10.1016/j.ecss.2021.107362
Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 110, 102817. https://doi.org/10.1016/j.ssresearch.2022.102817
Small, K., Kopf, R. K., Watts, R. J., & Howitt, J. (2014). Hypoxia, blackwater and fish kills: Experimental lethal oxygen thresholds in Juvenile Predatory Lowland River fishes. *PLoS ONE*, 9(4), e94524. https://doi.org/10.1371/journal.pone.0094524
Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9, 307. https://doi.org/10.1186/1471-2105-9-307
Sungmin, O., Dutra, E., & Orth, R. (2020). Robustness of process-based versus data-driven modeling in changing climatic conditions. *Journal of Hydrometeorology*, 21, 1929–1944. https://doi.org/10.1175/JHM-D-20-0072.1
Thakur, D. (2018). *LSTM and its equations*. Medium. https://medium.com/@divyanshu132/lstm-and-its-equations-5ee9246d04af
Vilone, G., & Longo, L. (2020). Explainable Artificial Intelligence: a Systematic Review.
Wang, J., Zhou, Y., Bai, X., & Li, W. (2022). Effect of algal blooms outbreak and decline on phosphorus migration in Lake Taihu, China. *Environmental Pollution*, 296. https://doi.org/10.1016/j.envpol.2021.118761
Wang, S., & Qin, B. (2025). Application of Optical Remote Sensing in Harmful Algal Blooms in Lakes: A Review. *Remote Sensing,* 17(8), 1381.https://doi.org/10.3390/rs17081381
Wang, W., Wang, G., Li, J., Chen, J., Gao, Z., Fang, L., Ren, S., & Wang, Q. (2025). Remote sensing identification and model-based prediction of harmful algal blooms in inland waters: Current Insights and Future Perspectives. *Water Research X*, 28, 100369. https://doi.org/10.1016/j.wroa.2025.100369
Wang, Y., Feng, M., Wang, J., Chen, X., Chen, X., Du, X., Xun, F., & Ngwenya, B. T. (2021). Algal blooms modulate organic matter remineralization in freshwater sediments: A new insight on priming effect. *Science of the Total Environment*, 784, 147087. https://doi.org/10.1016/j.scitotenv.2021.147087
Wells, M. L., Karlson, B., Wulff, A., Kudela, R., Trick, C., Asnaghi, V., Berdalet, E., Cochlan, W., Davidson, K., De Rijcke, M., Dutkiewicz, S., Hallegraeff, G., Flynn, K. J., Legrand, C., Paerl, H., Silke, J., Suikkanen, S., Thompson, P., & Trainer, V. L. (2020). Future HAB science: Directions and challenges in a changing climate. *Harmful Algae*, 91, 101632. https://doi.org/10.1016/j.hal.2019.101632
Yan, X., Xia, Y., Ti, C., Shan, J., Wu, Y., & Yan, X. (2024). Thirty years of experience in water pollution control in Taihu Lake: A review. *Science of the Total Environment*, 914, 169821. https://doi.org/10.1016/j.scitotenv.2023.169821
Yan, Z., & Alamdari, N. (2025). Projecting harmful algal blooms in North Biscayne Bay using machine learning and bias-corrected climate scenarios. *Journal of Hydrology: Regional Studies*, 62, 102865. https://doi.org/10.1016/J.EJRH.2025.102865
Zhao, Y., Liu, D., & Wei, X. (2020). Monitoring cyanobacterial harmful algal blooms at high spatiotemporal resolution by fusing Landsat and MODIS imagery. *Environmental Advances*, 2, 100008. https://doi.org/10.1016/j.envadv.2020.100008
Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. *Eco-Environment and Health*, 1(2), 107–116. https://doi.org/10.1016/j.eehl.2022.06.001

Acknowledgments

This research was supported by Ministry of Higher Education Malaysia through Fundamental Research Grant Scheme (FRGS/1/2023/TK06/UMT/02/1).