**Journal of Environmental & Earth Sciences**
https://journals.bilpubgroup.com/index.php/jees

**REVIEW**

# Water Quality Monitoring in the Era of Big Data: Sensors, AI, and Integrated Frameworks

***Peng Du***

*Beijing Institute of Mineral Geology, Beijing 101500, China*

## ABSTRACT

Water quality monitoring is rapidly evolving from traditional periodic sampling towards continuous, data-intensive, and intelligent systems. This review examines the current trends in the field of water quality monitoring within the context of big data, and specifically focuses on the intersection of sensor technologies, artificial intelligence (AI), and integrated monitoring systems. Originally, the article examines the current sensing methods, such as electrochemical sensors, optical and spectroscopic devices, biosensors, IoT-enabled networks, remote sensing, and mobile platforms, with an emphasis on their increased monitoring performance and their persistent shortcomings in the areas of fouling, calibration, selectivity, and field performance. Second, it addresses the distinguishing features of water quality big data and highlights the issues of heterogeneity, velocity, uncertainty, preprocessing, data fusion, interoperability, and governance. Third, the paper assesses the use of AI, with special focus on machine learning, deep learning, and water quality assessment, prediction, anomaly detection, and decision support, and critically discusses concerns related to transferability, interpretability, reproducibility, and quantifying uncertainty. Lastly, the article brings together the synthesis of how the intertwining of frameworks of sensing, data infrastructure, analytics, and management platforms can help to transform fragmented observations of water quality into proactive and adaptive control. This review uniquely adopts a systems-level perspective, treating sensors, big-data management, and AI as integrated layers of next-generation environmental intelligence.

***Keywords:*** Water Quality Monitoring; Big Data; Sensors; Artificial Intelligence; Integrated Frameworks

*CORRESPONDING AUTHOR:
Peng Du, Beijing Institute of Mineral Geology, Beijing 101500, China; Email: dododp@sina.com

ARTICLE INFO
Received: 1 April 2026 | Revised: 18 June 2026 | Accepted: 25 June 2026 | Published Online: 14 August 2026
DOI: https://doi.org/10.30564/jees.v8i8.13361

CITATION
Du, P., 2026. Water Quality Monitoring in the Era of Big Data: Sensors, AI, and Integrated Frameworks. Journal of Environmental & Earth Sciences. 8(8): 643–669. DOI: https://doi.org/10.30564/jees.v8i8.13361

# 1. Introduction

Water quality observation has been a key element in modern environmental management due to the increased urbanization, industrialization, agricultural intensification, and hydrological variability caused by climatic changes, which place varying and increasing stress on water systems [1,2]. There is an increase in the exposure of freshwater and coastal environments to over-enrichment of nutrients, heavy metals, emergent organic contaminants, pathogenic microorganisms, and complex combinations of pollutants whose occurrences vary spatially and temporally [3]. Here, the timely and reliable water quality data is no longer necessary to conduct the retrospective analysis, but rather it is required to perform the real-time risk identification, compliance with regulations, ecosystem, drinking water health, and real-time resource governance. Consequently, scientific and engineering communities are shifting towards a continuum of high-frequency, continuous, and distributed monitoring as opposed to the periodic and labor-intensive sampling.

Traditional water quality monitoring tools have offered the basis of assessment of the environment for decades, but their weaknesses are increasingly becoming clear in the light of dynamic and interconnected water systems [4]. The conventional methods are highly focused on grab sampling, lab tests, and fixed-point measurements. Although they can offer high levels of accuracy in analytical procedures, they have limited spatial coverage, low temporal resolution, delayed reporting, and high operational costs. These limitations mean that they may not be able to capture short-term contamination events, fast-changing algal blooms, intermittent discharge events, or non-linear interactions between physical, chemical, and biological variables. In practice, in most real-life scenarios, before laboratory results are made available, the critical window of opportunity to act might have closed. Such a poor fit between the rate at which the environment is changing and the rate of traditional monitoring has led to a high demand for more intelligent, automated, and scalable solutions [5].

The new growth of sensing technologies has begun filling some of the gaps. In situ electrochemical electrodes, optical instruments, biosensors, wireless sensor networks, unmanned platforms, and remote sensing systems can now provide volumes of water-related observations, never before attempted to be made, at a scale of multiple magnitudes [6,7]. Monitoring is also no longer confined to single-parameter gauges or remote stations; it is now more commonly made up of networks of sensors that can measure turbidity, pH, dissolved oxygen, conductivity, nutrients, chlorophyll, organic matter proxies, and, in other cases, biological contaminants in near real time. Simultaneously, the Earth observation data hydrometeorological records, the records of the treatment plants, and laboratory measurements are also adding extra layers of information [8]. Collectively, these trends have made water quality monitoring a common big-data challenge that is marked by high volume, high velocity, and high variety and has continued to face significant challenges of veracity and interoperability.

Nevertheless, big data creation does not necessarily lead to being put into action. The data of water quality is usually noisy, incomplete, multi-source, multi-scale, and highly dependent on the context [9]. Data reliability can be low due to sensor drift, fouling, missing records, communication problems, heterogeneous data format, variable calibration standards, and integration can be difficult because of this [10]. Besides, environmental variables often have nonlinear relationships that are driven by hydrological, meteorological, land-use, and anthropogenic factors. Such properties render water systems challenging to comprehend based on a set of traditional statistical operations. Therefore, the problem of water quality monitoring today is not only gathering larger volumes of data, but converting heterogeneous data flows into powerful assessment, prediction, and decision-making systems.

The solution to this challenge has come in the form of artificial intelligence. Deep learning and machine learning techniques can identify trends in complicated environmental data, identify anomalies, estimate unobservable parameters, predict pollution, and assist in optimization of operations [11]. It is demonstrated that AI-based models can be used to predict water quality indices, contamination state classification, algal bloom dynamics, and concealed relations among monitoring variables [12]. Data-driven models in comparison with purely mechanistic methods tend to be more appropriate when dealing with nonlinear processes and large-dimensional data. However, AI also brings about new challenges such as interpretability of models, site-to-site transferability, reproducibility, uncertainty estimation, and reliance on the quality of the data. Thus, AI cannot be regarded as a single analytical solution, but it is just a part of a bigger monitoring system that starts with sensing, is complemented by data governance and processing, and ends with decision support [13].

This system view encourages a change of focusing on sensors, data analytics, and management frameworks individually, and focusing on them as a single entity. **Figure 1** shows the conceptual map of the whole water quality monitoring in the age of big data, and the connections between sensing, data management, AI analytics, and decision support. The future water quality monitoring will rely on architectures where sensing technologies serve as the perception layer, big-data infrastructure will serve as the communication and storage layer, AI will become the inference layer, and integrated platforms will turn the output into early warning, visualization, and management [14]. These architectures could consist of edge cloud collaboration, cyber-physical waters, digital twins, and intelligent decision-support systems that have the ability to connect observations and simulation, prediction, and intervention. Significantly, these frameworks have value not just in technical sophistication, but in the ability of these frameworks to facilitate effective water governance by supporting drinking water systems, rivers, lakes, reservoirs, groundwater, coastal waters, and wastewater treatment plants.


**Figure 1.** Conceptual framework of water quality monitoring in the era of big data, showing the integration of sensor technologies, data infrastructures, AI analytics, and decision-support systems under multiple environmental pressures.

Despite the increased pace of literature growth on water quality monitoring, the existing reviews are still directed at discussing one aspect of the subject. A lot of literature discusses sensor technologies separately, with a focus on principles of detection and performance of devices [14]. Others look at applications of machine learning to prediction tasks without paying enough attention to the upstream problems of the quality of their data, the heterogeneity of their sensors, or operational deployment [15]. Other studies talk about smart water systems or IoT platforms, although there is a lack of synthesis of the interaction between sensing, AI, big-data management, and integrated governance frameworks as a comprehensive monitoring pipeline. Thus, there is still a necessity for a review that links these areas in a coherent and implementation-focused way.

It is in that integrative perspective that the novelty of this article is found. Instead of considering sensors, AI techniques, and monitoring systems as parallel subjects, this review presents them as mutually reinforcing layers of a water quality ecosystem based on the use of big data. It summarizes the development of modern sensing technologies first, not only as to the parameters that they detect, but also as to the nature and nature of the data that they produce [16]. Second, it gives heavy emphasis on data-centric issues, such as preprocessing, harmonization, missing-data treatment, interoperability, and governance, which are ignored even though they are decisive to model reliability [17]. Third, it critically discusses AI practices within the context of the real world of monitoring the environment, not only in terms of predictive performance, but also in terms of explainability, uncertainty, generalizability, and reproducibility [18]. Fourth, and above all, it promotes an integrated framework approach by commenting on how sensor networks, big-data infrastructures, intelligent analytics, and decision-support systems can be integrated to create operating monitoring architectures [19].

This review followed a systematic literature search strategy. We queried Web of Science, Scopus, and Google Scholar for peer-reviewed articles published between 2010 and 2026 using keyword combinations: (“water quality” OR “water monitoring”) AND (“sensor*” OR “IoT” OR “remote sensing”) AND (“big data” OR “machine learning” OR “artificial intelligence” OR “deep learning”). Additional papers were identified from reference lists of retrieved articles. Inclusion criteria: (i) English-language journal articles or conference proceedings; (ii) direct relevance to water quality monitoring technologies, data management, or AI applications; (iii) availability of full text. We excluded purely laboratory-based studies without field validation, and non-peer-reviewed reports. The final corpus comprised 126 primary sources, synthesised thematically rather than meta-analytically.

## 2. Sensor Technologies for Modern Water Quality Monitoring

### 2.1. Categories of Monitored Water Quality Parameters

The parameters to be measured by the system dictate the design and performance of the water quality monitoring systems [20]. The water quality indicators are usually categorized as physical, chemical, and biological in the modern monitoring practice, but in the actual aquatic systems, these categories are highly coupled and not independent. Physical parameters, i.e., temperature, turbidity, total suspended solids, color, and electrical conductivity, are often the initial alert on changing environmental parameters [21]. They can be measured reasonably easily and cheaply on a continuous basis and often form the basis of real-time monitoring networks. Their seeming simplicity is, however, deceptive. Physical indicators tend to be non-specific proxies: an increase in turbidity, e.g., can indicate sediment resuspension, inputs of runoff, biological growth, or anthropogenic disturbance, and cannot be interpreted without contextual data, provided by hydrological measurements and chemical measurements.

Chemical parameters offer a better perspective on processes of water quality and pollution levels. The pH, dissolved oxygen, oxidation-reduction potential, salinity, nitrate, ammonium, phosphate, dissolved organic carbon, heavy metals, and emergent contaminants are some of the key variables that are used to determine the health of the ecosystem and the performance of the treatment [21]. Such parameters can possess more diagnostically significant values than physical ones since the latter are more related to biogeochemical processes and sources of pollution. Simultaneously, they also prove harder to track with a high level of reliability. Chemical species can exist at low concentrations and can be rapidly changed, or can be affected by matrix effects, which affect sensor selectivity. Consequently, chemical surveillance continues to be a field where the trade-off between the level of analysis and field usability is especially high.

The biology and microbiology parameters are a bigger challenge. The pathogenic bacteria, viruses, algal toxins, chlorophyll-related indicators, and biomolecular markers presently gain importance since most water quality risks are biological in the end [22]. The standard laboratory analysis is highly specific to such targets, and the continuous in situ measurement is not as developed as in the case of physical variables and chemical variables. This difference shows a larger problem within the water quality monitoring: the simplest parameters to continuously monitor are not necessarily the most significant threats to human health or environmental risk. In turn, the current sensor development is not only motivated by the aspiration of capturing more variables, but also the necessity to overcome this disparity between the convenient operations of the environment and its environmental importance.

Another complication is that the significance of each parameter is based on the monitoring goal and the water system being discussed. The drinking water distribution systems place more importance on the residual disinfectants, microbial safety, and fast contamination reactions as compared to the lakes and reservoirs, which could be more concerned with nutrients, chlorophyll, and dissolved oxygen dynamics [23,24]. Process-control variables supplemented by ammonium, nitrate, chemical oxygen demand surrogates, and turbidity are some process-control variables in wastewater treatment systems, whereas salinity, chlorophyll, and eutrophication or harmful algal bloom indicators may be important process-control variables in coastal systems [25,26]. That is why the selection of the sensor cannot be considered a technological choice; it should be coordinated with the ecological, regulatory, and operational situation. This aspect has been largely overlooked in the literature, where sensors have been considered in some generalized manner without much regard to the decision environments in which they are supposed to operate.

## 2.2. In Situ Sensing Technologies

The current water quality monitoring has taken the form of in situ sensing, as it can provide continuous monitoring of the water quality that is directly in the water [27]. In situ sensors have significant benefits in terms of time resolution, early warning ability, and performance as compared to the laboratory-related approaches. They are especially useful in systems where the quick variations are frequent and irregular contamination occasions can be overlooked by the periodic sampling. But the real potential of in situ sensing is limited by the problems that are not necessarily reflected in laboratory performance indicators, such as fouling, drift, calibration liability, power requirements, and environmental disturbance. Critical analysis of in situ technologies should therefore extend past detection principles and examine long term robustness of the field and the reliability of data.

### 2.2.1. Electrochemical Sensors

One of the most common pieces of equipment in water quality monitoring is electrochemical sensors that have a comparatively low cost, are portable, and can be used on a continuous basis [28]. They are widely used in the determination of pH, dissolved oxygen, conductivity, redox potential, and selected ions using ion-selective electrodes [29]. What makes them successful is that a significant number of crucial water quality parameters are electrochemical or can be associated with genuine electrical reactions. Besides that, electrochemical devices are generally small and energy-saving, and that is why they are best suited for distributed monitoring networks and remote installations.

In spite of these advantages, electrochemical sensing has a number of existing drawbacks. The sensor drift also constitutes a significant problem, particularly when it is deployed over the long-term in natural waters or wastewater [30]. The degradation of the signal can be due to the aging of the membrane, the instability of the reference electrode, temperature fluctuation, or decontamination of the sensing surface. Biofouling and scaling are also major issues since they impact not only sensor sensitivity but also the maintenance frequency, which increases the operational cost [31]. There is also the problem of selectivity, particularly when ion-selective electrodes are being applied in more complex environmental matrices where the target signal can easily be contaminated by other ions. In such circumstances, nominally continuous monitoring can generate huge amounts of data, of which the apparent accuracy conceals the reality of uncertainty.

The value of electrochemical sensors is not reduced by these weaknesses, but it gives the impression that they should be used cautiously. Electrochemical measurements can frequently be most efficient when incorporated in larger monitoring structures that comprise automated calibration routines, quality-control programs, and cross-validation with complementary types of sensors [32]. Their future progress is expected not to rely so much on the creation of completely new principles of electrode functioning, but rather on the enhancement of anti-fouling measures, self-diagnostics, calibration stability, and incorporating them into intelligent data correction techniques. In this respect, the development of electrochemical sensing is indeed more and more reliant on system engineering than sensor chemistry.

### 2.2.2. Optical and Spectroscopic Sensors

Sensors based on optical and spectroscopy are increasingly popular due to their non-destructive, fast, and common multi-parameter measurement features [33]. Absorbance, fluorescence, and other spectroscopic principles are popular methods of determining the concentration of dissolved organic matter, chlorophyll, and nitrate, along with suspended particles and other optically active substances [34]. Their strengths are their ability to gain rich information with little reagent consumption and, in most situations, little physical contact with the target medium. This renders them desirable in the high-frequency monitoring and application in environments where reagent-based techniques are not feasible.

The strong point of optical methods lies in the fact that indirect information-dense signatures of water composition can be offered [35]. Rather than measuring only a single analyte using one highly specific reaction, optical sensors tend to measure a wider spectral or fluorescence signal that can be analyzed by either a calibration model or by multivariate analysis. This also qualifies them as befitting the revival of the big-data age, as their results can be merged with machine learning techniques to form pattern recognition and estimations of parameters. Practically, optical signals are now of particular significance to surrogate monitoring, where the hard-to-measure constituents are estimated using optical proxies whose values can be measured continuously.

This strength is, however, a weakness. Optical measurements tend to be model-dependent and are indirect, i.e., their accuracy is dependent on the conditions of calibration and environmental consistency [36]. Signal-response relationships can change due to changes in dissolved organic matter composition, particle size, ambient light conditions, or temperature, and decrease predictive robustness. There is also a great deal of uncertainty when it comes to optical fouling, window contamination, and signal attenuation in heavily turbid or colored water [37]. Besides, encouraging optical methods as multi-parameter solutions, there is a danger of exaggerating their analytical particularity. Not all spectrally rich data sets imply good quantification, particularly where one or more constituents share similar optical signals.

Thus, calibration transfer, adaptive modeling, and inherent accounting of uncertainty are essential to the long-term success of optical sensing in water quality monitoring [36]. The primary point that the sensors are not to be viewed as all-purpose substitutes for electrochemical methods, but as complementary technologies, which allow greater sensitivity, richer information, and integration of information with data-driven analytics. They are worth much in practice once spectral data is employed cleverly in larger schemes of analysis, as opposed to being considered as a direct alternative to traditional chemistry.

### 2.2.3. Biosensors and Lab-on-Chip Platforms

Biosensors and lab-on-chip sensors are a significant future in water quality monitoring since they seek to enhance sensitivity and selectivity to biologically and chemically complicated targets [38]. Biosensors can be used to detect certain toxins, pathogens, metabolites, or trace contaminants that are challenging to detect by traditional physical or electrochemical detection techniques, by the use of enzymes, antibodies, nucleic acids, aptamers, or living cells as recognition elements [39]. Lab-on-chip systems also aim at reducing the size of analytical processes by incorporating sampling, reaction, separation, and detection into miniaturized systems.

The potential of such technologies is high. They deal with one of the greatest flaws of the existing real-time monitoring systems: the inability to directly measure biologically meaningful hazards. As an example, biosensing technologies might be used to potentially assist quicker fecal pollution discovery or harmful algal toxin detection, or emergent micropollutant discovery, reducing the delay between environmental alteration and intervention action [40]. Miniaturized systems also present the opportunities of decentralized monitoring, less use of reagents, and field diagnostics carried out in the form of portable devices.

However, biosensors are a very lopsided type of technology. Most devices that have shown impressive performance at lab conditions face significant challenges in the field. The recognition elements can become inactive with time; non-specific binding can cause a decrease in selectivity and variable sample matrices can change the efficiency of the reactions [41]. Besides this, there have been delays in the development of proof-of-concept devices into rugged field systems than previously thought. Although laboratory-on-a-chip systems are beautiful in concept, they often have bottlenecks in terms of sample treatment, clogging, fluidic control, and long-term stability. These difficulties highlight a common problem in the development of water quality technology, namely that analytical newness does not always imply operational feasibility.

On this account, the future significance of biosensors and microfluidic platforms will be related to the possibility of overcoming highly demonstrative proofs and achieving stable integration into monitoring systems [42]. Initial strengths of theirs might be in hybrid systems where they complement but do not replace traditional sensors by offering validation-focused, event-driven, or high-value measurements of contaminants that otherwise cannot be monitored. Strategically, however, these technologies benefit optimally not in their ability to replicate the work of current sensors, but in their potential to extend the monitoring capacity to areas of analytical underservice.

## 2.3. Remote and Distributed Sensing Systems

The growth of distributed and remote sensing has led to the acknowledgment that monitoring of water quality needs to be more comprehensive in both heterogeneous and homogeneous environments [43]. Situated sensors have large temporal resolution on a local scale, whereas numerous water quality processes, including watershed runoff pulses, reservoir stratification patterns, estuarine mixing, and brood propagation, are spatially organized. Remote and distributed sensing systems aim to overcome this deficiency by making observations over larger spatial scales and also by connecting many nodes into synchronized networks of monitoring. Such systems are particularly relevant in the situation of big data since they can produce spatially large and, in many cases, constantly changing data that may be used by large-scale inference and prediction.

### 2.3.1. IoT-Enabled Sensor Networks

With the appearance of Internet of Things (IoT) architectures, water quality monitoring has evolved from an individual sensing view to a networked monitoring [44]. IoT-based systems combine sensors, communication devices, and data gateways with cloud computers and, in most cases, with automated analytics embedded into distributed monitoring systems [45]. Their greatest strength is in scalability. It is then possible to have cheaper nodes spread across rivers, reservoirs, urban drainage systems, treatment plants, or aquaculture systems, rather than having one or a few costly stations. This networked strategy enables higher spatial coverage, near real-time transmission, and centralized analysis.

Operationally, IoT systems are also especially appealing due to their ability to establish the framework of intelligent monitoring and not the mere collection of data [46]. Communication levels provide access to sensor data quickly, cloud computing platforms enable storage and visualization, and embedded software can trigger an alarm or preventive maintenance. Therefore, IoT is not a type of sensor itself, but a logic of a system that alters the organization and utilization of sensing.

Simultaneously, the practicality of the performance of the IoT-based water monitoring is usually limited by the weakest links in the system instead of the strongest. The low-cost sensor nodes can experience inconsistent reliability, wireless transmission can be unreliable in remote or hostile environments, and energy management can continue to be a bottleneck towards the autonomous deployment. In addition, the increase in networks can only increase data quality issues instead of addressing them. The increased number of nodes can result in a greater generation of information; however, it may also create a greater number of calibration inconsistencies, more missing data, and an increase in maintenance requirements. In this regard, the IoT designs face the risk of transforming the problem of water monitoring to distributed quality control instead of sparsely collected data [47].

This fact is significant as the rhetoric of smart monitoring is, at times, based on the assumption that the idea of connectivity is the inherent source of intelligence. As a matter of fact, connected sensor networks can only be smart with strict validation of data, robust communication design, and self-mending maintenance programs. The main problem of the IoT water monitoring is not simply to interconnect the devices, but to make sure that the resulting data streams will be trustworthy and useful over time [48].

### 2.3.2. Remote Sensing and Satellite Observations

Remote sensing has emerged as an effective instrument for observing water bodies at both regional and global levels [49]. Repeat images of surface water extent, chlorophyll-related signals, suspended matter, colored dissolved organic compounds, temperature variations, and bloom dynamics can be provided by satellite and airborne platforms that have multispectral, hyperspectral, or thermal sensors [50]. These are particularly useful in lakes, reservoirs, estuaries, and coastal areas that have a high level of spatial heterogeneity and limited ground-based monitoring. Big picture trends that cannot be measured by point sensors are visible with the help of remote sensing, and this makes remote sensors essential to synoptic evaluation and long-term trend analysis.

Its relevance in the era of big data is also based on its standardization and continuity [51]. Satellite programs produce long-term archives that have wide coverage in space, allowing retrospective analysis and combination with hydrological and climatic data. Remote sensing offers the sole possible method of constant monitoring of large or inaccessible water systems in several settings. It has thus become the focus of both scientific study and management uses, such as monitoring of blooms, watershed evaluation, and reservoir status [52].

However, remote sensing should be taken with caution. However, the majority of the satellite-based water quality products are based on indirect optical retrievals and thus the same calibration dependence and ambiguity as in the in situ optical sensing can be found in the satellite-based counterparts, only at larger scales and in more fluctuating atmospheric and hydrological conditions. Superficial measurements cannot reflect the conditions in the subsurface, and the performance of optical retrieval in turbid coastal waters, narrow rivers, or mixed waters may be degraded [35]. Interruptions in the temporal revisions and cloud coverage may occur at a time when fast events are taking place. Spatial resolution, which is continuously getting better, is inadequate for most small or highly divided water bodies.

These drawbacks do not reduce the significance of remote sensing, but they explain its proper place. Satellites work best when combined with in situ data rather than being substituted. They add a spatial context, generalizing patterns, and help to detect patterns/ anomalies that can direct the sampling. In that regard, remote sensing cannot be seen as an independent monitoring solution but as an essential component of multi-scale observation systems [53]. Its strength is its ability to be integrated.

### 2.3.3. Unmanned Aerial Vehicles and Mobile-Platform-Based Monitoring

Unmanned aerial vehicles (UAVS), autonomous surface vehicles, and other mobile platforms have gained increasing interest due to the difficulty they provide to bridge a significant scale gap between fixed in situ stations and large-scale satellite measurements. These systems provide adaptable, focused gathering of data, and may be implemented in locations that are challenging, perilous, or ineffective to be tracked manually. Optical, thermal, or multispectral sensors on UAVs can quickly monitor bloom conditions, sediment plumes, infrastructure-related anomalies, and robotic boats, as well as mobile sondes, which are able to make adaptive in-water measurements along horizontal and vertical gradients [54].

Responsiveness of mobile platforms can be termed as one of the biggest strengths. They have the advantage of being able to be oriented to areas of interest (as opposed to the situation with stationary stations), track moving records (such as active fronts), or even focus measurements when a pollution incident is suspected [55]. This is what makes them especially useful in event-based monitoring, responding during an emergency, and in mapping heterogeneous systems at high-resolution. Moreover, they offer a useful method of gathering calibration and validation data that can reinstate the satellite retrievals as well as the fixed monitoring networks.

Nevertheless, they are not as flexible as they have limitations. Battery life, payload capacity, flight regulations, and weather sensitivity are some of the constraints of UAV operations, whereas aquatic robots have problems associated with navigation, endurance, biofouling, and sensor stability [56]. Besides, mobile platforms can generate deep datasets in short-term campaigns, but they cannot be applied to continuous long-term monitoring except for being incorporated into automated systems of operation. Such a novelty can also run a risk of eclipsing practical questioning on continuity. High-resolution mapping is useful, although unless it is associated with regular data streams and decision processes, it might not be transformative as well as episodic.

Therefore, the most effective contribution of the UAVs and mobile sensing platforms can be in the form of adaptive elements of integrated monitors. They are especially effective in their application to supplement the fixed and remote systems to fill points of observation, confirm anomalies, or escalate sampling during critical times [57]. Their strategic importance is not about substituting other platforms, but making the system of overall monitoring more responsive and with greater spatial intelligence. The major sensor categories, their target parameters, strengths, limitations, and application scenarios are compared in **Table 1**.

**Table 1.** Major sensor technologies for water quality monitoring, including representative methods, target parameters, principal advantages, limitations, and typical application scenarios.

| Sensor Category | Representative Technologies | Typical Target Parameters | Main Advantages | Main Limitations | Typical Application Scenarios |
|---|---|---|---|---|---|
| Electrochemical sensors | pH electrodes, dissolved oxygen probes, conductivity sensors, ion-selective electrodes | pH, DO, EC, ORP, nitrate, ammonium, chloride | Low cost, real-time response, portable, suitable for continuous deployment | Drift, fouling, cross-sensitivity, frequent calibration | Rivers, drinking water networks, wastewater treatment, aquaculture |
| Optical and spectroscopic sensors | UV–Vis, fluorescence, turbidity meters, Raman/IR-based sensors | Turbidity, DOC proxies, chlorophyll, nitrate, suspended solids | Rapid, reagent-free, multi-parameter potential, high temporal resolution | Indirect measurement, matrix interference, optical fouling, calibration dependence | Reservoirs, lakes, treatment plants, online surrogate monitoring |
| Biosensors | Enzyme-, antibody-, aptamer-, and cell-based sensors | Pathogens, toxins, trace organics, algal toxins | High selectivity, potential for rapid biological hazard detection | Limited field robustness, bio-recognition instability, matrix effects | Drinking water safety, contamination screening, targeted hazard detection |
| Lab-on-chip systems | Microfluidic analytical platforms | Nutrients, pathogens, trace contaminants | Miniaturized, low reagent consumption, portable, integrated analysis | Clogging, pretreatment needs, limited long-term field reliability | Point-of-use analysis, portable diagnostics, event-triggered testing |
| IoT-enabled sensor nodes | Wireless multi-sensor platforms | Physical and chemical parameters in real time | Distributed coverage, scalable, remote access, data streaming | Power constraints, communication failure, maintenance burden | Watersheds, smart utilities, decentralized monitoring |
| Remote sensing platforms | Satellite multispectral/hyperspectral sensors | Chlorophyll, algal blooms, suspended solids, temperature, surface water dynamics | Large spatial coverage, repeated observation, regional assessment | Cloud interference, limited depth sensitivity, indirect retrieval | Lakes, estuaries, coastal waters, basin-scale assessment |
| UAV/mobile platforms | Drones, autonomous boats, robotic sondes | Surface optics, thermal anomalies, spatial gradients | Flexible deployment, high spatial detail, rapid targeted surveys | Limited endurance, payload restrictions, regulatory constraints | Bloom mapping, emergency response, inaccessible waters |

## 2.4. Performance Evaluation and Data Characteristics of Sensor Systems

One of the continued issues with literature on water quality sensing is that sensor performance is usually reported using mainly the analytical measures of sensitivity, detection limit, linear range, and response time [58]. Although these measures are imperative, they cannot be used to analyze sensors to be used in the actual environment monitoring. In use in the field, performance is determined by equally important factors such as durability, calibration stability, fouling resistance, energy requirement, maintenance load, and interoperability with data systems. A sensor that has high laboratory sensitivity can have little useful purpose when it has to be recalibrated on a regular basis, it makes unreliable measurements in changing water matrices, or it cannot be incorporated into a networkable monitoring system.

This more encompassing vision of performance is especially relevant in the big-data domain since the usefulness of any sensor is no longer determined simply by what it can measure, but also by the quality and format of the data it provides to a unified system [10]. Water quality sensors yield datasets varying in the frequency of time, spatial representativeness, signal-noise ratio, and the level of directness. Continuous and directly measured variables are some of the variables that are directly measured, whereas proxy relationships and model-based conversion are used in inference. Certain types of sensors can be interpreted to give scalar values, but others use more complicated spectral or multi-dimensional data that must undergo a lengthy preprocessing step before interpretation. These variations are important since they define the downstream analytics, define the data fusion approaches, and the accuracy of AI models trained on sensor data [59].

Practically, water quality surveillance has come to denote a diverse array of heterogeneous constellations of unequally potent and feeble devices [13]. This is not only an opportunity but also a challenge due to this heterogeneity. It provides an opportunity to have a richer and more robust observation by complementarity, but it makes the calibration, harmonization, and evaluation of uncertainty more complicated. The data from the low-cost sensor networks, lab measurements, optical systems, and remote platforms cannot easily be combined without considering a scale mismatch, measurement bias, and temporal misalignment. The effectiveness of integrated monitoring systems is thus not only a matter of innovative technology on the sensing side, but is also a matter of advanced development of good frameworks of sensor comparison, uncertainty propagation, and cross-platform validation [60].

One thing learned is that sensor systems are not to be considered in a vacuum but with respect to the decisions they are supposed to aid in monitoring. Timeliness and fault-tolerance could be of more value to early warning applications than analytical precision per se [61]. In the case of regulatory compliance, traceability and robustness in calibration can be considered first. Consistency and long-term continuity of the data might be more appropriate in predictive analytics than high single-point accuracy. This implies that no sensor technology is best for measuring the quality of water. Rather, the suitability of the sensing strategy is determined by the level at which its data characteristics match the targeted analytical and management purposes.

In this regard, the future of water quality sensing is not just in the advancement of individual devices, but in the creation of sensor ecosystems that are credible, interoperable, and analytically transparent [62]. The next step will probably improve data-centric design principles, such as automated quality control, uncertainty-conscious modeling, and dynamic system maintenance, and integrate hardware innovations better. That is, sensor-informed intelligence is becoming the main problem of concern rather than sensor deployment. This shift is the actual meaning of sensing technologies in the age of big data: they are no longer considered to be measurement tools, but rather building blocks of interactive environmental information systems.

# 3. Big Data Characteristics and Data Management in Water Quality Monitoring

## 3.1. Sources and Structure of Water Quality Big Data

The characteristics of big data used in water quality surveillance do not entail its size as much as the variety and the persistence of sources. The conventional models of monitoring were based primarily on periodical sampling and laboratory analysis, but the new models involve the combination of in situ sensors, laboratory records, satellite observations, hydrological and meteorological observations, treatment plant logs, and the land-use map [63]. This transition has resulted in transforming the data of water quality from a single measurement to a multi-source, multi-scale data information system.

These sources do not merely give further observations about the same process. Sensors have a high frequency time coverage, but with a small number of parameters [64]. Laboratory techniques offer increased specificity in the analysis, but with a lower frequency [38]. Remote sensing also helps in providing extensive spatial coverage, albeit indirect proxies [65]. Hydrometeorological information provides additional explanatory information on how the water quality changes are related to flow, rainfall, and climate driving forces [66]. Consequently, the water quality big data are always heterogeneous in style, magnitude, uncertainties, and senses.

One more important characteristic is the presence of historical archives and live streams. Long-term data are still needed to conduct trend analysis and model training, whereas streaming data are needed to conduct operational monitoring and early warnings. However, these records are hardly created using the same protocols or technologies. Water quality big data must then not be considered as a smooth data set, but as an irregular collection that must be interpreted carefully before being incorporated [67].

## 3.2. Characteristics of Big Data in Water Monitoring

The general structure volume, velocity, variety, veracity, and value come in handy in water quality monitoring, yet each dimension has a given environmental significance [68]. Volume has increased very fast with the use of sensors to record data at minute or second intervals to capture short-lived events as well as diel variability. Nevertheless, the abundance of data in itself is not a guarantee of greater comprehension, and without organization and edits, it might only become more complex.

Operational monitoring is a special consideration of velocity [69]. Data are increasingly being received in stream form and have to be inspected and processed almost in real time. This presents a conflict between timeliness and reliability, as there is not much use of rapid processing without screening for uncertainty in the first place. The first characteristic of water quality big data is perhaps variance, since various instruments are used to measure different properties of the system by fundamentally different mechanisms. To integrate them, it is not merely a matter of technical compatibility, but rather a need to know what each dataset is all about.

Truthfulness is essential, as an environmental choice can be based on the small variations that can be mixed with sensor drift, fouling, interference, or communication errors[70]. Here, veracity encompasses the history of calibration, uncertainty, as well as traceability, rather than nominal accuracy. Last but not least, value relies on data enhancement to explain, predict, and act. A smaller yet dependable data set can prove to be more effective, later on, than a considerably large data set that has not been well documented or is not linked to the requirements of management.

### 3.3. Data Acquisition, Transmission, and Storage Infrastructures

Sensors are not the only facilities required in water monitoring, but also the infrastructure to capture, transfer, and store information[71]. The sampling frequency, timing, and formatting are set by acquisition systems, whereas transmission systems influence the delivery of measurements in a timely and undamaged fashion. Storage platforms influence the accessibility and interoperability in the long run and for reuse. In huge data monitoring, these aspects are focal, not peripheral.

The aqua environments subject the procurement process to pragmatic restrictions such as fluctuating power, inclement weather, corrosion, and inaccessible maintenance. The sampling frequency should hence strike the right balance between the environmental relevance and durability of the system[72]. Monitoring at very high frequency could enhance event detection; however, it can raise noise, battery requirements, and the load on communication.

Cellular networks, telemetry, satellite links, and low-power wireless protocols will allow transmission systems that can be used to implement distributed monitoring, but there is uncertainty in communication failure[73]. Early warning and interpretation can be compromised by missing or delayed transmissions. Storage is also becoming complicated with the migration of data into cloud-based and distributed systems. Nevertheless, increased storage does not address issues of disjointed repositories, uneven metadata, or lack of discoverability. Interoperability is also required in case datasets of various instruments and institutions can be effectively integrated.

### 3.4. Data Preprocessing and Quality Control

Preprocessing is more critical as water quality data providers are becoming larger and more automated[74]. Raw data tend to have drift, noise, gaps, and artifacts due to both environmental variation and instrument limitations. Preprocessing is not only a technical cleanup process, but is in fact an interpretive process where the analysts make decisions about what is signal and what is error.

The generic statistical filtering is commonly not adequate for environmental monitoring. Extreme values can be taken to suggest actual instances of pollution, storm events, or even treatment failures, not erroneous measurements. On the other hand, seemingly steady data might hide issues with instruments. Quality control then needs the knowledge of the domain as well as contextual information, particularly when the information will be used in the future to generate predictive or automated warnings[75].

#### 3.4.1. Calibration, Denoising, and Anomaly Screening

The principle of calibration is that many sensors do not directly measure concentrations, but respond indirectly, and the response can change with time due to the fouling of the sensor, aging, or temperature, as well as matrix effects[76]. Field deployment Calibration is not a one-time process, though it is a continuous process.

High-frequency data requires denoising, which, however, is important to do sparingly. Temporary peaks can be a valued environmental feature, and excessive levels of smoothing will dampen the occurrences that continuous monitoring is supposed to record. This is the same case with anomaly screening. Statistical extremes may indicate the actual extremes in the environment other than mistakes. This is why anomaly detection must use the money rules with the contextual validation against the cross-sensor consonance, hydrologic effects, and repair history[77].

#### 3.4.2. Missing Data and Imputation

Missing data is common in water monitoring due to sensor malfunctions, lack of communication, maintenance, and adverse field conditions [78]. Gaps in continuous systems may negatively affect forecasting, anomaly detection, and trend interpretation, particularly in cases where the missingness is experienced during extreme events.

Imputation techniques, between interpolation and machine learning, may have the benefit of restoring continuity, yet confuse the observed and estimated values [79]. That difference is significant in the applications where threshold exceedance, rare events, or regulation is involved. Later missing data must then be treated in an open way, and the uncertainty well accepted. Repeated gaps are also, in some instances, indicative of weaknesses in the monitoring system, and not necessarily absent values.

#### 3.4.3. Data Fusion and Harmonization

One of the keys promises of big-data water monitoring is data fusion that enables complementary datasets to be merged to gain more comprehensive and more comprehensive insights [69]. A combination of point sensors, laboratory analyses, hydrometeorological drivers, and remote sensing can be more complete than any of them when used separately.

Fusion is, however, challenging due to the fact that datasets are rarely naturally aligned in terms of time, space, scale, and meaning [80]. It does not mean that measurements that look the same are not different in terms of representativeness and uncertainty. Harmonization thus takes more than file conversion; it entails having clear choices regarding comparability and scale. Fusion well done can enhance resilience and insight, but slew fusion can result in false coherence due to combining signals that are not compatible.

### 3.5. Data Governance, Standardization, and Real-Time Analytics

Due to the increased connectivity of monitoring systems, data governance is necessary. The governance encompasses standardization, metadata, ownership, access, security, and institutional arrangements in which data is shared and interpreted [81]. In the absence of it, integrated monitoring will be challenging even in cases where sensing technology is highly developed. **Table 2** provides a summary of the key big-data issues in water quality monitoring and the strategies applied to overcome the challenge.

**Table 2.** Major big-data challenges in water quality monitoring, their typical manifestations, consequences, and commonly adopted management strategies.

| Big-Data Issue | Typical Manifestation in Water Monitoring | Main Consequences | Common Mitigation Strategies |
|---|---|---|---|
| Volume | High-frequency sensor streams and long-term multi-site archives | Storage burden, processing bottlenecks, redundant information | Cloud storage, compression, time-series databases, scalable pipelines |
| Velocity | Near real-time streaming from sensor networks | Delayed validation may reduce early-warning value | Edge analytics, automated QC, low-latency communication systems |
| Variety | Sensors, lab data, remote sensing, hydrometeorology, operational logs | Difficult integration, inconsistent formats, semantic mismatch | Data standards, metadata frameworks, harmonization workflows |
| Veracity | Drift, fouling, missing values, transmission errors, inconsistent calibration | False alarms, unreliable models, weak confidence in decisions | Calibration management, anomaly screening, uncertainty tracking |
| Missing data | Sensor outage, maintenance, power loss, communication failure | Disrupted forecasting, reduced continuity, biased inference | Imputation, redundant sensing, fault diagnosis, resilient system design |
| Scale mismatch | Different temporal and spatial resolutions across datasets | Weak comparability, fusion errors, distorted interpretation | Temporal alignment, spatial resampling, multi-scale modeling |
| Interoperability | Different units, timestamps, naming rules, quality flags across platforms | Integration cost, poor reproducibility, limited transferability | Standardized schemas, APIs, shared ontologies, data governance protocols |
| Cybersecurity and access control | Connected monitoring linked to critical infrastructure | Vulnerability to tampering, privacy and security concerns | Authentication, encrypted transmission, role-based access control |
| Governance fragmentation | Data held by agencies, utilities, researchers, and vendors separately | Siloed datasets, weak coordination, poor emergency response | Cross-institutional protocols, open-data policies, governance agreements |

One of the most common issues is standardization since the datasets are frequently generated with various protocols, units, time, calibration procedures, and quality-control regulations [82]. This makes comparability less and model transfer difficult. Institutional fragmentation also affects governance by other agencies, utilities, companies, and researchers because data can be stored in various agencies, where the priorities differ. There are datasets that are siloed, and those that are shared openly without proper documentation may result in misunderstanding.

The issue of cybersecurity is gaining more importance with the integration of monitoring systems with vital water infrastructure[83]. There are also practical limits to real-time analytics. Algorithms that are fast are not sufficient; effective real-time monitoring needs to have a steady stream of data, automatic quality monitoring, results that can be understood, and organizations willing to take action to respond to the warning. The problem lies, after all, not in dealing with big data, but in developing credible, interoperable, and decision-relevant data systems. **Figure 2** summarizes the big-data workflow in water quality monitoring, from multi-source acquisition to preprocessing, harmonization, fusion, and analytics-ready outputs.

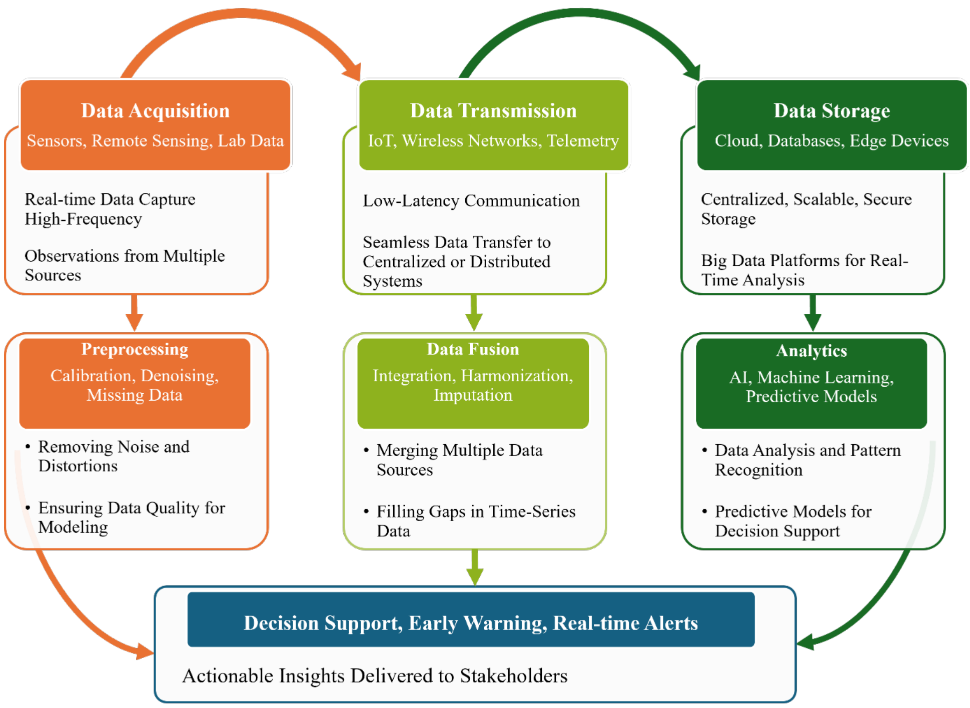


**Figure 2.** End-to-end big-data workflow for water quality monitoring, including data acquisition, transmission, storage, preprocessing, fusion, and downstream analytical and decision-support applications.

# 4. Artificial Intelligence for Water Quality Assessment and Prediction

## 4.1. The Role of AI in Modern Water Quality Monitoring

Artificial intelligence has gained more significance in water quality monitoring since the issue at hand is not just the acquisition of data anymore, but processing[84]. Contemporary surveillance systems produce high-volume data, which is heterogeneous and sometimes high-frequency data that cannot be analyzed easily with traditional statistical instruments. Correlations between water quality variables are seldom straightforward and constant. They are influenced by nonlinear interactions of hydrological states, pollutant inflow, biological activity, climatic forcing and infrastructure processes. In these circumstances, AI provides a versatile collection of methods in the extraction of patterns, estimation of hidden relations, and assisted prediction, which conventional linear models are not always capable of.

Complexity is a significant strength of AI in this field of interest. Multivariate dependencies, temporal autocorrelation, spatial heterogeneity, and incomplete observations are common quality characteristics of water quality datasets[85]. AI techniques are highly appropriate in such circumstances since they are capable of learning based on the data without having to define each of the processes a priori. It is particularly handy when the mechanistic understanding is incomplete, when there are many drivers in the environment, or when the goals of monitoring need quick prediction, as opposed to the extensive process simulation. Consequently, AI has been used in activities like water quality recognition, contamination, surrogate models, prediction of nutrient and oxygen dynamics, algal bloom, and optimization of the treatment process[13].

Nevertheless, the growing popularity of AI must not be confused with a solution that is universal. In most instances, predictive performance is given more attention than environmental interpretation. Models can be very useful using historical data, and at the same time, be very delicate when the conditions change, the observations are sparse, or when the sites change. Furthermore, it may appear that AI is sophisticated, which may conceal the fact that the quality of the model is strongly dependent on the quality of the data, preprocessing options, and problem definition. In water quality monitoring, where uncertainty is widespread, and the management decisions can be high-stakes, the quality of AI performance is not just related to the end product, which is accurate, but is also done in a manner that is transparent, robust, and operationally significant[14,86].

An important attitude is thus required. AI is not to be taken as a substitute for environmental reasoning; it is supposed to be a layer of analysis in a larger system of monitoring[87]. Its effectiveness lies in the extent to which it is connected to sensing systems, data management practices, and decision-support requirements. The key question is not whether it is possible to apply AI to the data on water quality, but in what circumstances it can really help to assess and predict it.

## 4.2. Machine Learning for Assessment, Classification, and Anomaly Detection

Machine learning is now the most popular field of AI in water quality research due to its ability to offer practical applications of regression, classification, clustering, and anomaly detection of various environmental data[15]. Machine learning algorithms have the benefit over traditional empirical models of being able to extend nonlinear relationships and interactions between predictors with flexibility, which is appealing to systems where water quality depends upon several coupled drivers. Machine learning is applied to a wide range of applications to estimate variables that are hard to measure with proxy signals, classify the conditions of water quality into the risk categories, identify the abnormal events, or the indices that are used to summarize the environmental status[88].

The ability of machine learning to adapt to both large and small data volumes, as well as its use in various contexts of monitoring, is one of its strengths[69]. It may be used on small datasets of sites, on larger regional databases, and even more on integrated sensor networks. Algorithms such as support vector machines, random forests, gradient boosting algorithms, k-nearest neighbors, and shallow artificial neural networks have all been applied to water quality studies since they offer predictive flexibility, but with relatively manageable levels of computational requirements. The models are particularly well-performing in cases when the relationships are not linear, yet the available data is not large enough to warrant more sophisticated deep-learning models.

However, the ubiquitous application of machine learning also demonstrates a number of drawbacks. To start with, performance is usually highly dependent on training data representativeness[89]. A model, which was trained in a different hydrological regime, pollution profile, or even seasonal pattern, would not produce good results when the conditions vary. Second, most studies emphasize a comparison of algorithms but do not pay a lot of attention to the environmental importance of predictors, uncertainty of measurements, or external validity of findings[90]. Third, the detection of anomalies in specific cases is not always the easiest since anomalies in the water systems do not necessarily constitute mistakes; they can be the most significant findings as far as the environment is concerned[11]. Therefore, machine learning can enrich the monitoring potential, though with one important condition: to carefully interpret what the data and projections mean.

### 4.2.1. Supervised Learning Models

The most common method of AI-based water quality analysis is supervised learning, since a large number of monitoring tasks may be converted into prediction problems with known answers[91]. Continuous variables that are estimated by the use of regression models include dissolved oxygen, nutrient concentration, biochemical oxygen demand surrogates, or water quality indices. Assigning samples or time periods to pre-existing categories like polluted and unpolluted, eutrophic and non-eutrophic, or safe after use and unsafe after use are classification models used to specify the samples or time periods. These activities are compatible with the format of supervised learning, whereby models are trained on labeled examples and applied to unseen cases.

The practical usefulness of supervised learning lies in the fact that it is able to use proxy relationships. Numerous variables of water quality of interest are costly, time-consuming, or immeasurable in practice. They can be determined by machine learning models based on more easily measured signals like temperature, turbidity, pH, conductivity, spectral measurements, and hydrometeorological drivers. This has rendered supervised learning especially useful with regard to operational monitoring, whereby continuous estimation can be used to prolong the usefulness of scarce sensing infrastructures [92]. It also assists in the process of bridging high-frequency but non-direct measurements and the lower-frequency laboratory analysis.

Simultaneously, supervised learning accepts the constraints of its labeling [93]. Training targets can also be uncertain, sparse, or varying in sites and studies. In environmental systems, labels would be frequently produced in accordance with certain protocols and would not necessarily represent process variations. A model that has been trained on past labels can thus replicate past measurement structures without necessarily making inferences to the future. Also, when datasets are imbalanced, temporally autocorrelated, or poorly independent between training and testing, high reported accuracy may be false. These problems imply that supervised learning cannot be considered on the basis of predictive measures, but also the realism of the validation strategy to field deployment conditions [91].

#### 4.2.2. Unsupervised Learning and Clustering

Unsupervised learning has a similar but equally significant role in monitoring water quality, especially when labels are not available, incomplete and unreliable [94]. These procedures are intended to find patterns, groupings, and latent structure in data instead of forecasting already predetermined outputs. Clustering and dimensionality reduction are widely applied in water quality studies to detect spatial similarity between monitoring sites, seasonal regimes, the pattern of pollution sources, or obscure relationships between variables. The methods particularly come in handy when analyzing the system on an exploratory basis, wherein one seeks to get familiar with how the system is organized, then build predictive models.

Unsupervised learning has the advantage of showing structure that cannot be seen visually [95]. Interacting gradients, as opposed to delimited categories, tend to control water quality systems, and clustering may assist in reducing this complexity to more understandable regimes. The visualization, screening of variables and noise reduction of high-dimensional data of high dimensional datasets (particularly generated by spectroscopic sensors or integrated monitoring platforms) can also be assisted by dimensionality reduction techniques.

Several limitations must however, be put on unsupervised learning since the clusters or latent components it detects are not necessarily environmental facts, but are mathematical structures [96]. Algorithms, or choices of parameters, can result in different groupings, and even visible patterns can be partly due to scaling decisions, sparseness in the data, or preprocessing decisions. Besides, the environmental sense of clusters is frequently retro-proscribed, which creates a threat of overinterpretation. Unsupervised approaches can thus be regarded as hypothesis-creating tools and not necessarily descriptions of the dynamics of water quality [96]. They are potentially of most use in informing the analysis that follows, uncovering latent heterogeneity, and informing a more specific monitoring design.

#### 4.2.3. Ensemble Learning Approaches

In the prediction of water quality, ensemble learning has attracted attention due to its ability to generate robustness when more than one model is used in water quality prediction, as opposed to using a single algorithm [97]. Random forests, gradient boosting, bagging-based frameworks, and stacked model combinations are methods that can, in many cases, obtain better predictive performance than the individual learners, particularly in nonlinear and noisy environmental data sets. It is especially helpful in water quality, where predictors and outcomes can be related differently depending on the season, location, and hydrological conditions [98].

The primary advantage of ensemble approaches lies in the fact that sensitivity to the weaknesses of an individual model is diminished [99]. They can be used to enhance generalization and decrease overfitting by combining several predictions. Practically, this renders them appealing to use in predicting water quality indices, in event classification, and in estimating contaminants through proxy variables. Variable importance measures are also provided by some ensemble methods, and can be somewhat more interpretable than more opaque models.

However, the performance may not necessarily imply a higher understanding. Computationally heavier and methodologically less transparent Ensemble models may increasingly become complex. Variable importance measures can indicate effect without showing the mechanisms of causation, and high prediction can be accompanied by low transferability when the training data is limited in nature. The literature also tends to take ensemble superiority in the benchmark comparisons as an adequate indication of more widespread applicability. As a matter of fact, the usefulness of ensemble learning is determined by the fact of whether enhanced prediction is translated into more confident and understandable assistance in monitoring decisions [100].

## 4.3. Deep Learning for Spatiotemporal Prediction

There is an increasing demand for water quality monitoring with deep learning due to its strong preference for large, complex datasets with both spatial and temporal dimensions [101]. Deep-learning architectures can find in raw or semi-processed data hierarchical representations which, unlike many traditional machine learning models, are appealing to sensor time series, image-based measurements, spectral information, and integrated multi-source information. In water quality, deep learning has found more and more applications in time-series forecasting, remote-sensing retrieval, anomaly detection, and event prediction tasks, particularly where the system dynamics are extremely nonlinear and determined by various interacting variables.

The primary strength it has is representational power [102]. Water quality signals usually have dependencies that are across time, space, and groups of variables that are not easily specified. Deep-learning models are able to learn such patterns more easily than shallow models, especially in the case of large datasets. This is useful in predicting the variability of such tasks as dissolved oxygen, changes in nutrient concentration, salinity, and the development of harmful algal blooms, where both the short-term variability and longer-term dependencies are important.

However, deep learning also aggravates the fundamental issues that exist in AI-based water quality analysis [103]. It normally takes a large amount of training data, requires a lot of computational time, and is more sensitive to tuning than standard machine learning. It is also not always very interpretable, and this can diminish the scientific trust and operational acceptance. In environmental monitoring, where datasets are often constrained, location-specific, or due to measurement error, the complexity of deep learning can imply improvements without necessarily having strength. Therefore, the question to answer is not whether deep learning can be more successful than simplistic models at specific scenarios, but whether the benefits of deep learning outweigh the intended complexity in the conditions of real monitoring [104]. A summary of the key AI approaches, their functions, their advantages, and disadvantages is given in **Table 3**.

**Table 3.** Main artificial intelligence methods applied in water quality monitoring, together with their typical analytical tasks, advantages, and key limitations.

| AI Category | Representative Methods | Main Tasks in Water Quality Monitoring | Strengths | Main Limitations |
|---|---|---|---|---|
| Supervised machine learning | Support vector machines, random forests, gradient boosting, shallow neural networks | Regression, classification, surrogate estimation, water quality index prediction | Good nonlinear fitting, flexible, effective with moderate datasets | Depends on labeled data, limited transferability, sensitive to data representativeness |
| Unsupervised learning | k-means, hierarchical clustering, PCA, t-SNE/UMAP | Pattern discovery, site grouping, source identification, dimensionality reduction | Useful when labels are scarce, supports exploratory analysis | Clusters may lack physical meaning, interpretation can be subjective |
| Ensemble learning | Random forest, XGBoost, stacking, bagging | Robust prediction, event classification, anomaly detection | Improved generalization, reduced overfitting, often strong predictive performance | More complex interpretation, may still lack transferability |
| CNN-based deep learning | 1D CNN, 2D CNN, hybrid CNN models | Remote sensing retrieval, spectral analysis, local pattern extraction | Strong feature learning from images and spectra | Data-hungry, calibration dependence, reduced interpretability |
| Sequence deep learning | RNN, LSTM, GRU, transformers | Time-series forecasting, pollution-event prediction, process dynamics modeling | Captures temporal dependencies and nonlinear dynamics | Computationally intensive, weak transparency, may fail under changing regimes |

**Table 3.** *Cont.*

| AI Category | Representative Methods | Main Tasks in Water Quality Monitoring | Strengths | Main Limitations |
|---|---|---|---|---|
| Explainable AI | SHAP, LIME, feature importance, partial dependence | Model interpretation, driver analysis, trust building | Improves transparency and scientific credibility | Explanations may be approximate and method-dependent |
| Uncertainty-aware AI | Bayesian models, prediction intervals, probabilistic forecasting | Risk-informed prediction and decision support | Supports confidence assessment and more reliable deployment | Often more difficult to implement and communicate |

#### 4.3.1. CNN-Based Methods

Convolutional neural networks are typically related to the analysis of images, whereas they can be used more widely in the context of water quality monitoring since a significant number of environmental datasets have spatial or locally organized details [105]. CNNs are especially applied in remote sensing to derive spatial features of multispectral or hyperspectral imagery to better estimate the surface water properties of chlorophyll concentration, suspended solids, or the extent of a bloom. They have strengths in learning spatial hierarchies directly out of image data, which may minimize the need to have hand-crafted features.

As well, CNNs may be trained to take one-dimensional time or spectral feeds where frequent local temporal or wavelength patterns are informative [106]. This allows them to be applied in the analysis of continuous sensors or spectroscopic measurements. Conventional architectures are able to identify local motifs, sudden transitions, and combinations of features that can be challenging to identify using simpler regression algorithms.

Nevertheless, the CNN performance is highly reliant on the availability and representativeness of the data [107]. Models trained in one region, type of sensor, or optical conditions might not be transferred easily to the other. They also share the shortcomings of satellite retrieval per se involved in remote sensing, such as atmospheric interaction, spatial resolution, and the indirectness of optical proxies. Thus, CNNs have the ability to enhance pattern extraction, but their result still has to be viewed in light of the constraints of the data source, both physical and observational.

#### 4.3.2. Recurrent, LSTM, and Transformer-Based Methods

One of the most crucial tasks in the field of water quality monitoring, time-series prediction has become a very significant aspect of AI usage, and recurrent architecture has consequently been extensively used [108]. RNNs and LSTM networks, in particular, are constructed to model the temporal relationships, and thus they are appropriate in predicting the variables of dissolved oxygen, pH, nutrient level, algal bloom signals, and performance of the treatment process. Their popularity is explained by the fact that the water quality is dynamic in nature, and most of the major variables do not only rely on the current situation, but also on the series of past states.

The LSTM-based models are especially useful when effects are delayed and persistent in time, as is the case with oxygen depletion after organic loading or bloom development in changing meteorological conditions [109]. Transformer-based models have been considered more recently as they are more efficient and flexible in representing long-range dependencies than more traditional recurrent structures. It provides new opportunities to forecast in systems with a complex temporal interaction and long memory.

However, time-based deep learning brings about a number of challenges. The accuracy of forecasting can reduce drastically in cases where the future is not the same as the training information. Sequences models may also be data-intensive and hard to interpret (e.g., when several predictors have time delays). Moreover, good short-term forecasting is not necessarily associated with environmental knowledge. A model can describe the regularities in time without modeling the processes that create them. This difference is important in water quality monitoring since a final aim may not just be to predict things, but also to take plausible action in the face of uncertainty [110].

### 4.4. Explainable AI, Uncertainty, and Interpretability

The demand for interpretability increases with the strength of AI models. Predictions have been common in water quality monitoring to assist in environmental management, treatment operation, or contamination warning [111]. In these circumstances, a statistically performing model that cannot be explained might not be easy to believe or accept. Explainable AI has thus become a significant field of research, in an effort to understand how models come to their conclusions and what inputs have the greatest impact on the prediction.

In environmental systems, interpretability is particularly valuable since it is hardly ever desired to achieve prediction [112]. It is common that researchers and practitioners must learn how to know the agents of change in water quality, review the physical plausibility of model behavior, and determine the ability to predict under extreme conditions. This can be addressed by the use of feature importance analysis, partial dependence methods, the use of local explanation tools, and the use of hybrid interpretable architectures. The techniques do not completely remove model opacities, although they are more transparent, and predictions can be evaluated against the knowledge about the domain.

Uncertainty is also essential [113]. Not only does model structure influence the prediction of water quality, but also noisy inputs, missing data, calibration drift, and changing environmental regimes. However, prediction of points is a common focus of AI research with little consideration of uncertainty limits or confidence estimates. It is a major shortcoming since when making decisions based on AI output, the expected conditions should not only be reflected, but also the level of confidence of the expected conditions. In practice, uncertainty-aware prediction can be useful even when average accuracy increases by a small margin, but there is no indication of reliability.

The more general difficulty is that explainability and uncertainty are frequently considered as secondary objectives to be added to the design instead of primary design objectives [13]. This is not adequate for water quality monitoring. Scientific and operational scientific models ought to be readable enough to foster trust and uncertain enough to foster responsible actions. The further applicability of AI in the discipline might rely not only on more powerful prediction, but also on more open and responsible modeling protocols.

### 4.5. Limitations, Transferability, and Methodological Challenges

Despite the fast development, AI in water quality monitoring continues to have significant methodological weaknesses [13]. Transferability is one of the most endemic ones. Most models are built and tested on data from one watershed, reservoir, treatment plant, or monitoring campaign. They tend to work under such local conditions but fail when used in other places. This is because the dynamics of water quality are determined by the site-based hydrology, climate, sources of pollution, the architectural design of infrastructures, and the surveillance measures. These patterns can be successfully learned by AI models; however, this does not necessarily imply that it will be generalized outside the contexts that it was trained in which they were trained.

Another significant limitation is the data limitations. Even though the concept of water quality monitoring is frequently referred to as a big-data field, numerous datasets are still small in the most important aspects regarding AI [69]. The labels can be sparse, the rare events could be underrepresented, and the long-term records could be characterized by inconsistent practices in measurement. The sensor streams with high frequencies can generate very large numbers of observations, which are, however, highly autocorrelated, and might not contain as much independent information as the volume of their data would suggest. This makes it difficult to train and evaluate such models, especially in instances where it is based on random data splitting, which overestimates prediction capability due to a lack of recreation of realistic deployment conditions.

Another issue is the increasing concern about reproducibility [114]. It is common to compare various algorithms in many studies that employ varying datasets, preprocessing pipelines, evaluation metrics, and validation schemes, and thus, direct comparison is challenging. Benchmarking can be highly fragmented, and trained models or code cannot always be published publicly. Consequently, the literature may suggest the impression of a quick-fix nature and provide little foundation upon which a cumulative evaluation would be possible. This issue is particularly acute when it comes to an interdisciplinary area, as environmental scientists, engineers, and data scientists might adhere to various conventions.

These constraints imply that the future phase of the AI advancement of water quality monitoring ought to put less emphasis on creating an increasingly intricate algorithm and more on increasing the methodological soundness. Improved benchmark data sets, cross-site validation, uncertainty reporting, explainable modeling, and integration with knowledge of the environmental processes will typically be more useful than marginal improvements in the isolated prediction metrics[115]. AI can change the way of water quality assessment and prediction, yet the contribution it will bring in the long term will be defined by whether it will become a future stable element of operational environmental intelligence or a collection of promising models.

# 5. Integrated Monitoring Frameworks and Practical Applications

## 5.1. From Fragmented Monitoring to Integrated Frameworks

The trend of water quality monitoring is to shift towards systems that are no longer solitary but instead systems that integrate sensing, data delivery, analytics and decision support into one working framework[19]. This shift is significant due to the fact that none of the cited elements, such as a high-performance sensor or a strong AI perspective, alone can provide credible and operational monitoring. The sensors produce observations, data structures store and transmit them, Artificial Intelligence derives trends and forecasts, and management applications convert them into alerts or actions. The actual worth of the contemporary monitoring is in the merit of the connection of these layers.

A collaborative framework is hence more than technical aggregation. It has a systems orientation whereby water quality monitoring is seen as a constant information flow and not as a series of independent activities[116]. These frameworks are essential, especially in complicated water settings where the environment has changed rapidly, the sources of pollution are dispersed, and the stakeholders are numerous and need combined efforts. In this respect, integration is not merely a technological aspiration, but it is a functional need.

## 5.2. Architectures of Intelligent Monitoring Systems

The modern integrated monitoring systems are usually structured around a multi-layer architecture[117]. Sensors, networks, and mobile platforms in the front-end capture physical, chemical, and in some cases biological signals. Those are being relayed to storage and processing systems, usually with the help of cloud or hybrid infrastructure via the communication networks. Based on this layer of data, AI models can be used to make an estimation, identify anomalies, predict, or recognize patterns. Lastly, decision-support interfaces, alert systems, and visualization tools transform the outputs of the analytics into a format that is usable by utilities, regulators, or environmental managers.

One of the key benefits of this architecture is that it enables the monitoring to change to active management rather than a passive one[118,119]. Rather than merely documenting the water quality state, the integrated systems may be used to detect unusual situations, anticipate probable degradation, and assist with quicker intervention. But system intelligence is reliant on the weak links of the chain. A very precise model would not be of much use when the sensor drift is not solved, and communication cannot be trusted, as well as the output is not understandable to the decision-makers. In this way, it is not merely a question of how to design every layer properly, but it is also necessary to make each of them functionally coherent with the others.

## 5.3. Edge-Cloud Collaboration and Digital Water Systems

The idea of edge-cloud collaboration has become a significant approach to smart water monitoring since it offers a combination of local responsiveness and high analytical capabilities[120]. The edge devices are able to handle the data close to the members of observation and can filter the data fast, screen anomalies, or send out local alerts with lower latency and bandwidth requirements. Cloud systems, on the other hand, are more suitable for large-scale storage, historical analysis, model training, and cross-site integration. They combine to provide responsive as well as scalable monitoring systems.

This architecture can be even more important when applied to digital water systems and digital twins. A digital twin may be seen as a dynamic virtual image of a physical water system, which is constantly fed with real-time information[121]. Such systems, in principle, enable the monitoring, simulation, prediction, and control to be integrated into a single platform. To manage water quality, this introduces the aspect of surpassing the reactive-based monitoring into the anticipatory-based control. However, its practical use is still difficult due to the need for stable data flows, dependable calibration, and a high level of adherence of data-driven and process-based models in digital twins. The potential is quite substantial, but their performance is subject to the maturity of the monitoring infrastructure on which they are built.

## 5.4. Decision Support and Early Warning Applications

The most obvious advantage of unified monitoring systems is that they can help in decision-making rather than data gathering. Integrated platforms may be employed in drinking water systems to detect the risks of contamination, reveal abnormal conditions of the distribution network, and assist in monitoring compliance [122]. They can enhance monitoring of eutrophication and algal blooms, sediment pulses, and episodic pollution in rivers, lakes, and reservoirs. They can be used in wastewater treatment and reuse systems to facilitate the optimization of processes, control of effluents, and resilience of treatment.

One of the most significant functions in this case is early warning [123]. A monitoring framework is operationally significant if it is capable of identifying emergent risk at an opportune time. Nevertheless, an efficient warning mechanism is tied not only to rapid analytics but also to the believability and decipherability of alerts. Excessively sensitive systems will produce false alarms, whereas overly conservative systems will fail to capture important events. The real issue of concern is then to come up with practical warning mechanisms that are responsive and trusted and the thresholds and outputs should be sensitive to the capacity by which an institution responds.

**Table 4** summarizes the representative areas of application of integrated monitoring frameworks and their objectives, data sources and challenges of implementation.

**Table 4.** Integrated water quality monitoring frameworks across major water environments, including monitoring objectives, typical data sources, AI functions, implementation challenges, and expected management benefits.

| Water Environment | Core Monitoring Objectives | Typical Data Sources | AI/Analytics Functions | Main Implementation Challenges | Expected Management Benefits |
|---|---|---|---|---|---|
| Drinking water systems | Contamination detection, residual disinfectant tracking, and distribution safety | Online sensors, lab tests, hydraulic data, operational logs | Anomaly detection, residual prediction, risk alerting | High reliability requirements, regulatory compliance, false alarm sensitivity | Faster warning, safer distribution management, compliance support |
| Rivers and streams | Pollution surveillance, runoff event detection, watershed assessment | In situ sensors, hydrometeorology, land-use data, grab samples | Event detection, source attribution, trend forecasting | High temporal variability, nonpoint-source complexity, site heterogeneity | Improved catchment management and pollution response |
| Lakes and reservoirs | Eutrophication tracking, bloom forecasting, stratification assessment | Buoy sensors, remote sensing, weather data, nutrient sampling | Bloom prediction, spatiotemporal mapping, regime classification | Spatial heterogeneity, seasonal dynamics, depth-related mismatch | Better ecosystem protection and water supply risk management |
| Groundwater systems | Contamination plume tracking, salinity intrusion, long-term quality change | Wells, lab analysis, hydrogeological data, sparse sensor networks | Trend analysis, interpolation, anomaly detection | Sparse observations, delayed response, subsurface uncertainty | Stronger aquifer protection and long-term planning |
| Coastal and estuarine waters | Salinity dynamics, nutrient loading, harmful algal bloom surveillance | Satellite data, buoys, tide and flow records, field sampling | Spatial mapping, bloom warning, multivariate forecasting | Mixing complexity, cloud cover, scale mismatch | Improved coastal risk control and ecosystem management |
| Wastewater treatment and reuse | Process optimization, effluent quality control, reuse safety | Online process sensors, SCADA data, laboratory measurements | Process prediction, control optimization, fault diagnosis | Sensor fouling, complex influent variability, operational integration | Smarter treatment control, lower cost, improved effluent reliability |

## 5.5. Barriers to Implementation and Future Directions

Nevertheless, even with recent substantial advances, integrated water quality monitoring continues to experience significant impediments to practice. Such technical issues as sensor stability, calibration effects, heterogeneity of data, and hard-to-learn transferability of AI models are troublesome [14]. The economic factors are still a major problem since installing, maintaining, communicating, and expert operation may be expensive, particularly in large systems or systems with limited resources. The barriers at the institutional level are also significant. Disjointed governance, poor standardization, and insufficient coordination among scientists, utilities, and regulators tend to obstruct complete integration of the entire system despite the availability of individual technologies [124].

The development in the future is thus expected to rely less on single-purpose innovation and more on system design. All these magic wizards will need tougher low-maintenance sensors, interoperable information fields, glorifiable AI structures, and clinical assistant devices better adjusted to operational realities [125]. The general trend is to independent, adaptive and predictive monitoring ecosystems where sensing, analytics, and control are components of an overarching environmental intelligence system. The importance of integrative frameworks is just in this shift: they transform the nature of water quality monitoring into the kind of practice, which is an action, rather than a measurement. **Figure 3** presents an integrated intelligent monitoring architecture and its future evolution toward predictive and adaptive water management.

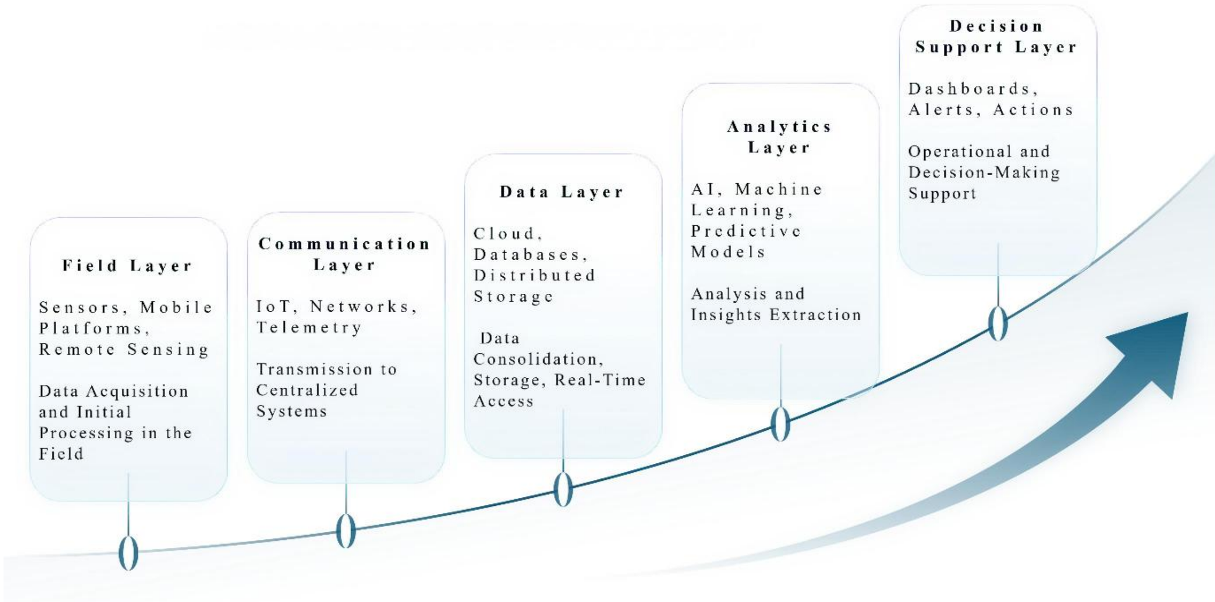


**Figure 3.** Integrated intelligent water quality monitoring framework linking sensing, communication, storage, AI analytics, digital twins, and decision support, with progression from conventional monitoring to predictive and adaptive management.

# 6. Conclusions

The implementation of water quality monitoring is in the middle of a paradigm shift from being measured periodically to being continuously measured, data-driven, and more intelligent. It has been demonstrated in this review that the emergence of big data is transforming the discipline not only by increasing sensing power but also by boosting the significance of data infrastructures, artificial intelligence, and trans-architectural monitoring systems. High-frequency and multi-scale observation of a variety of water environments is now possible thanks to modern sensor devices, and AI solutions offer new opportunities to assess, predict, and identify abnormalities and support operations. Simultaneously, these developments have led to the realization that an increase in data does not necessarily lead to improved monitoring. It can be said that the sensing devices are not the sole determinants of the reliability and usefulness of water quality information because calibration, preprocessing, interoperability, uncertainty management, and governance are also important determinants.

This review offers several distinctive contributions. First, it explicitly treats sensors, big-data infrastructure, and AI as mutually reinforcing layers of a single monitoring ecosystem, rather than as separate subjects. Second, it emphasises data-centric issues (preprocessing, harmonisation, interoperability) that are often overlooked despite their decisive impact on model reliability. Third, it critically assesses AI not only on predictive performance but also on explainability, uncertainty, generalisability, and reproducibility. Fourth, it promotes integrated framework architectures that connect sensor networks, data platforms, analytics, and decision-support systems for operational use. These frameworks can be particularly valuable in complex and dynamically evolving environments where the traditional methods of monitoring are too slow, too sparse, or too fragmented to facilitate effective intervention.

There are also some long-standing limitations, which have been pointed out in this review. The limitations to the dependability of real-time observation remain sensor drift, fouling, incomplete biological monitoring, and unevenness of field robustness. Heterogeneity, missing data, loose standardization, and fractured governance persist as challenges to big-data systems. Regardless of excellent predictive capabilities, AI models are frequently associated with the issues of transferability, interpretability, reproducibility, and uncertainty quantification. These problems suggest that technical innovation is not as much of a problem as before, but methodological and institutional integration. Even sophisticated monitoring systems will not be able to produce credible and effective results without clear standards, quality assurance, and operational preparedness.

In the future, the convergence of low-maintenance sensing technologies, interoperable data platforms, explainable and uncertainty-aware AI, and adaptive decision-support systems is likely to become the next step in the advancement. The new trends include edge-cloud collaboration, digital twins, autonomous monitoring platforms, and hybrid data-driven-process-based modeling that have the potential to provide new opportunities in more predictive and resilient water management. Nonetheless, to accomplish this vision, more effort will be needed in enhanced cooperation between environmental science, engineering, computer science, and policy. It will also necessitate a shift in the focus of research away, on one hand, to single instances of demonstration of performance, and, on the other hand, to scalable, transferable, and field-ready solutions.

To sum up, the surveillance of water quality in the age of big data should be perceived as a developing environmental intelligence system as opposed to a set of measurement instruments. Its long-term value is in facilitating the shift to the proactive and informed type of management instead of the reactive one. The future of monitoring systems is a more effective combination of sensors, AI, and operating structures that can help to maintain water safety, protect the ecosystem, and ensure the sustainable management of resources in a world that is getting more uncertain.

## Funding

This work received no external funding.

## Institutional Review Board Statement

Not applicable.

## Informed Consent Statement

Not applicable.

## Data Availability Statement

Not applicable. No new data was created and analyzed.

## Conflicts of Interest

The author declares no conflict of interest.

## AI Use Statement

The author declares that no artificial intelligence (AI) tools were used in the preparation of this manuscript.

## References

[1] Akhtar, N., Syakir Ishak M.I., Bhawani S.A., et al., 2021. Various natural and anthropogenic factors responsible for water quality degradation: A review. Water. 13(19), 2660.

[2] McGrane, S.J., 2016. Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: A review. Hydrological Sciences Journal. 61(13), 2295–2311.

[3] Mullungal, M.N., Peediyakkathodi S., Bibi S., et al., 2024. Nutrient contamination in marine environment. In: Alshemmari, H., Hashmi, M.Z., Kavil, Y.N., et al. (Eds.). Contaminated Land and Water: Remediation and Management. Springer: Cham, Switzerland. pp. 15–33.

[4] Ahmed, U., Mumtaz, R., Anwar, H., et al., 2020. Water quality monitoring: From conventional to emerging technologies. Water Supply. 20(1), 28–45.

[5] Borah, S.S., Khanal, A., Sundaravadivel, P., 2024.

Emerging technologies for automation in environmental sensing. Applied Sciences. 14(8), 3531.

[6] Jaywant, S.A., Arif, K.M., 2024. Remote sensing techniques for water quality monitoring: A review. Sensors. 24(24), 8041.

[7] Wang, Z.A., Moustahfid, H., Mueller, A.V., et al., 2019. Advancing observation of ocean biogeochemistry, biology, and ecosystems with cost-effective in situ sensing technologies. Frontiers in Marine Science. 6, 519.

[8] Chen, L., Wang, L., 2018. Recent advance in earth observation big data for hydrology. Big Earth Data. 2(1), 86–107.

[9] Fang, K., Wu, F., Gao, X., et al., 2026. Multi-source data fusion and heuristic-optimized machine learning for large-scale river water quality parameters monitoring. Remote Sensing. 18(2), 320.

[10] Teh, H.Y., Kempa-Liehr, A.W., Wang, K.I.-K., 2020. Sensor data quality: A systematic review. Journal of Big Data. 7(1), 11.

[11] Dogo, E.M., Nwulu, N.I., Twala, B., et al., 2019. A survey of machine learning methods applied to anomaly detection on drinking-water quality data. Urban Water Journal. 16(3), 235–248.

[12] Sheik, A.G., Kumar, A., Patnaik, R., et al., 2024. Machine learning-based design and monitoring of algae blooms: Recent trends and future perspectives—A short review. Critical reviews in environmental science and technology. 54(7), 509–532.

[13] Shende, A.D., Tekade, S.A., Deshmukh, A.A., et al., 2025. Water quality monitoring and control with AI. In: Ruiz-Vanoye, J.A., Díaz-Parra, O. (Eds.). Smart Water Technology for Sustainable Management in Modern Cities. IGI Global Scientific Publishing: Hershey, PA, USA. pp. 51–80.

[14] Zou, S., Ju, H., Zhang, J., 2025. Water quality management in the age of AI: Applications, challenges, and prospects. Water. 17(11), 1641.

[15] Ahmed, A.N., Othman, F.B., Afan, H.A., et al., 2019. Machine learning methods for better water quality prediction. Journal of Hydrology. 578, 124084.

[16] Abdulhussain, S.H., Mahmmod, B.M., Alwhelat, A., et al., 2025. A comprehensive review of sensor technologies in IoT: Technical aspects, challenges, and future directions. Computers. 14(8), 342.

[17] Elouataoui, W., 2024. AI-driven frameworks for enhancing data quality in big data ecosystems: Error_detection, correction, and metadata integration. arXiv preprint arXiv:2405.03870.

[18] Aderemi, I.A., Kehinde, T.O., Ugochukwu, D.O., et al., 2025. Beyond the black box: A systematic review of explainable AI for transparent and trustworthy water quality monitoring. IEEE Sensors Reviews. 2, 419–443.

[19] Miller, T., Durlik, I., Kostecka, E., et al., 2025. Integrating artificial intelligence agents with the internet of things for enhanced environmental monitoring: Applications in water quality and climate data. Electronics. 14(4), 696.

[20] Dong, J., Wang, G., Yan, H., et al., 2015. A survey of smart water quality monitoring system. Environmental Science and Pollution Research. 22(7), 4893–4906.

[21] Patil, P., Sawant D., Deshmukh, R., 2012. Physico-chemical parameters for testing of water—A review. International Journal of Environmental Sciences. 3(3), 1194.

[22] Verma, S., Verma, S., Ramakant, R., et al., 2025. Comprehensive assessment of physico-chemical and biological parameters in water quality monitoring: A review of contaminants, indicators, and health impacts. International Journal of Horticulture, Agriculture and Food Science. 9(2), 611019.

[23] Atique, U., An, K.-G., 2019. Reservoir water quality assessment based on chemical parameters and the chlorophyll dynamics in relation to nutrient regime. Polish Journal of Environmental Studies. 28(3), 1043–1061.

[24] Kennedy, L.C., Miller, S.E., Kantor, R.S., et al., 2021. Effect of disinfectant residual, pH, and temperature on microbial abundance in disinfected drinking water distribution systems. Environmental Science: Water Research & Technology. 7(1), 78–92.

[25] Parsa, Z., Dhib, R., Mehrvar, M., 2024. Dynamic modelling, process control, and monitoring of selected biological and advanced oxidation processes for wastewater treatment: A review of recent developments. Bioengineering. 11(2), 189.

[26] Pérez-Ruzafa, A., Campillo, S., Fernández-Palacios, J.M., et al., 2019. Long-term dynamic in nutrients, chlorophyll a, and water quality parameters in a coastal lagoon during a process of eutrophication for decades, a sudden break and a relatively rapid recovery. Frontiers in Marine Science. 6, 26.

[27] e Silva, G.M., Campos, D.F., Brasil, J.A.T., et al., 2022. Advances in technological research for online and in situ water quality monitoring—A review. Sustainability. 14(9), 5059.

[28] Alam, A.U., Clyne, D., Jin, H., et al., 2020. Fully integrated, simple, and low-cost electrochemical sensor array for in situ water quality monitoring. ACS Sensors. 5(2), 412–422.

[29] Uppuluri, K., 2023. Smart sensors for monitoring pH, dissolved oxygen, electrical conductivity, and temperature in water. In: Manjakkal, L., Lorenzelli, L., Willander, M. (Eds.). Sensing Technologies for Real Time Monitoring of Water Quality. Wiley: Hoboken, NJ, USA. pp. 3–23.

[30] Huang, Y., Wang, X., Xiang, W., et al., 2022. Forward-looking roadmaps for long-term continuous

water quality monitoring: Bottlenecks, innovations, and prospects in a critical review. Environmental Science & Technology. 56(9), 5334–5354.
[31] Delgado, A., Briciu-Burghina, C., Regan, F., 2021. Antifouling strategies for sensors used in water monitoring: Review and future perspectives. Sensors. 21(2), 389.
[32] Yaroshenko, I., Kirsanov, D., Marjanovic, M., et al., 2020. Real-time water quality monitoring with chemical sensors. Sensors. 20(12), 3432.
[33] Lal, K., Jaywant, S.A., Arif, K.M., 2023. Electrochemical and optical sensors for real-time detection of nitrate in water. Sensors. 23(16), 7099.
[34] Zhang, H., Tian, Z., Che, X., et al., 2025. Detection of water quality cod based on the integration of laser absorption and fluorescence spectroscopy technology. Water. 18(1), 93.
[35] Huang, C., Chen, Y., Zhang, S., et al., 2018. Detecting, extracting, and monitoring surface water from space using optical sensors: A review. Reviews of Geophysics. 56(2), 333–360.
[36] Kumar, M., Khamis, K., Stevens, R., et al., 2024. In-situ optical water quality monitoring sensors—Applications, challenges, and future opportunities. Frontiers in Water. 6, 1380133.
[37] Azil, K., 2021. Optical Methods for Detection and Analysis the Pollutants in Water [PhD Thesis]. Université Ferhat Abbas Sétif 1: Sétif, Algeria.
[38] Jang, A., Zou, Z., Lee, K.K., et al., 2011. State-of-the-art lab chip sensors for environmental water monitoring. Measurement Science and Technology. 22(3), 032001.
[39] Fdez-Sanromán, A., Bernárdez-Rodas, N., Rosales, E., et al., 2025. Biosensor technologies for water quality: Detection of emerging contaminants and pathogens. Biosensors. 15(3), 189.
[40] Ogwu, M.C., 2026. Emerging technologies for water quality monitoring. In Water Quality and Safety in the Global South: Challenges, Solutions and Future Directions. Springer Water: Cham, Switzerland. pp. 365–399.
[41] Zhang, Y., Hu, F., Zhou, R., et al., 2026. Enhancing selectivity in affinity biosensors through biorecognition-driven suppression of nonspecific binding. ACS Sensors. 11(2), 853–884.
[42] Aryal, P., Hefner, C., Martinez, B., et al., 2024. Microfluidics in environmental analysis: Advancements, challenges, and future prospects for rapid and efficient monitoring. Lab on a Chip. 24(5), 1175–1206.
[43] Chen, J., Chen, S., Fu, R., et al., 2022. Remote sensing big data for water environment monitoring: Current status, challenges, and future prospects. Earth's Future. 10(2), e2021EF002289.
[44] Ighalo, J.O., Adeniyi, A.G., Marques, G., 2020. Internet of things for water quality monitoring and assessment: A comprehensive review. In: Hassanien, A., Bhatnagar, R., Darwish, A. (Eds.). Artificial Intelligence for Sustainable Development: Theory, Practice and Future Applications. Studies in Computational Intelligence, Vol. 912. Springer: Cham, Switzerland. pp. 245–259.
[45] Zulkifli, C.Z., Garfan, S., Talal, M., et al., 2022. IoT-based water monitoring systems: A systematic review. Water. 14(22), 3621.
[46] Fang, S., Xu, L.D., Zhu, Y., et al., 2014. An integrated system for regional environmental monitoring and management based on internet of things. IEEE Transactions on Industrial Informatics. 10(2), 1596–1605.
[47] Pule, M., Yahya, A., Chuma, J., 2017. Wireless sensor networks: A survey on monitoring water quality. Journal of Applied Research and Technology. 15(6), 562–570.
[48] Miller, M., Kisiel, A., Cembrowska-Lech, D., et al., 2023. IoT in water quality monitoring—Are we really here? Sensors. 23(2), 960.
[49] Sogno, P., Klein, I., Kuenzer, C., 2022. Remote sensing of surface water dynamics in the context of global change—A review. Remote Sensing. 14(10), 2475.
[50] Yousef, F., 2013. Application of Remote Sensing in Aquatic Ecosystems [PhD Thesis]. Michigan Technological University: Houghton, MI, USA.
[51] Chi, M., Plaza, A., Benediktsson, J.A., et al., 2016. Big data for remote sensing: Challenges and opportunities. Proceedings of the IEEE. 104(11), 2207–2219.
[52] Chang, N.-B., Imen, S., Vannah, B., 2015. Remote sensing for monitoring surface water quality status and ecosystem state in relation to the nutrient cycle: A 40-year perspective. Critical Reviews in Environmental Science and Technology. 45(2), 101–166.
[53] Feng, L., You, Y., Liao, W., et al., 2022. Multi-scale change monitoring of water environment using cloud computing in optimal resolution remote sensing images. Energy Reports. 8, 13610–13620.
[54] Singh, J., Ahirwal, S., Ramteke, K., et al., 2025. Remote sensing techniques for monitoring aquatic ecosystems. In: Ganie, P.A., Posti, R., Pandey, P.K. (Eds.). Information Technology in Fisheries and Aquaculture. Springer: Singapore. pp. 71–107.
[55] Castell, N., Kobernus, M., Liu, H.-Y., et al., 2015. Mobile technologies and services for environmental monitoring: The Citi-Sense-MOB approach. Urban Climate. 14, 370–382.
[56] Mohsan, S.A.H., Khan, M.A., Noor, F., et al., 2022. Towards the unmanned aerial vehicles (UAVs): A comprehensive review. Drones. 6(6), 147.
[57] Zhang, Z., Zhu, L., 2023. A review on unmanned aerial vehicle remote sensing: Platforms, sensors,

data processing methods, and applications. Drones. 7(6), 398.

[58] Zainurin, S.N., Wan Ismail, W.Z., Mahamud, S.N.I., et al., 2022. Advancements in monitoring water quality based on various sensing methods: A systematic review. International Journal of Environmental Research and Public Health. 19(21), 14080.

[59] Mohan, S., Kumar, B., Nejadhashemi, A.P., 2025. Integration of machine learning and remote sensing for water quality monitoring and prediction: A review. Sustainability. 17(3), 998.

[60] Rahaman, T., 2025. Smart environmental monitoring systems for air and water quality management. American Journal of Advanced Technology and Engineering Solutions. 1(1), 1–19.

[61] Esposito, M., Palma, L., Belli, A., et al., 2022. Recent advances in internet of things solutions for early warning systems: A review. Sensors. 22(6), 2124.

[62] Mallick, R., Poddar S., 2025. Future directions in water quality management: Integrating advanced technologies and sustainable practices. In: Dubey, A.K., Sriastav, A.L., Kumar, A., et al. (Eds.). Computational Automation for Water Security: Enhancing Water Quality Management. Elsevier: Amsterdam, The Netherlands. pp. 215–227.

[63] Zessner, M., 2021. Monitoring, Modeling and Management of Water Quality. Water, 13(11), 1523.

[64] Adu-Manu, K.S., Tapparello, C., Heinzelman, W., et al., 2017. Water quality monitoring using wireless sensor networks: Current trends and future research directions. ACM Transactions on Sensor Networks (TOSN). 13(1), 1–41.

[65] Sun, X., Zhang, Y., Shi, K., et al., 2022. Monitoring water quality using proximal remote sensing technology. Science of the Total Environment. 803, 149805.

[66] Li, L., Wang, T., Li, G., et al., 2026. Hydro-meteorological factors and human activities contribute comparably to weekly water quality dynamics in the Yangtze River Basin. Journal of Hydrology. 135025.

[67] Ponce Romero, J.M., Hallett, S.H., Jude, S., 2017. Leveraging big data tools and technologies: Addressing the challenges of the water quality sector. Sustainability. 9(12), 2160.

[68] Postolache, O., Girão, P.S., Pereira, J.M.D., 2012. Water quality monitoring and associated distributed measurement systems: An overview. In: Voudouris, K., Vouts, D. (Eds). Water Quality Monitoring and Assessment. InTech: London, UK. pp. 25–64.

[69] Sun, A.Y., Scanlon, B.R., 2019. How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions. Environmental Research Letters. 14(7), 073001.

[70] Jesus, G., Casimiro, A., Oliveira, A., 2017. A survey on data quality for dependable monitoring in wireless sensor networks. Sensors. 17(9), 2010.

[71] Overmars, A., Venkatraman, S., 2020. Towards a secure and scalable IoT infrastructure: A pilot deployment for a smart water monitoring system. Technologies. 8(4), 50.

[72] Zhang, C., 2024. Fundamentals of Environmental Sampling and Analysis. John Wiley & Sons: Hoboken, NJ, USA.

[73] Olatinwo, S.O., Joubert, T.-H., 2019. Enabling communication networks for water quality monitoring applications: A survey. IEEE Access. 7, 100332–100362.

[74] Shim, J., Hong, S., Lee, J., et al., 2023. Deep learning with data preprocessing methods for water quality prediction in ultrafiltration. Journal of Cleaner Production. 428, 139217.

[75] Saab, C., Zéhil, G.-P., 2025. Statistical analysis techniques in water quality monitoring: A review. Stochastic Environmental Research and Risk Assessment. 39(9), 3723–3760.

[76] Sommer, S., 2019. What is sensor calibration and why is it important? Available from: https://www.realpars.com/blog/sensor-calibration (cited 30 March 2026).

[77] El-Shafeiy, E., Alsabaan, M., Ibrahem, M.I., et al., 2023. Real-time anomaly detection for water quality sensor monitoring based on multivariate deep learning technique. Sensors. 23(20), 8613.

[78] Du, J., Hu, M., Zhang, W., 2020. Missing data problem in the monitoring system: A review. IEEE Sensors Journal. 20(23), 13984–13998.

[79] Larson, D.M., Bungula, W., Lee, A., et al., 2023. Reconstructing missing data by comparing interpolation techniques: Applications for long-term water quality data. Limnology and Oceanography: Methods. 21(7), 435–449.

[80] Lahat, D., Adali, T., Jutten, C., 2015. Multimodal data fusion: An overview of methods, challenges, and prospects. Proceedings of the IEEE. 103(9), 1449–1477.

[81] Kumpel, E., MacLeod, C., Stuart, K., et al., 2020. From data to decisions: Understanding information flows within regulatory water quality monitoring programs. npj Clean Water. 3(1), 38.

[82] Campbell, J.L., Rustad, L.E., Porter, J.H., et al., 2013. Quantity is nothing without quality: Automated QA/QC for streaming environmental sensor data. BioScience. 63(7), 574–585.

[83] Bello, A., Jahan, S., Farid, F., et al., 2022. A systemic review of the cybersecurity challenges in Australian water infrastructure management. Water. 15(1), 168.

[84] Lowe, M., Qin, R., Mao, X., 2022. A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring. Water. 14(9), 1384.

[85] Jia, L., Yen, N., Pei, Y., 2023. Spatial and temporal water quality data prediction of transboundary watershed using multiview neural network coupling. IEEE Transactions on Geoscience and Remote Sensing. 61, 1–16.

[86] Nallakaruppan, M.K., Gangadevi, E., Shri, M.L., et al., 2024. Reliable water quality prediction and parametric analysis using explainable AI models. Scientific Reports. 14(1), 7520.

[87] Elshaikh, A., Mabrouki, J., Mohamed, A.A.O., 2024. The future of water management: Leveraging AI for effective decision support. In: Mabrouki, J., Zarour, M. (Eds.). Advancements in Climate and Smart Environment Technology. IGI Global Scientific Publishing: Hershey, PA, USA. pp. 213–225.

[88] Han, Z., Zhang, S., He, L., 2025. Predicting and investigating water quality index by robust machine learning methods. Journal of Environmental Management. 381, 125156.

[89] Clemmensen, L.H., Kjærsgaard, R.D., 2022. Data representativity for machine learning and AI systems. arXiv preprint. arXiv:2203.04706.

[90] Zhu, J.-J., Yang, M., Ren, Z.J., 2023. Machine learning in environmental research: Common pitfalls and best practices. Environmental Science & Technology. 57(46), 17671–17689.

[91] Ahmed, U., Mumtaz, R., Anwar, H., et al., 2019. Efficient water quality prediction using supervised machine learning. Water. 11(11), 2210.

[92] Verwoerd, J.C., 2026. Machine Learning Applications in Sensor-Based Infrastructure Monitoring. Preprints. DOI: https://doi.org/10.20944/preprints202601.0594.v1

[93] Arachie, C., Huang, B., 2021. Constrained labeling for weakly supervised learning. In: de Campos, C., Maathuis, M.H. (Eds.). Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence, Vol. 161. PMLR: Cambridge, MA, USA. pp. 236–246.

[94] Krishnaraj, A., Deka, P.C., 2020. Spatial and temporal variations in river water quality of the Middle Ganga Basin using unsupervised machine learning techniques. Environmental Monitoring and Assessment. 192(12), 744.

[95] Tyagi, K., Rane, C., Sriram, R., et al., 2022. Unsupervised learning. In: Pandey, R., Khatri, S.K., Singh, N.K., et al. (Eds.). Artificial Intelligence and Machine Learning for EDGE Computing. Elsevier: Amsterdam, The Netherlands. pp. 33–52.

[96] Greene, D., Cunningham, P., Mayer, R., 2008. Unsupervised learning and clustering. In: Cord, M., Cunningham, P. (Eds.). Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval. Springer: Berlin/ Heidelberg, Germany. pp. 51–90.

[97] Zheng, Y., Wei, J., Zhang, W., et al., 2024. An ensemble model for accurate prediction of key water quality parameters in river based on deep learning methods. Journal of Environmental Management. 366, 121932.

[98] Farzana, S.Z., Paudyal, D.R., Chadalavada, S., et al., 2024. Temporal dynamics and predictive modelling of streamflow and water quality using advanced statistical and ensemble machine learning techniques. Water. 16(15), 2107.

[99] Parker, W.S., 2013. Ensemble modeling, uncertainty and robust predictions. Wiley Interdisciplinary Reviews: Climate Change. 4(3), 213–223.

[100] Park, J., Lee, W.H., Kim, K.T., et al., 2022. Interpretation of ensemble learning to predict water quality using explainable artificial intelligence. Science of the Total Environment. 832, 155070.

[101] Zhi, W., Appling, A.P., Golden, H.E., et al., 2024. Deep learning for water quality. Nature Water. 2(3), 228–241.

[102] Payandeh, A., Baghaei, K.T., Fayyazsanavi, P., et al., 2023. Deep representation learning: Fundamentals, technologies, applications, and open challenges. IEEE Access. 11, 137621–137659.

[103] Chen, X.-W., Lin, X., 2014. Big data deep learning: Challenges and perspectives. IEEE Access. 2, 514–525.

[104] Li, X., Xiong, H., Li, X., et al., 2022. Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. Knowledge and Information Systems. 64(12), 3197–3234.

[105] Pu, F., Ding, C., Chao, Z., et al., 2019. Water-quality classification of inland lakes using Landsat8 images by convolutional neural networks. Remote Sensing. 11(14), 1674.

[106] Zhang, Y., Wu, L., Ren, H., et al., 2020. Mapping water quality parameters in urban rivers from hyperspectral images using a new self-adapting selection of multiple artificial neural networks. Remote Sensing. 12(2), 336.

[107] Krichen, M., 2023. Convolutional neural networks: A survey. Computers. 12(8), 151.

[108] Faruk, D.Ö., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Engineering applications of artificial intelligence. 23(4), 586–594.

[109] Zhang, C., Ding, W., Zhang, L., 2024. Impacts of missing buoy data on LSTM-based coastal chlorophyll—A forecasting. Water. 16(21), 3046.

[110] Beck, M.B., Van Straten, G., 2012. Uncertainty and Forecasting of Water Quality. Springer Science & Business Media: Berlin/Heidelberg, Germany.

[111] Altenburger, R., Ait-Aissa, S., Antczak, P., et al., 2015. Future water quality monitoring—Adapting tools to deal with mixtures of pollutants in water

resource management. Science of the Total Environment. 512, 540–551.
[112] Zheng, H., Liu, Y., Wan, W., et al., 2023. Large-scale prediction of stream water quality using an interpretable deep learning approach. Journal of Environmental Management. 331, 117309.
[113] Pourshahabi, S., Rakhshandehroo, G., Talebbeydokhti, N., et al., 2020. Handling uncertainty in optimal design of reservoir water quality monitoring systems. Environmental Pollution. 266, 115211.
[114] Lani, M.N., Ismail, A.M., Jusoh, M.Y.F., et al. Strengthening machine learning reproducibility to ensure water security in the long term. In: Kurniawan, T.A., Anouzla, A. (Eds.). Handbook of Digitalization and Big Data in the Water Sector. CRC Press: Boca Raton, FL, USA. pp. 168–193.
[115] Yan, J., Chen, X., Zhou, L., et al., 2025. Spatial transferability of machine learning models for water quality prediction: A progressive data integration approach. Journal of Water Process Engineering. 80, 109230.
[116] Dalcanale, F., Fontane, D., Csapo, J., 2011. A general framework for a collaborative water quality knowledge and information network. Environmental Management. 47(3), 443–455.
[117] Sudriani, Y., Sebestyén, V., Abonyi, J., 2023. Surface water monitoring systems—The importance of integrating information sources for sustainable watershed management. IEEE Access. 11, 36421–36451.
[118] Dogo, E.M., Salami, A.F., Nwulu, N.I., et al., 2019. Blockchain and internet of things-based technologies for intelligent water management system. In: Al-Turjman, F. (Ed.). Artificial Intelligence in IoT. Springer: Cham, Switzerland. pp. 129–150.
[119] Zhu, Z., Zhang, H., Zhang, S., et al., 2025. Intelligent monitoring and control systems for smart water management. In: Garg, M.C., Rajput, V.D., Minkina, T., et al. (Eds.). Nano-Solutions for Sustainable Water and Wastewater Management: From Monitoring to Treatment. Springer: Cham, Switzerland. pp. 369–390.
[120] Amanatidis, P., Lyratzis, E., Angelopoulos, V., et al., 2025. Intelligent water management through edge-enabled IoT, AI, and big data technologies. IoT. 7(1), 5.
[121] Wu, Z.Y., Chew, A., Meng, X., et al., 2023. High fidelity digital twin-based anomaly detection and localization for smart water grid operation management. Sustainable Cities and Society. 91, 104446.
[122] Palma, L., Hatam, F., Di Nardo, A., et al., 2024. Contaminations in water distribution systems: A critical review of detection and response methods. AQUA—Water Infrastructure, Ecosystems and Society. 73(6), 1285–1302.
[123] Quansah, J.E., Engel, B., Rochon, G.L., 2010. Early warning systems: A review. Journal of Terrestrial Observation. 2(2), 5.
[124] Rouse, M.J., 2013. Institutional Governance and Regulation of Water Services. IWA Publishing: London, UK.
[125] Dutta, P., Sarma, S., 2026. Smart water management: Role of IoT, AI, and machine learning in water quality monitoring. Water Environment Research. 98(2), e70249.