WATER QUALITY MONITORING USING MACHINE LEARNING AND IOT: A REVIEW

TAHSIN FUAD HASAN[2], NASSERELDIN AHMED KABASHI[1], TANVEER SALEH[2], MD ZAHANGIR ALAM[1], MOHD FIRDAUS WAHAB[1] AND ABDURAHMAN HAMID NOUR[3]

[1]Department of Chemical Engineering and Sustainability, International Islamic University Malaysia, Gombak, 53100 Kuala Lumpur, Malaysia.
[2]Department of Mechatronics Engineering, International Islamic University Malaysia, Gombak, 53100 Kuala Lumpur, Malaysia
[3]Faculty of Chemical and Process Engineering Technology, Universiti Malaysia Pahang, Malaysia
**Corresponding authors: nasreldin@iium.edu.my*

ABSTRACT: Water remains one of the most essential natural resources. With the ever-increasing population, the demand for water across various sectors, including agriculture, industry, and power, as well as the growing prevalence of pollution, has led to a significant strain on water supplies. The availability of fresh and usable water is becoming increasingly limited, making quality monitoring and analysis crucial for sustainable use and environmental protection. Traditional water quality monitoring techniques involve manual sampling, testing, and investigation, which may not always be reliable and are often inefficient in providing early warnings of water quality deterioration. However, with the emergence of machine learning (ML) and Internet of Things (IoT) technologies, the process of water quality monitoring and analysis has become more efficient, accurate, and cost-effective. ML algorithms can analyze large volumes of water quality data, enabling data-centric approaches to designing, supervising, simulating, assessing, and refining various water treatment and management systems. This review paper provides an overview of the past and current applications of machine learning and IoT in water quality monitoring and analysis. Long-term cost savings can be seen in different ways as reduced labor costs, lower operational costs, early detection and intervention prevent costly repairs and emergencies, minimized infrastructure costs, distributed IoT sensors reduce the need for extensive physical infrastructure, optimized resource allocation and efficiency improvements with IoT and Machine Learning in water quality monitoring can be highlighted in the following points, real-time monitoring: immediate data analysis allows for prompt adjustments and decision-making, enhanced accuracy, advanced sensors and algorithms improve data precision and reliability, scalability, systems can be easily expanded or adapted to meet evolving needs, predictive maintenance, automated systems proactively address issues before they escalate, reducing manual oversight. The paper explores various ML algorithms, including supervised and unsupervised learning and deep learning, along with their applications, and discusses the use of IoT sensors for real-time monitoring of water quality parameters such as pH, dissolved oxygen, temperature, and turbidity.

KEY WORDS: *Machine Learning, IoT (Internet of things), Smart Water Grid (SWG)*

1. INTRODUCTION

Water is universally recognized as one of the most essential resources for life, with its quality and availability being intrinsically linked to global living standards. The United Nations identifies the provision of clean water and sanitation as a core goal for global sustainable development, noting that over 3 billion people lack adequate monitoring, raising concerns about the quality of the water they rely on [1]. Similarly, the World Health Organization (WHO) estimates that approximately 829,000 diarrheal deaths each year can be attributed to microbiologically contaminated drinking water [2].

Water quality is increasingly compromised by excessive pollutants, primarily from human activities, including the over-exploitation of natural resources, industrialization, urbanization, agriculture, and population growth. In agricultural settings, fertilizers and pesticides can be washed into rivers by rain, leading to pollution. Industrial waste products, such as those from chemical factories, are often disposed of in rivers and lakes, further contaminating these water bodies, including open oceans [3]. Factories that use river water for power generation or machinery cooling can increase water temperature, reducing dissolved oxygen levels and disrupting aquatic ecosystems. Surface water bodies, particularly rivers, are highly susceptible to waste disposal [4,5].

This problem is exacerbated by the uneven distribution of rainfall, resulting in floods and droughts, and by negligence in water management, which further aggravates contamination. Additionally, the hydrochemistry of open water systems is influenced by a range of factors, including climatic conditions, soil-rock types, and human activities within watersheds, all of which contribute to the growing challenge of maintaining water quality [6].

To reduce water pollution, alleviate stress on water resources, and conserve these essential resources, real-time monitoring of water quality parameters has become increasingly vital. Water quality is assessed by measuring its physical, chemical, and biological conditions to determine how well it meets the needs of humans and ecosystems. Monitoring critical parameters helps identify deviations in water conditions and provides early warnings of emerging hazards [4, 7]. Traditional monitoring methods, which involved manual sampling, testing, and investigation, were limited by lengthy processes. These methods have evolved towards real-time data collection and subsequent analysis to enable prompt remedial action.

The evaluation of water quality can vary depending on the parameters considered, even when relevant standards are maintained. However, considering every parameter is not always viable due to cost constraints and technical challenges [8, 9]. In recent times, the development and widespread adoption of IoT and machine learning have emerged as substantive technological solutions for effective water quality monitoring and analysis.

With IoT, interconnectivity and the embedding of computing devices into everyday environments facilitate the seamless transaction and transfer of data. Machine learning, on the other hand, leverages data through algorithms to predict new information. The increased adoption of these technologies across various domains can be attributed to their ability to produce precise results and extend easily into customizable environments. In recent years, IoT and machine learning have shown remarkable adaptability in the fields of environmental science and engineering, offering promise for generating more accurate evaluation results, even when dealing with the complexities of water quality analysis and assessment [10]. This paper discusses the various ways in which IoT and machine learning have been implemented in different environments for water quality monitoring.

2. IOT (INTERNET OF THINGS)

The Internet of Things (IoT) represents a significant technological advancement, with extensive applications across various fields, including science, engineering, medicine, and technology. Its widespread adoption in industrial implementations is largely due to its ability to integrate communication and embedded technology into diverse applications. IoT functions by interconnecting physical computing devices within networks, enabling seamless data collection and transaction with minimal human intervention [11]. The capability of real-time data collection and reporting, along with the accessibility of this information on internet-connected devices, has revolutionized strategies and decision-making processes, leading to greater efficiency and impact. This technology has paved the way for the creation of automated and 'smart' systems across sectors, ranging from households and office spaces to transportation systems, infrastructure, healthcare, and water distribution systems.

An IoT system primarily comprises sensors, processors, connectivity, and a user interface. Wireless technologies such as Wi-Fi, Bluetooth, ZigBee, and RFID maintain interconnectivity between devices and the internet. Data is collected, stored, and analyzed using cloud services, while smartphones and computers function as the user interface and the central hub or remote control for IoT [12]. The architecture of IoT is typically divided into three layers: the physical layer (data collection subsystem) where sensors gather data from the environment, the network layer (data transmission subsystem) where data is converted into digital streams for processing, and the application layer (data management subsystem) that delivers specific services to users. Some publications further divide this architecture into four components, separating the network layer into network connectivity and cloud server [13, 11].

IoT communication can occur in two forms: device-to-device and device-to-cloud. One of the commonly used communication platforms is Wireless Sensor Networks (WSNs), which utilize self-sufficient, low-energy sensor nodes capable of measuring and recording environmental conditions. Each sensor node typically includes a power source, a microcontroller, a wireless radio transmitter, and a collection of environmental sensors (such as humidity, pressure, and temperature). Figure 1 illustrates the basic architecture of an IoT system.

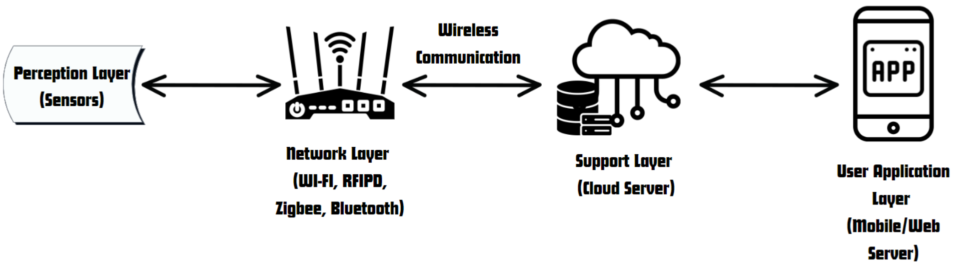

Fig.1: Basic architecture of an IoT system

While the initial setup costs for IoT and ML in water quality monitoring can be substantial, the long-term benefits far outweigh these expenses. The investment leads to significant cost savings through reduced labor and operational costs, minimized infrastructure needs, and optimized resource allocation. Enhanced accuracy, real-time monitoring, and predictive maintenance further contribute to long-term efficiency and financial savings, making IoT and ML technologies a valuable investment for sustainable water management. Automated systems predict potential issues before they escalate, reducing the need for manual oversight and extending the lifespan of equipment. This proactive approach prevents costly breakdowns and ensures system longevity.

2.1 IOT In Water Quality Monitoring

The application of IoT in water quality monitoring has significantly increased due to its efficiency and capability, with configurations varying based on environmental conditions and analysis requirements. A common application format is the Smart Water Grid (SWG), which integrates IoT technology into water distribution systems for comprehensive monitoring [14]. The development and implementation of SWG gained momentum in the 2000s, driven by global water-based companies seeking more sophisticated water management strategies [15]. SWG integrates smart water meters that enable remote readings of water consumption, replacing traditional water infrastructure. Sensor nodes are deployed along pipelines to detect leaks, while water quality sensor nodes are placed in tanks or along pipes to monitor chemical parameters such as pressure, flow, temperature, pH, conductivity, and turbidity. On the utility side, intelligent processes are employed to analyze and utilize the data collected by these sensing devices.

The SWG concept is closely related to the Smart Water Quality Monitoring System (SWQMS), which emphasizes the integration of intelligent water information systems through IT convergence into existing water infrastructure, resulting in an advanced smart management system. A similar concept is the Online Water Quality Monitoring (OWQM) system, which uses a network of online automatic monitoring devices, transmission networks, and business software for data analysis, forming the foundation of the original SWQMS concept. OWQM is designed to measure physicochemical parameters in real-time across various water sources, such as rivers, streams, lakes, oceans, groundwater, industrial wastewater, and urban drainage.

IoT applications in water quality monitoring can be tailored for specific purposes, such as creating a smart irrigation system that schedules irrigation based on environmental conditions or designing a robotic fish device to monitor debris in aquatic environments [16,17,18]. However, ongoing initiatives continue to focus on enhancing the monitoring process, improving information sharing, and refining decision-making processes.

The Internet of Things (IoT) introduces significant security and privacy challenges due to the vast number of interconnected devices, the diversity of those devices, and the sensitivity of the data they collect and transmit. IoT devices are often limited in computational power and storage, making it difficult to implement robust security measures.

2.1.1 Key security challenges include

Data Privacy: IoT systems often collect sensitive personal or environmental data, raising concerns about unauthorized access, misuse, or exposure of this data. Authentication and Authorization: Ensuring that only authorized users and devices can access the IoT network is critical, yet difficult due to the diversity and scale of IoT environments. Data Integrity: The integrity of data transmitted between devices must be protected to prevent tampering or corruption. Network Security: IoT systems are vulnerable to a range of network-based attacks, including Distributed Denial of Service (DDoS), man-in-the-middle (MitM) attacks, and eavesdropping. Physical Security: Many IoT devices are deployed in unsecured environments, making them susceptible to physical tampering. and privacy in IoT systems and propose advanced security measures for the preferred system.

The literature on IoT security highlights the significant challenges posed by the complexity and scale of IoT systems. While traditional security measures provide a foundation, they often fall short in addressing the unique demands of IoT environments. Advanced security strategies, such as lightweight cryptography, AI-driven anomaly detection, decentralized models, and privacy-preserving data analytics, offer promising solutions to enhance the security and privacy of IoT systems. By implementing these measures, IoT systems can achieve long-term resilience, ensuring that the benefits of IoT and machine learning in applications like water quality monitoring are fully realized while minimizing security risks.

2.2 Review Findings

Hamid et al. (2020) proposed a simplified architecture for a Smart Water Quality Monitoring System (SWQMS) designed to monitor and evaluate water quality in swimming pools, focusing on factors influencing pH and temperature [19]. The system utilizes a NodeMCU V3 processing unit with an ESP8266 Wi-Fi module, connected to a pH sensor and a DS18B20 temperature sensor, enabling real-time monitoring of pH and temperature. The data is monitored through the IoT cloud platform (Ubidots app). The study also investigated the significant factors influencing pH and temperature, revealing that the time of day did not affect pH but did influence temperature.

Similarly, Pasika et al. (2020) proposed a Water Quality Monitoring (WQM) system that measured the pH and turbidity of water, the water level in tanks, and the temperature and humidity of the atmosphere to assess water conditions in tanks [20]. Aiming for a low-cost architecture, the project selected an Arduino Mega MCU with an ESP8266 Wi-Fi module, along with pH, turbidity, ultrasonic, and DHT-11 (temperature and humidity) sensors. The ThinkSpeak mobile application was used for monitoring and cloud storage.

Geetha et al. (2016) summarized current developments in smart water quality monitoring and suggested an IoT-based approach that is both power- and cost-efficient for in-pipe water quality monitoring [13]. The proposed system includes sensors directly connected to a microcontroller with an integrated Wi-Fi module. The microcontroller analyzes the data sent to the cloud (Ubidots network) and notifies users of any deviations from the norm. Although power management is a concern, the system uses Wi-Fi for communication, given its existing infrastructure and intended use for monitoring home water quality. The system records data in the cloud for further analysis and monitors conductivity, pH, turbidity, temperature, and water level. Additionally, when parameters exceed a threshold limit based on WHO criteria, the cloud is designed to send alert SMS texts.

Gupta et al. (2018) introduced a smart water management system for housing societies, which uses an ultrasonic water level sensor and a turbidity sensor to monitor water levels and quality [10]. Residents can check the water level and quality in real-time via a smartphone app, accessing data broadcasted to the cloud by the sensors. The system also allows users to control the motor remotely via the app. The controller is a Raspberry Pi, and the data transfer protocol is MQTT (Message Queuing Telemetry Transport). The system is low-cost, easy to install, and reliable, making it suitable for use in older buildings without extensive modifications. It offers full automation and is a robust solution for smart water management.

Ranjan et al. (2020) leveraged IoT technology to develop a rainfall harvesting system that included both a collection or catchment area, such as a roof, and a storage system [21]. After analyzing existing systems, the authors concluded that users lacked awareness of rainfall, water quality, and water distribution. To address this, they proposed an IoT-based solution to establish a direct connection between users and the rainwater collection device. The model featured a building with two separate tanks for acidic and potable water, equipped with a raindrop detection sensor installed on the roof. A pH sensor measured the rainwater's acidity, and a servo motor on a hinge directed the water to the appropriate tank. The data was uploaded by a NodeMCU with a Wi-Fi module to a webpage created using HTML, CSS, and a PHP script, hosted by a free hosting service. The project aimed to ensure rainwater quality and provide users with essential data accessible via desktop or mobile devices.

Das and Jain (2017) developed a water quality monitoring system that used sensors to measure pH, conductivity, and temperature [12]. The system wirelessly transmitted data from the sensors to the microcontroller via a ZigBee module, which then sent the data to a smartphone or PC using a GSM module. Additionally, the system included proximity sensors that could notify authorities of water pollution via the GSM module. The microcontroller processed, analyzed, and transmitted the data, proving to be an efficient, low-cost, real-time water quality monitoring system. This system could help officials monitor water pollution and prevent waterborne diseases. It was easy to install, and the monitoring tasks could be performed by less-trained individuals.

Ramesh et al. (2017) developed an IoT-based system to detect environmental parameters and monitor water quality and contamination levels [22]. The system included sensors for hydrocarbons, chemicals, and metal content in a soil probe to monitor soil pollution, as well as pH, conductivity, dissolved oxygen, and turbidity sensors for water quality monitoring. This method could significantly impact land restoration projects in India and assist authorities in managing waste in affected areas. An IoT architecture was proposed to address cleanliness, waste management, and health concerns in a community. The platform featured three applications: real-time notifications for water quality, progress tracking of land recovery, and health statistics monitoring. Multiple sensors were placed in heavily polluted water resources, and the data collected was sent to a data aggregation system, which identified the safest water resources and alerted residents of potential risks. Similarly, the soil quality monitoring system measured the reduction of heavy metal content in the soil and notified the community. The system's capability for edge computing reduced bandwidth usage and computation overhead. An app was also implemented to transmit real-time health statistics from smartphones to servers, analyzing pollutants responsible for specific diseases. By integrating these three systems, the community could be informed about safe resources and health issues caused by polluted environments.

Maindalkar and Ansari (2015) proposed and discussed the design of a smartphone-based aquatic debris monitoring robot [23]. The robot integrates an Android smartphone with a robotic fish to monitor debris in various environments, accurately detecting debris while overcoming challenges such as wave impact, energy consumption, and irregular debris entrances. The paper presented lightweight computer vision algorithms for image processing, including image registration and adaptive background subtraction, to address these challenges. The robotic fish is powered by two NiMH batteries and communicates with a floating platform via a fiber-optic tether to relay camera, sensor, and control signals. Additionally, the paper explained the interfacing of sensors, DC motors, and Bluetooth with an Arduino (ATmega328) processor for real-time debris detection. The smartphone-based aquatic robot can adaptively configure the camera orientation and monitor the time interval for the next round using a coverage-based rotation scheduling algorithm.

Wireless sensor networks (WSNs) are commonly used alongside IoT in data acquisition and environmental monitoring systems due to their ease of installation, low cost, and easy maintenance [24]. Faustine et al. (2014) presented a WSN system prototype built for water quality monitoring in the Lake Victoria Basin [25]. The system uses an Arduino microcontroller, water quality sensors, and a wireless network connection module to detect and transmit real-time data on water temperature, dissolved oxygen, pH, and electrical conductivity. This data is made available to stakeholders through a website and mobile platforms in graphical and tabular formats. The core component of the system prototype, the WSN sensor node, is equipped with sensor and microcontroller units, a GPS receiver, a power supply, and an RF transceiver. The system uses four sensors to monitor different aspects of water quality but is expandable to accommodate additional sensors as needed. With a low-cost gateway module, the proposed prototype is suitable for long-term outdoor deployment and offers a software module that allows users to visualize WSN data without needing specific software installation.

Kamaludin et al. (2017) proposed an IoT-based water quality monitoring (WQM) system that combines a Radio Frequency Identification (RFID) system, a WSN platform, and Internet Protocol (IP) communication [26]. They utilized a 920MHz frequency for WSN communication in vegetation areas and measured pH levels and ambient temperature using analog sensors. The system uses the Digi Mesh protocol instead of the ZigBee protocol for better signal attenuation. The WSN platform allows RFID tags to communicate with the system gateway, powered by a mains-supplied power adapter. The sensor node, powered by Nickel Zinc (Ni-Zn) rechargeable batteries, includes a new circuitry design based on an Arduino Uno board with a double-layer PCB layout that measures pH levels and ambient temperature. The network gateway provides data to cloud storage via TCP/IP communication and is connected to the internet using an IoT module, Arduino Ethernet Shield. They also developed an Android OS mobile application for online monitoring, with an alarm-triggering system built in PHP to detect pH threshold values and generate alert sounds on users' mobile devices.

Myint et al. (2017) presented a smart water quality monitoring (SWQM) system for IoT environments, utilizing a reconfigurable sensor interface device [27]. The system collected real-time water data across five parameters from multiple sensors, which were computed on an FPGA board using VHDL and C programming languages. The data was then transmitted wirelessly to a monitoring PC through ZigBee communication and displayed using Python code on a Grafana dashboard. The proposed system included an RF module, an FPGA board, an ultrasonic sensor, a pH sensor, a digital temperature sensor, a turbidity sensor, and a CO2 sensor. The smart WQM system reduces power consumption, outperforming conventional microcontroller-based WSNs. The system demonstrated reliability and feasibility, with the potential to extend its coverage range in future WSN networks.

Beri (2015) outlines a low-cost wireless network system for automatically monitoring water quality using sensor technology, artificial intelligence techniques, and a database management system [14]. The system is scalable for public water distribution systems and adaptable for smaller settings like housing societies. The paper describes the use of a wireless sensor network (WSN) to collect real-time data on water quality parameters such as pH, temperature, dissolved oxygen, and conductivity. The system is powered by a PIC16F886 nano-watt MCU, with sensors sending data to the ADC, which is then transmitted via serial communication to a Zigbee modem and displayed on an LCD. The paper examines the challenges of detecting pH and the need for temperature correction and suggests using a single SIM card for monitoring, while also discussing potential issues with the GSM module.

Yasin et al. (2019) designed and implemented a new irrigation system using the Arduino Mega 2560 microcontroller and SIM900 GSM Shield [28]. This system allows for remote control and monitoring of the irrigation process. Moisture sensors placed in the soil automatically irrigate plants when the soil becomes dry, and the system can be controlled via SMS. In case of rain, a raindrop sensor module stops the irrigation process. The proposed system aims to promote plant growth while reducing water, labor, and time consumption, demonstrating a 60% reduction in water usage compared to conventional irrigation methods. The system is compatible with any mobile phone that supports SMS and allows for the easy addition of multiple phone numbers. However, the cost of purchasing, setting up, and maintaining the irrigation system's automatic equipment was noted to be high.

Using IoT and remote sensing (RS) technology, Prasad et al. (2015) developed a smart water quality monitoring system for Fiji [29]. The system uses RS technology to measure temperature, conductivity, pH, and oxidation-reduction potential (ORP). Anomalous measurements trigger an alert via IoT technology, indicating potential water pollutants. False positives are recorded but not treated as alerts. The system includes sensors, ADC, microcontroller, SD storage, and a GSM module. Data can be stored onboard or sent to a cloud server for analysis. Power conservation is critical, and the system design incorporates sleep mode and turns off idle modules to extend battery life. The system was tested on four different water sources to validate measurement accuracy, with results matching expectations. The system successfully used GSM technology to send alerts based on reference parameters to users for immediate action. The collected parameter references will be used to build classifiers for automated water analysis using neural network analysis. Overall, the system proved to be accurate, consistent, and an excellent contender for real-time water monitoring solutions.

Ali et al. (2022) designed a smart water grid (SWG) network capable of routing and monitoring water supply using fog computing, IoT, long-range wide-area network (LoRaWAN), and software-defined networking (SDN) [30]. The proposed architecture uses fog servers and controllers to collect and process data from sensors in the water grid, employing LoRaWAN technology for data communication to extend battery life. SDN is used within the LoRaWAN network to optimize the routing process. The architecture features a physically and logically distributed SDN approach, with controllers deployed at the fog layer for local control and a single controller for global control. The feasibility of the proposed architecture is evaluated using delay and network throughput metrics under the Mininet emulator, with experimental test-bed evaluation planned for future work. The paper highlights several advantages of the architecture over existing ones, including power consumption, security, privacy, and low-latency burst and leak detection. The use of the

LoRaWAN protocol reduces power consumption of SWG devices, enabling longer operation. Data is stored and analyzed at the fog server to preserve user privacy, with only critical events transmitted to the cloud server. Low-latency burst detection is achieved by processing data at the network's edges, providing low latency between SWG devices and the cloud.

Baanu et al. (2021) proposed an IoT-based system to monitor residual chlorine concentration in water distribution systems [11]. The study favored flow-through-type chlorine sensors for measuring residual chlorine and identified LoRa technology as ideal for long-range data communication. The paper also discussed various communication technologies suitable for real-time monitoring, including Wi-Fi, Zigbee, and LoRa, noting that Zigbee is preferred for short-range communication, while LoRa is better for remote monitoring over wide areas. Additionally, the paper explored optimal sensor placement, identifying three key locations for monitoring water quality: (i) where water exits the treatment facility, (ii) areas within the distribution system prone to contamination, such as corroded pipes or the ends of branch pipes, and (iii) points that are representative of overall water quality in the distribution system. The proposed system enhances timely decision-making, enables more efficient management of water resources, and acts as an early warning system.

In a review publication, Dong et al. (2015) surveyed research on Smart Water Quality Monitoring (SWQM) systems up to 2014 [15]. The authors examined three subsystems of SWQM: data management, data transfer, and data gathering. They discussed the selection of water quality parameters, monitoring technology, sampling sites, and frequency. Additionally, they explored network architecture and communication management for data transmission, as well as storage, analysis, and prediction for data management. The authors identified challenges and proposed future research directions for each subsystem, emphasizing the need for improved management strategies to develop reliable SWQM systems capable of monitoring large areas. The article also suggested different focuses for monitoring drinking water, wastewater, and environmental water quality.

Subsequently, Lalle et al. (2021) presented a survey of wireless communication technologies for Smart Water Grid (SWG) applications [31]. The authors noted that commonly used technologies such as cellular networks, ZigBee, 6LoWPAN, Bluetooth, and Wi-Fi suffer from issues related to power consumption, communication range, and penetration. To overcome these challenges, they recommended Low Power Wide Area Networks (LPWANs) due to their long-range communication, low power consumption, and excellent penetration capabilities. The article discussed the deployment of LPWANs in SWG applications such as water leak detection, water quality monitoring, and smart water metering. It also provided recommendations for advancing SWG, including addressing challenges and exploring research directions to enhance LPWAN performance.

Furthermore, Zainurin et al. (2022) conducted a review study on the overall development of water quality monitoring methodologies [32]. The study included a comparison of traditional methods with current innovations and reviewed regional variations in approach. Both within and beyond IoT, the study extensively examined various methods for monitoring water quality, including cyber-physical systems (CPS), electronic sensing, virtual sensing, and optical techniques. The study confirmed the relevance and suitability of CPS for water quality monitoring, highlighting its ability to connect the physical world (sensors, environment, humans) with the cyber world (software, data). This smart system allows real-time monitoring, early warnings for water quality issues, pollution detection, and improved sensitivity through potential future integration with advanced optical techniques.

Finally, Yasin et al. (2021) reviewed the use of IoT communication technology for water management and quality control [17]. The authors examined various components and techniques for implementing IoT in water management, including sensors, controllers, and IoT platforms. They compared different parameters used to measure water properties and evaluated the pros and cons of each technique. The review found that all the studies reviewed had achieved optimal solutions for reducing water waste in both private and public agricultural sectors by relying on IoT. The paper compared different studies based on microcontroller type, embedded programming language, sensors used, communication module, and protocol adopted. Researchers used a variety of microcontroller types, embedded programming languages, sensors, and communication modules, such as ZigBee, GSM, Raspberry Pi with built-in Wi-Fi, Arduino Ethernet Shield, and ESP8266. The paper concluded with recommendations for future research to enhance the performance of IoT-based water management systems.

3. MACHINE LEARNING (ML) TOOL

Machine Learning (ML), a crucial tool within the field of Artificial Intelligence (AI), has evolved into a powerful means of analysis, development, and implementation by leveraging Big Data [33]. ML excels at identifying significant patterns and correlations, making accurate predictions, and adapting independently as new data becomes available. Key steps before applying ML include data collection, algorithm selection, model training, and model validation.

Choosing the right algorithm is vital for any ML experiment. ML can be broadly categorized into two main types: supervised and unsupervised learning. Supervised learning involves a labelled dataset where the outputs are known, whereas unsupervised learning uses un-labelled data for training. Supervised learning is further divided into classification and regression. Classification is used for qualitative (categorical) datasets to assign labels, while regression deals with quantitative (continuous) data to estimate relationships between outputs and attributes for predictions.

The primary steps in an ML process include data processing, model training, and model evaluation. In unsupervised learning, the aim is to resolve various pattern recognition issues by categorizing data into distinct groups based on features, using techniques like dimensionality reduction and clustering. Unlike supervised learning, the number of groups and their significance in unsupervised learning are not predefined. Hybrid learning methods, such as semi-supervised learning, use both labelled and unlabelled data.
Common ML algorithms include, but are not limited to, Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), Artificial Neural Networks (ANN), and k-Nearest Neighbors (KNN).

In the context of water quality monitoring, ML is highly effective for analyzing large datasets to predict patterns and identify potential issues. Historical data analysis is crucial for forecasting water quality conditions and detecting problems. For example, predictive models can identify areas where water quality may be impacted by agricultural runoff or wastewater discharge, enabling targeted interventions and damage prevention. Additionally, ML models can facilitate real-time water quality monitoring, allowing for the rapid detection of parameter changes that can signal contamination or worsening water conditions. Overall, ML and AI play a significant role in advancing water quality monitoring and management.

Machine learning (ML) has emerged as a transformative tool in environmental monitoring, particularly in the assessment of water quality. The use of ML in this domain leverages the ability to process and analyze large datasets, providing predictive insights and enabling real-time monitoring. The integration of ML with water quality assessment promises to enhance the efficiency, accuracy, and timeliness of monitoring efforts. However, despite its potential, several challenges and limitations remain, which need to be critically examined.

3.1 Advantages Of ML In Water Quality Assessment

Machine learning offers several key advantages for water quality assessment:
Predictive Accuracy: ML models, particularly those based on deep learning, can provide high levels of predictive accuracy by identifying complex patterns in water quality data that traditional statistical methods may miss. Techniques like Random Forest, Support Vector Machines (SVM), and Artificial Neural Networks (ANNs) have been successfully used to predict various water quality parameters, such as pH, turbidity, and dissolved oxygen.

Data-Driven Insights: The ability of ML to analyze vast amounts of data from diverse sources—such as sensors, satellites, and historical records—enables the extraction of meaningful insights, leading to a better understanding of water quality dynamics. This is particularly beneficial in regions with limited access to real-time data.
Real-Time Monitoring and Decision-Making: The integration of ML with IoT systems allows for continuous monitoring and real-time analysis of water quality, enabling timely interventions and reducing the risk of pollution-related incidents.
Cost-Effectiveness: Over time, the automation and predictive capabilities of ML can reduce the need for extensive fieldwork and laboratory testing, leading to long-term cost savings.

3.2 Challenges And Limitations

Despite the significant potential of ML in water quality assessment, several challenges must be addressed:
Data Quality and Availability: The effectiveness of ML models is heavily dependent on the quality and quantity of the data used for training. In many regions, especially in developing countries, the lack of high-quality, comprehensive datasets poses a significant challenge. Data may be sparse, inconsistent, or biased, leading to inaccurate predictions and unreliable models.
Model Generalization: ML models trained on data from specific geographic locations or under certain conditions may not generalize well to other areas or different environmental conditions. This limits the applicability of ML in diverse and dynamic water systems.
Interpretability of Models: While ML models, especially deep learning models, can achieve high predictive accuracy, they often operate as "black boxes," making it difficult to understand the reasoning behind their predictions. This lack of interpretability can be a barrier to their adoption, particularly in regulatory or policy-making contexts where transparency is crucial.
Computational Resources: Training advanced ML models, particularly deep learning models, requires substantial computational power and resources. This can be a limiting

factor for organizations or regions with limited access to high-performance computing infrastructure.

Integration with Existing Systems: Integrating ML models into existing water quality monitoring frameworks can be complex. Legacy systems may not be compatible with the data formats or computational requirements of ML models, necessitating significant upgrades or redesigns.

3.3 Opportunities For Improvement

To overcome the challenges associated with the application of ML in water quality assessment, several opportunities for improvement can be explored:

Enhancing Data Collection: Efforts should be made to improve the quality and availability of water quality data. This could involve the deployment of more sophisticated sensors, the integration of satellite data, and the establishment of standardized protocols for data collection and reporting.

Hybrid Models: Combining ML with traditional modeling approaches or using ensemble methods can improve model generalization and robustness. Hybrid models that incorporate physical, chemical, and biological principles alongside data-driven insights could provide a more comprehensive understanding of water quality.

Improving Model Interpretability: Developing more interpretable ML models, such as decision trees or linear models, or incorporating techniques like SHAP (Shapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), can help bridge the gap between model accuracy and interpretability.

Accessible Computational Resources: The rise of cloud computing and the availability of AI-as-a-service platforms can democratize access to the computational resources needed for ML, making it easier for organizations of all sizes to implement advanced models.

Cross-Disciplinary Collaboration: The successful application of ML in water quality assessment requires collaboration between data scientists, environmental scientists, engineers, and policymakers. Cross-disciplinary partnerships can help ensure that ML models are not only technically sound but also practically relevant and aligned with environmental goals.

4 WATER QUALITY MONITORING

Chen et al. (2020) analyzed extensive data from major rivers and lakes in China between 2012 and 2018 [34] to assess the performance of ten machine learning models in predicting water quality. They evaluated the models using precision, recall, F1-score, weighted F1-score, and key water quality factors. The results showed that large datasets significantly improved the accuracy of water quality predictions.

The study included ten machine learning models: seven widely used ones—Logistic Regression (LR), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Decision Tree (DT), Completely Random Tree (CRT), Naive Bayes (NB), and k-Nearest Neighbors (KNN)—and three newly developed ensemble learning models—Random Forest (RF), Completely Random Tree Forest (CTF), and Deep Cascade Forest (DCF). Among these, DT, RF, and DCF exhibited superior performance, particularly when trained with specific datasets for pH, Dissolved Oxygen (DO), Chemical Oxygen Demand (CODMn), and Ammonia Nitrogen (NH3-N). The study identified two critical sets of water parameters that could enhance the prediction of water quality, highlighting DT, RF, and DCF as the most effective models for future monitoring and early warning systems. The results indicated that increasing the training data from 1% to 10% significantly improved model performance, emphasizing the importance of large datasets and key water parameters in enhancing prediction accuracy.

In Lu and Ma's (2020) study [35], two novel hybrid decision tree-based models were proposed to improve water quality predictions. These models combined Extreme Gradient Boosting (XGBoost) and Random Forest (RF) with Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), an advanced data denoising method. The models were applied to 1,875 hourly data points from the Gales Creek site in the Tualatin River, known for its high pollution levels, to predict indicators such as temperature, dissolved oxygen, pH, specific conductance, turbidity, and fluorescent dissolved organic matter.

The study introduced two hybrid models: CEEMDAN-XGBoost and CEEMDAN-RF, which utilized CEEMDAN to preprocess raw data with large fluctuations, enhancing the prediction performance of XGBoost and RF. Performance was evaluated using six error metrics and compared with four conventional models. The CEEMDAN-RF model excelled in predicting water temperature, dissolved oxygen, and specific conductance, with Mean Absolute Percentage Errors (MAPEs) of 0.69%, 1.05%, and 0.90%, respectively. The CEEMDAN-XGBoost model performed best for pH, turbidity, and fluorescent dissolved organic matter, with MAPEs of 0.27%, 14.94%, and 1.59%, respectively. The average MAPEs for these models were the lowest, indicating superior overall prediction performance. The stability of both hybrid models was higher compared to benchmark models. Despite high prediction accuracy, future research should consider additional factors affecting water quality and explore parallel computing to address the high demand for short-term predictions.

Solanki et al. (2015) developed a water quality prediction model using deep learning techniques [36]. Their study utilized data from the Chaskaman River near Nasik, Maharashtra, India, which was analyzed using the WEKA tool. The research found that unsupervised learning techniques, specifically denoising autoencoders and deep belief networks, were more effective at predicting variable data compared to supervised learning techniques. Accuracy was assessed using criteria such as mean absolute error and mean square error. The data showed significant fluctuations in turbidity, pH, and dissolved oxygen, with turbidity exhibiting the greatest variation during the monsoon season. Data were categorized into three seasonal groups—winter, summer, and monsoon—using clustering techniques. Missing values were replaced with the mean of available values through data cleaning. Traditional techniques, including Multi-layer Perceptron and Linear Regression, were compared with the deep learning approach of Deep Belief Networks. The study concluded that unsupervised learning methods could accurately predict variable data, with turbidity showing the highest variation during the monsoon season. pH exhibited minimal variation, and dissolved oxygen showed slight variation during the summer. The water quality prediction model can be employed for continuous monitoring and to address uncertain conditions.

Kim et al. (2013) evaluated the efficacy of three machine learning techniques—Random Forest, Cubist, and Support Vector Regression (SVR)—using Geostationary Ocean Colour Imager (GOCI) satellite data to estimate chlorophyll-a (chl-a) and suspended particulate matter (SPM) concentrations in two regions on South Korea's west coast [37]. Due to the limited number of samples, the effectiveness of the models was assessed using leave-one-out cross-validation (CV) and in situ measurements collected over four days in 2011 and

2012. The results indicated that SVR outperformed the other techniques. The study also highlighted the importance of discussing the spatiotemporal distributions of water quality metrics in relation to tidal phases using hourly GOCI images.

Khan and See (2016) proposed a model using Machine Learning techniques to predict future water quality trends based on current data [38]. They employed Artificial Neural Networks (ANN) with Nonlinear Autoregressive (NAR) time series analysis for efficient prediction and analysis. Four water quality metrics—chlorophyll, specific conductance, dissolved oxygen, and turbidity—were measured. The goal was to develop models that forecast future values using current parameter values. Performance metrics such as regression, mean squared error (MSE), and root mean square error (RMSE) were used to evaluate four ANN models. The results demonstrated the viability of the proposed ANN-NAR model, showing enhanced prediction accuracy.

Haghiabi et al. (2018) assessed the effectiveness of artificial intelligence techniques, including ANN, Group Method of Data Handling (GMDH), and Support Vector Machine (SVM), for predicting various components of water quality in the Tireh River, located in southwest Iran [39]. The study tested various transfer and kernel functions, leading to the development of ANN and SVM models. Results showed that both models performed as expected, with the radial basis function (RBF) and tansig functions yielding the best results among those examined. While the GMDH model performed adequately, it was less accurate compared to ANN and SVM. All models exhibited some overestimation, but the SVM model proved to be the most accurate. The study provided insights into the internal relationships between water quality components, with the ANN model utilizing two hidden layers and the SVM model employing RBF and tansig functions.

Guo et al. (2014) developed two machine learning models—Artificial Neural Network (ANN) and Support Vector Machine (SVM)—to forecast the effluent total nitrogen (T-N) concentration at a wastewater treatment plant in Ulsan, Korea [9]. They optimized model parameters and evaluated performance using pattern search methods and sensitivity analysis, incorporating daily water quality and meteorological data as input parameters. The results showed that both models could accurately predict the effluent's T-N concentration over a 1-day interval. While the SVM model demonstrated superior prediction accuracy, the sensitivity analysis revealed that the ANN model was more reliable in understanding the cause-and-effect relationship between T-N concentration and input values for integrated food waste and wastewater treatment. Consequently, the ANN model was deemed more suitable for decision-making and process control. The study suggests that machine learning models can serve as reliable tools for early warning and water quality control in wastewater treatment. Future research could enhance the accuracy of ANN and SVM models by incorporating long-term data sampling.

Li et al. (2020) evaluated the effectiveness of ANN and SVM models in predicting Total Nitrogen (TN) and Total Phosphorus (TP) levels in an agricultural drainage river in eastern China [40]. The study aimed to examine the relative importance of input variables and discuss strategies for improving water quality. Sensitivity analyses were performed on both models using monthly, bimonthly, and trimonthly datasets. The findings indicated that SVM models outperformed ANN models in forecasting precision and generalization ability. The study recommends SVM models as a potent alternative for more accurate and effective water quality predictions in agricultural watersheds. Sensitivity analyses for SVM and ANN models can help managers quickly identify spatiotemporal water quality fluctuations due to natural and anthropogenic changes in agricultural drainage rivers.

Chen et al. (2023) proposed a technique for accurately estimating urban river water quality using remote sensing data from multiple sources, even with limited sample availability [41]. Their goal was to address scale inconsistencies among remote sensing datasets and achieve efficient, large-scale water quality inversion. To tackle the complex nonlinear relationships between ground point data and remote sensing data, they suggested a self-optimizing machine learning approach that automatically finds optimal model parameters from a small number of samples, thereby reducing training time. The researchers used feature enhancement and spatial mapping methods to ensure consistency in water quality information. The results demonstrated that their method accurately estimated chlorophyll a, turbidity, and ammonia nitrogen from UAV and satellite images. The study introduces a novel technique for integrating air-space-ground monitoring of urban inland rivers. However, monitoring accuracy is limited by data availability, and further research is needed to address potential errors in spatial mapping. The researchers recommend expanding monitoring frequency and range to include seasonal and annual assessments of urban river water quality.

Imani et al. (2020) developed an application for predicting water quality resilience using ANN and the Fuzzy Analytic Hierarchy Process [42]. The model accurately forecasts resilience, identifying vulnerable areas for improved water management. The Bayesian Regularization algorithm demonstrated superior performance in predicting water quality resilience. The study proposes integrating resilience mapping into the annual report of São Paulo state's environmental agency for more effective planning. This approach could support water supply maintenance and be enhanced by incorporating real-time data monitoring systems for a more dynamic resilience prediction system.

Ahmed et al. (2022) proposed an enhanced water quality index (WQI) method using a semi-supervised machine learning technique to assess water quality. This approach addresses the limitations of traditional methods, which are often time-consuming, expensive, biased toward physico-chemical parameters, and reliant on a large number of parameters [43]. The proposed method involves parameter selection, weight assignment, sub-index calculation, sub-index aggregation, and classification. For the Rawal watershed in Pakistan, data on physical-chemical, atmospheric, meteorological, and hydrological topography parameters were collected. The new technique achieved a 100% classification rate, eliminating the need to include all criteria for classification. The study demonstrated that this method, which incorporates a broad range of parameters and machine learning techniques, accurately classified the stream network. It assigned high scores to variables such as electrical conductivity, Secchi disc depth, dissolved oxygen, lithology, and geology, using feature tree-based techniques like LightGBM, Random Forest, CatBoost, AdaBoost, and XGBoost. The findings suggest that this improved method can reduce the uncertainties associated with previous approaches, contribute to global water management planning, and warrant further investigation for other water bodies.

In a review paper, Zhu et al. (2022) discussed the application of machine learning algorithms in assessing water quality across various contexts, including drinking water, sewage, ocean, and surface and groundwater [8]. The review examined the performance of machine learning in different aquatic environments, highlighting the benefits and limitations of commonly used methods. While machine learning has proven effective in predicting water quality, optimizing resource allocation, and managing water shortages, challenges remain in fully leveraging these techniques due to difficulties in obtaining accurate data and the complexity of real-world water treatment and management systems. The review suggests overcoming these challenges by developing more advanced sensors, enhancing the feasibility and reliability of algorithms, and training interdisciplinary professionals to advance machine learning techniques and their application in engineering practices.

Hassan and Woo's systematic review in 2021 aimed to evaluate the usefulness of machine learning (ML) approaches for assessing water quality indicators from satellite data [44]. The study reviewed data from Scopus, Web of Science, and IEEE citation databases, identifying 113 qualifying studies from an initial search of 1796 publications. The review found that the most commonly used ML models for retrieving water quality parameters included ANN, RF, SVM, regression, Cubist, genetic programming (GP), and DT. Typical indicators of water quality identified were turbidity, temperature, salinity, colored dissolved organic matter, and chlorophyll-a. The review concluded that ML can effectively monitor water quality, enabling researchers to predict and learn from natural environmental processes and assess human impacts on ecosystems. These insights can support policymakers and water resource managers in preventing water pollution and ensuring compliance with environmental regulations.

4.1 Projects Using IOT And Machine Learning

Various projects and studies have explored the simultaneous use of IoT and machine learning (ML) in water quality monitoring and related applications. Initially, ML was primarily viewed as a tool for generating predictive models for wireless sensor networks (WSNs) and IoT systems. However, as ML applications expanded, it became evident that ML could offer significant benefits when applied to WSNs or IoT [45].

Adeleke et al. (2023) sought to develop and assess the effectiveness of ML and IoT in water storage stations [46]. They created a system prototype and evaluated its performance using classification and reliability metrics. The study analyzed physical and chemical water parameters such as temperature, pH, turbidity, dissolved oxygen, total dissolved solids, oxidation-reduction potential, and electrical conductivity to assess water pollutants in drinking water. ANN and SVM machine learning algorithms were employed to predict the impurity levels in the water based on sensor data. An automated water treatment method was also introduced to address specific contamination levels. The study found that the ANN models outperformed the SVM models. The research concluded that combining AI and IoT is effective for remote monitoring of water conditions and that automated water treatment systems offer significant advantages in mitigating water pollution.

Jha et al. (2020) proposed a two-phase approach to develop a framework for cloud-based water quality monitoring [47]. In the first phase, they surveyed existing water monitoring systems, and in the second phase, they designed a framework to evaluate groundwater quality in communal or overhead tanks. Sensors monitored parameters such as turbidity, TDS, conductivity, BOD, nitrate, fecal coliform, and pH. The sensor data was analyzed in a cloud-based environment called Ubidots using machine learning methods. The decision tree classifier achieved a classification accuracy of 84% on a dataset of 307 records. The study suggested extending the research using big data stream processing in a Spark framework for distributed contexts. They also recommended a microcontroller-based system connected to display systems and mobile devices via GSM and Bluetooth to predict water quality. The aim was to prevent health issues caused by contaminated water.

Chowdhury et al. (2019) proposed a sensor-based water quality monitoring system utilizing Wireless Sensor Network (WSN) components, including a microcontroller for processing, a communication system, and various sensors [48]. The system leveraged remote monitoring and IoT technology for real-time data access. Sensor data was analyzed and compared to benchmark values using Spark streaming analysis, Spark MLlib, deep learning neural network models, and the Belief Rule Based (BRB) system. Automated SMS alerts were sent if the measured values exceeded threshold limits. The system's high frequency, mobility, and low power consumption were notable features. The goal was to continuously monitor river water quality in off-grid locations with minimal cost and energy consumption while maintaining high detection accuracy. The study emphasized using Big Data Analytics and IoT for real-time monitoring and suggested that the system could be expanded to include other parameters such as total dissolved solids, chemical oxygen demand, and dissolved oxygen.

Wu et al. (2020) aimed to classify water images into "clean" and "polluted" categories for a water pollution monitoring system that utilized IoT technology to capture water images [49]. The authors identified challenges in water image classification due to low inter-class and high intra-class variability. To enhance feature representation, they proposed an attention neural network that encoded channel-wise and multi-layer properties. They constructed a hierarchical attention neural network using a channel-wise attention gate structure and conducted comparative experiments on an image dataset related to water surfaces. The proposed neural network was integrated into a water image-based pollution monitoring system for real-time monitoring and immediate response. The authors also aimed to improve the network by incorporating the ability to handle mixed pollutants and developing a lightweight version for low-resource platforms.

Pappu et al. (2017) monitored water quality in residential storage tanks [50]. Their system used a pH sensor and TDS meter to measure water quality parameters, employing K-Means clustering to predict water quality based on trained datasets from various water samples. Implemented with low-cost embedded devices like Arduino Uno and Raspberry Pi 3, the system analyzed sensor data using the K-Means clustering algorithm. The advantages of this algorithm included faster processing, tighter clusters, and relative efficiency. The system used Arduino as the microcontroller and Raspberry Pi 3 as the processing unit, with pH and TDS sensors deployed in the water and connected to the Arduino microcontroller. Results were updated on a cloud server. The system was fully automated and used IoT technologies for device communication and water quality prediction. It could be extended to ponds, rivers, and water pipes, though data security and integrity must be ensured during transmission for analysis and control of the water tank valve and storage area.

Sagan et al. (2020) demonstrated that machine learning could significantly optimize water quality monitoring by combining sensor data from real-time monitoring with satellite data [51]. Models such as partial least squares regression, support vector regression, and deep neural networks showed higher accuracy compared to traditional models. However, certain water quality variables, such as pathogen concentration, cannot be directly measured through remote sensing due to their non-optical nature or lack of high-resolution hyperspectral data, though they can be inferred using other measurable data.

Mustafa et al. (2020) reviewed research published from 2014 to 2020 on the use of artificial neural networks (ANNs) in hydrology [52]. Their review highlighted that ANNs are a powerful and effective tool for predicting and monitoring water quality parameters, yielding satisfactory outcomes. The article discussed various ANN algorithms, their recent applications, advantages, and limitations in hydrology. It also emphasized the integration of neural networks with other technologies such as ANN-based hindcast models, geographical information systems (GIS), and wireless sensor networks. The review suggested employing multiple AI models for water quality prediction and monitoring. For model validation, the authors used various numerical indicators, including R, R^2, RMSE, NSC, and PCC. Future hydrology research should explore other soft computing technologies, such as deep learning tools, genetic algorithms, random forests, and extreme learning machines. Compared to current laboratory-based methods, the review found that utilizing soft computing and communication technologies for water system monitoring offers quicker, more effective, environmentally friendly alternatives that enhance real-time public health security.

5. CRITICAL REVIEW

The literature reviewed provides a comprehensive overview of the integration of Internet of Things (IoT) and machine learning (ML) technologies in water quality monitoring. It highlights the evolution of these technologies from traditional methods and their current applications, benefits, and limitations.

5.1 Strengths

Advancement of Technologies: The review acknowledges the significant advancements in IoT and ML technologies, which have expanded their applications in water quality monitoring. The shift from traditional methods to these advanced technologies reflects a broader trend towards more efficient and accurate data collection and analysis.

Diverse Applications: The literature covers a wide range of applications, demonstrating the versatility of IoT and ML in different contexts. For instance, it includes studies on real-time water quality monitoring, predictive modeling, and automated water treatment systems. This variety underscores the potential of these technologies to address various aspects of water management.

Integration of IoT and ML: The review effectively illustrates how IoT and ML can complement each other in water quality monitoring. IoT provides the infrastructure for data collection, while ML offers advanced analytical capabilities to interpret this data, thus enhancing the overall monitoring process.

Identification of Key Challenges: The literature review does a commendable job of identifying the challenges associated with implementing IoT and ML technologies. Issues such as the need for advanced sensors, data quality, hardware and software constraints, and the complexity of real systems are crucial considerations for further development.

5.2 Weaknesses

Limited Discussion on Data Quality and Management: While the review mentions the importance of advanced sensors and data quality, it lacks a detailed discussion on the specific challenges related to data management and quality control. For instance, the impact of data noise, missing values, and the need for data preprocessing in machine learning models are not thoroughly explored.

Insufficient Focus on Interdisciplinary Collaboration: The review briefly touches upon the need for interdisciplinary talent but does not delve deeply into how effective collaboration between different fields (e.g., environmental science, data science, and engineering) can be fostered. The integration of domain-specific knowledge with technological expertise is crucial for the successful implementation of IoT and ML in water quality monitoring.

Lack of Comparative Analysis: The review does not provide a comparative analysis of different IoT and ML approaches in water quality monitoring. While individual studies are highlighted, a synthesis of their findings to compare the effectiveness of various methods or technologies could offer more actionable insights.

Future Directions: The review suggests that further development is needed in areas such as sensor technology and algorithm improvement. However, it does not provide concrete recommendations or potential research directions for overcoming the identified challenges. More detailed guidance on future research areas could enhance the utility of the review.

6. CONCLUSION

As a fundamental life source, water quality and condition must be preserved and maintained to meet even the most basic human needs. Traditional methods of water quality monitoring are no longer the most effective means of conservation, as advancements in IoT and machine learning (ML) have addressed previous limitations. IoT and its associated services are increasingly integrated into our daily lives, work processes, and business operations. Significant ongoing research aims to develop essential components and models to support the next generation of internet services, facilitated by numerous interconnected devices. Meanwhile, ML remains a powerful tool for harnessing information and data to generate predictions and trends, enabling a comprehensive understanding and solution to complex problems and systems.

This paper provides a brief literature review and analysis of research and projects related to water quality monitoring using IoT technologies and machine learning algorithms. IoT has been utilized in water quality monitoring to collect data from various sensors, analyze this data using machine learning algorithms, and provide real-time information for efficient water management. However, challenges identified in the literature highlight the need for advanced sensors to collect high-quality data and for selecting hardware and software configurations that provide necessary feedback while adhering to cost and environmental constraints, as well as ensuring ease of application and accessibility for all communities.

Machine learning is increasingly employed in water environments for various purposes, including predicting water quality and managing water resources. Nevertheless, its full potential is constrained by challenges such as data availability, the complexity of real systems, and the need for specialized knowledge and curated algorithms. To address these challenges, there is a need to develop advanced sensors for more accurate data collection, improve algorithms and models for broader application, and train interdisciplinary talent in advanced machine learning techniques for engineering practices.

The long-term benefits of using IoT and machine learning for water quality monitoring include significant cost savings and efficiency improvements. These technologies enable real-time monitoring and analysis, reducing the need for manual sampling and laboratory testing, which lowers labor and operational costs. Early detection and automated intervention help prevent costly repairs and environmental damage by addressing issues before they escalate. Additionally, IoT systems reduce the need for extensive physical infrastructure, minimizing infrastructure costs. The scalability and adaptability of these systems allow for efficient resource allocation and continuous optimization, further enhancing the overall effectiveness and sustainability of water quality management.

ACKNOWLEDGEMENT

The author wishes to acknowledge the support received from the KOE and the CHES department, as well as express gratitude to all the technicians and colleagues involved. Special thanks are extended to RMC IIUM for enabling this grant under Project ID: IUMP-SRCG22-005-0005 and Project Title: "Investigation of Water Quality Monitoring Using IoT and Machine Learning Techniques: Sungai Pusu River IIUM Gombak Case Study."

REFERENCES

[1] Goal 6 | Department of Economic and Social Affairs. Retrieved May 7, 2023, from https://sdgs.un.org/goals/.

[2] Prüss-Ustün, A., Wolf, J., Bartram, J., Clasen, T., Cumming, O., Freeman, M. C., & Johnston, R. (2019). Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: an updated analysis with a focus on low-and middle-income countries. International journal of hygiene and environmental health, 222(5), 765-777.

[3] Wang, J., Liu, X. D., & Lu, J. (2012). Urban River Pollution Control and Remediation. Procedia Environmental Sciences, 13(2011), 1856–1862. https://doi.org/10.1016/j.proenv.2012.01.179.

[4] Landrigan, P. J., Fuller, R., Acosta, N. J. R., Adeyi, O., Arnold, R., Basu, N. (Nil), Baldé, A. B., Bertollini, R., Bose-O’Reilly, S., Boufford, J. I., Breysse, P. N., Chiles, T., Mahidol, C., Coll-Seck, A. M., Cropper, M. L., Fobil, J., Fuster, V., Greenstone, M., Haines, A., Zhong, M. (2018). The Lancet Commission on pollution and health. The Lancet Commissions, 391(10119), 462–512. https://doi.org/10.1016/S0140- 6736(17)32345-0.

[5] Kothari, V., Vij, S., Sharma, S., & Gupta, N. (2021). Correlation of various water quality parameters and water quality index of districts of Uttarakhand. Environmental and Sustainability Indicators, 9, 100093. doi:10.1016/j.indic.2020.100093.

[6] Akhtar, N., Syakir Ishak, M.I., Bhawani, S.A., & Umar, K. (2021). Various Natural and Anthropogenic Factors Responsible for Water Quality Degradation: A Review. Water.

[7] Wang, J., Liu, X. D., & Lu, J. (2012). Urban River Pollution Control and Remediation. Procedia Environmental Sciences, 13(2011), 1856–1862. https://doi.org/10.1016/j.proenv.2012.01.179

[8] Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. Eco-Environment & Health.

[9] Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y. M., Park, J. P., & Cho, K. H. (2015). Prediction of effluent concentration in a wastewater treatment plant using machine learning models. Journal of Environmental Sciences, 32, 90-101.

[10] Gupta, K., Kulkarni, M., Magdum, M., Baldawa, Y., & Patil, S. (2018). Smart Water Management in Housing Societies using IoT. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT). doi:10.1109/icicct.2018.8473262.

[11] Baanu, Bharani & Babu, K.s. Jinesh. (2021). Smart water grid: a review and a suggestion for water quality monitoring. Water Supply. 22. 10.2166/ws.2021.342.

[12] B. Das and P. C. Jain, “Real-time Water Quality Monitoring System using Internet of Things,” 2017 International Conference on Computer, Communications and Electronics, COMPTELIX 2017, pp. 78–82, 2017.

[13] Geetha S, Gouthami S. (2016). Internet of things enabled real time water quality monitoring system. Smart Water.;2:1-19.

[14] Beri, N. N. (2015). Wireless sensor network based system design for chemical parameter monitoring in water. International Journal of Electronics, Communication & Soft Computing Science and Engineering, 3(6), 56.

[15] Dong J, Wang G, Yan H, Xu J, Zhang X. (2015) A survey of smart water quality monitoring system. Environ Sci Pollut Res Int. Apr; 22(7):4893-906. doi: 10.1007/s11356-014-4026-x. Epub 2015 Jan 6. PMID: 25561262.

[16] Parra, L., Rocher, J., Escrivá, J., & Lloret, J. (2018). Design and development of low cost smart turbidity sensor for water quality monitoring in fish farms. Aquacultural Engineering, 81, 10-18.

[17] Yasin, H. M., Zeebaree, S. R., Sadeeq, M. A., Ameen, S. Y., Ibrahim, I. M., Zebari, R. R., & Sallow, A. B. (2021). IoT and ICT based smart water management, monitoring and controlling system: A review. Asian Journal of Research in Computer Science, 8(2), 42-56.

[18] Maindalkar, A. A., & Ansari, S. M. (2015). Design of Robotic Fish for Aquatic Environment Monitoring. International Journal of Computer Applications, 117(17).

[19] Hamid, S.A., Rahim, A.M., Fadhlullah, S.Y., Abdullah, S.B., Muhammad, Z., & Leh, N.A. (2020). IoT based Water Quality Monitoring System and Evaluation. 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 102-106.

[20] Pasika, S., & Gandla, S. T. (2020). Smart water quality monitoring system with cost-effective using IoT. Heliyon, 6(7), e04096. doi:10.1016/j.heliyon.2020.e04096.

[21] Ranjan, V., Reddy, M. V., Irshad, M., & Joshi, N. (2020). The Internet of Things (IOT) Based Smart Rain Water Harvesting System. 2020 6th International Conference on Signal Processing and Communication (ICSC). doi:10.1109/icsc48311.2020.9182767.

[22] Ramesh, M. V., Nibi, K. V., Kurup, A., Mohan, R., Aiswarya, A., Arsha, A., & Sarang, P. R. (2017). Water quality monitoring and waste management using IoT. 2017 IEEE Global Humanitarian Technology Conference (GHTC). doi:10.1109/ghtc.2017.8239311.

[23] Maindalkar, A. A., & Ansari, S. M. (2015). Design of Robotic Fish for Aquatic Environment Monitoring. Int. J. Comput. Appl, 117, 31-34.

[24] Li, L. (2014). Software development for water quality's monitoring centre of wireless sensor network. Computer Modeling New Tech, 132-136.

[25] Faustine, A., Mvuma, A. N., Mongi, H. J., Gabriel, M. C., Tenge, A. J., & Kucel, S. B. (2014). Wireless sensor networks for water quality monitoring and control within lake victoria basin: Prototype development. Wireless Sensor Network, 6(12), 281.

[26] Kamaludin, K. H., & Ismail, W. (2017). Water quality monitoring with internet of things (IoT). 2017 IEEE Conference on Systems, Process and Control (ICSPC). doi:10.1109/spc.2017.8313015.

[27] Myint, C. Z., Gopal, L., & Aung, Y. L. (2017). Reconfigurable smart water quality monitoring system in IoT environment. 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). doi:10.1109/icis.2017.7960032.

[28] Yasin, H. M., Zeebaree, S. R., & Zebari, I. M. (2019, April). Arduino based automatic irrigation system: Monitoring and SMS controlling. In 2019 4th Scientific International Conference Najaf (SICN) (pp. 109-114). IEEE.

[29] Prasad, A. N., Mamun, K. A., Islam, F. R., & Haqva, H. (2015). Smart water quality monitoring system. 2015 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE). doi:10.1109/apwccse.2015.7476234.

[30] Ali, A. A., Saadi, S. M., Mahmood, T. M., & Mostafa, S. A. (2022). A smart water grid network for water supply management systems. Bulletin of Electrical Engineering and Informatics, 11(3), 1706-1714.

[31] Lalle, Y., Fourati, M., Fourati, L. C., & Barraca, J. P. (2021). Communication technologies for Smart Water Grid applications: Overview, opportunities, and research directions. Computer Networks, 190, 107940. doi:10.1016/j.comnet.2021.107940.

[32] Zainurin, S. N., Wan Ismail, W. Z., Mahamud, S. N. I., Ismail, I., Jamaludin, J., Ariffin, K. N. Z., & Wan Ahmad Kamil, W. M. (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.
[33] Ma, C., Zhang, H. H., & Wang, X. (2014). Machine learning for Big Data analytics in plants. Trends in Plant Science, 19(12), 798–808. doi:10.1016/j.tplants.2014.08.004.
[34] Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R., & Ren, H. (2020). Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. Water research, 171, 115454.
[35] Lu, H., & Ma, X. (2020). Hybrid decision tree-based machine learning models for short-term water quality prediction. Chemosphere, 249, 126169.
[36] Solanki, A., Agrawal, H., & Khare, K. (2015). Predictive analysis of water quality parameters using deep learning. International Journal of Computer Applications, 125(9), 0975-8887.
[37] Yong Hoon Kim, Jungho Im, Ho Kyung Ha, Jong-Kuk Choi & Sunghyun Ha (2014) Machine learning approaches to coastal water quality monitoring using GOCI satellite data, GIScience & Remote Sensing, 51:2, 158-174, DOI: 10.1080/15481603.2014.900983.
[38] Khan, Y., & See, C. S. (2016). Predicting and analyzing water quality using Machine Learning: A comprehensive model. 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT). doi:10.1109/lisat.2016.7494106.
[39] Haghiabi, A. H., Nasrolahi, A. H., & Parsaie, A. (2018). Water quality prediction using machine learning methods. Water Quality Research Journal, 53(1), 3-13.
[40] Li, Y., Wang, X., Zhao, Z., Han, S., & Liu, Z. (2020). Lagoon water quality monitoring based on digital image analysis and machine learning estimators. Water research, 172, 115471.
[41] Chen, P., Wang, B., Wu, Y., Wang, Q., Huang, Z., & Wang, C. (2023). Urban river water quality monitoring based on self-optimizing machine learning method using multi-source remote sensing data. Ecological Indicators, 146, 109750.
[42] Imani, M., Hasan, M. M., Bittencourt, L. F., McClymont, K., & Kapelan, Z. (2021). A novel machine learning application: Water quality resilience prediction Model. Science of The Total Environment, 768, 144459. doi:10.1016/j.scitotenv.2020.144459.
[43] Ahmed, M., Mumtaz, R., & Anwar, Z. (2022). An Enhanced Water Quality Index for Water Quality Monitoring Using Remote Sensing and Machine Learning. Applied Sciences, 12(24), 12787.
[44] Hassan, N., & Woo, C. S. (2021, August). Machine learning application in water quality using satellite data. In IOP Conference Series: Earth and Environmental Science (Vol. 842, No. 1, p. 012018). IOP Publishing.
[45] Messaoud, S., Bradai, A., Bukhari, S. H. R., Qung, P. T. A., Ahmed, O. B., & Atri, M. (2020). A Survey on Machine Learning in Internet of Things: Algorithms, Strategies, and Applications. Internet of Things, 100314. doi:10.1016/j.iot.2020.100314.
[46] Adeleke, I. A., Nwulu, N. I., & Ogbolumani, O. A. (2023). A hybrid machine learning and embedded IoT-based water quality monitoring system. Internet of Things, 22, 100774.
[47] Jha, B. K., Sivasankari, G. G., & Venugopal, K. R. (2020). Cloud-based smart water quality monitoring system using IoT sensors and machine learning. International Journal of Advanced Trends in Computer Science and Engineering, 9(3).
[48] Chowdury, M. S. U., Emran, T. B., Ghosh, S., Pathak, A., Alam, M. M., Absar, N., & Hossain, M. S. (2019). IoT based real-time river water quality monitoring system. Procedia computer science, 155, 161-168.
[49] Wu, Y., Zhang, X., Xiao, Y., & Feng, J. (2020). Attention Neural Network for Water Image Classification under IoT Environment. Applied Sciences, 10(3), 909. doi:10.3390/app10030909.
[50] Pappu, S., Vudatha, P., Niharika, A. V., Karthick, T., & Sankaranarayanan, S. (2017). Intelligent IoT based water quality monitoring system. International Journal of Applied Engineering Research, 12(16), 5447-5454.
[51] Sagan, V., Peterson, K. T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B. A., & Adams, C. (2020). Monitoring inland water quality using remote sensing: Potential and limitations of

spectral indices, bio-optical simulations, machine learning, and cloud computing. Earth-Science Reviews, 205, 103187.
Mustafa, H. M., Mustapha, A., Hayder, G., & Salisu, A. (2021). Applications of IoT and Artificial Intelligence in Water Quality Monitoring and Prediction: A Review. 2021 6th International Conference on Inventive Computation Technologies (ICICT). doi:10.1109/icict50816.2021.9358675.