World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Review Article)

Advances in AI and optimization for water quality index prediction and integrated water resources assessment: A review

Sara Bouziane [1,*], Badraddine Aghoutane [1], Aniss Moumen [2], Anas El Ouali [3], Ali Essahlaoui [4] and Abdellah El-Hmaidi [4]

[1] *Department of Computer Science, Informatics and Applications Laboratory, Faculty of Sciences, Moulay Ismail University, Meknes, Morocco.*
[2] *Department of Computer Science, National School of Applied Sciences, Ibn Tufail University, Kenitra, Morocco.*
[3] *Department of Geology, Scientific Institute, Mohammed V University, Rabat, Morocco.*
[4] *Department of Geology, Faculty of Sciences, Moulay Ismail University, Meknes, Morocco.*

World Journal of Advanced Research and Reviews, 2025, 28(03), 2093-2107

Publication history: Received on 11 November 2025; revised on 28 December 2025; accepted on 30 December 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.3.4212

Abstract

Accurate water quality assessment is critical for sustainable water resources management under growing environmental pressures. The Water Quality Index (WQI) provides a practical framework for summarizing complex water quality data into a single indicator. This review examines recent advances in artificial intelligence and optimization techniques for WQI prediction, with a focus on machine learning, ensemble models, deep learning, and hybrid approaches. Existing studies demonstrate strong predictive capabilities but remain largely model-centric and limited by localized datasets and weak system integration. This review identifies methodological limitations and outlines key components required for future integrated monitoring frameworks, including data acquisition, model interpretability, and uncertainty-aware decision support. The findings provide guidance for advancing toward scalable and transparent water quality assessment systems.

Keywords: Water Quality; Integrated System; Water Quality Index; Artificial Intelligent; Explainability XAI

1. Introduction

The sustainable management of freshwater resources is a critical global challenge in the face of accelerating urbanization, industrial development, and climate variability. Water quality monitoring plays a fundamental role in safeguarding ecosystems, protecting public health, and ensuring the availability of clean water for diverse human and ecological needs [1]. The Water Quality Index (WQI) has become a widely used tool for simplifying complex physicochemical, biological, and ecological measurements into a single interpretable score. By translating raw measurements into a composite index, the WQI provides an accessible framework for both policymakers and the public to assess water suitability for domestic, agricultural, and industrial purposes [2, 3].

Despite its usefulness, traditional WQI calculation methods suffer from significant limitations. They rely on fixed parameter weights and static thresholds that fail to capture seasonal variability, local conditions, and nonlinear dynamics of water chemistry [4]. These rigid approaches limit their sensitivity to emerging contamination issues and reduce their capacity for adaptive water management. Recent advances in artificial intelligence (AI) and machine learning (ML) have demonstrated the ability to model nonlinear interactions and temporal patterns, with studies reporting high predictive accuracy and robustness across multiple contexts [5, 6]. Complementary optimization techniques, such as Genetic Algorithms (GA) and Particle Swarm Optimization (PSO), further refine predictive models

* Corresponding author: Sara Bouzianea

by optimizing parameter selection and enhancing efficiency. Yet, despite these promising advances, current AI-based WQI solutions remain fragmented, often tied to case-specific datasets and rarely embedded within real-time monitoring infrastructures such as IoT sensor networks or cloud platforms [7]. This lack of integration hinders scalability and operational adoption at regional or national levels.

The increasing maturity of AI and optimization techniques, coupled with the proliferation of IoT and cloud technologies, presents a timely opportunity for synthesis. While individual reviews exist on AI models or WQI methods, few studies have systematically examined how predictive modeling, optimization strategies, and system architectures can be combined into end-to-end frameworks for water quality assessment. A comprehensive review is therefore needed to consolidate methodological advances, identify limitations, and chart a pathway toward scalable, real-time, and explainable monitoring systems. This review contributes to the broader challenge of water resource management by framing WQI prediction not only as a modeling problem but as part of a full-stack architecture that includes data acquisition, processing, analytics, and decision support.

The objectives of this review are threefold. First, it aims to synthesize recent developments in AI-based techniques for Water Quality Index (WQI) prediction, with particular emphasis on deep learning, hybrid models, and ensemble learning approaches. Second, it seeks to evaluate the role of optimization algorithms, including Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Multi-Gene Genetic Programming (MGGP), in enhancing both the performance and the practical deplorability of predictive models. Finally, the review examines emerging architectural frameworks that integrate IoT sensing, edge and cloud processing, explainable AI (XAI), and decision-support systems to enable real-time water quality monitoring. By addressing these objectives, this review aims to clarify the current state of research, highlight persistent challenges, and provide a roadmap for future work toward modular, scalable, and interoperable architectures for intelligent water resource management.

2. Materials And Methods

This review provides a systematic overview of artificial intelligence (AI) and optimization techniques applied to Water Quality Index (WQI) prediction and their role in broader water resource monitoring architectures. The methodology combined structured literature retrieval, screening, and qualitative synthesis with supplementary bibliometric and architectural analysis.

2.1. Literature search strategy

Relevant studies were retrieved from major scientific databases, including Scopus, Web of Science, IEEE Xplore, ScienceDirect, and SpringerLink. The search covered the years 2000–2025, with emphasis on the last 15 years to capture recent advances. Keywords and Boolean combinations included: “Water Quality Index”, “WQI prediction”, “Artificial Intelligence”, “Machine Learning”, “Optimization algorithms”, “IoT water monitoring”, “cloud computing”, and “Explainable AI”.

2.2. Inclusion and exclusion criteria

Studies were included if they applied AI or optimization techniques to WQI prediction or related indices, and reported sufficient methodological detail (e.g., model description, dataset characteristics, and evaluation metrics such as RMSE, R2, or MAE). Works ad- dressing the integration of WQI models into broader architectures (e.g., IoT-enabled monitoring, edge computing, or cloud-based frameworks) were also retained. By contrast, studies relying solely on deterministic or statistical approaches without AI or optimization components were excluded. Papers lacking methodological transparency, performance outcomes, or peer-reviewed publication status were also omitted.

2.3. Study selection process

The initial database search yielded 150 records. After removing duplicates and screening titles and abstracts, 141 articles remained. A further abstract assessment excluded 29 studies that did not meet the inclusion criteria, such as relying solely on deterministic models or lacking methodological transparency. Ultimately, 112 studies were retained for qualitative synthesis and comparative analysis. Table. 1 illustrates the study selection process.

Table 1 Flow diagram of the study selection process

Step	Description	Number (n)
1	Records identified through database searching	150
2	Duplicates removed	9
3	Records screened	141
4	Records excluded after abstract assessment	29
5	Studies included in bibliometric/comparative analysis	112

2.4. Data extraction and synthesis

From each eligible study, bibliographic details, study context, and methodological characteristics were systematically extracted. Key descriptors included publication year, geographical focus, type of water body, and the AI or optimization techniques employed. Where available, performance indicators such as accuracy, RMSE, MSE, and R2 were recorded. Extraction followed a standardized form to ensure comparability across studies. The synthesized data were analyzed narratively to identify methodological trends, recurring challenges, and opportunities for integration into full-stack monitoring systems.

2.5. Bibliometric and comparative analysis

A bibliometric overview was conducted using metadata (titles, abstracts, and keywords) to identify research hotspots and emerging themes (e.g., optimization, IoT, explainability). Word clouds and clustering analysis were used to highlight patterns, while comparative figures evaluated the strengths, weaknesses, and performance of AI and optimization techniques across different case studies.

2.6. Architecture analysis

In parallel, system-level architectures were examined to assess how AI models were integrated into operational monitoring frameworks. Specific components analyzed included IoT sensor networks, edge/cloud computing layers, data pipelines, and decision-support interfaces. These architectural features were compared against common limitations identified in the literature, such as fragmentation, interpretability, and scalability.

2.7. Conceptual framework development

Finally, insights from the review were synthesized into a conceptual end-to-end pipeline for WQI prediction. The proposed framework emphasizes real-time data acquisition, cloud/edge-based processing, AI-driven optimization, explainability, and decision-support mechanisms, highlighting the need for modular and interoperable architectures in future water resource monitoring.

3. Results

3.1. Global keyword trends

To obtain an overall picture of the thematic orientation of the reviewed literature, a word cloud was generated using titles, abstracts, and keywords of all retained studies (Fig. 1). Generic terms such as water, quality, and index were excluded to highlight more specific technical and methodological concepts. The visualization reveals the prominence of terms such as prediction, system, monitoring, optimization, IoT, networks, cloud computing, and explainable, which collectively underline the intersection of artificial intelligence, optimization strategies, and real-time architectures in water quality monitoring.

Figure 1 Global word cloud summarizing the most frequent terms from titles, abstracts, and keywords in the reviewed corpus

3.2. Overview of the selected literature

The systematic search yielded a total of 150 records, of which 112 were retained after screening and eligibility checks (Table. 1). These studies span the period 2000–2025, with a noticeable increase in publications after 2015, reflecting the growing interest in applying artificial intelligence (AI) and optimization to water quality monitoring. Fig. 2 illustrates the distribution of publications per year, showing a marked acceleration in the last decade.

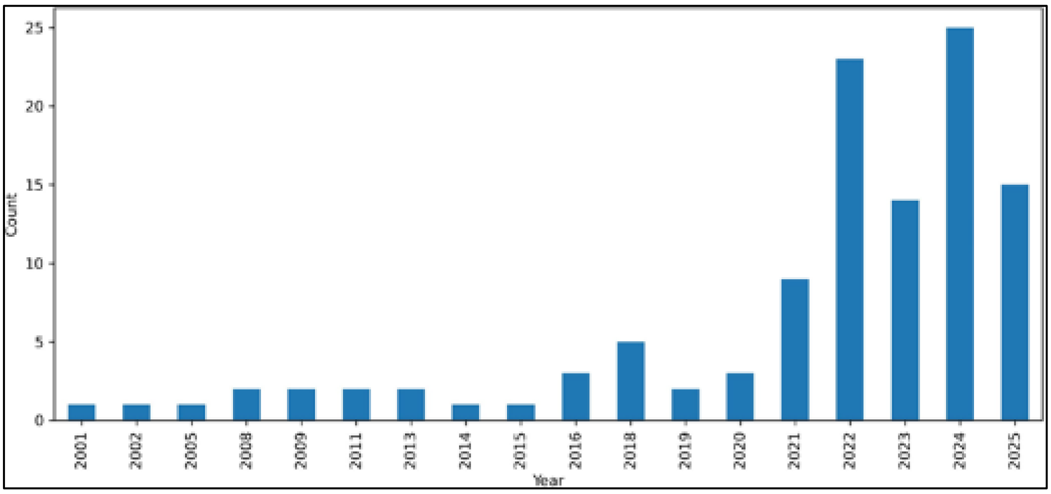

Figure 2 Publications per year in AI-based WQI and optimization studies (2000–2025)

The corpus includes diverse types of contributions. The majority of references are journal papers, followed by books chapters, conference papers, thesis, and technical reports (Fig. 3). This distribution highlights the predominance of peer-reviewed journal research, but also underscores the role of conference proceedings in reporting emerging AI methods before journal publication. Publication sources are also concentrated in a limited number of outlets. Fig. 4 summarizes the top contributing journals and, showing that environmental and computational venues dominate the dissemination of this research.

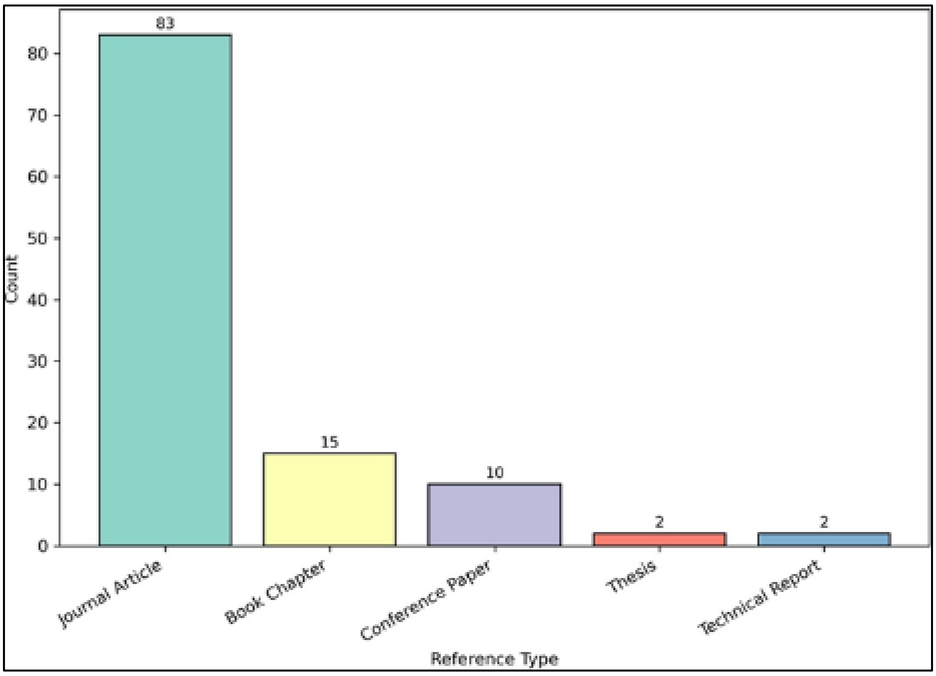

Figure 3 Distribution of reference types in the reviewed corpus

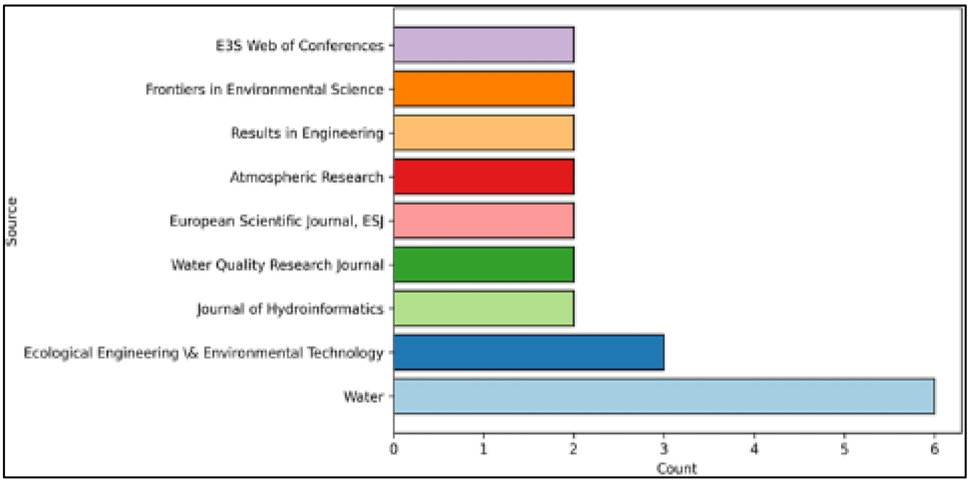

Figure 4 Top publication sources of the selected studies

3.3. Thematic clustering of the literature

To uncover thematic structures within the reviewed corpus, a clustering analysis of titles and keywords was performed using TF–IDF vectorization and Keans grouping. Six clusters emerged, each representing distinct but complementary research streams in the field of AI-driven water quality assessment. Word clouds were generated to visualize the most frequent terms in each cluster (Fig. 5).

Figure 5 Word cloud representation of each Cluster

Cluster 0 (IoT and Edge Computing): This cluster is dominated by terms such as IoT, edge, computing, smart, sensor, wireless, network. It reflects research on real-time monitoring platforms, emphasizing distributed architectures where edge devices handle data preprocessing before transmission to cloud servers. Such approaches address latency and efficiency challenges in large-scale monitoring deployments.

Cluster 1 (Prediction and Explainable AI): Characterized by terms such as pre- diction, rainfall, ensemble, neural, SVR, explainable, SHAP (Shapley Additive explanations), this cluster represents work on predictive modeling and model interpretability. The prominence of explainability-related keywords indicates a growing emphasis on transparent and trustworthy AI for water quality applications, especially in hydrological forecasting contexts.

Cluster 2 (Environmental Monitoring and Pollution Studies): With terms such as pollution, Morocco, river, irrigation, management, optimization, this cluster includes regional applications of AI to pollution assessment and water quality indices. Optimization appears here as a supporting component in case studies where model calibration and feature selection are required for localized datasets.

Cluster 3 (Water Treatment and Smart Systems): This cluster is defined by terms including treatment, smart, technology, management, optimization, highlighting AI-based frameworks for water treatment, aquaculture, and municipal systems. Optimization is particularly visible in this cluster, reflecting its role in designing efficient control strategies and improving operational reliability.

Cluster 4 (Environmental Engineering and Regional Indices): Terms such as environmental, basin, resilience, entropy, Maaouya, indices reveal a focus on basin-specific studies and advanced index development. This cluster illustrates how AI models are ap- plied in conjunction with novel indices to capture spatiotemporal dynamics and support resilience-oriented water management.

Cluster 5 (Hybrid and Genetic Programming Approaches): Featuring terms like monitoring, IoT, fuzzified, multi, gene, genetic, reservoir, aquaculture, this cluster reflects studies employing metaheuristic and hybrid optimization methods. Multi-Gene Genetic Programming (MGGP), genetic algorithms, and fuzzified approaches are recurrent themes, particularly in aquaculture and recirculating water systems. This confirms that optimization techniques, while less frequent overall, are crucial in niche applications requiring computational efficiency and adaptability.

Overall, the clustering results demonstrate that while predictive modeling and IoT monitoring dominate the literature, optimization terms are consistently associated with hybrid, treatment, and aquaculture-focused studies (Clusters 2, 3, and 5). This reinforces the conclusion that optimization plays a targeted but significant role in improving the adaptability and performance of AI-based water quality monitoring systems.

3.4. AI techniques in WQI calculation

The reviewed literature demonstrates the increasing application of artificial intelligence methods to Water Quality Index (WQI) prediction over the last decade, moving from fuzzy logic approaches to advanced deep learning. Early work by researchers proposed a hybrid probabilistic WQI model for the Jajrood River in Iran, integrating probabilistic neural networks (PNNs), Bayesian networks (BNs), and fuzzy inference systems (FIS) [8]. This framework addressed uncertainty and expert judgment, offering more flexible and robust evaluations compared to deterministic indices.

Later, another study evaluated the Adaptive Neuro-Fuzzy Inference Systems (ANFIS) on the Yamuna River, India, comparing them against conventional Multi- Linear Regression (MLR) [9]. Both ANFIS and ANN models outperformed MLR by up to 10% during validation, while showing negligible differences between ANN and ANFIS, confirming their reliability for WQI prediction.

ANN-based models remained a focus in subsequent research. A study applied a forward and backpropagation ANN to the Brahmani River, achieving regression coefficients greater than 0.9 across stations. The study demonstrated the strong generalization of neural networks when trained on large datasets with appropriate data partitioning [10]. In the same year, another research advanced this line of work with a hybrid Random Tree–ANN (RT-ANN) model for North Pakistan, which combined decision trees with ANN learning. The hybrid approach improved predictive performance, reaching RMSE = 2.319 and NSE = 0.945, outperforming standalone ANN or tree models [11].

Parallel developments also explored Support Vector Machines (SVM). A previous work enhanced SVM by combining it with the Differential Evaluation and Gray Wolf Optimization (DE-GWO) algorithm, significantly improving predictive accuracy [12]. More recently, researchers demonstrated that SVM combined with sensitivity analysis remains effective even with missing water quality parameters, reinforcing the robustness of SVM-based frameworks [13].

Also, a study applied Deep Capsule Crystal Edge Graph neural networks to WQI modeling, reporting MSE = 6.7 and 99% accuracy [14]. Similarly, other work employed deep neural networks (DNNs) to predict multiple water quality indices, achieving correlation coefficient R = 0.9994 and RMSE = 0.0020. These results under- line the increasing potential of deep learning to capture nonlinearities in complex environmental datasets, enabling highly accurate real-time assessments [6]. A summary of the main AI-based techniques reviewed, their features, and performance is presented in Table. 2.

Table 2 Performance evaluation of AI-based WQI prediction approaches

Model	**Key Features**	**Performance Metrics**	**Reference**
Fuzzy/Probabilistic (PWQI)	Combines FIS, BN and PNN	Real-time probabilistic assessment	[8]
ANFIS	Neuro-fuzzy learning	+10% accuracy vs. MLR	[9]
ANN	Forward/ Backpropagate training	$R^2 > 0.9$	[10]
RT-ANN (Hybrid)	Random Tree + ANN	RMSE = 2.319, NSE= 0.945	[11]
SVM	DE-GWO optimization, sensitivity analysis	RMSE = 0.0020, R = 0.9994	[12,13]
Deep Learning (Capsule GNN, DNN)	Optimized activation	MSE = 6.7, Accuracy = 99%	[6,14]

Overall, the reviewed AI-based models illustrate a clear chronological trajectory of increasing sophistication, from early probabilistic frameworks (2011) to neuro-fuzzy systems (2020), hybrid ANN-based models (2022), enhanced SVMs (2022–2023), and deep learning architectures (2024). Each stage builds on prior approaches, progressively im- proving predictive accuracy and model adaptability. This trend suggests that future re- search will need to balance accuracy with interpretability and efficiency.

3.5. Optimization techniques in WQI Calculation

Alongside the development of AI models, optimization algorithms have become essential for tuning parameters, selecting features, and improving computational efficiency. Genetic Algorithms (GA) were among the first widely applied methods. Authors combined GA with Multiple Linear Regression (GA–MLR) to estimate the Comprehensive

Pollution Index (CPI) for the Shatt Al-Arab River in Iraq [15]. Their GA–MLR framework optimized the selection and weighting of predictor variables, achieving a minimum CPI of 0.3777 and highlighting GA's role in refining predictions through evolutionary search.

Around the same time, Particle Swarm Optimization (PSO) was explored as a complementary technique. Additionally, another study tested PSO in combination with machine learning classifiers, developing a PSO–NBC model that reached 92.8% accuracy, outperforming a PSO–SVM model (77.6%) [16]. These results demonstrated the ability of PSO-enhanced learning models to improve predictive performance, particularly when paired with lightweight classifiers.

Hybrid optimization methods were subsequently proposed to combine the strengths of multiple approaches. In Malaysia, authors applied the Hybrid Particle Swarm Optimization and Genetic Algorithm (HPSOGA) to WQI prediction in the Klang River [17]. The hybrid outperformed standalone algorithms, illustrating how multi-strategy optimization increases precision and efficiency in complex modeling tasks.

More recently, researchers demonstrated the potential of Multi-Gene Genetic Programming (MGGP) integrated with IoT-enabled aquaculture systems. Their model achieved R2 = 0.9112 and RMSE = 0.6441 while remaining computationally efficient for deployment on constrained IoT platforms [18]. This case shows how optimization contributes not only to accuracy but also to practical feasibility in real-time monitoring systems. A comparative overview of the reviewed optimization techniques is provided in Table. 3.

Table 3 Comparison of optimization techniques applied in WQI prediction

Method	Application	Performance Metrics	Reference
Genetic Algorithm (GA)	GA–MLR, CPI estimation (Shatt Al-Arab River)	Min. CPI = 0.3777	[15]
Particle Swarm Optimization (PSO)	PSO–NBC and PSO–SVM	Accuracy: (NBC), (SVM) 92.8% 77.6%	[16]
Hybrid PSO + GA (HPSOGA)	Klang River prediction	Outperformed standalone methods	[17]
MGGP	IoT-enabled aquaculture systems	R2 = 0.9112, RMSE = 0.6441	[18]

The reviewed optimization studies reveal the complementary role of metaheuristic techniques in strengthening WQI models. Genetic Algorithms have proven effective for feature selection and parameter calibration, while Particle Swarm Optimization enhances classifier performance, particularly for lightweight models such as NBC. Hybrid approaches like HPSOGA outperform individual methods, suggesting that multi-strategy optimization can further boost accuracy and efficiency. Finally, MGGP demonstrates how optimization can be adapted for IoT-based real-time monitoring, combining accuracy with computational feasibility. Taken together, these methods illustrate that optimization is less about replacing AI models and more about ensuring their adaptability, stability, and deployability in real-world systems.

3.6. Architectural frameworks for water resource monitoring

Beyond individual models, recent studies emphasize the importance of system-level architectures for water quality monitoring. The integration of IoT sensors, satellite data, wireless networks, and cloud–edge infrastructures enables the deployment of optimized AI models in real time, moving from isolated case studies toward scalable environmental observatories. In this subsection, we synthesize the architectural frameworks reviewed in the literature, organized into functional layers.

3.6.1. Data acquisition layer

The foundation of any monitoring system is the continuous collection of reliable environmental data. IoT-enabled sensors deployed in rivers, reservoirs, and treatment facilities measure critical parameters such as pH, turbidity, dissolved oxygen, and flow rate at high temporal resolutions [19, 20]. Remote sensing platforms, including satellites and UAVs, extend this capacity by providing basin-scale perspectives and supporting hydrological modeling [21]. Wireless Sensor Networks (WSNs), based on Lo Ra WAN, Zigbee, or NB-IoT, provide distributed and fault-tolerant data collection, ensuring resilience even in remote areas [22]. A notable implementation is the Online Monitoring System (OMS), which combines low-power WSNs with alerting mechanisms (SMS alarms) to authorities when thresholds are exceeded [23].

3.6.2. Data transmission and integration

Collected data must flow securely and efficiently from field nodes to processing systems. Edge devices are increasingly used to perform preliminary preprocessing and com- pression, reducing latency and bandwidth use [24]. Depending on local infrastructure, transmission employs LPWAN, 3G/4G/5G, or Wi-Fi protocols [25]. Middleware and standardized APIs are essential to integrate heterogeneous data streams, promoting interoperability and scalability across devices and systems [26].

3.6.3. Data storage and processing

This layer ensures that raw data are cleansed, transformed, and prepared for modeling. Automated pipelines handle outlier detection, normalization, and synchronization [27]. Stream-processing tools such as Apache Kafka and Flink support real-time data handling, while batch frameworks like Spark provide efficient scheduled processing [28]. Hybrid storage solutions balance scalability and latency: MongoDB or S3 for cloud repositories, and InfluxDB at the edge for fast queries [29]. A previous study proposed a three-tier IoT–edge–cloud framework that dynamically assigns processing tasks based on network conditions, computational load, and privacy needs [30].

3.6.4. Analytics and AI integration layer

At the analytics layer, AI models transform processed data into actionable insights. Classical machine learning (ANNs, SVMs, Random Forests) and deep learning (CNNs, LSTMs, Capsule Graph Networks) have been deployed to capture nonlinear dynamics of WQI [10, 12,14]. Hybrid and optimized models (ANFIS, RT-ANN, MGGP) further enhance adaptability through metaheuristics such as GA and PSO [11, 15, 16,18]. Deployment strategies increasingly rely on containerized microservices (e.g., Docker) to enable real-time retraining and version control in cloud dashboards [31]. An illustrative pipeline is described by a study that integrated SVM with sensitivity analysis for robust prediction despite missing inputs [13].

3.6.5. Explainability layer

To address the opacity of complex AI models, explainable AI (XAI) has been introduced. Techniques such as SHAP Shapley Additive explanations, LIME (Local Interpretable Model-agnostic Explanations), and feature importance plots clarify how physicochemical parameters (e.g., $NO3-$, $NH4+$, $PO43-$) drive predictions, thus improving transparency and trust [32]. Fig. 6 illustrates a comparison between black-box and explainable frameworks [33].

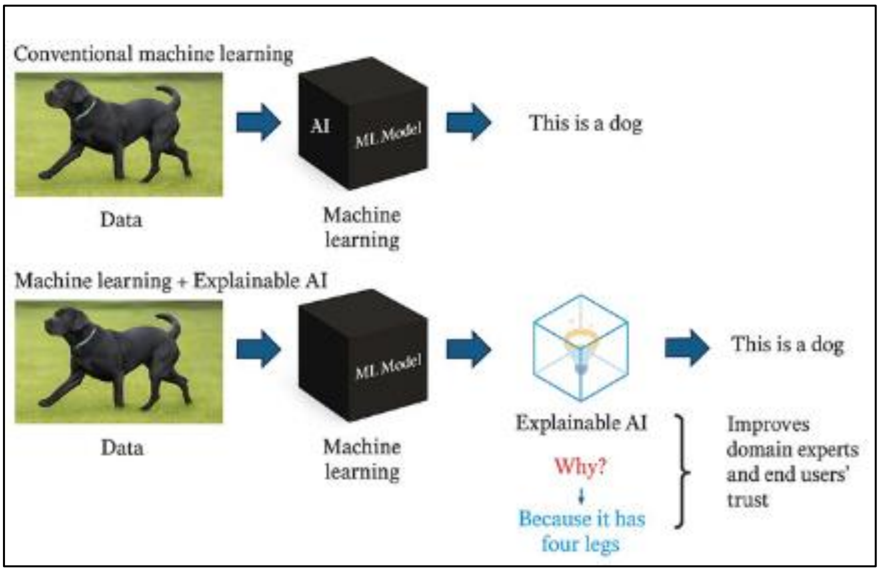

Figure 6 Illustrative comparison between the concept of conventional ML and XAI approach

3.6.6. Decision support and visualization

The final layer translates analytical outputs into actionable decisions. Interactive dashboards, GIS-integrated maps, and automated alerts provide real-time visualization and early warning for policymakers [34]. Scenario modeling supports adaptive governance, allowing managers to test interventions virtually before implementation. Feedback loops enhance resilience by updating models and thresholds dynamically, aligning monitoring with Sustainable Development Goals on clean water and climate adaptation.

4. Discussion

The synthesis of 112 studies highlights both significant progress and persistent barriers in AI-driven Water Quality Index (WQI) prediction. By organizing the results into three main dimensions; AI models, optimization techniques, and architectural frameworks; this review contributes a structured perspective on the state of the art. In this section, we critically interpret these findings in light of existing literature and outline remaining challenges and future directions.

4.1. From local model optimization to integrated architectures

While individual AI models achieve remarkable predictive accuracy, they remain largely confined to local or case-specific studies. Neural networks, support vector machines, and neuro-fuzzy hybrids consistently outperform traditional regression methods, with deep learning achieving near-perfect accuracy in some cases [6,14]. Yet their deployment is rarely linked to real-time infrastructures or standardized pipelines, limiting transferability across regions. This echoes earlier critiques in environmental informatics, where methodological advances outpace practical system integration [35, 36].

Optimization techniques further improve predictive accuracy and adaptability. Genetic Algorithms and Particle Swarm Optimization enhance parameter tuning and feature selection, while hybrid approaches (HPSOGA, MGGP) demonstrate superior flexibility [15–18]. However, their benefits remain tied to isolated prediction tasks rather than embedded within operational frameworks. The IoT–MGGP system is one of the few examples linking optimization to live monitoring workflows [18]. This suggests that optimization is not an endpoint but a facilitator of deployable AI, provided it is integrated into end-to-end architectures.

4.2. Infrastructural and data limitations

Despite technical advances, persistent infrastructural challenges hinder the scalability of AI-based WQI systems. Data availability and quality are an important concern. AI models require large volumes of high-resolution labeled data, but monitoring net- works are often sparse, sensor calibrations inconsistent, and data formats fragmented [5,37]. Missing values and seasonal gaps introduce bias and reduce generalizability, especially in underrepresented or remote regions. These issues confirm earlier observations that the" data bottleneck" is the main barrier to operational AI in environmental monitoring [38,39].

Transparency and interpretability also remain critical limitations. Complex architectures such as deep neural networks and ensembles often operate as" black boxes", making it difficult for stakeholders and regulators to validate predictions [37]. Without explainability mechanisms, trust and adoption remain limited.

Computational costs compound these challenges. High-performing models, particularly deep and hybrid architectures, often require GPUs and continuous cloud connectivity. This creates barriers for agencies in resource-constrained settings, reinforcing digital inequalities [6,18].

Finally, the field lacks standardized protocols for preprocessing, feature engineering, and benchmarking. Variability in metrics and the scarcity of open-source implementations make reproducibility and cross-study comparisons difficult [37]. Addressing these gaps is vital for building globally credible tools.

4.3. Toward standardized, full-stack frameworks

The reviewed evidence highlights the need to move beyond isolated models toward standardized, full-stack architectures for water resource assessment. As illustrated in Fig. 7, a robust pipeline should integrate real-time data acquisition (IoT sensors, satellite imagery), edge and cloud-based processing, optimized and explainable AI analytics, and decision-support dashboards for policymakers. Recent proposals for IoT–cloud integration frameworks [30,40] confirm that modular layered designs offer scalability, resilience, and adaptability, though few have been applied specifically to WQI.

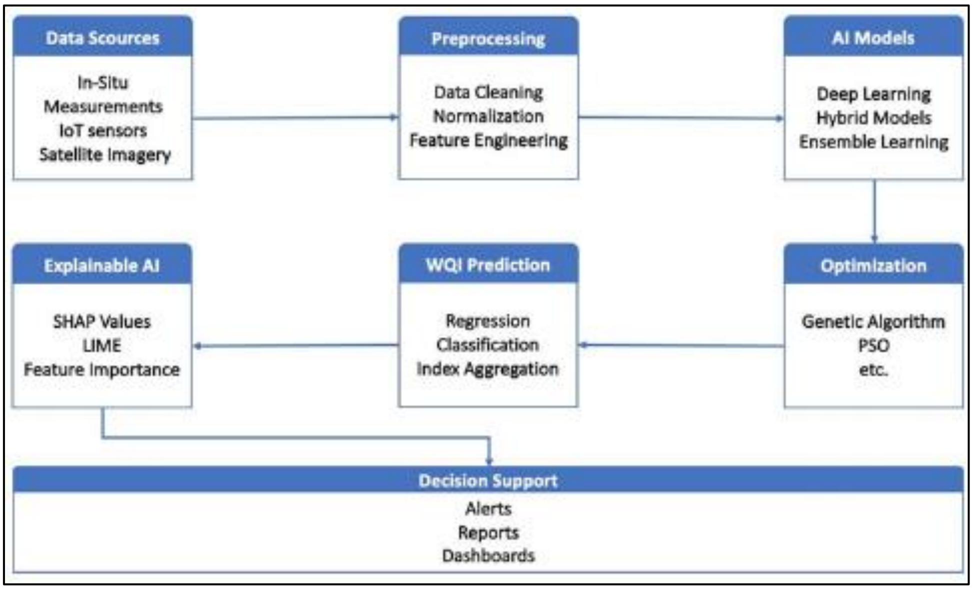

Figure 7 Proposed AI and Optimization Pipeline for Water Resource Assessment

Explainable AI (XAI) tools such as SHAP and LIME can bridge the interpretability gap, improving regulatory acceptance and stakeholder trust [32]. Decision-support dashboards connected to these systems facilitate adaptive governance, enabling early warning and scenario-based management [41].

Such architectures not only enhance technical reproducibility but also align with the Sustainable Development Goals (SDG 6: Clean Water, SDG 13: Climate Action) by ensuring scalable, transparent, and adaptive water quality monitoring.

4.4. Research gaps and future directions

This review identifies several research gaps. First, few studies incorporate spatiotemporal analytics and multi-source fusion (e.g., combining IoT, meteorological, and satellite data) for basin-scale WQI prediction. Second, interoperability standards remain underdeveloped, limiting data exchange across agencies. Third, computational efficiency, particularly edge-enabled lightweight models, needs greater emphasis to enable deployment in regions with constrained resources. Finally, benchmarking protocols and open-source toolkits are urgently required to improve reproducibility and comparability [37].

Future research should therefore prioritize modular, open, and interoperable architectures that embed optimized AI within continuous monitoring workflows. Pilot implementations in diverse hydrological contexts will be essential to validate real-world scalability. Addressing these gaps would transform AI-driven WQI prediction from an academic focus into an operational backbone of water resource management.

4.5. Threats to validity

As with any review study, certain limitations and potential biases may affect the validity of our findings. First, the search strategy, while comprehensive, was restricted to major databases (Scopus, Web of Science, IEEE Xplore, ScienceDirect, SpringerLink). Relevant studies indexed elsewhere or published in non-English sources may have been excluded, introducing a degree of publication bias.

Second, although we applied clear inclusion and exclusion criteria, the heterogeneity of reporting practices across studies meant that some methodological details (e.g., dataset size, preprocessing steps, hyperparameter settings) were either incomplete or inconsistent. This may have influenced our ability to compare performance metrics directly.

Third, bibliometric and clustering analyses were based on metadata (titles, abstracts, and keywords), which may not fully capture the conceptual depth of each article. While these methods provide valuable insights into research trends, they cannot substitute for detailed content analysis.

Finally, as the field of AI and water resource monitoring is rapidly evolving, recent advances published after our cut-off date (June 2025) may not be reflected in this synthesis. This inherent lag between literature review and publication is a known challenge in fast-moving domains.

Despite these threats to validity, the methodological triangulation employed, combining structured literature review, bibliometric mapping, and comparative syn- thesis, provides confidence that the patterns and gaps identified in this study are representative of the state of the art.

5. Conclusion

This review has synthesized recent advances in artificial intelligence (AI) and optimization techniques for Water Quality Index (WQI) prediction. Models such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and hybrid frameworks have demonstrated strong predictive performance, often surpassing traditional regression-based methods. Optimization approaches including Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Multi-Gene Genetic Programming (MGGP) further enhance accuracy by supporting feature selection and parameter tuning. Together, these advances highlight the growing maturity of AI-driven water quality assessment.

Nevertheless, most existing solutions remain fragmented, confined to localized studies, and rarely embedded within real-time monitoring infrastructures. This limits their scalability and applicability in large-scale water resource management. A critical finding of this review is the need to move beyond isolated model development toward standardized, modular system architectures that integrate AI models with IoT sensors, edge computing, and cloud platforms. Such end-to-end frameworks would enable continuous monitoring, adaptive feedback, and early warning systems across diverse hydrological settings.

Compliance with ethical standards

Acknowledgments

During the preparation of this work the authors used ChatGPT in order to reformulate, check grammar, and structure ideas. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Disclosure of conflict of interest

The authors declare that they have no conflicts of interest related to this study.

Author Contributions Statement

All authors made substantial contributions to this study. Sara Bouziane conceptualized and designed the study, curated the dataset, conducted the investigation and analyses, prepared visualizations, and drafted the original manuscript. Badreddine Agouties and Abdellah El Hmaidi supervised the research, managed the project administration, and contributed to critical review and editing of the manuscript. Aniss Moumen performed formal analyses, contributed to data visualization, and participated in manuscript revision. Anas El Ouali contributed to validation, provided resources, and reviewed and edited the manuscript. Ali Sahraoui contributed to the methodological design, validation of results, and manuscript review and editing. All authors reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

References

[1] Mohsin, M.: Sustainable water management solutions for urban areas. International Journal for Research in Applied Science and Engineering Technology, 12, 538–540 (2024) https://doi.org/10.22214/ijraset.2024.64528

[2] Dwivedi, R., Dubey, A.: A review of water quality index (wqi) assessment methods. International Journal of Chemical Research and Development, 5, 24–27 (2023) https://doi.org/10.33545/26646552.2023.v5.i2a.51

[3] Dubey, S., Dubey, S., Raghuwanshi, K.: Unlocking iot and machine learning's potential for water quality assessment: An extensive analysis and future directions. Water Conservation Science and Engineering, 10, 18 (2025) https://doi. org/10.1007/s41101-025-00342-7

[4] Kumar, D., Kumar, R., Sharma, M., Awasthi, A., Kumar, M.: Global water quality indices: Development, implications, and limitations. Total Environment Advances, 9, 200095 (2024) https://doi.org/10.1016/j.teadva.2023.200095

[5] Frincu, R.M.: Artificial intelligence in water quality monitoring: A review of water quality assessment applications. Water Quality Research Journal, 60, 164–176 (2025) https://doi.org/10.2166/wqrj.2024.049

[6] Toumi, S., Lekmine, S., Touzout, N., Moussa, H., Elboughdiri, N., Boudraa, R., Benslama, O., Kebir, M., Danish, S., Zhang, J., Amrane, A., Tahraoui, H.: Harnessing deep learning for real-time water quality assessment: A sustainable solution. Water, 16, 3380 (2024) https://doi.org/10.3390/w16233380

[7] Mahule, A., Roy, K., Sawarkar, A.D., Lachure, S.: Enhancing Environmental Resilience, pp. 48–74. CRC Press, Boca Raton, USA (2024). https://doi.org/10. 1201/9781032683805-4

[8] Nikoo, M.R., Kerachian, R., Malakpour-Estalaki, S., Bashi-Azghadi, S.N., Azimi- Ghadikolaee, M.M.: A probabilistic water quality index for river water quality assessment: a case study. Environmental Monitoring and Assessment, 181, 465– 478 (2011) https://doi.org/10.1007/s10661-010-1842-4

[9] Gaya, M.S., Abba, S.I., Abdu, A.M., Tukur, A.I., Saleh, M.A., Esmaili, P., Wahab, N.A.: Estimation of water quality index using artificial intelligence approaches and multi-linear regression. IAES International Journal of Artificial Intelligence (IJ-AI), 9, 126 (2020) https://doi.org/10.11591/ijai.v9.i1.pp126-134

[10] Ghadai, M., Satapathy, D.P., Krishnasamy, S., Ramalingam, M., Sreelal, G.P., Dhilipkumar, B.: Artificial neural network and weighted arithmetic indexing approach for surface water quality assessment of the brahmani river. Global Nest Journal, 24, 562–568 (2022) https://doi.org/10.30955/gnj.004414

[11] Aslam, B., Maqsoom, A., Cheema, A.H., Ullah, F., Alharbi, A., Imran, M.: Water quality management using hybrid machine learning and data mining algorithms: An indexing approach. IEEE Access, 10, 119692–119705 (2022) https://doi.org/ 10.1109/ACCESS.2022.3221430

[12] Xia, J., Zeng, J.: Environmental factors assisted the evaluation of entropy water quality indices with efficient machine learning technique. Water Resources Management, 36, 2045–2060 (2022) https://doi.org/10.1007/s11269-022-03126-z

[13] Mamat, N., Razali, S.F.M., Hamzah, F.B.: Enhancement of water quality index prediction using support vector machine with sensitivity analysis. Frontiers in Environmental Science, 10 (2023) https://doi.org/10.3389/fenvs.2022.1061835

[14] Nanjappachetty, A., Sundar, S., Vankadari, N., Bapu, T.B.B.R., Shanmugam, P.: An efficient water quality index forecasting and categorization using optimized deep capsule crystal edge graph neural network. Water Environment Research, 96 (2024) https://doi.org/10.1002/wer.11138

[15] Abdulkareem, I., Abbas, A., Dawood, A.: Modeling pollution index using artificial neural network and multiple linear regression coupled with genetic algorithm. Journal of Ecological Engineering, 23, 236–250 (2022) https://doi.org/10.12911/ 22998993/146177

[16] Agrawal, P., Sinha, A., Kumar, S., Agarwal, A., Banerjee, A., Villuri, V.G.K., Annavarapu, C.S.R., Dwivedi, R., Dera, V.V.R., Sinha, J., Pasupuleti, S.: Exploring artificial intelligence techniques for groundwater quality assessment. Water, 13, 1172 (2021) https://doi.org/10.3390/w13091172

[17] Chia, S.L.: Sustainable management of river water quality using artificial intelligence optimisation algorithms. PhD thesis (2021)

[18] Palconit, M.G.B., Bautista, M.G.A.C., II, R.S.C., Alejandrino, J.D., Evangelista, I.R.S., Alajas, O.J.Y., Vicerra, R.R.P., Bandala, A.A., Dadios, E.P.: Multi-gene genetic programming of iot water quality index monitoring from fuzzified model for oreochromis niloticus recirculating aquaculture system. Journal of Advanced Computational Intelligence and Intelligent Informatics, 26, 816–823 (2022) https://doi.org/10.20965/jaciii.2022.p0816

[19] Ramadh, B.: IoT-Enabled Smart Water Management: Reducing Waste and Enhancing Efficiency (2025). https://doi.org/10.2139/ssrn.5146167

[20] Singh, Y., Walingo, T.: Smart water quality monitoring with iot wireless sensor networks. Sensors, 24, 2871 (2024) https://doi.org/10.3390/s24092871

[21] Verma, T.S.: Water quality analysis through satellite images. International Journal of Scientific Research In Engineering And Management, 08, 1–5 (2024) https://doi.org/10.55041/IJSREM33992

[22] Suhonen, J., Kohvakka, M., Kaseva, V., Hamalainen, T.D., Hannikainen, M.: Low-Power Wireless Sensor Network Platforms, pp. 381–419. Springer, New York, USA (2013). https://doi.org/10.1007/978-1-4614-6859-2 13

[23] Chowdury, M.S.U., Emran, T.B., Ghosh, S., Pathak, A., Alam, M.M., Absar, N., Andersson, K., Hossain, M.S.: Iot based real-time river water quality monitoring system. Procedia Computer Science, 155, 161–168 (2019) https://doi.org/10. 1016/j.procs.2019.08.025

[24] Hidayati, D., Andriyansah, A., Cesna, G.P., Fauzi, A.Y., Apriliasari, D., Rahardja, U.: Building efficient iot systems with edge computing integration. International Journal of Cyber and IT Service Management, 4 (2024) https://doi.org/10.34306/ijcitsm.v4i2.163

[25] Donta, P.K., Dustdar, S.: Towards intelligent data protocols for the edge. In: 2023 IEEE International Conference on Edge Computing and Communications (EDGE), pp. 372–380. IEEE, Chicago, USA (2023). https://doi.org/10.1109/ EDGE60047.2023.00060

[26] Kolapo, R., Kawu, F.M., Abdulmalik, A.D., Edem, U.A., Young, M.A., Mordi, E.C.: Edge computing: Revolutionizing data processing for iot applications. International Journal of Science and Research Archive, 13, 023–029 (2024) https://doi.org/10.30574/ijsra.2024.13.2.2082

[27] Bala, B., Behal, S.: A brief survey of data preprocessing in machine learning and deep learning techniques. In: 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 1755–1762. IEEE, Dharan-, 8, Sunsari, Nepal (2024). https://doi.org/10.1109/I-SMAC61858.2024.10714767

[28] Jandhyala, V.S.V.: Mastering data pipelines for al: A beginner's guide to building efficient workflows. International Journal for Multidisciplinary Research, 6 (2024) https://doi.org/10.36948/ijfmr.2024.v06i05.29550

[29] Gupta, S., Sharma, H.K., Kapoor, M.: Artificial Intelligence-Based Cloud Storage for Accessing and Predication, pp. 157–168. Springer, New York, USA (2023). https://doi.org/10.1007/978-3-031-18896-1 13

[30] Reddy, P.: Integrating edge computing with advanced cloud computing: A paradigm shift for iot applications. World Journal of Advanced Research and Reviews, 2, 037–049 (2019) https://doi.org/10.30574/wjarr.2019.2.1.0036

[31] Vasques, X.: Machine Learning in Production, pp. 375–464. Wiley, Hoboken, USA (2024). https://doi.org/10.1002/9781394220649.ch6

[32] Hellen, N., Sabuj, H.H., Alam, M.A.: Explainable AI and Ensemble Learning for Water Quality Prediction, pp. 235–250 (2023). https://doi.org/10.1007/ 978-981-19-7528-8 19

[33] Dharmarathne, G., Abekoon, A.M.S.R., Bogahawaththa, M., Alawatugoda, J., Meddage, D.P.P.: A review of machine learning and internet-of-things on the water quality assessment: Methods, applications and future trends. Results in Engineering, 26, 105182 (2025) https://doi.org/10.1016/j.rineng.2025.105182

[34] Dhumvad, A., Prabhu, S., Silva, S.F.D., Simu, S., Padiyar, P., Turkar, V., Salgaonkar, V.: Water pollution monitoring and decision support system. In: 2022 3rd International Conference for Emerging Technology (INCET), pp. 1–6. IEEE, Belgaum, India (2022). https://doi.org/10.1109/INCET54531.2022.9824110

[35] Yan, X., Zhang, T., Du, W., Meng, Q., Xu, X., Zhao, X.: A comprehensive review of machine learning for water quality prediction over the past five years. Journal of Marine Science and Engineering, 12, 159 (2024) https://doi.org/10.3390/ jmse12010159

[36] Dandekar, P., Thakre, V., Sharma, K., Kushwaha, A.: A predictive model for water quality index assessment by machine learning approach. In: 2024 2nd Inter- national Conference on Computer, Communication and Control (IC4), pp. 1–6. IEEE, Indore, India (2024). https://doi.org/10.1109/IC457434.2024.10486535

[37] Lowe, M., Qin, R., Mao, X.: A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring. Water, 14, 1384 (2022) https://doi.org/10.3390/w14091384

[38] Mustafa, H.M., Mustapha, A., Hayder, G., Salisu, A.: Applications of iot and artificial intelligence in water quality monitoring and prediction: A review. In: 2021 6th International Conference on Inventive Computation Technologies (ICICT), pp. 968–975. IEEE, Tamil Nadu, India (2021). https://doi.org/10.1109/ ICICT50816.2021.9358675

[39] Jenila, C., Kadambarajan, J., Vardhan, V.H., Reddy, S.V.V., Krishna, V.R., Vardhan, K.V.: Ai and iot-assisted water quality monitoring and prediction. In: 2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 309–314. IEEE, Tamil Nadu, India (2024). https://doi.org/10.1109/ICECA63461.2024.10801021

[40] Leonila, T., Senthil, G.A., Geerthik, S., Sowmiya, R., Nithish, J.: Dynamic water quality monitoring via iot sensor networks and machine learning technique. In: 2024 International Conference on Communication, Computing and

Internet of Things (IC3IoT), pp. 1–6. IEEE, Chennai, India (2024). https://doi.org/10.1109/IC3IoT60841.2024.10550224

[41] Raman, R., Martin, N.: Iot-enabled water pollution detection for real-time monitoring and pollution source identification with mqtt protocol. In: 2024 Inter- national Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), pp. 1–6. IEEE, Chennai, India (2024). https://doi.org/10.1109/ADICS58448.2024.10533607