Proceedings of the 7th African International Conference on Industrial Engineering and Operations Management
Cape Town, South Africa, May 13-15, 2026

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/AF07.20260436
Published: May 13, 2026

Adoption of Machine Learning–Driven Predictive Maintenance for Optimal Operational Reliability in Water Infrastructure Systems in Developing Countries: A Bibliometric Review

Gift Laolang
Faculty of Engineering, Built Environment & Information Technology
Walter Sisulu University, Buffalo City Campus
East London, Eastern Cape, South Africa
glaolang@wsu.ac.za

Dr Motheo Meta Tjebane (PhD)
Faculty of Engineering and the Built Environment
Durban University of Technology
Durban, KwaZulu-Natal, South Africa
motheometa@gmail.com

Abstract

Traditional maintenance strategies, typically reactive or schedule based, often fail to prevent water infrastructure failures and operational inefficiencies. In response, machine learning–driven predictive maintenance has emerged as a promising approach for enhancing water infrastructure reliability through early fault detection and data-driven decision-making. This study presents a bibliometric review of 29 peer-reviewed, Scopus-indexed publications to examine global research trends, thematic developments, and geographical contributions within this domain. The findings reveal that India emerges as the most influential contributor with eight publications indicating significant advancements in predictive maintenance and machine learning applications. In contrast, African representation remains limited, with only South Africa and Sudan appearing in the dataset, highlighting a critical regional research gap. The most prolific journals include Water (Switzerland) and Sensors, reflecting a strong focus on water infrastructure and sensor based technologies. Thematic analysis identifies four dominant research streams: machine learning models and advanced artificial intelligence techniques, water infrastructure applications, IoT-enabled decision-support systems, and infrastructure management systems. Furthermore, density visualisation reveals a highly centralised knowledge structure characterised by distinct clustering patterns, suggesting both specialisation and limited cross-cluster integration. The study emphasises the need to expand research and implementation efforts in African regions to support the adoption of machine learning–driven predictive maintenance, thereby enhancing the sustainability, resilience, and operational efficiency of water infrastructure systems.

Keywords
Machine Learning, Predictive Maintenance, Water Infrastructure Systems, Developing Countries

1. Introduction
Water infrastructure systems are critical to socio-economic development, public health, and environmental sustainability (Valero et al., 2025; Jerolleman et al. 2026). These systems including water supply networks, treatment facilities, and distribution pipelines form the backbone of urban and rural service delivery (Eissa & Nasr, 2025).

However, Ángel and Tekinerdogan (2024) stressed that water supply networks are increasingly under strain due to ageing infrastructure, rapid urbanisation, climate variability, and growing demand for reliable water services. These challenges are particularly pronounced in developing countries, where water infrastructure systems are often characterised by frequent failures, high levels of non-revenue water, and inadequate maintenance regimes, ultimately undermining operational reliability and service efficiency (Adeniran, 2022; Yazdekhasti et al., 2020; Bulti & Amelo, 2023). Traditionally, maintenance practices in water infrastructure have relied on reactive and preventive approaches (Horbatuck & Beruvides, 2019; Chiadighikaobi et al., 2025). In developing countries such as South Africa, the reliance on reactive maintenance has led to significant backlogs, high emergency repair costs, and poor infrastructure management (Ndlovu et al., 2025). Moreover, Geisbush and Ariaratnam (2023) highlighted that reactive maintenance addresses failures after they occur, often resulting in service disruptions, increased operational costs, and potential environmental and public health risks. On the other hand, Alhamad et al. (2022) argue that preventive maintenance, while more proactive than reactive approaches, and is often based on predetermined schedules rather than the actual condition of infrastructure assets. As a result, this approach may lead to inefficient resource allocation, either by performing unnecessary maintenance or by failing to address emerging issues on time. As a result, these approaches often lead to inefficiencies, including unnecessary maintenance interventions or delayed responses to critical failures. Collectively, these limitations highlight the inadequacy of conventional maintenance strategies in addressing the dynamic and complex nature of modern water infrastructure systems (Okeola et al., 2025).

In response to these challenges, the integration of digital technologies, particularly machine learning (ML), is transforming infrastructure management by enabling predictive, efficient, and sustainable solutions (Ganga et al., 2025). Furthermore, Pani, et al. (2024) asserted that ML-driven predictive maintenance leverages historical and real-time data to detect patterns, predict system failures, and optimise maintenance decision-making. Rathore et al. (2024) also added that this data-driven approach has the potential to enhance operational reliability, reduce downtime, and improve resource allocation. However, Rahim et al. (2025) stressed that, despite these advantages, the adoption of ML-driven predictive maintenance in water infrastructure systems especially in developing countries, specifically African countries, remains limited. Key barriers include insufficient data availability, limited technical capacity, high implementation costs, and institutional constraints (Tlabu et al., 2022). This gap restricts a comprehensive understanding of research trends, thematic developments, and existing knowledge gaps, thereby constraining informed decision making and effective implementation. This study aims to systematically analyse and map the existing body of knowledge on machine learning–driven predictive maintenance for enhancing operational reliability in water infrastructure systems in developing countries. By employing a bibliometric review approach, the study seeks to provide a structured synthesis of research trends, identify key thematic areas, and highlight critical gaps.

1.1 Objectives

To systematically analyse and map the existing body of knowledge on machine learning driven predictive maintenance and its role in enhancing the operational reliability of water infrastructure systems in developing countries using bibliometric techniques. To identify research themes, gaps and propose future research directions for improving water infrastructure systems reliability.

2. Literature Review

Embarki et al. (2024) asserted that the concept of predictive maintenance has gained increasing attention as a strategic approach to improving infrastructure performance. Unlike traditional maintenance methods, predictive maintenance leverages data analytics and advanced computational techniques to anticipate system failures and optimise maintenance interventions (Dhanya et al., 2025). In recent years, machine learning has emerged as a key enabler of predictive maintenance, offering the capability to process large datasets, identify complex patterns, and generate accurate predictions. In the context of infrastructure systems, ML-driven predictive maintenance has been applied across various sectors, including transportation, energy, and manufacturing (Frost et al., 2021; Hundekari et al., 2025). These applications demonstrate significant improvements in asset performance, cost efficiency, and system reliability (Rihan et al., 2025). Existing studies highlight the potential of machine learning (ML) techniques such as neural networks, decision trees, and support vector machines in detecting anomalies, predicting pipe failures, and optimising maintenance schedules. These approaches enable condition-based monitoring and facilitate proactive decision-making. Nevertheless, Bui et al. (2024) expressed that, machine learning (ML) techniques implementation is often constrained by challenges related to data quality, interoperability of systems, and limited digital infrastructure. Furthermore, the adoption of ML-driven solutions in developing countries is influenced by socio-economic and institutional factors (Phuong & Huong, 2025; Srivastava & Dixit, 2023). Additionally, limited financial resources, lack of skilled personnel, and weak regulatory frameworks often hinder the integration of advanced technologies into existing infrastructure systems (Ismail et al., 2024; Toumba et al., 2025; Ntanis et al., 2025).

In the water sector, predictive maintenance applications are increasingly being explored, particularly in areas such as pipe failure prediction, leak detection, pump performance monitoring, and water quality assessment (Ortiz et al., 2023; Amorelli et al., 2025; Madala, 2025; Shashikala et al.. 2025). Studies have revealed that these models leverage historical failure data and pipeline attributes (e.g. age, material, environmental factors) to estimate failure probabilities and identify high risk areas (Zhang & Ye, 2020). Additionally, Sahoo et al. (2024) enunciates that predictive models have been applied to monitor pump operations and detect anomalies that may indicate impending failures. In the United States, Cao et al. (2024) found that the application of machine learning predictive models to forecast influent characteristics allows for proactive operational adjustments, resulting in improved efficiency, optimised resource utilisation, and reduced operational costs. Joseph et al. (2025) in Australia, looked at early leak and burst detection in water pipeline networks using machine learning (ML) approaches. The study finds that machine learning significantly improves leak detection in water distribution networks compared to traditional methods, particularly in handling complex data patterns. Through a comparative analysis of fourteen algorithms, models such as Random Forest, K-Nearest Neighbours, and Decision Tree demonstrated superior performance across key metrics, including accuracy and precision. In India, Rafi et al. (2025) study demonstrates that integrating machine learning and predictive analytics into web/mobile-based tools significantly enhances the management of water supply networks. The findings show that the system improves the accuracy and dynamism of network mapping, enables real-time monitoring and rapid issue detection, and supports predictive maintenance through advanced analytics. By applying techniques such as LSTM for time-series forecasting, linear regression, clustering (K-means), and support vector machines, the tool effectively analyses diverse data sources to forecast potential failures and optimise maintenance scheduling.

Despite these advancements, the adoption of ML-driven predictive maintenance in water infrastructure systems remains uneven, particularly in developing countries in Africa. Odoi (2025) stressed that while predictive maintenance holds transformative potential for African water sectors, its adoption is hindered by data scarcity, financial constraints, and technical challenges. Rubaba and Walingo (2025) found several critical gaps in water quality monitoring within eThekwini in South Africa, including limited real-time surveillance, uneven data management systems, budgetary constraints, and cybersecurity vulnerabilities. Further highlights disparities between rural and urban access, slow commercialisation of academic innovations, policy misalignment, and insufficient technical capacity, all of which collectively hinder effective and sustainable water quality management. A similar study conducted by Ngwenya et al. (2025) in South Africa, found that the need for further refinement of ML models, limited monitoring of critical parameters such as nitrogen, phosphorus, and total oxidised nitrogen, and insufficient use of standardised evaluation frameworks, such REFORMS checklist. Additionally, the findings highlight minimal attention to groundwater monitoring, inadequate model prototyping, and a heavy reliance on battery-powered sensors with limited exploration of power-harvesting alternatives. The study also points to restricted open access to ambient water quality data, which constrains model development and broader research advancement.

In light of these gaps, this study adopts a bibliometric approach to systematically analyse the existing literature on the adoption of machine learning–driven predictive maintenance for water infrastructure systems in developing countries. By mapping publication trends, thematic clusters, and geographic distribution of research, the study seeks to provide a comprehensive understanding of the current state of knowledge.

3. Methods

To achieve the research objectives, this study adopts a quantitative bibliometric approach supported by scientometric analysis. A systematic and structured procedure was followed to identify, screen, and select relevant publications. The research process comprised two main stages: (i) data collection and refinement, and (ii) scientometric analysis and visualisation. Bibliometric analysis is widely recognised as a robust method for mapping and evaluating scientific research outputs, enabling the identification of publication trends, influential contributions, and knowledge structures within a given field (Stefanis et al., 2025). Accordingly, this approach is well-suited to the present study as it provides an objective, transparent, and evidence-based means of analysing the existing body of knowledge. Furthermore, it facilitates the identification of under-researched areas and emerging themes, thereby offering a foundation for future research directions (Kumar, 2025). Scientometric analysis, as a subset of bibliometric, was adopted to quantitatively examine relationships within the literature, including co-authorship networks, keyword co-occurrence, and thematic clustering. Unlike traditional narrative reviews, this approach enables the systematic exploration of patterns, trends, and intellectual structures within the research domain, enhancing the rigor and reproducibility of the study.

4. Data Collection

Data collection for this study followed a structured and replicable process to ensure the reliability, transparency, and validity of the bibliometric analysis. The Scopus database was selected due to its extensive coverage of high-quality, peer-reviewed publications across multiple disciplines (Baas et al., 2020). The database was systematically queried in March 2026 using predefined search strings aligned with the study objectives. The initial search yielded 58 documents. These records were subsequently refined through a series of filtering stages to ensure relevance and quality. Filters were applied to include only peer-reviewed journal articles, publications written in English, and subject areas directly related to the research domain. This process reduced the dataset to 43 eligible publications. To enhance the robustness of the dataset, an additional quality assessment and source ranking process was undertaken. This involved evaluating the relevance and contribution of each publication to the research domain. Based on these criteria, 29 documents met the required threshold and were selected for inclusion. The final dataset of 29 peer-reviewed articles formed the basis for the subsequent scientometric analysis. Chen et al. (2025) concur that smaller datasets can yield meaningful insights, particularly when the research domain is narrowly defined and the data are carefully curated for relevance. Similarly, Rather et al. (2024) argue that smaller datasets are well suited for analytical tasks such as thematic evolution analysis, keyword co-occurrence mapping, and the identification of emerging research trends. This dataset enabled the systematic mapping of research trends, identification of thematic clusters, and exploration of the knowledge structure within the field. The full search code is as follows: TITLE-ABS-KEY (machine AND learning AND predictive AND maintenance AND water AND infrastructure AND systems AND developing AND countries) AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND (LIMIT-TO (SUBJAREA , "ENGI") OR LIMIT-TO (SUBJAREA , "ENVI")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp")) AND (LIMIT-TO (LANGUAGE , "English")).

The refined records were exported in comma-separated values (CSV) format and prepared for further screening and analysis. The CSV format ensured compatibility with VOSviewer, the tool used for the scientometric analysis. The VOSviewer tool is widely used to visualise and analyse these clusters, facilitating clearer interpretation of relationships between publications, authors, and keywords (Jie, 2022). A preliminary screening was conducted in Microsoft Excel, in which titles, abstracts, and keywords were systematically reviewed. Only studies explicitly related to Machine Learning–Driven predictive maintenance for operational reliability in water infrastructure systems for further analysis.

5. Results and Discussion

5.1 Annual Publications Per Year (2022 – 2025)

The output of journal publications is presented in Figure 1, which illustrates the annual number of publications indexed in Scopus. In 2020 and 2021, only one publication was recorded in each year, indicating the early stage of research development in this field. A moderate increase to three publications in 2022 suggests emerging scholarly interest, likely driven by advancements in digital technologies and predictive analytics. However, this growth was not sustained in 2023, when publications declined to one, reflecting the fragmented nature of the research area. Research activity regained momentum in 2024, with four publications, then surged to 19 in 2025. This sharp increase highlights a rapid expansion of academic interest and emphasise the growing recognition of machine learning–driven predictive maintenance as a critical approach for enhancing operational reliability in water infrastructure systems.

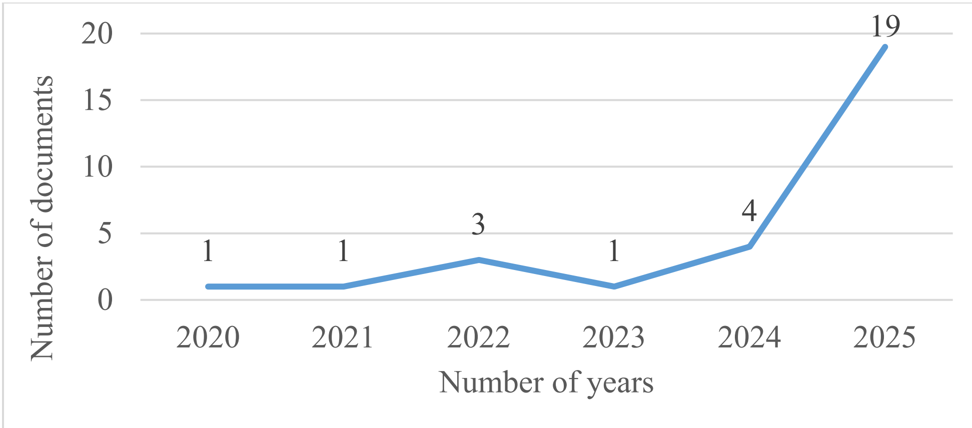

Figure 1. Annual publications from 2020 to 2025

5.2 Analysis of Countries by Documents and Citations

Country contributions were assessed based solely on the number of documents produced per country (Table 1) and Figure 2 (world map). India leads in research output with 8 publications and 31 citations, demonstrating both high productivity and scholarly impact. This is followed by the United States (4 publications; 3 citations) and China, which also shows a notable contribution with 3 publications, although with comparatively high citation impact (25 citations). United Kingdom (2 publications; 27 citations), Iraq (2 publications; 11 citations), Malaysia (2 publications; 3 citations) and Spain (2 publications; 5 citations), indicating strong influence despite relatively fewer publications. Research and implementation of machine learning in water sectors are dominated by countries such as India, the United States, China, the United Kingdom, Iraq, Malaysia and Spain, with limited contributions from Africa developing countries. Representation of developing African countries remains significantly low, with only South Africa (1 publication; 1 citation) and Sudan (1 publication; 5 citations). This highlights a critical research gap, particularly given the pressing need for improved water infrastructure systems across the continent. The comparatively low citation impact further suggests the limited visibility and influence of African research in this domain.

Table 1. Analysis of countries by documents and citations

Country	Documents	Citations
India	8	31
United States	4	3
China	3	25
United Kingdom	2	27
Iraq	2	11
Malaysia	2	3
Spain	2	5
Kuwait	1	5
Saudi Arabia	1	5
Sudan	1	5
Uzbekistan	1	5
Iran	1	6
Brazil	1	6
Finland	1	7
Netherlands	1	16
Romania	1	4
South Africa	1	1
Turkey	1	1

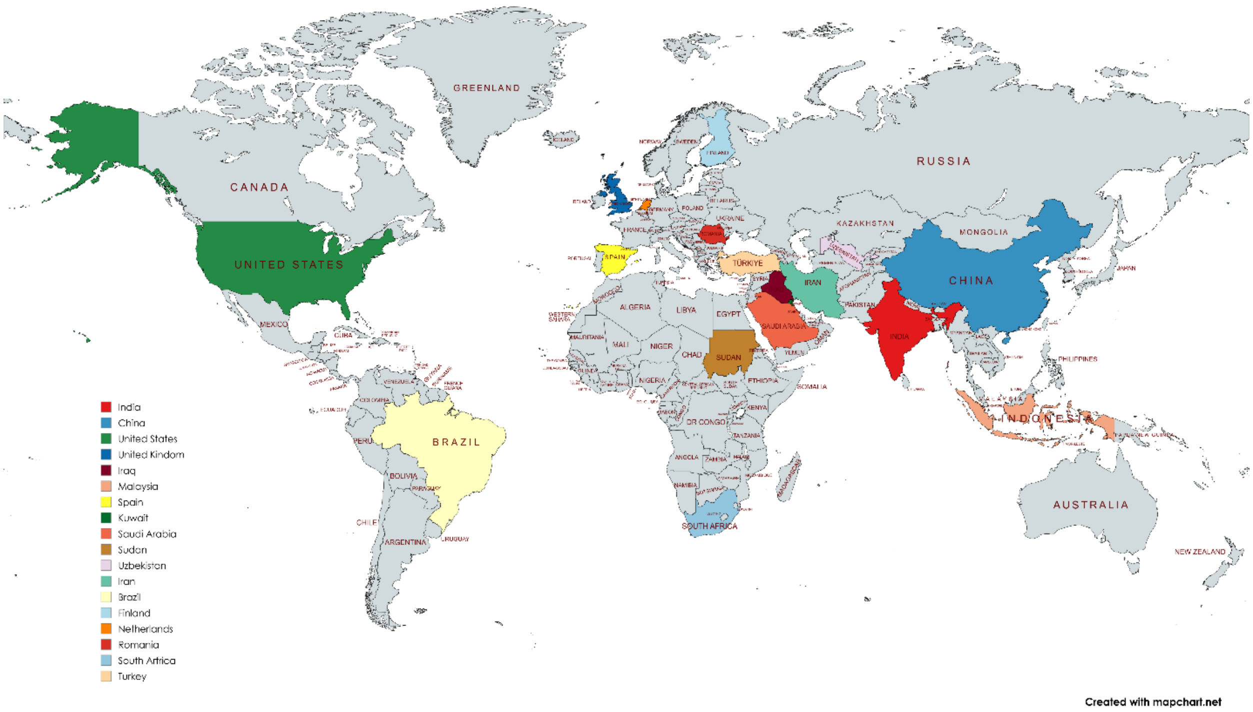

Figure 2. World map of all documents distributed per country

5.4 Prolific Journals

The analysis of publication sources in Table 2 indicates that research on machine learning–driven predictive maintenance in water infrastructure systems is dispersed across a diverse range of journals and conference proceedings, reflecting the interdisciplinary nature of the field. The most prolific journals are Water (Switzerland) and Sensors, each contributing 2 documents with 9 citations each. These journals reflect the strong influence of water management and sensor-based technologies in advancing research on infrastructure monitoring and predictive systems. The presence of Sensors (Switzerland) with 1 document but a relatively high citation count (26 citations) indicates that sensor-based methodologies are particularly influential and widely referenced in the literature.

Table 2. Top ten contributing Journals

Journal	Documents	Citations
Water (Switzerland)	2	9
Sensors	2	9
Sensors (Switzerland)	1	26
Water Research	1	16
2nd IEEE International Conference on Networks, Multimedia and Information Technology (NMITCON)	1	16
Engineering Research Express	1	15
Applied Chemical Engineering	1	11
Computational Methods in Applied Science	1	7
Water Conversation and Management	1	6
World Environmental and Water Resource Congress	1	6

5.5 Co-occurring Keywords

The keyword co-occurrence analysis provides insight into the dominant research themes and intellectual structure of the field. The results indicate a strong concentration around machine learning applications and predictive maintenance, confirming the growing interest in data-driven approaches for water infrastructure management. The most prominent keyword is “Machine Learning” with 19 occurrences and the highest total link strength (50), indicating its central role in connecting multiple research themes. Closely related is “Predictive Maintenance” (17 occurrences; link strength 45), which confirms that failure prediction and maintenance optimisation are core research priorities within water infrastructure systems.

Table 3. Top ten co-occurring keywords

Keyword	Occurrences	Total Link Strength
Machine Learning	19	50
Predictive Maintenance	17	45
Learning Systems	12	44
Machine-Learning	13	42
Predictive Analysis	6	26
Water Distribution Systems	6	25
IoT (Internet of Things)	4	22
Internet of Things	4	22
Water Management	5	21
Water Supply	5	19

5.5 Cluster Analysis

Cluster analysis is a powerful bibliometric technique that enables researchers to uncover hidden patterns, assess research performance, and identify emerging trends (Muhamad, 2023). It is commonly applied to detect and label emerging research areas by grouping publications through methods such as bibliographic coupling, co-citation, or co-word analysis (Carrión, 2022). Density visualizations were adopted to display areas of high and low concentration of prominent themes or topics within a field. This scientometric literature review identified four distinct clusters, respectively. Table 4 and Figure 3 illustrate the keywords network according to the clusters.

Table 4. Keywords co-occurrence network

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Machine learning	Deterioration	Decision making	Asset management
Machine-learning	Forecasting	Internet of things	Information management
Predictive maintenance	Learning systems	Internet of things	Learning algorithms
Maintenance	Predictive analytics	Internet of things (IoT)	
Artificial intelligence	Water supply		
Artificial neural networks	Water management		
Deep learning	Water distribution networks		
Graph neural networks	Water distribution systems		
Article	Decision trees		
Sustainability	Support vector machines		

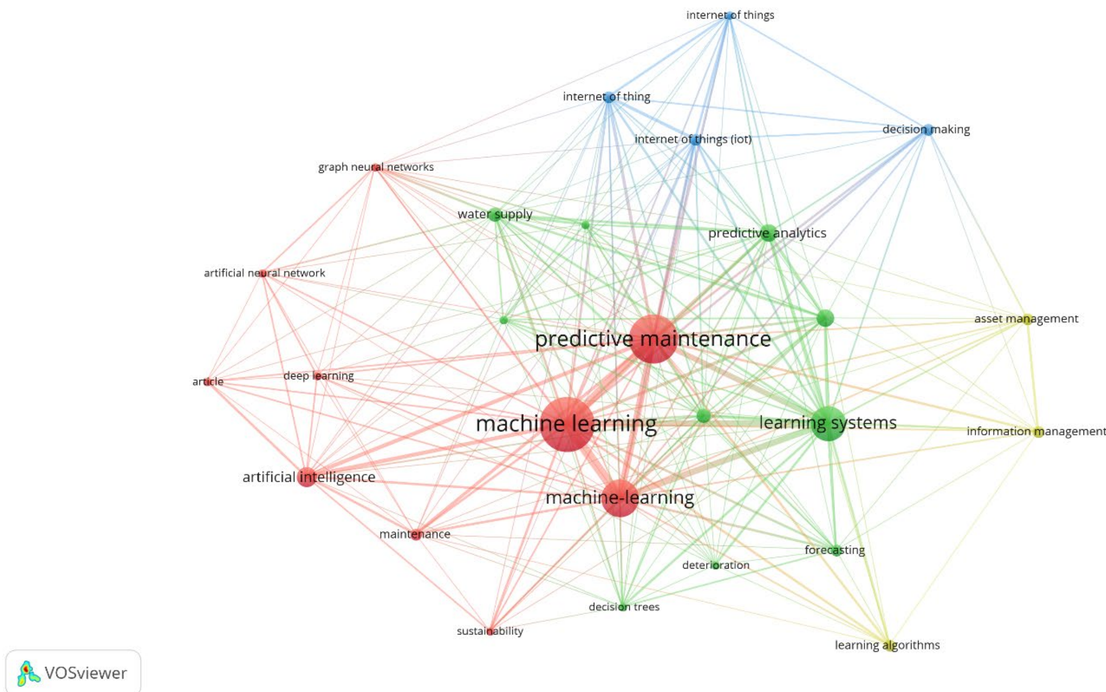

Figure 3. Keywords co-occurrence network

5.5.1 Cluster 1 (Red): Machine Learning and Advanced AI Techniques

This cluster focuses on ML-driven predictive maintenance by providing the core modelling techniques required to detect patterns, anomalies, and failure risks. AI models achieve high fault detection accuracy and reduce false alarms, enabling timely interventions (Balakrishnan & Subramani, 2025). Supervised, unsupervised, and deep learning models are widely used for fault detection, anomaly detection, and predictive analytics in water systems (Balaraman et al., 2025). Techniques such as XGBoost, autoencoders, Long Short-Term Memory (LSTM) and Graph neural networks have shown high accuracy in detecting faults and predicting failures (Saranya et al., 2025). Moreover, Hu et al. (2025) demonstrate that comparative evaluations of Graph Neural Network (GNN) architectures, including Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), and GraphSAGE, indicate that GraphSAGE achieves superior predictive performance, significantly outperforming traditional machine learning methods. Xiong (2025) presents a comparative evaluation of multiple machine learning models for water quality prediction using regional monitoring data, highlighting that the Random Forest model is highly robust to overfitting, making it particularly suitable for handling complex datasets. Tiquio and Villanueva (2024) concurred that, for pipeline leak classification using machine learning ensemble methods and hyper parameter tuning, the Artificial Neural Network (ANN) achieved the highest accuracy (92%), followed closely by Long Short-Term Memory (LSTM) at 91% and Support Vector Machine (SVM) at 90%, while the Random Forest model recorded a comparatively lower accuracy of 86%.

5.5.2 Cluster 2 (Green): Water Infrastructure and Predictive Applications

This cluster highlights the practical application of solutions within water infrastructure systems. It focuses on deterioration modelling, system monitoring, and failure prediction, effectively linking artificial intelligence techniques with infrastructure performance. Ndlovu et al. (2025), in the South African context, demonstrate this application through pipeline failure prediction. Their study shows that machine learning models, such as extreme gradient boosting and survival analysis, can effectively predict the probability of pipeline failures. By leveraging historical data, these models enable the prioritisation of maintenance based on pipeline condition, performance, and criticality. This approach supports municipalities in transitioning from reactive to proactive maintenance strategies, ultimately improving the reliability and sustainability of water infrastructure systems. A similar study by Zuo et al. (2025), conducted in Sudan, developed a machine learning optimised multi-generation energy system designed to simultaneously produce electricity, freshwater, and hydrogen for residential applications. The findings indicate substantial performance gains, with exergy efficiency improving from 48% to 60%, power output increasing from 6.1 MW to 11.6 MW, and the levelised cost of electricity decreasing from 8 to 3 cents/kWh. In addition to energy generation, the system delivers co-benefits through the integrated production of freshwater and hydrogen, demonstrating the potential of intelligent, multi-output infrastructure systems.

5.5.3 Cluster 3 (Blue): IoT (Internet of Things) and Decision-Making Systems

This cluster reflects digital integration within water infrastructure systems, centred on the Internet of Things, data-driven decision-making, and predictive analytics. It signifies a broader transition from conventional, reactive management approaches toward intelligent and interconnected systems capable of real-time monitoring, analysis, and response. Letsoalo et al. (2026) highlighted that, the Internet of Things (IoT) enabled sensors provide continuous data streams for real-time monitoring of water quality, distribution efficiency, and leak detection. This allows for proactive interventions, reducing risks and improving resource utilisation. Trejo et al. (2025) supported smart water systems integrate IoT with other technologies like wireless sensor networks (WSNs) and remote sensing to enhance urban water governance and decentralized water management. Haripriya et al. (2025) highlight that in India, IoT sensors are deployed to monitor water levels, gas concentrations, and flow obstructions, while machine learning algorithms analyse the collected data to forecast blockages and recommend timely interventions. Field simulations demonstrate a 40% reduction in emergency maintenance events, alongside improved response times by municipal authorities. More so, Saranya et al. (2025) in Malaysia, focusing on efficient smart water distribution optimisation through the integration of the Internet of Things and Graph Neural Networks, demonstrates the growing effectiveness of advanced digital technologies in water infrastructure systems. Using publicly available water distribution network (WDN) datasets, the study reports a leak detection accuracy of 95.8% and a flow prediction RMSE of 0.34 L/s, outperforming traditional machine learning models such as Random Forests and Support Vector Machines. The findings highlight the capacity of IoT-enabled data streams, when coupled with network-aware learning models like GNNs, to significantly enhance predictive maintenance capabilities, improve fault detection, and optimise hydraulic performance.

5.5.4 Cluster 4 (Yellow): Management systems

This cluster focuses on asset management, information management, and learning algorithms, emphasising the strategic, organisational, and institutional dimensions of the field. It emphasises how institutional adoption facilitates the effective integration of technological innovations into core infrastructure governance processes, particularly in asset lifecycle planning, maintenance prioritisation, and budgeting and investment decision making. Effective asset management frameworks focus on optimizing investments across the asset lifecycle, reducing costs, and improving resource allocation Vekaria et al. (2022). Tools like PIPEiD provide a unified, web-based platform for managing water pipeline data, addressing strategic, tactical, and operational levels of asset management (Sinha, 2019).Moreover, Samra et al. (2018) highlighted that, the integrated asset management systems, such as those applied in Kelowna, Canada, enable organisations to balance lifecycle costs, user costs, and physical state, leading to significant cost savings and improved service levels. Maumela et al. (2025), in the South African context, propose multi-institutional framework approaches that emphasise strategic alignment and value-driven planning. These approaches are designed to address fragmented governance structures and enhance coordination across institutions, thereby supporting the delivery of a more sustainable and reliable water supply system.

5.6 Density visualisation Analysis

The density visualisation (Figure 4) highlights the intensity and concentration of research themes within machine learning–driven predictive maintenance for water infrastructure systems. The colour gradient from yellow to green and blue represents the frequency and co-occurrence strength of keywords. The **high-density core (yellow region)** is dominated by Machine Learning and Predictive Maintenance, indicating that these are the most prominent and interconnected themes. This suggests a strong research emphasis on data-driven and predictive approaches to maintenance. The **medium-density zone (green region)** comprises application oriented themes such as water management, learning systems, predictive analytics, water distribution systems, forecasting, and deterioration. These themes demonstrate how machine learning techniques are being operationalised within water infrastructure, with increasing focus on performance prediction, condition assessment, and system optimisation. In contrast, the **low-density areas (blue/green regions)** include emerging and less frequently studied topics such as Internet of Things, decision making, asset management, information management, and advanced techniques such as graph neural networks. Despite their lower occurrence, these themes are strategically significant, reflecting a shift toward smart infrastructure, real-time monitoring, and digital integration. The density map reveals a centralised research structure with a strong methodological core. There is a clear diffusion from core machine learning techniques to applied water infrastructure contexts, and further toward management and digital integration layers. While the application of machine learning in water systems is relatively well established, its integration with IoT enabled systems and asset management frameworks remains an emerging area with significant potential for future research.

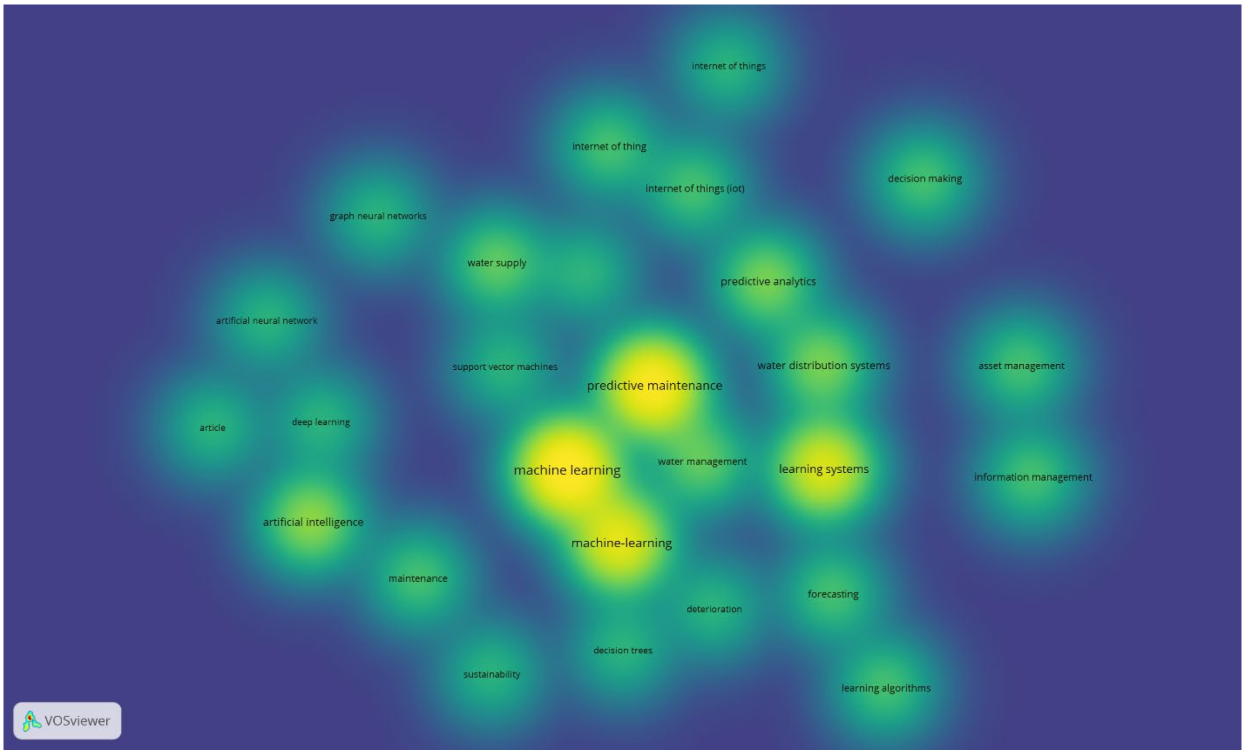

Figure 4. Density visualisation

6. Conclusions and recommendations

This bibliometric review analysed 29 peer-reviewed, Scopus-indexed publications on machine learning–driven predictive maintenance to optimise operational reliability in water infrastructure systems. The findings demonstrate a clear and accelerating growth trajectory in scholarly output, with a sharp surge in publications reaching 19 in the year 2025. This rapid increase reflects the rising academic and practical recognition of machine learning–based predictive maintenance as a critical innovation for enhancing operational reliability, efficiency, and resilience in water infrastructure systems. Geographically, the analysis reveals a strong dominance of developing countries, particularly India and China, which collectively lead in research productivity and citation impact. India emerges as the most influential contributor with 8 publications and 31 citations, followed by China with 3 publications and 25 citations. In contrast, representation from African developing countries remains significantly limited, with only South Africa and Sudan appearing in the dataset. This imbalance highlights a critical regional research gap, despite the increasing vulnerability and infrastructure challenges faced by water systems across Africa. The limited scholarly visibility of African countries in this domain not only reflects disparities in research output but also signals a potential disconnect between global technological advancements and local infrastructure realities. The most prolific journals in this research area are Water (Switzerland) and Sensors, indicating that studies on water systems, monitoring technologies, and sensor-driven analytics are primarily concentrated within these publication outlets. Thematic analysis identifies four dominant research clusters: machine learning and advanced artificial intelligence techniques, water infrastructure and predictive applications, IoT-enabled decision-support systems, and infrastructure management systems. Density visualisation further indicates a highly centralised knowledge structure, with strong methodological clustering across high-, medium-, and low-density cores. While machine learning applications in water systems are relatively well established, their integration with IoT-enabled frameworks and asset management systems remains an emerging. The study recommends that future research and practice should focus on increasing African participation in machine learning–driven predictive maintenance studies, as current output is dominated by countries such as India and China. There is a need to develop machine learning–driven predictive models that address the unique challenges of African water infrastructure systems, including ageing assets, limited data availability, and resource constraints. Stronger international and intra-African collaborations should be encouraged to promote knowledge sharing and capacity building in this field. In addition, greater integration of machine learning with IoT technologies and water infrastructure asset management systems is essential to improve real-time monitoring and predictive capabilities. Investment in digital data infrastructure and accessible datasets should also be prioritised to support advanced analytics. Strengthen interdisciplinary collaboration between water experts, and infrastructure managers is necessary to ensure that predictive maintenance solutions are practical, scalable, and effectively implemented in real-world water systems.

References

Adeniran, A., What is the state of water infrastructure governance research in Nigeria? A review, *Water International*, vol. 47, no. 4, 2022.

Alhamad, Y. Alkhezi, and M. F. Alhajri, "Nonlinear integer programming for solving preventive maintenance scheduling problem for cogeneration plants with production," *Sustainability*, vol. 15, no. 1, 2022.

Amorelli, F. Termine, G. Pau, F. Arena, V. M. Salerno, and M. Collotta, "Predictive maintenance for water supply networks: Advanced expert system models for enhanced water resource management and monitoring," in *IEEE International Workshop on Metrology for Living Environment (MetroLivEnv),* 2025.

Ángel and B. Tekinerdogan, Urban Water Distribution Networks: Challenges and Solution Directions. Elsevier, 2024.

Balakrishnan and M. Subramani, "AI-driven predictive maintenance and downtime optimization in water treatment plants," in *Handbook of AI for Clean Water*, 2025.

Balaraman, K. Patel, P. K. R. Mahendran, and N. R. Kasarla, "AI-driven predictive maintenance in water and wastewater systems: Enhancing efficiency, reliability, and sustainability," in *IEEE Control and System Graduate Research Colloquium (ICSGRC),* 2025.

Baas, M. Schotten, A. Plume, G. Côté, and R. Karimi, "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies," *Quantitative Science Studies*, vol. 1, no. 1, pp. 377–386, 2020.

Bui, H. Y. Godoy, and S. M. Elachachi, "Assessment of artificial intelligence applications in urban water networks," *Journal of Pipeline Systems Engineering and Practice*, vol. 16, no. 1, 2024.

Bulti and G. A. Y. Amelo, "Water infrastructure resilience and sanitation challenges in developing countries," AQUA – *Water Infrastructure Ecosystems and Society*, vol. 72, no. 6, 2023.

Cao, P. Paudel, J. Baldwin, and F. Rabbi, "Data centric smart digital platform for wastewater treatment optimization," *Water Environment Federation Proceedings*, 2024.

Carrión, "Kubernetes as a standard container orchestrator: A bibliometric analysis," *Journal of Grid Computing*, vol. 20, no. 4, 2022.

Chen, S. Chen, Z. Chen, L. Xiao, and J. Hu, "How much data is sufficient for reliable bibliometric domain analysis? A multi-scenario experimental approach," Scientometrics, vol. 130, no. 5, pp. 2923–2946, 2025.

Chiadighikaobi, Q. A. A. Qais, O. C. Onuoha, R. I. O. Isiaka, A. O. Joel, and U. J. Erumu, "Optimizing preventive maintenance strategies for civil engineering systems," *NIPES Journal of Science and Technology Research*, vol. 7, no. 2, 2025.

Dhanya, R. Bhavish, B. R. Gowda, H. Reddy, and H. Reddy, "Enhancing industrial equipment reliability through predictive maintenance using machine learning," in *IEEE International Conference on Communication, Automation, Management and Security (ICCMS)*, 2025.

Embarki, A. El Kihel, and B. El Kihel, "Artificial intelligence methods for predictive maintenance decision-making," in *Sustainable Civil Infrastructures*, 2024.

Frost, N. Moser, J. Nöcker, and L. Zhukov, "Getting value from predictive maintenance," in *Mechanisms and Machine Science*, 2021.

Geisbush and S. T. Ariaratnam, "Failure prevention in large-diameter water pipelines using reliability-centered maintenance," *Water*, vol. 15, no. 24, 2023.

Haripriya, S. S. Hammed, C. Preethi, S. Pavalarajan, R. Padmakumar, and M. M. Surya, "Automated drainage management system using AI and IoT," in *IEEE ICIMA*, 2025.

Hu, Y. Zhang, W. Liu, Z. Song, H. Ji, and F. Wang, "Predicting water pipe failures with graph neural networks: Integrating coupled road and pipeline features," *Water*, vol. 17, no. 9, pp. 1307–1307, 2025.

Ismail, A. H. Muse, M. A. Hassan, Y. H. Muse, and S. Nadarajah, "Machine learning for water source quality analysis," *Water*, vol. 16, no. 20, 2024.

Jie, "VOSviewer application status and its knowledge base," *DOAJ (Directory of Open Access Journals),* 2022.

Joseph et al., "Early leak and burst detection in water pipelines using machine learning approaches," *Water*, vol. 17, no. 14, 2025.

Kumar, "Bibliometric analysis: Comprehensive insights into tools, techniques, applications, and solutions for research excellence," *Spectrum of Engineering and Management Sciences*, vol. 3, no. 1, pp. 45–62, 2025.

Letsoalo et al., "Digital approaches in water quality management," in Handbook of Digitalization and Big Data in the *Water Sector*, vol. 2, 2026.

Madala, "Predictive maintenance in water treatment plants using machine learning," in *IEEE International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2025.

Maumela, D. Mathaba, and M. Kao, "Integrated framework for urban water infrastructure planning and management in Gauteng Province," *Water*, vol. 17, no. 15, 2025.

Muhamad, "Bibliometric computational mapping analysis using VOSviewer," *Jurnal Manajemen Dan Teknologi Rekayasa*, vol. 2, no. 1, 2023.

Ngwenya, T. Paepae, and P. N. Bokoro, "Monitoring water quality using machine learning and IoT," *Journal of Water Process Engineering*, vol. 73, 2025.

Ortiz et al., "Failure modelling for drinking water pipes," *Applied Water Science*, vol. 13, no. 11, 2023.

Rathore et al., "Machine learning in predictive maintenance for industrial equipment," in *IEEE ICAC2N*, 2024.

Rather, S. Kumar, and A. H. Gandomi, "Breaking the data barrier: A review of deep learning techniques for democratizing AI with small datasets," *Artificial Intelligence Review*, vol. 57, no. 9, 2024.

Rihan et al., "Industrial predictive maintenance for sustainable manufacturing," in *Artificial Intelligence and Machine Learning for Industry 4.0*, 2025.

Rubaba and T. Walingo, "AI-based water quality monitoring in eThekwini," *Water*, vol. 17, no. 22, 2025.

Sahoo, A. Singh, and M. Kumari, "Machine learning anomaly detection in water pumps," in *IEEE ICDSIS*, 2024.

Samra, M. Ahmed, A. Hammad, and T. Zayed, "Multiobjective framework for municipal infrastructure management," *Journal of Construction Engineering and Management*, vol. 144, no. 1, 2018.

Srivastava and G. Dixit, "AI and big data in developing country context," in *IFIP Advances in Information and Communication Technology*, 2023.

Stefanis et al., "Innovative methodological pillars for bibliometric studies: AI screening, data normalization and dual-tool analysis," *Discover Applied Sciences*, vol. 7, no. 9, 2025.

Tlabu, A. Telukdarie, and B. G. Mwanza, "Maintenance 4.0 for water pumping infrastructures," *in IEEE IEEM*, 2022.

Tiquio and A. Villanueva, "Pipeline leak classification using machine learning ensemble methods and hyperparameter tuning," in *2024 4th International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, 678–683, 2024.

Valero, E. Pummer, V. Heller, M. Kramer, D. B. Bung, S. Mulligan, and S. Erpicum, "The unspoken value of water infrastructure," *Renewable and Sustainable Energy Reviews*, vol. 212, 2025.

Zhang and Z. Ye, "Water pipe failure prediction using AutoML," *Facilities*, 2020.

Xiong, "Comparative analysis of machine learning models for water quality prediction using regional monitoring data," *Informatics*, vol. 49, no. 16, 2025.

Biographies

Gift Laolang (PhD candidate), is a Lecturer in the Department of Built Environment (Construction Management and Quantity Surveying) at Walter Sisulu University. He is currently pursuing a PhD in Sustainable Housing at the University of KwaZulu-Natal. He holds a Master of Science (MSc) in Construction Management (with distinction), a BTech in Construction Management, and a National Diploma in Building Science from the University of Johannesburg. His academic work includes contributions to peer-reviewed journals and national and international conference proceedings. His research expertise focuses on sustainable water infrastructure, with a strong interest in the integration of digital technologies within the built environment. His work is driven by a commitment to advancing sustainable development in alignment with the United Nations Sustainable Development Goals (SDGs), particularly in infrastructure efficiency, resilience, and innovation.

Dr Motheo Meta Tjebane (PhD), is a Lecturer at the Durban University of Technology (DUT), specialising in Construction Management, Quantity Surveying, and Digital Education. She holds a PhD in Construction Management, a Master's degree in Construction Management (with distinction), a BTech in Quantity Surveying, a Postgraduate Diploma in Digital and Online Education, and a Diploma in Building Science. Her academic focus lies in integrating technology and education within the construction industry to enhance learning experiences and improve real world application. She's currently supervises PhD, Master's, and Honours level students in Construction Management and

Quantity Surveying. Scholarly contributions include peer-reviewed journal articles, conference proceedings, and book chapters at both national and international levels. Her research expertise centres on the adoption of artificial intelligence in the construction industry, as well as AI applications for asset management in healthcare facilities. Her interdisciplinary work spans, Higher Education, Civil Engineering, Engineering Education, Safety Engineering, and Educational Technology.