DMPedia Lecture Notes in Multidisciplinary Research,
Volume 1, pp.784-802, 2026

International Conference on Multidisciplinary Perspectives in Advanced Computing and Technology (IMPACT 2026)
G. B. Pant University of Agriculture and Technology, Uttarakhand, India. Jan. 10-11, 2026

Survey on Water Disease Detection Using Deep Learning

Yash Pratap Singh, Prinsvil Yadav, Kusum Lata
Department of Computer Science and Engineering, Sharda School of Engineering and Technology, Greater Noida, India

singhyash0207@gmail.com, prinsvil@ug.sharda.a.c.in, kusumlata.1@sharda.ac.in

Abstract:

Waterborne infections are a major public health threat, especially in developing countries where access to clean water is poor. Conventional detection techniques, such as culture-based and chemical methods, are accurate but often time-consuming, expensive, and not ideal for online monitoring. Recent breakthroughs in artificial intelligence, particularly deep learning, open new opportunities for rapid and precise determination of water quality. This paper summarises different detection methods, rule-based, statistical, and deep learning, and highlights how models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Physics-Informed Neural Networks (PINNs) improve pathogen detection using image, sensor, and genomic information. The integration of AI with IoT-enabled devices and intelligent dashboards also facilitates real-time monitoring and predictive maintenance of domestic water supplies. Though challenges such as data scarcity and model interpretability persist, deep learning-based solutions are highly promising for revolutionising water safety management and mitigating the outbreak of waterborne diseases.

Keywords: CNN, RNN, PINN, Water Disease Detection, Deep Learning

1. Introduction

Water is one of the most essential elements for sustaining life on Earth. Approximately 71% of the Earth's surface is covered with water, yet only a small fraction is suitable for human consumption or use. With a growing population and industrialisation, the quality of freshwater sources is under constant threat from various contaminants, including industrial effluents, agricultural runoff, and untreated sewage [1]. The World Health Organisation (WHO) estimates that unsafe water causes more than 485,000 deaths each year due to diarrhoea alone. These alarming statistics underscore the critical need for advanced and efficient water quality monitoring systems that can detect disease-causing agents early and accurately. Traditional methods of water quality monitoring typically rely on laboratory testing and chemical analysis, which are not only time-consuming and costly but also require specialised personnel and equipment. Moreover, by the time the analysis results are available, the contaminated water may already have been consumed or used, leading to outbreaks of waterborne disease. In this context, there is a growing need for real-time, scalable, and cost-effective water-quality assessment and disease-detection systems. The integration of advanced computational technologies, such as artificial intelligence (AI), particularly deep learning, offers promising solutions to these challenges [1]. Deep learning, a subfield of machine learning, mimics the human brain's ability to learn patterns and make decisions. It has revolutionised various domains, including image recognition, speech processing, medical diagnosis, and, more recently, environmental monitoring.

2. Application of Deep Learning in Water Disease Detection

Why do we require deep learning in water disease detection when we already have standard laboratory procedures? Deep learning is needed because there are limitations to conventional methods. Pathogen detection by tests in a laboratory is typically time-consuming, manpower-heavy, and skilled-personnel-dependent [1]. Deep learning, however, is capable of analysing microscopic images or genomic sequences in real-time and detecting harmful microorganisms automatically, without human error [2]. One question remains: how can deep learning assist in averting outbreaks of waterborne diseases at large scales? This is where predictive modelling comes into play. Predictive models like recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have the capability to analyse past data on water quality, rainfall, pollution, and previous cases of disease to predict cholera, dysentery, or typhoid outbreaks in the future. Is continuous monitoring of water quality possible with deep learning? Yes, it is. Classic methods of monitoring depend on periodic sampling, which can fail to detect sudden contamination events. With the combination of deep learning and IoT-enabled sensors, parameters such as pH, turbidity, temperature, and dissolved oxygen can be measured in real-time. The models flag water as safe or unsafe and alert when irregularities indicate a disease risk. This real-time monitoring is particularly important for rural and underdeveloped areas where waterborne disease is prevalent [2].

3. State-of-the-Art in Deep Learning for Water Disease Detection

The cutting-edge in deep learning largely exploits models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These models are applied to the analysis of different types of data, ranging from microscopic images of a pathogen to sensor data in real-time. The current advances aim at designing real-time, automated detection systems that are fast and highly precise. The combination of deep learning with technologies such as the Internet of Things (IoT) and spectroscopy is facilitating holistic water quality monitoring and early warning systems for public health.

3.1 Image-Based Detection

Current research indicates that convolutional networks are particularly effective at detecting waterborne dangers directly from images at various sizes. At the microscope level, CNN/instance-segmentation workflows (e.g., Faster/Mask R-CNN variants) enhance accuracy and resilience for bacterial detection compared to traditional image analysis, allowing for automated pathogen screening from a microscopic field of view. Deep models work on both SAR and optical satellite imagery for harmful algal blooms (HABs). Improved CNN pipelines that have been trained on optical imagery (such as case studies of Noctiluca scintillans) enhance HAB detection under changing water colour, whereas U-shaped segmentation networks (U-Net families) map bloom extent for taking action [3].

3.2 Sensor-Data (Time-Series) Analysis

For real-time monitoring in surface waters and water distribution networks, LSTM-family models prevail because they can learn the nonlinear long-range dependencies within multi-parameter streams (DO, pH, turbidity, conductivity, nutrients)[4]. One of the most referenced stacked-LSTM works for Yangzhou (China) IoT stations demonstrated robust six-month predictions and beat ARIMA/SVR baselines—proof that sequence deep models can handle longer horizons required for planning. Reviews across inland waters support LSTMs' better performance than conventional models, and hybrid CNN-LSTM models enhance short-term anomaly detection as well as denoising. Large syntheses point out that deep learning is capable of filling spatial–temporal gaps, blending in-situ and remote signals, and enabling hypothesis testing for water-quality–disease associations when the ground truth is lacking—an important facilitator for public-health surveillance[4].

3.3 Genomic/Metagenomic Detection

For rapid, culture-free identification of pathogens and AMR markers from water samples, deep models (1D-CNNs, RNNs, Transformers) speed up read-level classification, binning, and novel-taxa discovery. Current reviews consolidate how DL now facilitates most metagenomic steps—from quality control and misassembly detection to taxonomic profiling—rendering sequence-based surveillance more actionable for water safety[5].

4. Navigating the Complex Landscape of Waterborne Pathogen Detection: A Technical Review

Detection of waterborne pathogens is an imperative yet challenging endeavour because of the variety of microorganisms, their low density in water, and the intricate composition of environmental matrices. A technical review of the area points towards a transition from conventional culture-based techniques to sophisticated molecular and sensor technologies. This change is prompted by the pressing requirement for quicker, more precise, and on-site detection to avert and contain waterborne disease outbreaks.

4.1 Executive Summary: A Retrospective and Prospective View of Waterborne Pathogen Detection

Waterborne pathogens continue to present a substantial global public health challenge, with a significant impact on morbidity, mortality, and economic stability. This report provides a comprehensive review of the evolution of water disease detection technologies, analysing the trajectory from traditional, culture-based methodologies to the latest advancements in biosensing and data-driven systems[6]. The analysis is framed within the context of global health equity and the United Nations Sustainable Development Goals (SDGs 3, 6, 11, 13, and 17), which are closely interrelated with water safety and disease prevention.1

4.2 Presenting the Global Imperative and the Evolving Landscape

The economic and health implications of waterborne diseases worldwide are staggering and profoundly far-reaching. Pathogens in water supplies, usually a result of human actions and poor infrastructure, cause a range of diseases such as cholera, typhoid, and gastroenteritis[6]. The overall global health burden of these diseases has been calculated as around three million disability-adjusted life years (DALYs) lost per year, accompanied by a directly related loss of almost $12 billion per year. This is not evenly distributed. Individuals living in developing nations are disproportionately impacted, suffering from a chronic lack of proper supplies of clean water and limited financial and technological resources for establishing effective monitoring programs.

To meet this pressing worldwide challenge, a paradigm shift in water safety management and pathogen detection is needed. This report maps the "complex landscape of waterborne disease research" by systematically evaluating the development of detection technologies[17]. The report moves from an analysis of conventional detection techniques to a close examination of advanced molecular, genomic, and sensor-based technologies. The report concludes with a critical assessment of the ongoing challenges and a future vision for directions in research, highlighting the imperative need to enhance the connection between laboratory research and in-practice management practices.

4.3 Foundational Methods: The Golden Standard and Its Inherent Limitations

The gold standard for the detection of waterborne pathogens has traditionally been culture-based techniques, in which microorganisms from a water sample are grown up in a controlled laboratory environment. This gives immediate evidence of live organisms and is usually necessary for regulatory purposes. It is, however, subject to several important deficiencies.

4.3.1 Culture-Based and Indicator-Based Detection

For many years, culture-based techniques have been the "gold standard" for the detection of waterborne pathogens because of their comparative simplicity and affordability. The method depends upon the isolation and growth of microorganisms on individual growth media. Indicator organisms are one such common practice, which are substitute bacteria that indicate the probable presence of a larger class of pathogens[8].

4.3.2 A Critical Assessment of Drawbacks

In spite of their root position, conventional techniques are plagued by serious limitations. The most essential flaw is that the process takes so long; the detection of coliforms takes from 18 to 72 hours, a considerable delay that is especially unfavourable in the event of an outbreak of a disease[9]. Conventional techniques are labour-intensive and do not have the sensitivity needed to identify pathogens at low levels of concentration. In addition, they cannot detect non-culturable pathogens, a significant deficiency in holistic water safety testing.

A closer look, however, divulges a more essential scientific issue with the use of indicator organisms. Though standardised, their effectiveness as a surrogate for water pollution is something debated consistently. The EPA, for instance, observes that although indicator bacteria like Appendix**: Comprehensive Review of Select Papers on Water Disease Detection.**

Table 1 provides a summary of key research papers, serving as a foundational resource for further study on waterborne pathogen detection.

Table 1: Waterborne Pathogen Detection

Author	Methods	Key Findings	Technical Terms Used	Limitations
Usisipho Feleni,	This review evaluates traditional culture-based	The review highlights that while traditional culture-	Chemiluminescence, Biosensors,	Traditional methods are time-consuming

et al.2025 [36]	and molecular techniques against modern biosensor technologies, specifically detailing electrochemical biosensors and chemiluminescence systems.[30]	based methods are simple and cost-effective, they are slow and have low sensitivity, which is a major drawback during disease outbreaks.[33]	Impedimetric Immunosensor, CFU (Colony-Forming Units), SPR (Surface Plasmon Resonance), Microfluidic Chip, PCR (Polymerase Chain Reaction).[30]	(18–72 hours), laborious, and have low sensitivity for pathogens at low concentrations. They also cannot detect non-culturable pathogens.[30]
S. Saleem, et al.2023 [33]	This review examines the evolution of waterborne pathogenic bacteria detection from traditional culture-based methods to modern molecular and biosensor techniques. It specifically discusses various forms of PCR, NGS, and the integration of AI.[33]	The review found that advancements in digital droplet PCR (ddPCR), NGS, and biosensors have significantly improved detection sensitivity and specificity. It also notes that integrating AI with these technologies can enhance detection accuracy and enable real-time data analysis.	Polymerase Chain Reaction (PCR), quantitative real-time PCR (qPCR), digital droplet PCR (ddPCR), NGS (Next-Generation Sequencing), Artificial Intelligence (AI).[33]	On-site practical implementation of molecular-based methods and biosensors remains a challenge. There is a need for robust, cost-effective, and user-friendly techniques for routine monitoring.
Z. Li, et al.2025 [34]	This systematic review, conducted in accordance with PRISMA guidelines, comprehensively surveys advancements in Metal Nanocluster (MNC)-based biosensors for detecting viral and bacterial pathogens. It focuses on optical (colourimetric and fluorescence) and electrochemical platforms.[34]	The paper demonstrates that MNC-based biosensors offer high sensitivity, specificity, portability, and cost-efficiency. The integration of nanotechnology with biosensing platforms enables real-time and point-of-care diagnostics for applications in clinical diagnostics, environmental monitoring, and food safety.[34]	Metal nanocluster-based biosensors, nanobiosensors, optical biosensors, electrochemical biosensors, AgNCs (Silver Nanoclusters), AuNCs (Gold Nanoclusters), CuNCs (Copper Nanoclusters), PRISMA guidelines.	The need for enhanced stability, multiplex detection capability, and clinical validation of the technology is highlighted as a key limitation and future direction.
C. Li, et al.2025 [32]	This study developed a cost-effective, citizen science-based approach to monitor the tap water microbiome using low-biomass sampling and 16S rRNA gene metabarcoding.[32]	The study found a diverse microbiome in household tap water and detected opportunistic pathogens such as *Mycobacterium*, *Acinetobacter*, and *Legionella[32]*.	Microbiome assessment, 16S rRNA gene metabarcoding, Amplicon Sequence Variants (ASVs), citizen science[32].	The study was based on a limited number of sequenced samples[32]. The efficacy of indicator organisms in representing the presence of

				pathogens is still debated[32].
J.-W. Lee, et al.2022 [40]	The paper reviews molecular diagnostic methods for assessing microbial water quality. These include conventional PCR, qPCR, multiplex qPCR (mqPCR), ddPCR, and next-generation sequencing. It also discusses the use of electrochemical biosensors and wastewater-based epidemiology (WBE)[40].	The paper highlights that molecular methods are more rapid, sensitive, specific, and reproducible than traditional culture-based methods, particularly for viable but not culturable (VBNC) microbes[40].	End-point PCR, DNA microarray, Real-time quantitative PCR (qPCR), Multiplex qPCR (mqPCR), Digital droplet PCR (ddPCR), High-throughput next-generation DNA sequencing (HT-NGS), Antibiotic-resistance genes (ARGs), Viable but not culturable (VBNC) microbes, Wastewater-based epidemiology (WBE)[40].	Limitations include the high cost of specialized instruments and reagents, the need for trained personnel, and a lack of standardized protocols. Conventional PCR is also noted to only provide qualitative (yes/no) results, and issues with specificity and cross-contamination can occur[40].
(Systematic Review) 2025[39]	This review evaluates the role of AI in water monitoring, focusing on the integration of IoT-sensor networks and satellite remote sensing. It discusses various models like ANFIS and deep neural networks (DNNs) and their use in predictive analysis and real-time anomaly detection.[43]	The review demonstrates that AI-driven systems can achieve high prediction accuracy (94%) and reduce field sampling costs. These systems enable continuous, in-situ monitoring and can fuse multi-source data to improve reliability.[43]	Artificial Intelligence (AI), Internet of Things (IoT), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), YOLOv5 algorithm, Principal Component Analysis (PCA)[39].	Key challenges include issues with data standardization, model interpretability, and ethical governance[43] The performance of these systems can also be significantly influenced by environmental factors such as lighting conditions and water turbidity.[44]
(Review Article) [39]	This review focuses on molecular analytical techniques for pathogenic bacteria in wastewater, including PCR-based methods and DNA sequencing, and their application in	The review highlights that molecular methods are gradually becoming the mainstream detection technique due to their high accuracy and specificity[42]. WBE is identified as a powerful	Wastewater-based epidemiology (WBE), Polymerase chain reaction (PCR), Quantitative or real-time PCR (qPCR), Digital PCR (dPCR), Viable but	A key challenge is the low concentration of pathogenic bacteria in raw wastewater, which requires highly sensitive detection methods.

	wastewater-based epidemiology (WBE).[42]	tool for the early warning of infectious disease outbreaks, as it can detect pathogens from asymptomatic individuals in a community[42].	nonculturable (VBNC) cells[42].	[42]Traditional methods are also noted for their limitations, including time-consuming procedures and the inability to detect nonculturable cells[42].

5. Overview of Waterborne Diseases and Their Causative Agents

Waterborne diseases occur when waterborne pathogens infect drinking or recreational water, which spreads by ingestion, contact with the skin, or inhalation of aerosols[11]. Pathogens are from various biological groups, and each group of diseases has characteristic manifestations.

5.1 Categories of Pathogens

Waterborne pathogens can be broadly classified into the following groups:

- **Bacteria**
  - Examples: Vibrio cholerae, Escherichia coli (E. coli O157:H7), Salmonella typhi, Shigella spp., Leptospira.
  - Typically cause acute gastrointestinal illness, systemic infections, or enteric fever.
- **Viruses**
  - Examples: Hepatitis A virus (HAV), Hepatitis E virus (HEV), Rotavirus, Norovirus, Enteroviruses, Adenoviruses.
  - Causes a wide range of illnesses, particularly viral gastroenteritis and hepatitis.
- **Protozoa (Parasites)**
  - Examples: Giardia lamblia, Cryptosporidium parvum, Entamoeba histolytica.
  - Notable for cyst or oocyst stages resistant to chlorination, making them difficult to remove from treated water.
- **Helminths (Parasitic Worms)**
  - Examples: Dracunculus medinensis (Guinea worm), Schistosoma spp.
  - Infections are often associated with skin penetration or ingestion of contaminated water containing larval forms[12].
- **Fungi (Occasionally Waterborne)**
  - Examples: Candida spp., Aspergillus spp. (in immunocompromised individuals).
  - Less common, but may be opportunistic in hospital water systems.

5.2 Major Diseases

The following are key waterborne diseases along with their causative agents:

- **Cholera**

  - Causative Agent: *Vibrio cholerae*
  - Characterised by severe watery diarrhoea, dehydration, and outbreaks linked to unsafe water supplies[13].
- **Typhoid and Paratyphoid Fever**
  - Causative Agent: *Salmonella enterica* serovars Typhi and Paratyphi
  - Transmitted via faecal contamination, leading to prolonged fever, abdominal pain, and systemic illness.
- **Bacillary Dysentery (Shigellosis)**
  - Causative Agent: *Shigella* spp.
  - Symptoms include bloody diarrhoea, cramps, and fever.
- **Hepatitis A and E**
  - Causative Agents: Hepatitis A virus (HAV) and Hepatitis E virus (HEV)
  - Spread by ingestion of contaminated water; causes acute liver inflammation and jaundice
- **Gastroenteritis (Viral/Bacterial/Protozoal)**
  - Causative Agents: Norovirus, Rotavirus, *E. coli* O157:H7, *Giardia lamblia*, *Cryptosporidium parvum*
  - Presents with diarrhoea, vomiting, nausea, and abdominal cramps.
- **Amoebic Dysentery (Amoebiasis)**
  - Causative Agent: *Entamoeba histolytica*
  - Causes chronic diarrhoea, abdominal pain, and occasionally liver abscesses.
- **Leptospirosis**
  - Causative Agent: *Leptospira interrogans*
  - Acquired from contaminated water (often floodwaters); symptoms include fever, headache, muscle pain, and jaundice[14]**.**
- **Guinea Worm Disease (Dracunculiasis)**
  - Causative Agent: *Dracunculus medinensis*
  - Infection occurs when people drink stagnant water containing copepods with larvae, which leads to painful skin blisters.
- **Schistosomiasis (Bilharzia)**
  - Causative Agent: *Schistosoma* spp.

- Infection from skin contact with freshwater infested with cercariae; causes urinary or intestinal disease.

6. Traditional Methods for Water Quality Analysis

Maintaining safe water quality has hitherto depended on lab-based methods that measure both the chemical content of water and microbial contaminants. Such methods set a standard against which recent, technology-centric approaches like deep learning can be measured.

6.1 Physicochemical Parameters

Physicochemical analysis is the backbone of water quality evaluation, measuring physical and chemical characteristics that directly impact potability and the health of ecosystems [14]. The most frequently analysed parameters are pH, turbidity, colour, temperature, dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), total dissolved solids (TDS), and nutrient concentrations (nitrates, phosphates, ammonia). All these parameters determine if water is acceptable for drinking, irrigation, or aquatic life.

6.2 Microbiological Analysis

Whereas physicochemical analyses reflect the overall quality, microbiological examination determines the presence of disease-causing waterborne pathogens directly. Conventional microbiology is based on culture methods, in which water samples are incubated on selective growth media to isolate and identify bacteria such as Escherichia coli, Salmonella, Shigella, or Vibrio cholerae [16]. The presence of indicator organisms, especially total coliforms and E. coli, is central to microbiological analysis. These organisms are indicative of faecal contamination and act as surrogates for the possible presence of pathogens. Other more sophisticated but still conventional equipment includes membrane filtration, where bacteria are retained on a filter and cultured on nutrient agar, and microscopy for the detection of protozoan cysts such as Giardia or Cryptosporidium[16]. Although they are reliable, they have limitations: they take time (24–72 hours with cultures), human effort, and could fail to identify viable but not culturable organisms. They also need sterile methods and skilled microbiologists, which makes them less suitable in low-resource laboratories or in real-time monitoring.

7. Advanced Technologies for Pathogen Detection

Conventional microbiological techniques, while accurate, tend to be slow and cumbersome. To resolve such issues, some new technologies have been developed that detect pathogens in water more rapidly, sensitively, and specifically[17].

7.1 Molecular Methods

Molecular methods target the identification of microorganisms' genetic material, which enables quick and highly specific detection. Polymerase Chain Reaction (PCR) and its modifications (qPCR, RT-PCR, and multiplex PCR) are the most prevalent, facilitating amplification of pathogen-specific DNA or RNA sequences in a matter of hours. This method is particularly good for detecting viable but non-culturable (VBNC) microbes like Cryptosporidium or Giardia that tend to evade conventional culture-based detection. Next-generation sequencing (NGS) and metagenomics take things one step further by examining the entire microbial community in water samples[18]. While highly sensitive and capable of handling large datasets, molecular techniques are frequently dependent on specialised equipment, high expense, and trained staff, making them difficult to deploy in field settings.

7.2 Immunological Assays

Immunological techniques are based on antigen-antibody specific binding for the detection of pathogens or their toxins. Enzyme-Linked Immunosorbent Assay (ELISA) is one of the most popular tools for detecting bacterial toxins, viral antigens, or protozoal cyst proteins with high specificity. Lateral Flow Immunoassays (LFIA), as with home pregnancy tests, are quick, point-of-care devices that do not require advanced instrumentation.

7.3 Biosensors

Biosensors are a novel technology that combines a biological recognition component (e.g., enzymes, antibodies, or nucleic acids) with a physical transducer to yield measurable signals For instance, electrochemical biosensors were fabricated to detect E. coli in drinking water with high sensitivity at minute scales[19]. However, biosensors are plagued by instability, poor reproducibility, and issues with large-scale commercialisation. Robustness and cost reduction are areas of active research to facilitate practical implementation.

8. Water Disease Detection Approaches

Waterborne diseases remain a serious threat to global public health, especially in developing countries where access to safe drinking water is restricted[20]. Generally, these methods can be classified into three categories: rule-based, statistical, and deep learning. Each of these groups possesses specific strengths and weaknesses, and the suitability often lies in the deployment scenario, data availability, and nature of waterborne disease in question.

8.1 Rule-Based Approaches

Rule-based systems are among the oldest methods used to detect waterborne disease. Such systems rely on pre-specified rules, thresholds, and expert judgment to determine water quality parameters and detect possible risk factors. For example, the World Health Organisation (WHO) and local regulatory agencies commonly establish threshold values for indicators such as pH, turbidity, total dissolved solids (TDS), biological oxygen demand (BOD), nitrate content, and the presence of pathogens such as E. coli. In a rule-based system, if the measured water parameters exceed these levels, the system flags the sample as contaminated and possibly harmful. Rule-based methods have several drawbacks, though. Firstly, they are highly dependent on domain knowledge in establishing thresholds, which can vary by geography due to ecological and environmental variation. For example, a turbidity threshold for indicating pollution in one category of water body may not hold in another. Secondly, these systems may fail to detect intricate interactions among various water parameters [21].

8.2 Statistical Approaches

With advancements in computational power, statistical and machine learning methods were brought in to address the inflexibility of rule-based systems. Statistical methods analyse large datasets of water quality parameters to identify correlations, trends, and probabilistic patterns that may signal the occurrence of waterborne diseases. Methods like regression analysis, decision trees, support vector machines (SVMs), random forests, and Bayesian networks are widely used.One of the key strengths of statistical approaches is their capacity to capture variability in water quality. Statistical models, in contrast to rule-based systems that operate on individual parameters, can assess the impact of multiple factors on each other's contributions to environmental contamination risk [22]. A model developed using SVM, for instance, might be trained on past data in which turbidity, nitrate concentration, and the number of microbes are associated with reported cases of cholera or typhoid outbreaks. After training, the model can predict disease probability from new input data. These are not the only strengths of statistical methods. They tend to need large, high-quality datasets for training, which might not be available in low-resource areas. Also, although they can learn linear and certain non-linear relationships, their performance can decrease in very complex and dynamic aquatic ecosystems where variable interactions are complicated. Another limitation is that they need feature engineering, i.e., domain knowledge, to determine which variables and data transformations are most useful. This is a time-consuming process involving technical and subject-matter skills.

8.3 Deep Learning Approaches

The latest developments in water disease detection leverage deep learning, a branch of artificial intelligence that uses multi-layer neural networks to automatically learn hierarchical features from raw data. Compared to statistical models that require engineer-designed features, deep learning algorithms can directly handle raw data such as time-series sensor readings, microbial genome sequences, and even satellite images of water bodies to identify patterns of contamination. The major advantage of deep learning methods lies in their ability to capture highly sophisticated, nonlinear associations [22]. For instance, a CNN trained on microscopic images of water samples can accurately differentiate between clean and contaminated samples more effectively than conventional methods. In the same vein, deep learning models can fuse multimodal data, e.g., water sensor data, weather variables, and past outbreak reports, to provide comprehensive insights into the likelihood of disease outbreaks.

Nonetheless, deep learning methods are not without drawbacks. They demand large quantities of labelled data for training, which can be time-consuming and costly to obtain in the water application domain. In addition, these models are computationally intensive and may be prohibitively expensive for real-time detection in limited-resource environments. Another key issue is the lack of interpretability; while deep learning models tend to be more accurate than traditional approaches, they are generally considered "black boxes," making it hard for health authorities to understand the rationale behind predictions. This transparency hurdle can impede adoption and trust in public health settings.

8.4 Comparative Insights

Overall, rule-based methods are simple and interpretable but have low scope and flexibility. Statistical methods are more flexible and robust, as they model probabilistic relationships, but are driven by data availability and feature engineering [22]. Deep learning techniques represent unmatched accuracy and the capability to identify complex relationships but are hampered by the need for enormous datasets, the expense of high computation, and the absence of interpretability. In reality, most practical water disease detection systems in the real world are moving toward hybrid solutions, integrating the interpretability of rules, flexibility of statistics, and potency of deep learning to create strong and scalable systems.

9. Water Disease Detection Dataset

Building a reliable deep learning model for detecting water disease depends on the availability of varied and high-quality datasets. They are categorically divided into two major groups: those with physicochemical parameters and those with microscopic images. Such data are essential for training and testing models that can precisely predict and detect contamination.

9.1 Water Quality Datasets

The papers cite several comprehensive and publicly available datasets. These datasets often include a variety of physicochemical parameters and are crucial for training and validating models.

- **Multipurpose Water Distribution Network (WDN) Datasets:** A collection of small- and medium-sized WDNs is available, including datasets like Anytown, Modena, Balerma, and C-Town[23]. These datasets provide extensive time-series data on parameters such as pressure, flow rate, and contamination events, totalling over 1.3 million hours for network analysis tasks such as anomaly detection and leak identification.
- **Global River Water Quality Archive (GRQA):** This archive provides water quality data from various sources globally. The data is often used for broad-scale water quality analysis and modelling. Some studies have noted that while valuable, the publicly available portion of GRQA may have limited daily observations and temporal coverage in some regions.
- **China Water Quality Datasets:** An extensive spatiotemporal dataset for China covers four decades (1980–2022) with over 330,000 observations across daily, weekly, and monthly records[23]. It spans 18 distinct indicators from 2,384 monitoring sites, providing a comprehensive resource for analysing surface water quality.
- **Kaggle Water Quality Datasets:** Various datasets from sources like North Queensland, Australia, are available on platforms like Kaggle. These include real-time, hourly measurements of water level, temperature, electrical conductivity, turbidity, and nitrate concentrations. These datasets are often pre-processed to remove outliers and are used for forecasting, imputation, and anomaly detection.
- **Water Quality Portal (WQP):** This is a collaborative service in the United States that integrates publicly available water quality data from the U.S. Geological Survey (USGS), the Environmental Protection Agency (EPA), and over 400 other agencies. It contains a vast amount of discrete water-quality data, making it a premiere source for nationwide research.

9.2 Microorganism and Image-Based Datasets

These datasets are specifically used for computer vision and image-based deep learning models to identify pathogens or other visual indicators of contamination.

- **Environmental Microorganism Image Dataset (EMDS):** This dataset is a crucial resource for training CNNs to detect microorganisms in water samples. Researchers have used versions like EMDS-6 and EMDS-7 for classification tasks[23]. It is noted that these datasets were supplemented with images of common pathogenic microorganisms from the web to create a more comprehensive set for training models to distinguish between harmful and non-harmful bacteria.
- **Stagnant Water Image Datasets:** An image dataset of stagnant water and wet surfaces, consisting of over 1,900 labelled images, was created for a study on stagnant water detection. The dataset is useful for deep learning models focused on disease control by identifying potential breeding sites for mosquitoes and other disease vectors in a given area.

9.3 PINNs

A Physics-Informed Neural Network (PINN) model is a specialised type of deep learning that integrates physical laws directly into the learning process. It is primarily used to solve forward and inverse problems involving partial differential equations (PDEs) and other physical constraints. Unlike traditional data-driven models that rely solely on large datasets, PINNs can leverage both observed data and the governing physical equations[24]. The loss function in a PINN is composed of two main parts: a data-fitting term (to fit the available data) and a physics-informed term (to enforce the PDE constraints). Because of this unique hybrid approach, PINNs do not necessarily require large, pre-labelled datasets in the conventional sense. Instead, they often use a small amount of scattered data points, or even no data at all, and rely on the physical laws as a form of self-supervision to guide the learning process. The provided documents mention several types of datasets and data sources that are suitable for PINNs:

- **Simulated Data:** One common approach is to use data generated from traditional numerical methods, such as Finite Element Method (FEM) or Finite Difference (FD) schemes, which can be used to validate the PINN's output[24]. For example, a study on groundwater flow in unconfined aquifers compared the results of a PINN model with those from the finite element model COMSOL and found good agreement.
- **Sensor and Observational Data:** In water management, PINNs can be used with real-world sensor data from the field. One study used data from four study locations in the Sacramento–San Joaquin Delta of California, including historical daily outflow values and simulated daily salinity values from 1991 to 2015, to train and test PINN models.
- **Hybrid Datasets:** Some models, such as Fourier Neural Operators (FNOs), use a blend of observed data from physical systems with data from simulations to train an emulator that can solve PDEs with greater efficiency.

9.4 Data from Specific Studies and Contexts

Some papers mention datasets generated for specific research projects or case studies, which, while not always public, illustrate the kind of data used.

- **Wind Turbine and Electrical Load Data:** For applications in predictive maintenance, models are trained on real-time sensor data from wind turbines or electrical grids to predict failures or anomalies[24]. This time-series data includes operational parameters and sensor readings to ensure reliability in critical infrastructure.

- **Clinical and ECG Data:** In a medical context, such as atrial fibrillation (AF) detection, datasets include ECG signals collected from various devices like resting ECGs, Holter monitors, or wearable patches. This data is used to train models to identify abnormal heart rhythms that may indicate a health issue.

Conclusion

Waterborne diseases continue to pose a significant threat to public health, especially in communities relying on stored household water tanks. Traditional methods of water quality assessment and disease detection are often reactive, expensive, and inaccessible to many households[26]. This paper presents a novel approach that combines Physics-Informed Neural Networks (PINNs) with a smart dashboard interface to provide a proactive, accurate, and real-time solution for monitoring water quality and predicting disease risks in household water storage systems. By integrating deep learning with known physical and biological principles governing water contamination, PINNs offer an interpretable and robust framework for early disease detection. The proposed system not only identifies potential diseases that may be caused by current water conditions but also predicts when the tank needs cleaning. This level of predictive maintenance and health risk alerting has the potential to greatly reduce the incidence of preventable waterborne illnesses at the domestic level.

The accompanying dashboard software ensures that this advanced technology remains accessible, user-friendly, and actionable for everyday users. Real-time insights, visual analytics, cleaning schedules, and disease warnings empower users to take control of their household water hygiene, thus bridging the gap between advanced AI models and public health outcomes. In the future, the system can be scaled to integrate community-level data, support remote rural monitoring, and contribute to centralised water quality management systems. With continued development, sensor integration, and community adoption, the proposed solution could become a cornerstone in the fight against waterborne diseases and a model for intelligent water management in homes across the world.

Challenges and Future Scope

Deep learning (DL) and artificial intelligence (AI) have demonstrated vast potential for transforming water management, but their mass adoption is hampered by significant challenges. Data inequity and a lack of data are key challenges. The majority of AI use is concentrated in data-rich, temperate countries, while arid and poor countries, where water problems are often most extreme, face a huge disparity. This is further exacerbated by the fact that water system historical databases are generally too small to adequately train DL algorithms, particularly for identifying infrequent abnormal events such as pipe bursts.[27] In addition, acquiring and normalising this data is difficult, labour-intensive, and costly. Hydrochemistry in water systems is controlled by many variables, including climate, soil types, and anthropogenic activities, which complicate the acquisition and analysis of data. This dependence on high-quality, big, and properly labelled datasets implies that models can be significantly impaired in cases of limited data or non-standard environmental scenarios.

The second important challenge is the "black box" aspect of most sophisticated DL models. Though these models may achieve superior predictive accuracy, decision-making processes might lack explainability and intuitiveness. This lack of transparency undermines stakeholder trust, a vital consideration in key decisions on public health and critical infrastructure. Without explanations of a model's output, policymakers and water managers may be reluctant to rely entirely on AI-based systems. Such complexity also raises scalability and computational cost concerns, where deployment at larger scales is problematic, particularly for resource-constrained organisations. The future of AI in water management is enormous, with an emphasis on tackling current issues to build better, more reliable systems. One of the main directions is the creation of hybrid models that combine DL with established physical and hydrological models. This aims to bridge the data-driven predictive capabilities of AI with the well-understood laws of physics, balancing predictability and interpretability. This can aid in overcoming data paucity by enabling models to train on limited amounts of data, while remaining consistent with the laws of physics. Work is also shifting toward building reliable AI models by focusing on quantitative measures of aspects such as fairness, uncertainty, and interpretability [28]. XAI development is important to make models not only trustworthy but also explainable, enabling improved human monitoring and informed decision-making.

Moreover, future research will emphasise the use of multi-source data fusion, integrating data from in-situ sensors, remote sensing, and other sources into comprehensive models that are better equipped to represent intricate environmental dynamics. Also a priority is the creation of low-cost, light-weight AI hardware and software solutions in order to advance general adoption and make AI accessible to all parts of the world, not only those with rich resources. Lastly, the creation of clear ethical and governance procedures is considered paramount for the assurance that AI will be utilised to foster sustainable, equitable, and inclusive water management for everyone.

References

[1] Wiese, T. L. (2024). Predictive maintenance using artificial intelligence in critical infrastructure: A decision-making framework. *International Journal of Engineering, Business and Management (IJEBM)*, *8*(4), 1-4.

[2] Attia, M., Driss, Z., Abd-Elhamid, H. F., & Ennetta, R. (2025). Enhanced water quality prediction: Application of deep neural networks and adaptive neuro-fuzzy inference systems to assess calcium concentration. *Desalination and Water Treatment*, *322*, 101193.

[3] Huang, H., Wang, P., Pei, J., Wang, J., Alexanian, S., & Niyato, D. (2025). Deep learning advancements in anomaly detection: A comprehensive survey. *IEEE Internet of Things Journal*.

[4] Humnabadkar, A., Karve, A., Shivbhakta, B., & Kokate, A. A. (2024). Water quality monitoring and measures of water usage in homes. *Int. J. Multidisciplinary Res.*, *6*(5), 1-9.

[5] Rizwan, A., Zoha, A., Mabrouk, I. B., Sabbour, H. M., Al-Sumaiti, A. S., Alomainy, A., ... & Abbasi, Q. H. (2020). A review on the state of the art in atrial fibrillation detection enabled by machine learning. *IEEE reviews in biomedical engineering*, *14*, 219-239.

[6] Moreno-Rodenas, A., Verbist, K., Mertens, A., Gerritsma, I., Deng, J., Haag, A., ... & Amarnath, G. (2025). Applications of AI for water management.

[7] Chandrasekaran, R., & Paramasivan, S. K. (2022). A state-of-the-art review of time series forecasting using deep learning approaches. *International Journal on Recent and Innovation Trends in Computing and Communication*, *10*(12), 92-105..

[8] Ed-Dehbi, W., Ahlaqqach, M., & Benhra, J. (2025). Artificial Intelligence for Optimal Water Resource Management: A Literature Review. *Engineering Proceedings*, *97*(1), 52.

[9] Kularbphettong, K., Waraporn, P., Raksuntorn, N., Vivhivanives, R., Sangsuwon, C., & Boonseng, C. (2023, December). Water Quality Index (WQI) Prediction Using Machine Learning Algorithms. In *2023 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 383-387). IEEE.

[10] Chowdhury, A. H., & Rahman, M. S. (2025). Machine learning and spatio-temporal analysis of meteorological factors on waterborne diseases in Bangladesh. *PLoS Neglected Tropical Diseases*, *19*(1), e0012800.

[11] Virnodkar, S. S., Pachghare, V. K., Patil, V. C., & Jha, S. K. (2020). Remote sensing and machine learning for crop water stress determination in various crops: a critical review. *Precision Agriculture*, *21*(5), 1121-1155.

[12] Fritz, B., & Fritz, J. (2022). Artificial intelligence for MRI diagnosis of joints: a scoping review of the current state-of-the-art of deep learning-based approaches. *Skeletal Radiology*, *51*(2), 315-329.

[13] Chen, Z., Pawar, K., Ekanayake, M., Pain, C., Zhong, S., & Egan, G. F. (2023). Deep learning for image enhancement and correction in magnetic resonance imaging—state-of-the-art and challenges. *Journal of Digital Imaging*, *36*(1), 204-230.

[14] Nyakuri, J. P., Nkundineza, C., Gatera, O., & Nkurikiyeyezu, K. (2024). State-of-the-art deep learning algorithms for internet of things-based detection of crop pests and diseases: A comprehensive review. *IEEE Access*, *12*, 169824-169849.

[15] Wang, B., Dai, L., & Liao, B. (2023). System architecture design of a multimedia platform to increase awareness of cultural heritage: A case study of sustainable cultural heritage. *Sustainability*, *15*(3), 2504.

[16] Toumi, S., Lekmine, S., Touzout, N., Moussa, H., Elboughdiri, N., Boudraa, R., ... & Tahraoui, H. (2024). Harnessing deep learning for real-time water quality assessment: a sustainable solution. *Water*, *16*(23), 3380.

[17] Zhao, S., Liu, R., Liu, Y., Zeng, T., Chen, C., & Xu, L. (2025). A water quality prediction model based on modal decomposition and hybrid deep learning models. Water, 17(2), 184.

[18] Olufemi, O. D., Ejiade, A. O., Ogunjimi, O., & Ikwuogu, F. O. (2024). AI-enhanced predictive maintenance systems for critical infrastructure: Cloud-native architectures approach. *World Journal of Advanced Engineering Technology and Sciences*, *13*(02), 229-257.

[19] Matamoros, N., Puchulu, M. B., Lerner, J. E. C., Maury-Sintjago, E., López, J. L., Sosio, V., ... & Cormick, G. (2024). Feasibility of increasing calcium content of drinking tap water following quality regulations to improve calcium intake at population level. *Gates open research*, *8*, 5.

[20] Cormick, G., Matamoros, N., Romero, I. B., Perez, S. M., White, C., Watson, D. Z., ... & Garitta, L. (2022). Testing for sensory threshold in drinking water with added calcium: A first step towards developing a calcium fortified water. *Gates open research*, *5*, 151..

[21] Rathor, S., & Kumari, S. (2023, July). Use of machine learning & IoT for water resources management. In *2ND INTERNATIONAL CONFERENCE ON FUTURISTIC AND SUSTAINABLE ASPECTS IN ENGINEERING AND TECHNOLOGY: FSAET-2021* (Vol. 2721, No. 1, p. 040014). AIP Publishing LLC.

[22] Kumar, R., & Tung, S. (2023, July). A comparative experimental study on concrete usingstone dust and river sand as fine aggregates. In *2nd International Conference On Futuristic and Sustainable Aspects in Engineering and Technology: FSAET-2021* (Vol. 2721, No. 1, p. 020002). AIP Publishing LLC.

[23] Qin, Z., Wu, D., Xiao, Z., Fu, B., & Qin, Z. (2018). Modeling and analysis of data aggregation from convergecast in mobile sensor networks for industrial IoT. *IEEE Transactions on Industrial Informatics*, *14*(10), 4457-4467.

[24] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

[25] Hussain, M., Cifci, M. A., Sehar, T., Nabi, S., Cheikhrouhou, O., Maqsood, H., ... & Mohammad, F. (2023). Machine learning based efficient prediction of positive cases of waterborne diseases. *BMC medical informatics and decision making*, *23*(1), 11.

[26] Fang, X., Liu, W., Ai, J., He, M., Wu, Y., Shi, Y., ... & Bao, C. (2020). Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. *BMC infectious diseases*, *20*(1), 222.

[27] Wang, Y., Sun, S., Chen, X., Zeng, X., Kong, Y., Chen, J., ... & Wang, T. (2021). Short-term load forecasting of industrial customers based on SVMD and XGBoost. *International Journal of Electrical Power & Energy Systems*, *129*, 106830.

[28] Vijay Anand, M., Sohitha, C., Saraswathi, G. N., & Lavanya, G. V. (2023, May). Water quality prediction using CNN. In *Journal of Physics: Conference Series* (Vol. 2484, No. 1, p. 012051). IOP Publishing.

[29] Khant, N. A., Lumongsod, R. M., San, A., Moon, J., Namkoong, S., & Kim, H. (2025). Navigating the complex landscape of waterborne disease research. *Journal of Water and Health, 23(2), 168-189.*

[30] Feleni, U., Morare, R., Masunga, G. S., Magwaza, N., Saasa, V., Madito, M. J., & Managa, M. (2025). Recent developments in waterborne pathogen detection technologies. *Environmental Monitoring and Assessment, 197(3), 233.*

[31] Alhamlan, F. S., Al-Qahtani, A. A., & Al-Ahdal, M. N. A. (2015). Recommended advanced techniques for waterborne pathogen detection in developing countries. *The Journal of Infection in Developing Countries, 9(02), 128-135.*

[32] Wen, X., Fang, C., Huang, L., Miao, J., & Lin, Y. (2025). Mapping total microbial communities and waterborne pathogens in household drinking water in China by citizen science and metabarcoding. *Frontiers in Microbiology, 16, 1609070.*

[33] Oon, Y. L., Oon, Y. S., Ayaz, M., Deng, M., Li, L., & Song, K. (2023). Waterborne pathogens detection technologies: advances, challenges, and future perspectives. *Frontiers in Microbiology, 14, 1286923.*

[34] Shahrashoob, M., Dehshiri, M., Yousefi, V., Moassesfar, M., Saberi, H., Molaabasi, F., ... & Rhee, K. Y. (2025). Optical and electrochemical biosensors for detection of pathogens using metal nanoclusters: A systematic review. *Biosensors, 15(7), 460.*

[35] Kidanemariam, A., & Cho, S. (2025). Metal–organic-framework-based optical biosensors: recent advances in pathogen detection and environmental monitoring. *Sensors (Basel, Switzerland), 25(16), 5081.*

[36] Feleni, U., Morare, R., Masunga, G. S., Magwaza, N., Saasa, V., Madito, M. J., & Managa, M. (2025). Recent developments in waterborne pathogen detection technologies. *Environmental Monitoring and Assessment, 197(3), 233.*

[37] Glasgow, H. B., Burkholder, J. M., Reed, R. E., Lewitus, A. J., & Kleinman, J. E. (2004). Real-time remote monitoring of water quality: a review of current applications, and advancements in sensor, telemetry, and computing technologies. *Journal of experimental marine biology and ecology, 300(1-2), 409-448.*

[38] Pan, D., Deng, Y., Yang, S. X., & Gharabaghi, B. (2025). Recent advances in remote sensing and artificial intelligence for river water quality forecasting: A review. *Environments, 12(5), 158.*

[39] Oon, Y. L., Oon, Y. S., Ayaz, M., Deng, M., Li, L., & Song, K. (2023). Waterborne pathogens detection technologies: advances, challenges, and future perspectives. *Frontiers in Microbiology, 14, 1286923.*

[40] Paruch, L. (2022). Molecular diagnostic tools applied for assessing microbial water quality. *International Journal of Environmental Research and Public Health, 19(9), 5128.*

[41] Banakar, M., Hamidi, M., Khurshid, Z., Zafar, M. S., Sapkota, J., Azizian, R., & Rokaya, D. (2022). Electrochemical biosensors for pathogen detection: an updated review. *Biosensors, 12(11), 927.*

[42] Zhang, S., Li, X., Wu, J., Coin, L., O'brien, J., Hai, F., & Jiang, G. (2021). Molecular methods for pathogenic bacteria detection and recent advances in wastewater analysis. *Water, 13(24), 3551.*

[43] Sharma, R., Satapathy, A., Srivastava, V., & Saxena, R. (2025). Revolutionizing water quality management the impact of machine learning and artificial intelligence. *Computational Automation for Water Security, 27-42*.

[44] Ighalo, J. O., Adeniyi, A. G., & Marques, G. (2021). Artificial intelligence for surface water quality monitoring and assessment: a systematic literature analysis. *Modeling Earth Systems and Environment, 7(2), 669-681*.