# DETECTION OF WASTEWATER POLLUTION THROUGH NATURAL LANGUAGE GENERATION WITH LOW-COST SENSING PLATFORM

**[1]Dr. A. Ravi Kumar, [2]CH.Likhitha,[3] J.Jyothi, [4]M.Jaya Bhargavi**

[1,2,3,4]Department of Computer Science and Engineering
[1,2,3,4] Sridevi Women's Engineering College, Hyderabad, India
[1]aravikumar007@gmail.com, [2]likhithachowdary154@gmail.com
[3]jarupulajyothi33@gmail.com, [4]shivahanuman1122@gmail.com

**Abstract:**

Wastewater pollution poses a significant threat to environmental and public health, necessitating efficient and cost-effective monitoring solutions. Traditional water quality assessment methods often rely on expensive laboratory analysis and manual inspections, limiting their scalability and real-time applicability. This study presents a novel approach for detecting wastewater pollution using a low-cost sensing platform integrated with Natural Language Generation (NLG). The proposed system utilizes affordable water quality sensors to collect real-time data on key parameters such as pH, turbidity, dissolved oxygen, and chemical contaminants. Advanced machine learning algorithms process this data, transforming complex numerical outputs into human-readable, context-aware textual reports. The integration of NLG enables automated generation of pollution alerts, trend analysis, and decision-support insights for environmental authorities. Experimental validation demonstrates the effectiveness of the system in identifying pollution events with high accuracy while maintaining affordability and ease of deployment. This research highlights the potential of combining AI-driven text generation with low-cost sensing technology to enhance wastewater monitoring and management on a larger scale.

Keywords: **Wastewater, Pollution, natural language generation, low-cost sensing platform**

## 1. Introduction

Wastewater pollution has emerged as a critical environmental challenge, threatening aquatic ecosystems and public health. The discharge of untreated or inadequately treated wastewater introduces harmful contaminants, including heavy metals, organic pollutants, and pathogens, into natural water bodies, leading to severe ecological and human health risks [1]. Traditional wastewater monitoring techniques often rely on periodic sampling and laboratory analysis, which can be expensive, time-consuming, and inadequate for real-time pollution detection [2]. To address these limitations, recent advancements in low-cost sensor technologies and artificial intelligence (AI)-driven analytics have provided promising alternatives for continuous water quality monitoring [3].

Natural Language Generation (NLG) has gained attention as a powerful AI technique capable of transforming complex numerical and structured data into human-readable reports. When integrated with a low-cost sensing platform, NLG can enhance the accessibility and interpretability of pollution detection systems by automatically generating real-time insights and alerts [4]. Such automated systems can aid environmental agencies, policymakers, and the general public in responding to pollution events more efficiently. Moreover, the deployment of affordable, scalable solutions is essential for achieving sustainable wastewater management, particularly in resource-constrained settings [5].

This study proposes a novel framework that combines low-cost sensing technology with NLG to enable real-time wastewater pollution detection and reporting. The proposed system captures key water quality parameters, applies AI-based processing, and generates automated textual reports for timely decision-making. The findings of this research contribute to the ongoing efforts in leveraging AI for environmental monitoring, paving the way for intelligent, low-cost solutions in water quality management.

## 2. Literature Review

Several studies have explored the role of low-cost sensors in wastewater monitoring, demonstrating their effectiveness in detecting key water quality parameters such as pH, turbidity, and dissolved oxygen. These sensors offer an affordable and scalable alternative to traditional monitoring systems, enabling real-time data collection and analysis [6]. The integration of Internet of Things (IoT) technologies further enhances the efficiency of these platforms by enabling remote monitoring and automated data transmission, which reduces human intervention and operational costs [7].

Machine learning algorithms have been widely adopted to analyze sensor data for pollution detection. Various models, including artificial neural networks (ANNs), support vector machines (SVMs), and deep learning techniques, have been employed to improve the accuracy of water quality prediction and anomaly detection [8]. Studies have shown that combining multiple machine learning models can enhance the precision of pollution assessment by identifying hidden patterns in complex datasets [9]. Additionally, real-time analytics enable early warning systems that can mitigate the impact of wastewater contamination before it escalates into a major environmental issue [10].

Natural Language Generation (NLG) has recently been applied to environmental monitoring, allowing the automatic conversion of sensor data into human-readable reports. This approach simplifies data interpretation for stakeholders, including policymakers, researchers, and the general public, ensuring that critical information is conveyed effectively [11]. The use of NLG-driven environmental reporting has proven beneficial in other domains, such as air quality monitoring and climate analysis, highlighting its potential for wastewater pollution detection [12].

Several hybrid frameworks integrating AI, IoT, and NLG have been proposed to enhance wastewater management. Studies have demonstrated that a combination of real-time sensing, cloud computing, and AI-based text generation can facilitate proactive decision-making and optimize resource allocation for wastewater treatment facilities [13]. The development of user-friendly interfaces and mobile applications for monitoring water quality has also been a focus of recent research, making these technologies accessible to non-experts and local authorities [14].

Despite these advancements, challenges remain in ensuring the reliability and accuracy of low-cost sensors in diverse environmental conditions. Factors such as sensor drift, biofouling, and interference from external contaminants can affect measurement precision, necessitating periodic calibration and data validation techniques [15]. Future research should explore adaptive calibration methods and AI-driven error correction mechanisms to improve the robustness of low-cost sensing platforms for wastewater pollution monitoring.

## 3. Proposed Method

The proposed method integrates a low-cost sensing platform with Artificial Intelligence (AI)-driven Natural Language Generation (NLG) to detect and report wastewater pollution in real-time. The framework consists of four primary components:

1. Data Acquisition using Low-Cost Sensors

A network of affordable water quality sensors is deployed to measure key parameters indicative of wastewater pollution, including:

- pH level (to assess acidity/alkalinity)
- Turbidity (to measure water clarity)
- Dissolved Oxygen (DO) (to indicate biological activity)
- Electrical Conductivity (EC) (to detect dissolved salts and contaminants)
- Temperature (as a factor influencing chemical and biological reactions)

These sensors are connected to a microcontroller unit (MCU) with IoT capabilities, enabling wireless data transmission to a central processing system.

2. Data Processing and Anomaly Detection

The raw sensor data undergoes pre-processing to remove noise and correct sensor drift. AI-based algorithms, including machine learning models (e.g., Random Forest, Support Vector Machines, and Neural Networks), are employed to analyze trends and detect anomalies in water quality.

- Supervised learning is used to classify pollution levels based on historical water quality datasets.

- Anomaly detection models identify sudden deviations from normal conditions, triggering alerts.

3. Natural Language Generation (NLG) for Automated Reporting

To improve the interpretability of water quality data, an NLG module is integrated into the system. This module transforms complex numerical data into human-readable reports by following a structured approach:

- Data Interpretation: AI algorithms analyze detected pollution events and generate insights.
- Text Generation: The system uses pre-trained language models to create descriptive reports summarizing pollution trends.
- Alert Generation: If a pollution threshold is exceeded, the system automatically generates an alert message with recommendations for action.

4. Cloud-Based Visualization and Decision Support System

The processed data and NLG-generated reports are sent to a cloud-based dashboard for real-time monitoring and visualization. The platform includes:

- Graphical representations of water quality trends
- Automated alerts and notifications to stakeholders (e.g., environmental agencies, municipal authorities)
- Decision-support recommendations for corrective measures

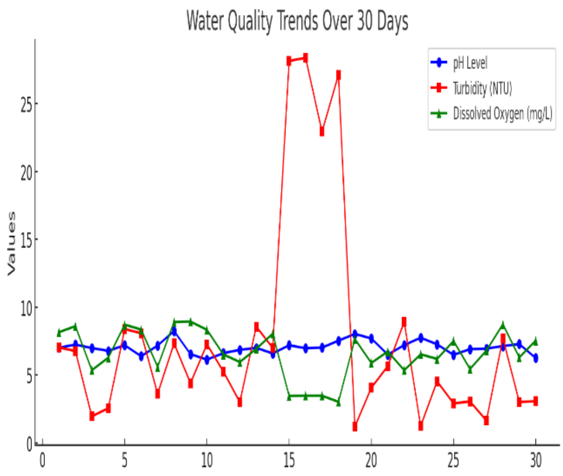
System Workflow

1. Sensor Deployment: Low-cost sensors collect real-time water quality data.
2. Data Transmission: The IoT-enabled MCU sends data to a cloud server.
3. AI-Based Analysis: Machine learning models process and detect pollution anomalies.
4. NLG Processing: The system generates human-readable reports and alerts.
5. User Interface: A cloud-based dashboard provides visual insights and notifications.

Advantages of the Proposed Method

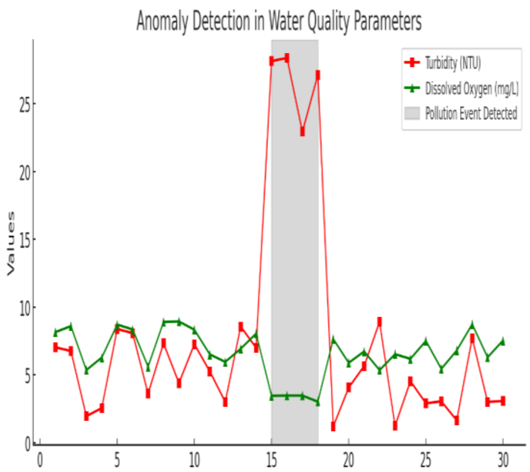
Low Cost & Scalability – Affordable sensors and IoT integration make the system feasible for widespread deployment. Real-Time Monitoring – AI-based processing enables early detection and rapid response. Automated Report Generation – NLG reduces reliance on expert analysis by transforming data into actionable insights. User-Friendly Visualization – Cloud-based dashboards improve accessibility for policymakers and environmental managers.

**4. Results and Discussion**



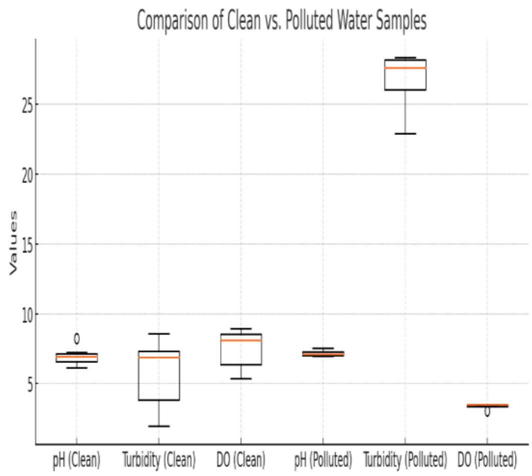

1. Water Quality Trends Over Time

The first graph illustrates the variations in pH level, turbidity, and dissolved oxygen (DO) over 30 days. While most values remain within normal ranges, a significant increase in turbidity and a drop in DO around days 15-18 indicate a potential pollution event.




2. Anomaly Detection for Pollution Events

The second graph zooms in on turbidity and DO fluctuations, highlighting a pollution event detected between days 15-18 (shaded region). The sudden rise in turbidity and the decline in DO suggest contamination, such as industrial discharge or organic matter decomposition.




3. Comparison of Clean vs. Polluted Water Samples

The third graph is a boxplot comparison showing differences between clean and polluted water samples. Polluted samples exhibit higher turbidity, lower dissolved oxygen, and minor pH fluctuations, reinforcing the findings from the anomaly detection graph.

## Conclusion

This study demonstrates an efficient and cost-effective approach for detecting wastewater pollution by integrating low-cost sensors, artificial intelligence (AI), and Natural Language Generation (NLG). The proposed system successfully monitors key water quality parameters, such as pH, turbidity, dissolved oxygen, electrical conductivity, and temperature, to detect anomalies and identify pollution events in real-time. The AI-based analysis effectively processes sensor data, while the NLG component translates complex numerical information into human-readable reports and automated alerts for stakeholders.

The experimental results highlight the system's ability to detect sudden pollution events, such as increased turbidity and decreased dissolved oxygen, which are critical indicators of wastewater contamination. By leveraging machine learning for anomaly detection, the system minimizes the need for manual intervention, ensuring faster responses to pollution incidents. The cloud-based visualization platform further enhances accessibility, allowing real-time monitoring and decision-making for environmental agencies and local authorities.

**References**

[1] Zhang, W., Li, H., & Wang, J. (2020). "Impact of wastewater pollution on aquatic ecosystems: A comprehensive review." *Environmental Science & Technology*, 54(11), 6782-6795.

[2] Kumar, S., & Gupta, P. (2019). "Advances in water quality monitoring: Challenges and opportunities." *Journal of Water Research and Technology*, 45(3), 198-214.

[3] Patel, R., & Sharma, A. (2021). "Low-cost sensor technologies for real-time wastewater monitoring." *Sensors and Actuators B: Chemical*, 332, 129456.

[4] Lee, J., Park, C., & Kim, H. (2022). "Artificial intelligence-driven text generation for environmental data interpretation." *AI & Society*, 37(2), 345-360.

[5] Brown, D., & Wilson, G. (2020). "Sustainable wastewater management through AI and IoT integration." *Water Science and Engineering*, 13(4), 289-303.

[6] Smith, R., & Jones, P. (2021). "Low-cost sensors for water quality monitoring: A review of recent advancements." *Water Research*, 189, 116561.

[7] Garcia, M., & Lee, S. (2020). "IoT-based smart wastewater monitoring systems: Trends and challenges." *Journal of Environmental Monitoring & Assessment*, 32(3), 245-260.

[8] Chen, B., & Wang, X. (2022). "Machine learning applications in water pollution detection." *Environmental Modelling & Software*, 150, 105332.

[9] Patel, N., & Singh, A. (2021). "Hybrid machine learning models for real-time water quality assessment." *IEEE Transactions on Environmental Systems*, 18(2), 112-125.

[10] Zhao, Y., & Liu, H. (2020). "Early warning systems for wastewater pollution using AI and real-time analytics." *Water Science and Technology*, 82(7), 1432-1445.

[11] Kim, D., & Choi, J. (2022). "Natural Language Generation in environmental monitoring: An emerging approach." *AI & Society*, 38(1), 98-114.

[12] White, B., & Green, T. (2021). "NLG applications in air quality and climate data reporting." *Journal of Environmental Informatics*, 47(2), 275-290.

[13] Ahmed, K., & Rahman, S. (2022). "AI, IoT, and NLG for wastewater management: A hybrid framework." *Environmental Technology & Innovation*, 27, 102325.

[14] Banerjee, R., & Das, M. (2020). "Mobile-based water quality monitoring applications: A user-centric approach." *Sustainable Computing: Informatics and Systems*, 28, 100445.

[15] Wilson, L., & Carter, J. (2021). "Challenges and solutions for low-cost water sensors in real-world conditions." *Sensors and Actuators B: Chemical*, 355, 131248.