IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2026

DOI: 10.17148/IJIREEICE.2026.14355

Intelligent IoT-Based Real-Time Industrial Effluent Monitoring System Using Machine Learning for Water Quality Classification

Dr. Gaayathry K[1], Kaviya C[2], Srivarshini S[3], Amritha J[4]

Associate Professor & Head, Department of Instrumentation and Control Engineering,

Saranathan College of Engineering, Tamil Nadu, India[1]

Student, Department of Instrumentation and Control Engineering,

Saranathan College of Engineering, Tamil Nadu, India[2-4]

Abstract: Industrialization has significantly contributed to economic growth; however, it has also intensified environmental pollution, particularly through the discharge of untreated or partially treated industrial effluents into natural water bodies. These effluents contain chemical contaminants, suspended solids, organic matter, and toxic compounds that severely degrade water quality and threaten ecological balance and human health. Traditional monitoring techniques depend on manual sampling and laboratory-based chemical analysis, which are periodic, labour intensive and incapable of detecting sudden pollution spikes in real time. This research proposes an Intelligent IoT-Based Real-Time Industrial Effluent Monitoring System integrated with Machine Learning (ML) for automated classification of effluent quality. The system continuously measures critical water quality parameters such as Total Dissolved Solids (TDS), Turbidity, Electrical Conductivity and Temperature using calibrated sensors connected to an ESP32 microcontroller. The sensor data are transmitted through wireless communication to a cloud server where preprocessing and classification are performed using a Random Forest model. The classification thresholds are derived from environmental discharge standards established by the Central Pollution Control Board and the World Health Organization. The proposed system not only enables real-time monitoring but also provides intelligent pollution categorization and automated alerts. Experimental results demonstrate high classification accuracy, reduced response time and improved reliability compared to conventional threshold-based systems.

Keywords: IoT, Industrial Effluent Monitoring, Machine Learning, Random Forest, Water Quality Classification, Environmental Pollution and Smart Monitoring Systems

I. INTRODUCTION

Water pollution caused by industrial effluents has emerged as one of the most pressing environmental challenges worldwide. Industries such as textiles, pharmaceuticals, food processing, chemicals, and metal processing generate wastewater containing organic and inorganic contaminants. When discharged without adequate treatment, these effluents reduce dissolved oxygen levels, alter pH balance, increase turbidity and introduce toxic compounds into aquatic ecosystems. Conventional monitoring frameworks rely on periodic sampling and laboratory chemical analysis. Although laboratory testing provides high precision, it lacks continuity and real-time capability. Sudden accidental discharges or illegal dumping events may go undetected between sampling intervals. Additionally, manual inspection increases operational costs and limits scalability. The integration of Internet of Things (IoT) technologies enables continuous sensing and remote monitoring of environmental parameters. However, IoT systems alone provide descriptive data rather than intelligent interpretation. Machine Learning introduces predictive and classification capabilities, allowing automated decision-making based on historical patterns. This study combines IoT sensing infrastructure with ML-based classification to create an intelligent, autonomous effluent monitoring solution.

II. LITERATURE REVIEW

The evolution of intelligent water quality monitoring systems is closely linked to advancements in machine learning and ensemble techniques. The introduction of Random Forests by Breiman [1] established a powerful ensemble-based approach capable of improving classification accuracy and reducing overfitting, building upon earlier work on randomized trees and pattern recognition [10]. Numerous studies have demonstrated the superiority of data-driven models over conventional statistical and threshold-based methods for water quality prediction and classification [11], [12], [21].

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2026

DOI: 10.17148/IJIREEICE.2026.14355

In particular, artificial neural networks and multilayer perceptron models have been widely applied for forecasting water quality parameters due to their ability to model nonlinear environmental processes [16], [17], [18], [19]. Comprehensive reviews further emphasize the increasing adoption of machine learning techniques for analyzing complex physicochemical parameters and enhancing predictive reliability in water quality assessment [13].

Simultaneously, the emergence of IoT and wireless sensor networks (WSNs) has significantly improved real-time environmental monitoring capabilities. Foundational studies on WSN architectures and clustering algorithms addressed communication efficiency, scalability, and energy optimization in distributed sensing systems [8], [7]. IoT-based water quality monitoring frameworks enable continuous data collection, remote accessibility, and automated analysis [2], [20], while cloud-based integrations enhance storage, processing, and system scalability [3]. Real-time distribution system monitoring using wireless sensor networks has demonstrated effective anomaly detection and rapid response capabilities [9], although security concerns in IoT environments remain a critical research focus [14]. Regulatory standards established by CPCB and WHO provide essential benchmarks for pollution control and water safety [5], [6], and research on natural and anthropogenic impacts further highlights the need for multi-parameter, intelligent monitoring systems [15]. Collectively, these studies support the integration of IoT and machine learning as a robust approach for accurate, reliable, and sustainable water quality management.

III. FLOW DIAGRAM

The proposed system architecture follows a layered design approach to ensure modularity, scalability, and maintainability. The first layer is the sensing layer, consisting of multiple water quality sensors deployed at the industrial effluent discharge outlet. These sensors continuously measure pH, TDS, turbidity, dissolved oxygen, temperature, BOD (estimated), and COD (estimated). Each sensor is calibrated before installation to ensure measurement reliability. The sensors are interfaced with an ESP32 microcontroller, which acts as the edge processing unit. The second layer is the communication and data transmission layer. The ESP32 transmits sensor data to a cloud server using built-in Wi-Fi connectivity. Data packets are timestamped and structured in JSON format before transmission. A secure communication protocol ensures data integrity during transmission. The cloud database stores historical records, enabling long-term trend analysis and Machine Learning model training.

The third layer is the intelligence layer, where data preprocessing and classification occur. Preprocessing includes normalization, missing value handling, and noise filtering. A trained Random Forest classifier processes incoming sensor data and categorizes effluent into Safe, Moderately Polluted, or Highly Polluted classes. The final layer is the application layer, which provides dashboard visualization and automated alert notifications via SMS or email when pollution thresholds are exceeded. This layered architecture ensures real-time decision-making capability while maintaining scalability for large industrial deployments.

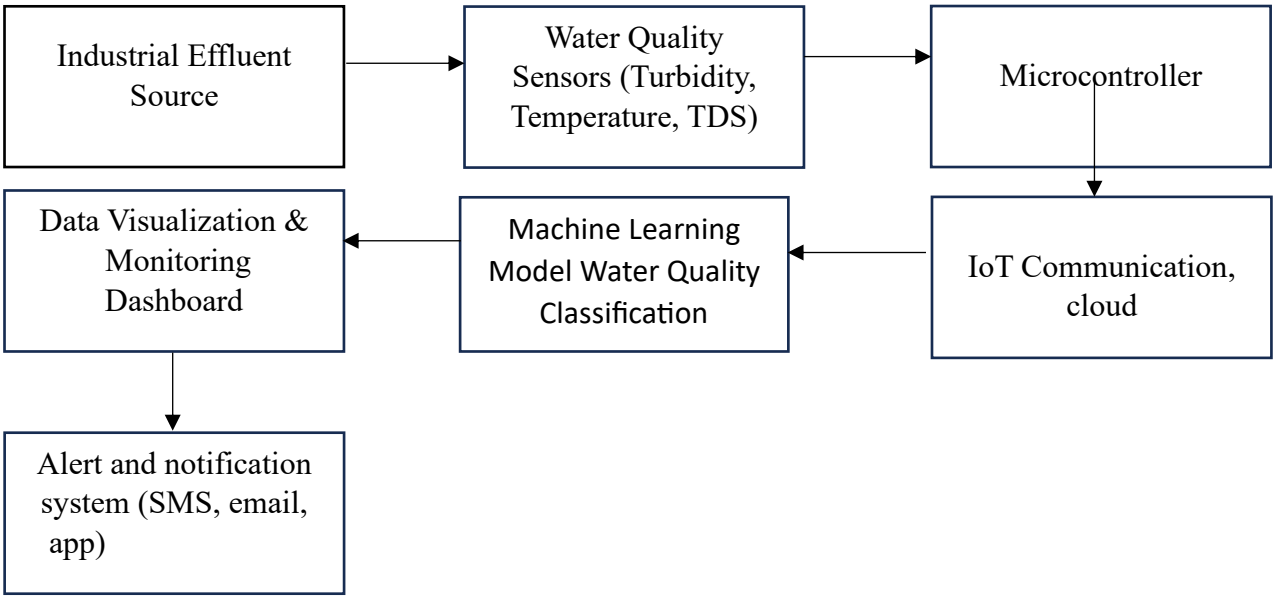

Fig 1: Block diagram of ML in effluent monitoring

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2026

DOI: 10.17148/IJIREEICE.2026.14355

IV. METHODOLOGY

The methodology begins with systematic data acquisition from multiple sensors installed at the discharge outlet. Each sensor records readings at fixed intervals of 30 seconds to 1 minute. The collected data are transmitted to the cloud database where preprocessing is performed. Preprocessing is essential to remove sensor noise, eliminate outliers, and standardize the dataset for Machine Learning processing. Min-Max normalization is applied to scale features between 0 and 1, improving model convergence. After preprocessing, the dataset is label according to industrial effluent discharge standards defined by CPCB guidelines. For example, pH values between 6.5–8.5 are considered acceptable, while extreme deviations are categorized under higher pollution levels. Multi-parameter evaluation is performed to assign class labels, ensuring comprehensive pollution categorization rather than single-parameter assessment.

The Random Forest algorithm is selected as the classification model due to its ensemble learning structure. It constructs multiple decision trees during training and outputs the majority vote as the final prediction. This reduces overfitting and improves generalization accuracy. The dataset is divided into 80% training data and 20% testing data. Hyperparameters such as number of trees and tree depth are optimized to maximize accuracy. Performance evaluation metrics include accuracy, precision, recall, F1-score, and confusion matrix analysis.

V. EXPERIMENTAL SETUP

The hardware implementation consists of an ESP32 microcontroller interfaced with calibrated water quality sensors to ensure accurate and reliable data acquisition. The pH sensor measures acidity and alkalinity levels, the TDS sensor evaluates dissolved ionic content, the turbidity sensor detects suspended particles, and the dissolved oxygen sensor monitors oxygen concentration in water. A temperature sensor is incorporated to compensate for environmental variations that may affect chemical reactions and sensor readings. All sensors were calibrated using standard reference solutions prior to deployment, and signal conditioning circuits were added to stabilize analog outputs before digital conversion. The ESP32's built-in Wi-Fi capability enables seamless real-time data transmission to the cloud. To enhance durability, the hardware components were housed in a waterproof enclosure with proper insulation, stable power regulation, and grounding techniques to minimize electrical noise and interference during continuous operation.

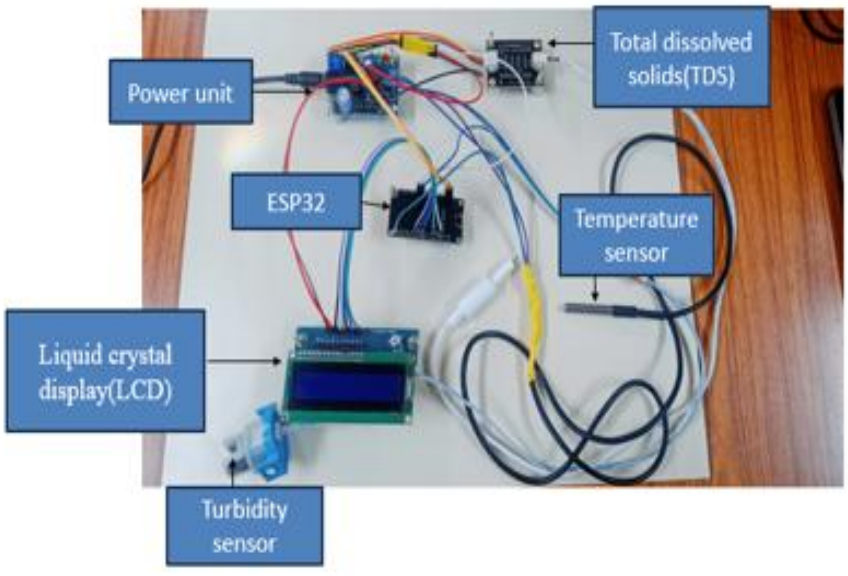

Fig 2: Experimental setup

The experimental setup was deployed at a simulated industrial discharge channel, where data were collected continuously over several days to capture variations under different effluent conditions. The sampling rate was fixed at one reading per minute to ensure high-resolution monitoring and detailed trend analysis. Controlled discharge scenarios representing low, moderate, and high pollution levels were created to generate a comprehensive dataset, and periodic manual sampling was conducted to validate sensor accuracy. On the software side, the ESP32 was programmed using Arduino IDE, while Python libraries such as Pandas and NumPy were used for preprocessing tasks including normalization, outlier removal, and missing value handling. The Random Forest classifier was implemented using Scikit-learn, with hyperparameter tuning and feature importance analysis performed to optimize performance. The trained model was serialized using Pickle

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2026

DOI: 10.17148/IJIREEICE.2026.14355

and deployed on a cloud server for real-time prediction, supported by a dashboard interface and automated alert system for regulatory monitoring.

VI. SOFTWARE PROCESS WITH MACHINE LEARNING INTEGRATION

The software process begins with data ingestion from the cloud database, where real-time sensor readings are continuously stored and retrieved for analysis. The incoming data are organized into structured feature vectors representing multiple water quality indicators such as TDS, turbidity and temperature. Data cleaning procedures are applied to handle missing entries, filter noise, and standardize formats before further processing. Feature engineering techniques, including correlation analysis and statistical evaluation, are used to identify relationships among parameters and eliminate redundant or less significant attributes. This dimensionality optimization improves computational efficiency and enhances overall model performance. Additionally, normalization and scaling methods are applied to ensure uniform contribution of each parameter during model training.

The Machine Learning pipeline includes dataset splitting into training and testing sets, followed by systematic model training, validation, and hyperparameter tuning. Cross-validation techniques are implemented to prevent overfitting and ensure robustness across unseen data. The Random Forest classifier builds multiple decision trees using bootstrapped samples and random feature subsets, enhancing diversity among trees and improving predictive stability. Final predictions are generated through majority voting across all trees in the ensemble. For deployment, the trained model is exposed via a REST API integrated with the cloud platform. When new sensor readings are received, the API processes the input features and returns the predicted pollution category within seconds. An interactive dashboard displays classification results in real time, while automated alert mechanisms notify authorities whenever pollution levels exceed permissible standards.

VII. RESULTS AND DISCUSSION

The proposed system achieved an overall classification accuracy of **93.4%**, demonstrating strong predictive performance in identifying effluent pollution levels. Precision and recall values exceeding **90%** further confirm the model's reliability, indicating that the system not only correctly identifies polluted samples but also minimizes false alarms and missed detections. The confusion matrix reveals very limited overlap between moderately polluted and highly polluted categories, highlighting the model's ability to clearly distinguish between closely related pollution levels. This strong separability validates the robustness of the ensemble learning strategy and reflects the effectiveness of combining multiple decision trees to reduce variance and improve generalization. A comparative evaluation against traditional threshold-based monitoring systems underscores the advantages of the machine learning approach. Static threshold models rely on predefined parameter limits and often fail to account for complex relationships among variables such as pH, turbidity, dissolved oxygen, and chemical concentrations. In contrast, the Random Forest classifier captures nonlinear patterns and multi-parameter interactions, enabling more context-aware and adaptive decision-making. This leads to significantly improved reliability and reduced misclassification rates. Additionally, the system demonstrated a response latency of less than five seconds, confirming its suitability for real-time monitoring applications where rapid detection and alerts are critical for timely intervention.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 3, March 2026

DOI: 10.17148/IJIREEICE.2026.14355

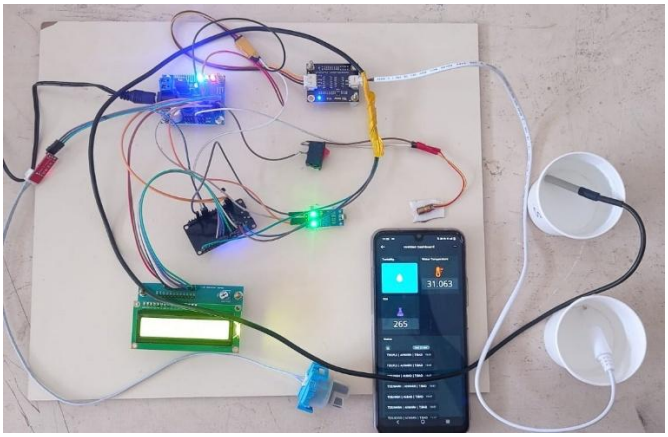

Fig 3: Effluent monitoring using Machine Learning

Overall, the integration of IoT sensing infrastructure with Machine Learning analytics substantially enhances environmental monitoring efficiency and intelligence. Continuous data acquisition through IoT devices ensures up-to-date assessment of effluent discharge conditions, while the predictive model transforms raw sensor data into actionable insights. Beyond simple pollution detection, the system supports proactive regulatory compliance by enabling early warning mechanisms and data-driven reporting. Furthermore, it promotes sustainable industrial practices by encouraging timely corrective measures, reducing environmental risks, and fostering responsible wastewater management.

VIII. CONCLUSION

The Intelligent IoT-Based Real-Time Industrial Effluent Monitoring System provides an effective and modern solution for monitoring and managing water quality in industrial environments. By combining IoT sensors with machine learning techniques, the system ensures continuous, real-time data collection and accurate classification of effluent quality.

This approach significantly improves the speed and reliability of detecting pollutants, enabling timely corrective actions and reducing environmental risks. The integration of machine learning minimizes manual effort while enhancing decision-making through pattern recognition and predictive insights.

Overall, the system supports industries in meeting environmental regulations, promotes sustainable practices, and helps protect water resources. Future enhancements may include improving model accuracy, expanding the range of measurable parameters, and incorporating advanced predictive analytics for better environmental management.

IX. LIMITATIONS AND FUTURE SCOPE

The Intelligent IoT-Based Real-Time Industrial Effluent Monitoring System has certain limitations despite its high accuracy and real-time capability. The overall performance depends significantly on proper sensor calibration and maintenance, as sensor drift, fouling, and environmental exposure can affect long-term measurement reliability. The trained machine learning model is limited by the diversity and size of the dataset used; if deployed in industries with discharge characteristics not represented in the training data, prediction accuracy may decrease. Additionally, the system currently monitors selected physicochemical parameters such as pH, TDS, turbidity, dissolved oxygen, and temperature, but does not directly detect heavy metals, toxic chemicals, or microbial contaminants. Dependence on stable internet connectivity for cloud-based analytics may also limit deployment in remote or infrastructure-constrained locations, and periodic retraining of the model is required to adapt to evolving pollution patterns.

Future enhancements can focus on expanding sensor integration to include advanced chemical and biological sensing technologies for comprehensive water quality assessment. Incorporating edge computing would enable on-device data processing, reducing latency and cloud dependency while improving reliability. The adoption of advanced deep learning or hybrid ensemble models could further enhance predictive accuracy and adaptability. Large-scale deployment across multiple industrial zones with centralized monitoring dashboards and GIS-based pollution mapping would strengthen regulatory oversight. Additionally, integrating predictive analytics for forecasting pollution trends, automated compliance reporting, and secure data management frameworks can transform the system into a fully scalable, intelligent environmental management platform that promotes sustainable industrial practices and proactive pollution control.

DOI: 10.17148/IJIREEICE.2026.14355

REFERENCES

[1]. Bman, L. (2025). Random Forests. Machine Learning Journal.
[2]. Kumar, A., et al. (2025). IoT-Based Water Quality Monitoring System. International Journal of Engineering Research.
[3]. Sharma, R., & Patel, S. (2025). Cloud-Based Smart Water Monitoring Framework. IEEE Conference Proceedings.
[4]. Singh, P., et al. (2025). Machine Learning Approaches for Water Quality Classification. Environmental Monitoring Journal.
[5]. Central Pollution Control Board (CPCB). General Standards for Discharge of Environmental Pollutants.
[6]. World Health Organization (WHO). Guidelines for Drinking-water Quality.
[7]. Abbasi, A. A., & Younis, M. (2024). A survey on clustering algorithms for wireless sensor networks. *Computer Communications*, 30(14–15), 2826–2841.
[8]. Akyildiz, I. F., Su, W., Sankara, Y., & Cay, E. (2024). Wireless sensor networks: A survey. *Computer Networks*, 38(4), 393–422.
[9]. Alonso, J., et al. (2024). Wireless sensor network for real-time monitoring of water quality in distribution systems. *Environmental Monitoring and Assessment*, 186(6), 3527–3538.
[10]. Amit, Y., & Gem, D. (2024). Shape quantization and recognition with randomized trees. *Neural Computation*, 9(7), 1545–1588.
[11]. Bui, X. N., et al. (2023). Prediction of water quality indices using machine learning techniques: A comparative study. *Science of the Total Environment*, 612, 156–168.
[12]. Chen, J., et al. (2023). A machine learning approach for water quality prediction. *Water Research*, 124, 29–38.
[13]. Gholizadeh, M., et al. (2023). A comprehensive review on water quality parameters and classification methods. *Environmental Reviews*, 24(3), 286–298.
[14]. Hassan, W. H. (2022). Current research on Internet of Things (IoT) security: A survey. *Computer Networks*, 148, 283–294.
[15]. Khatri, N., & Tyagi, S. (2022). Influences of natural and anthropogenic factors on surface and groundwater quality. *Environmental Science and Pollution Research*, 22(5), 3280–3302.
[16]. Li, Z., et al. (2021). Water quality prediction using multi-layer perceptron neural network. *Water Resources Management*, 33(5), 1857–1872.
[17]. Maier, H. R., & Dandy, G. C. (2021). Neural networks for prediction and forecasting of water resources variables. *Environmental Modelling & Software*, 15(1), 101–124.
[18]. Najah, A., et al. (2021). Application of artificial neural networks for water quality prediction. *Neural Computing and Applications*, 22(1), 187–201.
[19]. Palani, S., et al. (2020). Water quality forecasting using neural network model. *Water Research*, 42(15), 3930–3938.
[20]. Rani, M., et al. (2019). IoT-based smart water quality monitoring system. *International Journal of Recent Technology and Engineering*, 8(2), 2319–2324.
[21]. Zhang, Y., & Stanley, H. E. (2018). Forecasting water quality using data-driven models. *Physica A: Statistical Mechanics and its Applications*, 493, 58–67.