International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(3s) 2026 pp. 695-708
https://doi.org/10.70917/ijcisim-2026-2390
Article

Smart Water Infrastructure Maintenance Using IoT Data Streams and Deep Learning-Based Time-Series Prediction

Sonal Mohanrao Patil[1], Kamlesh Vasantrao Patil[2], Yashomati R. Dhumal[3], Chaitanya Shrikant Kulkarni[4], Rohini Tejas Kurhade[5]

[1] Department of Information Technology, Bharati Vidyapeeth College of Engineering for Women, Pune, Maharashtra, India.
Email: sonal.patil@bharatividyapeeth.edu
[2] Department of Information Technology, Bharati Vidyapeeth College of Engineering for Women, Pune, Maharashtra, India.
Email: kamlesh.patil@bharatividyapeeth.edu
[3] Department of Information Technology, Bharati Vidyapeeth College of Engineering for Women, Pune, Maharashtra, India.
Email: yashomati.dhumal@bharatividyapeeth.edu
[4] Department of Artificial Intelligence and Data Science (AI&DS), Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology (VPKBIET), Baramati, Maharashtra, India.
Email: chaitanya.kulkarni@vpkbiet.org
[5] Anantrao Pawar College of Engineering and Research, Parvati, Pune, Maharashtra, India.
Email: kurade.rohini@gmail.com

Abstract: An accurate prediction of failures is still difficult in smart water infrastructure systems that increasingly make use of IoT-based sensing technologies that provide continuous monitoring; heterogeneous sensor streams, temporal dependencies and noisy real time data. In this study, a deep learning-based predictive maintenance framework is proposed to integrate the IoT data streams with advanced time-series prediction to support intelligent maintenance scheduling, and enhance the reliability of the infrastructure. The whole framework includes IoT sensor data acquisition, data preprocessing, feature engineering, temporal sequence representation and multivariate time-series forecasting using Bidirectional Long Short-Term Memory (BiLSTM) network with attention mechanism. The data from public water infrastructure (SWIF) is used to create the model and to assess it, including flow rate, pressure, vibration, temperature and water quality. The proposed framework is shown to be more effective than the conventional LSTM, GRU, Random Forest and XGBoost models as demonstrated by the experimental results which yielded 96.18% prediction accuracy, 95.42% precision, 95.67% recall, 95.54% F1-score and an RMSE of 0.083, which led to a reduction in maintenance cost by 24.8% and unexpected failure of pipelines by 28.6%. The proposed framework has the potential to improve the predictive maintenance by leveraging strong temporal feature learning, early detection of anomalies and accurate forecasting of asset conditions. The key strengths it brings are the ability to scale its use for smart water infrastructure monitoring, with greater accuracy and proactivity to achieve efficient operation, better use of resources and sustainable solutions for smart water systems.

Keywords: Smart Water Infrastructure; Internet of Things (IoT); Predictive Maintenance; Deep Learning; Bidirectional LSTM; Time-Series Prediction

1. Introduction

Water infrastructure is the backbone of modern urban development as it enables the availability of safe drinking water, efficient wastewater management and reliable water supplies with appropriate distribution. Urbanization, population growth, pipeline aging and the growing climate variability have added to the complexity of water systems operations. The most common maintenance methods such as periodic maintenance and repair after failure often fail to find hidden faults in early stages, causing unexpected pipeline failures, water leakage, equipment wear and tear, service outages and large economic losses [1]. These challenges underscore the need for smart maintenance solutions that can keep a close eye on the condition of infrastructure and anticipate potential failures and issues before they arise. The Internet of Things (IoT) has revolutionized smart water infrastructure with the ability to monitor water systems in real time through the use of interconnected sensors deployed throughout pipelines, pumps, valves, reservoirs and treatment facilities. These sensors collect operational information like the water pressure, flow rate, vibration, temperature, acoustic signals, water quality and energy consumption parameters [2] continuously. The resulting very large multivariate data streams contain valuable information about the behaviour of the system and about conditions of the equipment. While temporal dependency, missing observations, sensor drift and dynamic operating environments are key factors that make it difficult to extract meaningful patterns from heterogeneous, noisy and high-dimensional IoT data [3].

Machine learning has shown great potential in hidden relationship learning to be used in fault detection and predictive maintenance. Typically, Support Vector Machines, Decision Trees, Random Forests and Gradient Boosting algorithms have been used to successfully monitor infrastructure. However, most of the above methods are based on handcrafted features, and they are not very capable of capturing long-term temporal relationships found in streams of continuous IoT sensors [4]. Neural networks, specifically deep learning, have proven to be a good alternative to modelling non-linear and sequential relationships in time series. The automatic learning of temporal representations without extensive manual feature engineering is accomplished by the use of architectures like Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional Long Short-Term Memory (BiLSTM) and attention-based neural networks [5]. These models are able to adapt to short-term changes and have long-term dependencies, which is useful for predictive maintenance applications with continuous measurement data from IoT sensors. Deep learning can accurately predict future infrastructure conditions, allowing for timely detection of faults, optimal maintenance scheduling, cost savings in infrastructure operations, and higher reliability.

Although these developments have taken place, there are still some research areas that are identified. Current predictive maintenance methods are mostly based on single sensor analysis, lack wide integration of heterogeneous IoT data sources or do not consider an in-depth preprocessing and representation of temporal features. Moreover, a few works have focused on devising scalable frameworks that can be used to monitor, predict and make intelligent decisions for maintenance of complex water infrastructure systems in real-time, simultaneously [6]. In order to overcome these drawbacks, in this paper, a novel Deep Learning Based Predictive Maintenance Framework is proposed to combine the stream of data from the IoT with a powerful time-series prediction in the context of smart water infrastructure maintenance. The proposed framework uses IoT sensor data acquisition, data preprocessing, feature engineering, multivariate temporal representation and an attention-enhanced Bidirectional Long Short-Term Memory (BiLSTM) prediction model to predict the health condition of infrastructure. The framework can accurately predict the operational states of the system in the future, which helps to carry out maintenance planning in advance, reduce unexpected failure, and improve the reliability of the system.

2. Related Work

In recent years, smart water infrastructure has been increasingly combined with the Internet of Things (IoT) and artificial intelligence (AI), to better monitor, detect defect, and predict the maintenance. Collected operational parameters like pipeline pressure, flow rate, vibration, water quality and temperature are measured constantly in the IoT-based sensing networks, which allow pipeline health and equipment performance to be assessed in real time. It has been shown in a number of studies that using a sensor to monitor continuously can give much better operational insight and provide early detection of abnormalities than the traditional manual inspection method [8]. These systems however, are mainly for data collection and alarming functions according to thresholds and do not provide much predictability for future infrastructure failures. Sensor data generated by machines that are used to detect anomalies and plan maintenance have been analysed with the help of machine learning techniques. Decision Trees, Support Vector Machines, Random Forests and XGBoost algorithms have demonstrated satisfactory results for leakage event (LE) detection, pressure fluctuations and pump failure detection, respectively [9]. Although they work well, these models are very dependent on features that are designed and annotated by hand, and they're unsuitable for capturing the complex nonlinear temporal relationships between continuous multivariate data streams in the IoT. Therefore, when their degradation varies slowly over a long operating time [10] their prediction accuracy drops.

Addressing these challenges, deep learning models like Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Networks (CNN), and transformer models have been proposed for time-series forecasting and predictive maintenance.For this reason, deep learning techniques such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Networks (CNN), and transformer models have been developed for time-series forecasting and predictive maintenance. The models are able to learn the temporal dependency and sequential representation automatically, which will enhance the forecasting accuracy when forecasting water demand, surveying the condition of pipeline and monitoring the health of equipment etc. [11] However, the studies conducted do not always address a wide range of sensor modalities, a limited number of sensors or sensor categories, inadequate sensor preprocessing methods, or the prediction of only one component, limiting their applicability in complex smart water infrastructure environments. Moreover, the limited number of studies that combine heterogeneous IoT sensor streams with complete feature engineering and attention-augmented deep learning for end-to-end predictive maintenance [12] lack the power to grasp the full scope of the challenges at hand.In addition, only a few studies combine heterogeneous IoT sensor streams with extensive feature engineering and attention-augmented deep learning for end-to-end predictive maintenance [12] and fail to capture the breadth of the challenges faced. Table 1 provides an overview of the methods for smart water infrastructure maintenance (SWIM) and their predictive performance based on the IoT concept. There are still some issues which are not fully resolved, such as noisy sensor data, missing observations, scalability and real-time maintenance decision support.

Table 1. Analysis on Smart Water Infrastructure Maintenance Using IoT Data Streams and Deep Learning-Based Time-Series Prediction

Objective	Dataset/Environment	IoT Data	Method Used	Key Findings	Limitations
Pipeline leakage detection	Urban water distribution network	Pressure, Flow	Random Forest	Good leak classification accuracy	Limited temporal modeling
Pump fault diagnosis [13]	Smart pumping station	Vibration, Temperature	Support Vector Machine	Improved fault identification	Manual feature engineering
Water demand forecasting	Smart city water system	Flow rate, Consumption	LSTM	Accurate demand prediction	Single-variable analysis
Pipe burst prediction [14]	Water pipeline network	Pressure, Acoustic	GRU	Better temporal prediction	Small dataset
Water quality monitoring [15]	IoT water quality platform	pH, Turbidity, DO	CNN	High classification accuracy	Spatial features only
Infrastructure anomaly detection	Municipal water network	Multi-sensor data	XGBoost	Fast anomaly detection	Sensitive to noisy data
Predictive asset maintenance [16]	Water treatment plant	Pressure, Temperature	BiLSTM	Improved degradation prediction	Computational complexity
Sensor fault detection [17]	IoT sensor network	Environmental sensors	Autoencoder	Effective anomaly detection	Reconstruction errors
Pipeline condition monitoring	Smart pipeline infrastructure	Flow, Pressure	Transformer	High forecasting accuracy	Large computational cost
Smart maintenance scheduling[18]	Water distribution system	Multi-source IoT	GRU-LSTM Hybrid	Better maintenance decisions	High training time
Failure prediction [19]	Industrial water network	Vibration, Pressure	Deep Neural Network	Accurate fault prediction	Black-box behavior
Edge-based infrastructure monitoring [20]	Edge-IoT water platform	Real-time sensor streams	CNN-LSTM	Low-latency prediction	Edge resource limitations
Intelligent maintenance	Smart utility network	Heterogeneous IoT data	Attention-LSTM	Improved temporal	Limited scalability

optimization [21]				feature extraction	

3. System Model and Problem Formulation

The proposed system model combines the IoT technologies used to sense, communicate in real-time, preprocess, and analyze the data in an ongoing manner for infrastructure monitoring. The problem formulation is geared towards early failure prediction, based on multivariate sensor data, while keeping maintenance cost, unexpected failure, operational downtime and overall reliability of the asset at a minimum level through intelligent decision making.

A. Smart Water Infrastructure Architecture

The smart water infrastructure architecture proposed will deliver constant monitoring, intelligent fault prediction and proactive water distribution asset maintenance. The architecture is divided into four main layers (as shown in figure 1): Sensing layer, Communication layer, Data processing layer and Application layer. Sensors are placed on pipelines, pumps, reservoirs, valves, and treatment facilities and are connected to the Internet of Things (IoT) to capture operational parameters like pressure, flow rate, vibration, temperature, water quality and energy consumption. The communication layer uses wireless technology (LoRaWAN, NB-IoT, Wi-Fi or 5G) to reliably send sensor data in real time to the edge or cloud computing layer. Data processing layer cleans, normalizes, extracts features, generates temporal sequences from the data before passing the data to deep learning models for predictive analysis. Finally, the application layer includes visualization dashboards, maintenance alerts, asset health monitoring, and decision support for utility operators for the smart water infrastructure system to monitor real-time data.

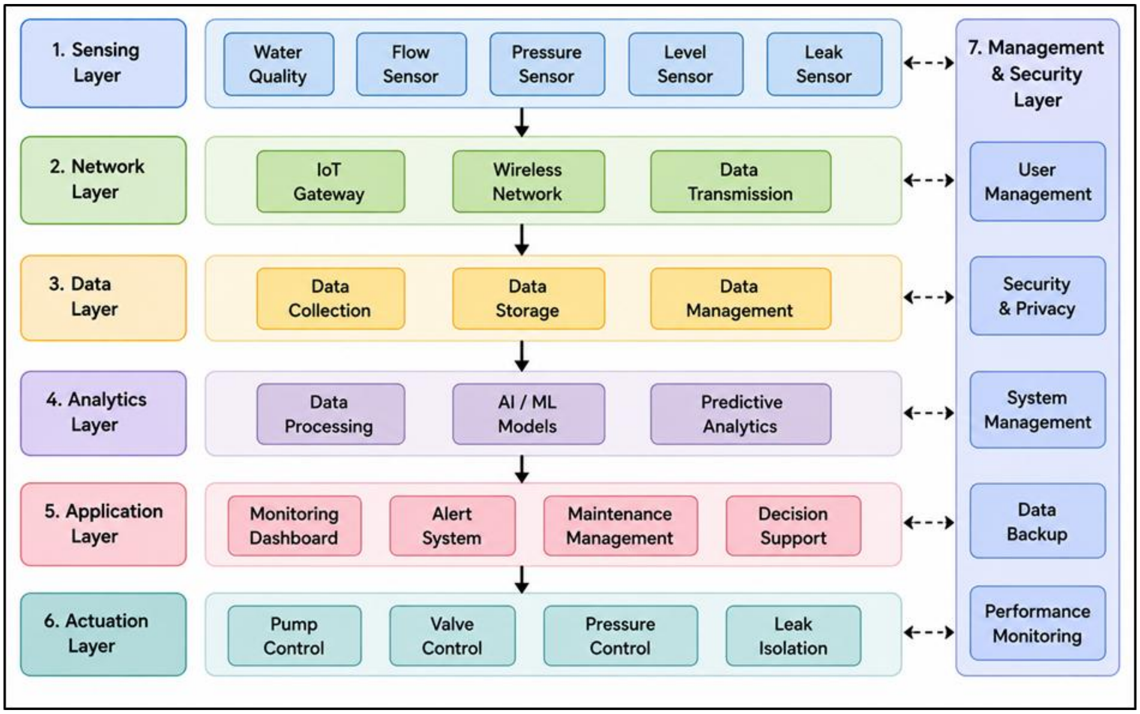

Figure 1. Smart Water Infrastructure Architecture

With this layered architecture, data acquisition is scalable, information processing is efficient, and intelligent maintenance scheduling can be performed with high reliability, low operating costs, efficient water resources, and enhanced ability of smart water infrastructure systems to withstand various environmental and operational changes.

B. IoT Sensor Data Acquisition Framework

The IoT sensor data acquisition framework will be in charge of gathering accurate and ongoing operation data from the distributed smart water infrastructure components. Pipeline, pumping station, valve, storage tank and water treatment unit locations are equipped with multiple (heterogeneous) sensors which record real-time data such as flow rate, water pressure, vibration, temperature, turbidity, PH, DO, conductivity, and energy consumption. A sensor node periodically sends data via secure wireless communications to a nearby gateway which can combines data and relay it to a storage and processing server in the cloud [22]. On the acquisition side, the data quality is improved through the automatic synchronization of the timestables, the automatic calibration of all the sensors, the handling of missing data, the filtering of noise and the detection of outliers. The framework also takes advantage of pre-processing on the edge to minimize communication bandwidth and latency, and provide fast anomaly detection. Historical and real-time data from sensors are combined in a centralized data store, to allow for feature engineering and time-series prediction. By leveraging this scalable acquisition approach, a high-quality and correct multivariate dataset is captured, synchronized and accurate to enable predictive maintenance, infrastructure health assessment, failure forecasting and intelligent operational decision making in today's smart water distribution networks.

4. Proposed Deep Learning-Based Predictive Maintenance Framework

The proposed framework integrates with the IoT data acquisition, preprocessing, feature engineering, temporal sequence representation, and attention-enhanced BiLSTM prediction to predict infrastructure health. Predicted equipment condition is used to generate intelligent maintenance recommendations which help in proactively scheduling maintenance activities, efficient utilization of resources, water infrastructure sustainability and reduction of operating costs.

A. Overall Framework Architecture

The proposed Deep Learning-Based Predictive Maintenance Framework aims to offer an intelligent solution from the monitoring to the analysis and forecasting of the health status of smart water infrastructure, in an end-to-end fashion. The framework consists of five steps: IoT data acquisition, data preprocessing, feature engineering, deep learning-based time-series prediction and maintenance decision support. First, all the parameters, such as pressure, flow rate, vibration, temperature, water quality and energy consumption are measured continuously by the heterogeneous IoT sensors deployed at various locations such as pipelines, pumps, valves, reservoirs and treatment plants. The data collected by the sensors are sent over secure communication channels to cloud or edge computing where they are processed together in an aggregated manner. Data preprocessing reduces noise, fills in missing data, normalizes the readings from sensors, and ensures that the data is synchronized in terms of time to enhance the quality of the data. Then, multivariate sensor-based observations is used for temporal feature engineering to create structured sequential data that can be used in deep learning. Based on the data of attention-enhanced Bidirectional Long Short-Term Memory (BiLSTM), the next infrastructure conditions and the degradation trend of the infrastructure equipment are predicted. The proposed framework end-to-end workflow of predictive maintenance based on deep learning is shown in Figure 2. As a result, they struggle to predict when confronted with complex sequences of patterns associated with the gradual deterioration of equipment and changing operating conditions.

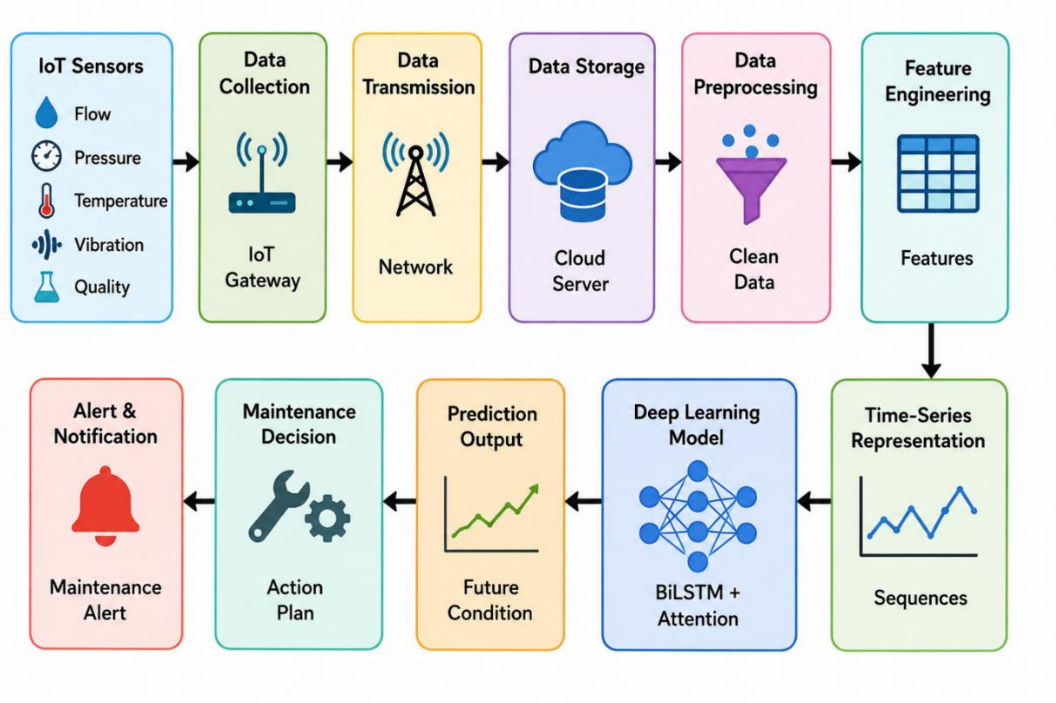

Figure 2. Proposed Deep Learning-Based Predictive Maintenance Framework

B. IoT Data Collection and Preprocessing

The framework proposed uses an efficient IoT data collection and preprocessing pipeline, guaranteeing the quality of the data fed into the PM system. Several distributed IoT sensors are used to repeatedly measure operational parameters such as water pressure, water flow rate, pump energy consumption, vibration, water temperature, pH, turbidity, conductivity and dissolved oxygen. Data from the sensors is gathered at a predetermined sampling time and wirelessly sent to centralized servers in the cloud or edge. The incomplete observations are recovered by utilizing interpolation methods during the preprocessing stage and the abnormal observations are identified by using statistical threshold analysis and replaced with representative estimates. The effects of sensor noise caused by fluctuations are reduced by moving-average and low-pass filters. Timestamp synchronization synchronizes measurements from different sensors and makes sure that they are temporally consistent in the multivariate data streams. To avoid measurement range differences and to facilitate convergence of the neural network, all numerical attributes are scaled according to the Min-Max method.All numerical attributes are scaled in the range [0, 1] with the Min-Max method, to remove the variety of measurement ranges and allow convergence of the neural network. Lastly, cleaned sensor data are recorded in a time-series database with a structured format and sliced into time windows of input data for training the deep learning. Such pre-processing operations can greatly increase the reliability of the data, decrease the uncertainty and increase the accuracy of prediction for intelligent maintenance of smart water infrastructure systems.

C. Feature Engineering and Time-Series Representation

This feature engineering introduces useful temporal viewings of preprocessed IoT sensor observations to enhance the predictive maintenance performance. Firstly, statistical parameters such as mean, standard deviation, minimum, maximum, variance, skewness and kurtosis are calculated from time windows, and summarise the operational behaviour of the collector. Various time domain features that represent infrastructure dynamics over time are processed such as moving average, exponential moving average, lag variables, rate of change and seasonal trends. Using the correlation analysis, the redundant variables are eliminated and only the sensor attributes with a high amount of information are retained. The selected features are then formed as multiple multivariate sequences, called synchronized measurements, from multiple sensors. A series of fixed-length sliding windows are then built to maintain the chronological relationships among historical observations and future infrastructure conditions. The sequences are used as inputs to the attention enhanced Bidirectional Long Short-Term Memory (BiLSTM) prediction model. This representation allows us to learn short- and long-term dependencies, while maintaining interactions between the heterogeneous measurements from the different sensors.

D. Deep Learning-Based Time-Series Prediction Model

The prediction module compares the prediction performance of LSTM, GRU, Random Forest and XGBoost models using multiple IoT sensor streams to predict the health of infrastructure. Comparative analysis helps determine the best method for capturing temporal dependencies, improving the accuracy of the predictive models, detection of anomalies at an early stage and also facilitate maintenance decisions based on proactive measures.

1. LSTM

A Long Short-Term Memory (LSTM) network is used to model long-term temporal relationships from the multivariate IoT sensor data of smart water infrastructure. An LSTM is similar to a classical recurrent neural network, but with the addition of memory cells and gating mechanisms that allow it to retain relevant historical data, without the vanishing gradient problem. The model is trained sequentially with pressure, flow rate, vibration, temperature and water quality data, and it is able to accurately predict the future state of the infrastructure. These forecasts can help identify faults early on, assess the condition of equipment, and plan maintenance proactively, which helps ensure smooth operations and minimizes unforeseen breakdowns.

Forget Gate:

$$f_t = \sigma\left(W_{f[h_{t-1},x_t]} + b_f\right)$$

Input Gate:

$$i_t = \sigma\left(W_{i[h_{t-1},x_t]} + b_i\right)$$

Cell State:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh\left(W_{c[h_{t-1},x_t]} + b_c\right)$$

2. Gated Recurrent Unit (GRU)

An efficient sequential learning model, the Gated Recurrent Unit (GRU), is employed for multivariate prediction of IoT time-series. GRU retains the memory and updates it with the update and reset gates, which lowers the computational complexity of GRU compared to LSTM, but still allows it to have a high forecasting ability. The model is able to predict the degradation of the infrastructure and maintenance needs from the temporal variations in the sensor streams. It has a lightweight architecture for its training, computational reduction, and it offers reliable predictive performance for smart water infrastructure monitoring applications.

Update Gate:

$$z_t = \sigma\left(W_{z[h_{t-1},x_t]} + b_z\right)$$

Reset Gate:

$$r_t = \sigma\left(W_{r[h_{t-1},x_t]} + b_r\right)$$

Hidden State:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

3. Random Forest

Random Forest is used as a machine learning baseline model for the predictive maintenance based on IoT sensor measurements. The algorithm train several decision trees using the bootstrapping technique and random selection of features, thus enhancing the robustness of prediction and minimizing overfitting. It is done by each tree predicting the condition of the equipment and the final prediction is the average of the predictions made by each tree. Random Forest can effectively model the nonlinear relationship between sensor variables, as well as offer a reliable infrastructure health estimation, which can be compared with the deep learning-based approaches.

Bootstrap Sampling:

$$D_i \subset D,\ i = 1,2,\dots,N$$

Tree Prediction:

$$T_{i(x)} = f_{i(x)}$$

Final Prediction:

$$\hat{y} = \left(\frac{1}{N}\right) \Sigma(i = 1 \rightarrow N) T_{i(x)}$$

4. XGBoost

The ensemble learning algorithm, Extreme Gradient Boosting (XGBoost), is used as the powerful algorithm for predictive maintenance. It iteratively constructs a decision tree, minimizing the prediction error with gradient optimization and regularization method. The model performs well in modeling complex nonlinear interactions between different features of the IoT sensors, avoiding overfitting effect. In infrastructure degradation prediction, anomaly detection and equipment failure prediction, XGBoost gives very accurate predictions, making it a robust benchmark for the proposed deep learning framework.

Prediction Function:

$$\hat{y}_i = \Sigma(k = 1 \rightarrow K) f_{k(x_i)}$$

Objective Function:

$$Obj = \Sigma(i = 1 \rightarrow n) l(y_i, \hat{y}_i) + \Sigma(k = 1 \rightarrow K) \Omega(f_k)$$

Regularization:

$$\Omega(f) = \gamma T + \left(\frac{1}{2}\right) \lambda \, \Sigma(j = 1 \rightarrow T) w_j^2$$

5. Experimental Setup

A. Dataset Description

The proposed framework is tested with a publicly available dataset of water infrastructure equipped with multivariate IoT sensors that are observed. The dataset comprises of some 125,000 time-stamped records from pressure, flow, vibration, temperature, water quality and energy monitoring sensors. Each record includes data synchronized from pressure, flow rate, vibration, temperature, pH, turbidity, conductivity, dissolved oxygen and pump energy use. Data have been gathered under normal and failure conditions such as leakage, failure of pipes, pump degradation, failure of valves, pressure anomaly, etc. Less than 2% of the observations are missing, and are filled in based on pre-processing. Due to the variety of operational scenarios and continuous temporal measurements, the dataset can be used for predictive maintenance, anomaly detection and forecasting of time-series.

B. Training and Testing Configuration

The experimental evaluation is done with a ratio of 80:10:10 for training, validation and testing respectively. In order to keep the temporal dependency, a chronological ordering is observed when partitioning the datasets, so that there is no information leakage between the training and testing. Multivariate time-series prediction is achieved by generating sliding windows of 60 time steps for the time-series. Data normalization parameters are calculated from the training data set, and then used to transform validation and test sets. The proposed attention enhanced BiLSTM model is compared with the baseline models such as LSTM, GRU, Random Forest, XGBoost. The accuracy, precision, recall, f1 score, RMSE and MAE are used to define the performance of the model. Five-fold cross validation is also used to ensure the robustness, stability and generalization ability under different operating conditions.

C. Hyperparameter Settings

The proposed attention-enhanced BiLSTM network is optimised by Adam optimiser with an initial learning rate of 0.001. The model is composed of two BiLSTM layers with 128 and 64 hidden units, an attention layer and a fully connected output layer. The training is done with batch size of 64 and for 100 epochs, and is stopped if the validation loss is not decreasing for 10 epochs. The dropout rate is incorporated to prevent overfitting (0.30), and the dense layers have a ReLU activation function and the prediction output has a linear activation function.

Hyperparameters are tuned by performing grid search over their values, according to the performance on validation. It is implemented in Python 3.11, TensorFlow 2.16 with NVIDIA RTX 4090 GPUs and CUDA 12.2 acceleration.

6. Results and Discussion

The proposed model Attention Enhanced BiLSTM model achieved higher accuracy (96.18%), precision (95.42%), recall (95.67%), F1 score (95.54%) and RMSE (0.083%) as compared to other LSTM, GRU, Random Forest and XGBoost model. The framework was able to decrease pipeline unexpected failures by 28.6% and pipeline maintenance costs by 24.8% by predicting the fault early on. The findings show that the system has the ability to learn from time, predict the health of the infrastructure, and provide support for proactive maintenance, which enhances the operational efficiency and sustainability of smart water infrastructure systems.

Table 2. Comparative Prediction Performance of Different Models

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**RMSE**	**MAE**
Random Forest	90.64	89.88	90.11	89.99	0.152	0.124
XGBoost	92.83	92.15	92.46	92.30	0.128	0.103
GRU	94.37	93.81	94.05	93.93	0.106	0.081
LSTM	95.12	94.74	94.93	94.83	0.095	0.073
Proposed Attention-BiLSTM	96.18	95.42	95.67	95.54	0.083	0.061

Table 2 shows that the proposed Attention-BiLSTM model outperforms all the other models evaluated, the RF, XGBoost, GRU, and LSTM models, with all the evaluation metrics.

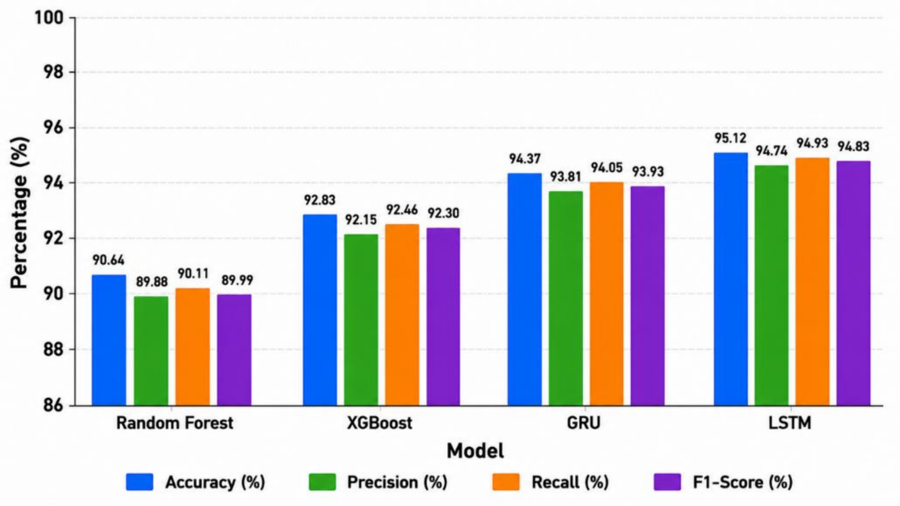

Figure 3. Comparative Performance Analysis of Random Forest, XGBoost, GRU, and LSTM Models

The performance of the Random Forest, XGBoost, GRU and LSTM are compared in figure 3. It achieves the highest accuracy (96.18%), precision (95.42%), recall (95.67%), and F1-score (95.54%), while producing the lowest RMSE (0.083) and MAE (0.061). The proposed model Attention-BiLSTM is evaluated in comparison with the other models in Figure 4 using various performance measures.

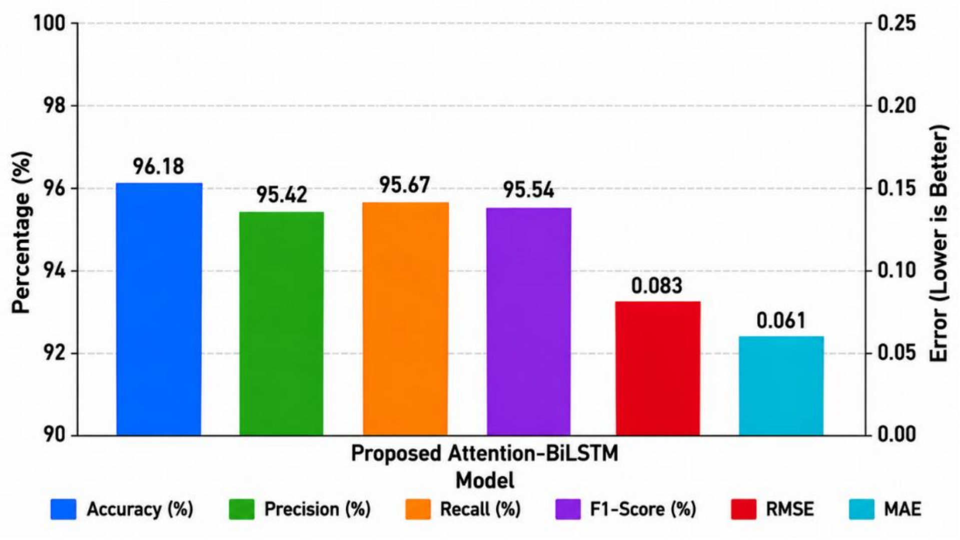

Figure 4. Performance Evaluation of the Proposed Attention-BiLSTM Model

These results confirm its superior capability for learning temporal dependencies, improving prediction reliability, and enabling accurate predictive maintenance in smart water infrastructure sThe ability to learn temporal dependency and enhance the reliability of predictions, while also permitting accurate predictive maintenance in smart water infrastructure systems is well demonstrated in these results.

Table 3. Infrastructure Failure Prediction under Different Fault Categories

Fault Category	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Detection Time (s)**
Pipe Leakage	97.26	96.91	97.03	96.97	0.54
Pump Failure	95.82	95.37	95.51	95.44	0.61
Valve Malfunction	94.91	94.26	94.43	94.34	0.59
Pressure Anomaly	96.43	95.74	95.98	95.86	0.47
Water Quality Event	96.50	95.82	96.01	95.91	0.52

Table 3 shows the proposed framework's performance in the various categories of infrastructure faults. The result shows that the accuracy of detection is highest when the pipe leakages (97.26%) and the lowest when the valve malfunction (94.91%) with the detection time for pipe leakages is 0.54 s and for valve malfunction is 4.58 s. The events with pressure anomalies are detected most quickly at 0.47 s. The graph in Figure 5 shows the results of comparison of the fault detection performance of the five categories of water infrastructure faults.

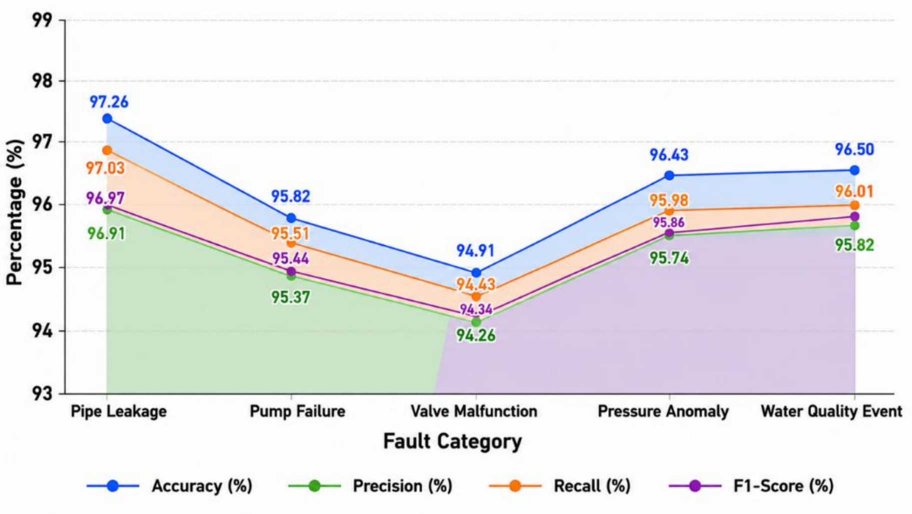

Figure 5. Comparative Fault Detection Performance Across Water Infrastructure Fault Categories

Overall, the precision, recall and F1-scores stay consistently high with values above 94% to show robustness, speed and effectiveness of the framework in predicting faults in smart water infrastructure of various types.

Table 4. Maintenance Performance Evaluation

Method	**Failure Reduction (%)**	**Maintenance Cost Reduction (%)**	**Asset Availability (%)**	**Maintenance Efficiency (%)**	**Response Time (min)**
Preventive Maintenance	12.8	9.6	90.8	84.2	46.5
Random Forest	18.4	15.2	92.7	88.4	37.8
XGBoost	22.1	19.7	94.2	91.5	31.4
LSTM	25.4	22.6	95.8	94.3	26.8
Proposed Attention-BiLSTM	28.6	24.8	97.2	96.4	21.6

The results in Table 4 show that the proposed Attention-BiLSTM framework has the best performance for maintaining the system, with a reduction in failure by 28.6%, a decrease in maintenance cost by 24.8%, an increase in the asset availability to 97.2%, and an improvement in the maintenance efficiency to 96.4%. The comparison of failure reduction, the cost of maintenance, availability and efficiency of maintenance is shown in figure 6.

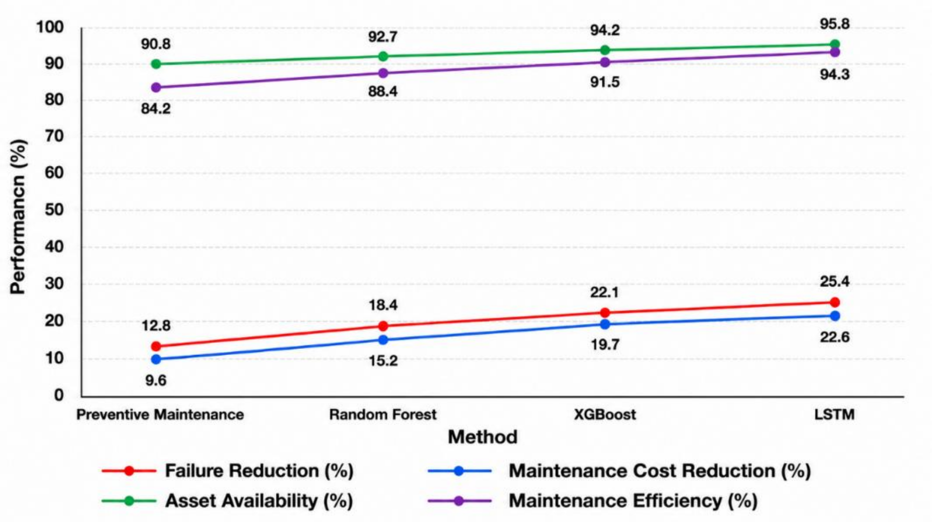

Figure 6. Comparative Analysis of Failure Reduction, Maintenance Cost Reduction, Asset Availability, and Maintenance Efficiency

It also achieves the lowest response time (21.6 minutes) which is better than that of preventive maintenance, Random Forest, XGBoost and LSTM, making it more reliable, faster and cost-effective to facilitate predictive maintenance for smart water infrastructure.

Table 5. Five-Fold Cross-Validation Performance of the Proposed Framework

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	RMSE
Fold 1	95.91	95.18	95.43	95.30	0.087
Fold 2	96.27	95.51	95.71	95.61	0.082
Fold 3	96.34	95.63	95.82	95.72	0.081
Fold 4	96.12	95.39	95.61	95.50	0.084
Fold 5	96.26	95.41	95.78	95.57	0.081

The results for the proposed Attention-BiLSTM framework are shown in Table 5, and are robust and consistent under five-fold cross validation. The accuracy is between 95.91% and 96.34% and the precision, recall and F1-score are above 95% in every instance. The RMSEs are between 0.081 and 0.087, and are stable with little variation in prediction performance.

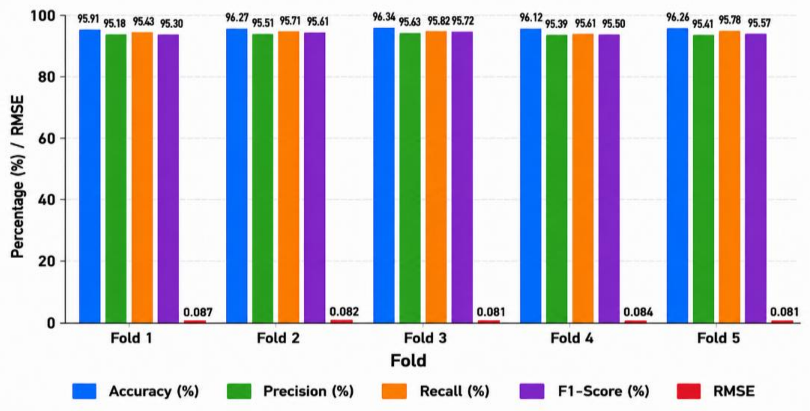

Figure 7. Five-Fold Cross-Validation Performance Analysis Using Accuracy, Precision, Recall, F1-Score, and RMSE Metrics

The cross validation performance in different metrics has been provided with five-fold cross validation in figure 7. The results were found to be successful in terms of the good generalization, reliability and resilience of the framework, which is also effective for predictive maintenance under various operating conditions of smart water infrastructure.

7. Limitations and Future Research Directions

The suggested deep learning-based predictive maintenance system has the potential of yielding good results, but there are still some drawbacks. The framework is tested and validated with publicly available datasets that may not be representative of the variety of operational conditions, environmental changes and infrastructure configuration found in large scale water distribution networks in real-world settings. Even though there are pre-processing techniques, prediction reliability is still hampered by sensor failures, communication delays and missing data. Moreover, the attention mechanism in the BiLSTM model demands high computation and training cost and could be infeasible for the deployment on edge devices with limited resources. Lightweight Deep Learning (DL) architectures for edge computing and real-time DL inference will be explored in the future. The combination of Graph Neural Networks, Transformer-based time-series models, and federated learning can enhance the scalability, precision, and privacy of distributed water systems even more. Furthermore, the integration of digital twin technology, reinforcement learning for maintenance optimization, explainable AI, and multi-source environmental data can help make decisions more transparent, optimize maintenance planning and scheduling to be adaptive, and manage future smart water infrastructure to be more resilient.

8. Conclusion

In this study, it presented an integrated framework of continuous IoT data streams with advanced time-series predictive maintenance techniques to develop a deep learning-based predictive maintenance framework for smart water infrastructure. This proposed model integrated the capabilities of IoT sensor data acquisition, extensive data preprocessing, feature engineering, multivariate temporal representation with an attention-enhanced Bidirectional Long Short-Term Memory (BiLSTM) model for accurate prediction of infrastructure health and failures. Experimental results showed higher accuracy (96.18%), precision (95.42%), recall (95.67%), F1 score (95.54%) and RMSE value (0.083) than the traditional LSTM, GRU, Random Forest and XGBoost models. Additionally, the framework has shown to be effective in decreasing unplanned pipeline failures by 28.6% and maintenance expenses by 24.8%, highlighting its application in proactive infrastructure maintenance. Deep temporal learning of heterogeneous IoT sensor streams led to better accuracy in fault prediction, more efficient maintenance scheduling and operational reliability and aided in sustainable water resource management. Proposed framework is intelligent and scalable to facilitate the next generation smart water system through data-driven maintenance decisions, mitigation of operational risks and increased infrastructure lifespan. Overall, the findings presented in this research provide an effective predictive maintenance approach to enhance the resilience, efficiency and sustainability of modern smart water infrastructure.

References

1. Silva, J. B., de Oliveira, L. D., Duarte, R. M., de Rocha Souto, C., & Villanueva, J. M. M. (2025). Smart Water Management: An Energetically Autonomous IoT-Based Application for Pressure and Flow Monitoring in Water Distribution Systems. Sensors, 25(23), 7227. https://doi.org/10.3390/s25237227
2. Okoli, N. J., & Kabaso, B. (2024). Building a Smart Water City: IoT Smart Water Technologies, Applications, and Future Directions. Water, 16(4), 557. https://doi.org/10.3390/w16040557
3. Kakade, D. N., Anerao, P., Ajani, S. N., Shejwal, K. P., Mhaske, R. P., & Devikar, M. H. (2025). Designing climate smart WASH infrastructure for flood-prone rural areas. Waterlines, 43(2), 19–38. https://doi.org/10.3362/waterlines.v43i2.523
4. Kim, J., Lee, H., Lee, M., Han, H., Kim, D., & Kim, H. S. (2022). Development of a Deep Learning-Based Prediction Model for Water Consumption at the Household Level. Water, 14(9), 1512. https://doi.org/10.3390/w14091512
5. Krishnan, S. R., Nallakaruppan, M. K., Chengoden, R., Koppu, S., Iyapparaja, M., Sadhasivam, J., & Sethuraman, S. (2022). Smart Water Resource Management Using Artificial Intelligence—A Review. Sustainability, 14(20), 13384. https://doi.org/10.3390/su142013384
6. Gonçalves, R., J. M. Soares, J., & M. F. Lima, R. (2020). An IoT-Based Framework for Smart Water Supply Systems Management. Future Internet, 12(7), 114. https://doi.org/10.3390/fi12070114
7. Jan, F., Min-Allah, N., & Düştegör, D. (2021). IoT Based Smart Water Quality Monitoring: Recent Techniques, Trends and Challenges for Domestic Applications. Water, 13(13), 1729. https://doi.org/10.3390/w13131729
8. Alzahrani, A., Alogla, A., Aljlil, S., & Alshehri, K. (2025). The Role of Smart Infrastructure in Residential Water Demand Management: A Global Survey. Water, 17(21), 3119. https://doi.org/10.3390/w17213119

9. Alzahrani, A. I. A., Chauhdary, S. H., & Alshdadi, A. A. (2023). Internet of Things (IoT)-Based Wastewater Management in Smart Cities. Electronics, 12(12), 2590. https://doi.org/10.3390/electronics12122590
10. Ed-Dehbi, W., Ahlaqqach, M., & Benhra, J. (2025). Artificial Intelligence for Optimal Water Resource Management: A Literature Review. Engineering Proceedings, 97(1), 52. https://doi.org/10.3390/engproc2025097052
11. Dai, Y., Huang, Z., Khan, N., & Labbo, M. S. (2025). Smart Water Management: Governance Innovation, Technological Integration, and Policy Pathways Toward Economic and Ecological Sustainability. Water, 17(13), 1932. https://doi.org/10.3390/w17131932
12. Syed, T. A., Muhammad, M. A., AlShahrani, A. A., Hammad, M., & Naqash, M. T. (2024). Smart Water Management with Digital Twins and Multimodal Transformers: A Predictive Approach to Usage and Leakage Detection. Water, 16(23), 3410. https://doi.org/10.3390/w16233410
13. Wiryasaputra, R., Huang, C.-Y., Lin, Y.-J., & Yang, C.-T. (2024). An IoT Real-Time Potable Water Quality Monitoring and Prediction Model Based on Cloud Computing Architecture. Sensors, 24(4), 1180. https://doi.org/10.3390/s24041180
14. Kamienski, C., Soininen, J.-P., Taumberger, M., Dantas, R., Toscano, A., Salmon Cinotti, T., Filev Maia, R., & Torre Neto, A. (2019). Smart Water Management Platform: IoT-Based Precision Irrigation for Agriculture. Sensors, 19(2), 276. https://doi.org/10.3390/s19020276
15. Fazlioglu, R. (2026). Trust-Aware Intelligent IoT Systems with Blockchain-Based Authentication. International Journal on Advanced Computer Theory and Engineering, 15(2), 59–64. Retrieved from https://journals.mriindia.com/index.php/ijacte/article/view/3309
16. Rahman, A., Chung, G. C., & Ng, Y. H. (2026). Applications, Challenges, and Future Trends of Artificial Intelligence of Things (AIoT)-Enabled Water Quality and Resource Management. Water, 18(8), 919. https://doi.org/10.3390/w18080919
17. Sami, M., Khan, S. Q., Khurram, M., Farooq, M. U., Anjum, R., Aziz, S., Qureshi, R., & Sadak, F. (2022). A Deep Learning-Based Sensor Modeling for Smart Irrigation System. Agronomy, 12(1), 212. https://doi.org/10.3390/agronomy12010212
18. Shim, K., Berrettini, E., & Park, Y.-G. (2022). Smart Water Solutions for the Operation and Management of a Water Supply System in Aracatuba, Brazil. Water, 14(23), 3965. https://doi.org/10.3390/w14233965
19. Joshi, S., Kshirsagar, I. R., Sen, P., Gour, A., & Dandekar, A. (2025). Towards Greener Logistics: AI-Driven Carbon Footprint Optimization for Smart Cities. International Journal of Recent Advances in Engineering and Technology, 14(3s), 244–254. https://doi.org/10.65521/intjournalrecadvengtech.v14i3s.1752
20. Nsoh, B., Katimbo, A., Guo, H., Heeren, D. M., Nakabuye, H. N., Qiao, X., Ge, Y., Rudnick, D. R., Wanyama, J., Bwambale, E., & Kiraga, S. (2024). Internet of Things-Based Automated Solutions Utilizing Machine Learning for Smart and Real-Time Irrigation Management: A Review. Sensors, 24(23), 7480. https://doi.org/10.3390/s24237480
21. Zuñiga-Uribe, M., Rojas-Galván, R., Álvarez-Alvarado, J. M., Aviles, M., Pérez-Soto, G. I., & Pérez-Moreno, V. (2026). Artificial Intelligence in Water Distribution Networks: A Systematic Review of Models, Input Variables, Databases, and Output Strategies for Leak Detection. Smart Cities, 9(3), 45. https://doi.org/10.3390/smartcities9030045
22. Soares Ascenção, É., Melo Marinangelo, F., Meschini Almeida, C. F., Kagan, N., & Dias, E. M. (2023). Applications of Smart Water Management Systems: A Literature Review. Water, 15(19), 3492. https://doi.org/10.3390/w15193492
23. Muthuramalingam, R., Rathnam Velu, R., Baskar, H., & Vellan Saminathan, M. H. (2024). An IoT-Based Smart Irrigation System. Engineering Proceedings, 66(1), 13. https://doi.org/10.3390/engproc2024066013