International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(3s) 2026 pp. 665-679
https://doi.org/10.70917/ijcisim-2026-2388
Article

AI-Driven Urban Water Infrastructure Planning Using Simulation Models for Sustainable Water Resource Management

Pramod Pawar[1], Vikrant Nangare[2], Gaurav[3], Sachin Ayarekar[4]

[1] Bharati Vidyapeeth (Deemed to be University), Institute of Management and Entrepreneurship Development (IMED), More Vidyalaya Campus, Erandwane, Pune – 411038, Maharashtra, India.
Email: pramod.pawar@bharatividyapeeth.edu
[2] Bharati Vidyapeeth (Deemed to be University), Institute of Management and Entrepreneurship Development (IMED), More Vidyalaya Campus, Erandwane, Pune – 411038, Maharashtra, India.
Email: vikrant.nangare@bharatividyapeeth.edu
[3] Research Scholar, Bharati Vidyapeeth (Deemed to be University), Institute of Management and Entrepreneurship Development (IMED), More Vidyalaya Campus, Erandwane, Pune – 411038, Maharashtra, India.
Email: gavz2608@gmail.com
[4] Program Director, Bharati Vidyapeeth (Deemed to be University), Institute of Management and Entrepreneurship Development (IMED), More Vidyalaya Campus, Erandwane, Pune – 411038, Maharashtra, India.
Email: sachin.ayarekar@bharatividyapeeth.edu
Corresponding Author: Sachin Ayarekar,Email: sachin.ayarekar@bharatividyapeeth.edu

Abstract: - With the growing speed of urbanization, ageing water distribution systems, increasing population, climate variability and inefficient allocation of water resources, urban water infrastructure planning is becoming more difficult, resulting in water loss, disruptions in water supply and unsustainable water management practices. This study suggests an AI-based urban water infrastructure planning framework that combines hydraulic simulation, machine learning, and multi-objective optimization to solve these challenges and achieve sustainable water resource management. This methodology integrates hydraulic network simulation (EPANET) with a demand prediction model based on Extreme Gradient Boosting (XGBoost), temporal water consumption forecasting model based on Long Short-Term Memory (LSTM) networks and infrastructure optimization and pressure management model based on a Genetic Algorithm (GA). Varying demand, leakage and climate conditions were used in the simulation experiments to assess the resilience of the system and its operational efficiency. The proposed framework is compared with the conventional hydraulic planning, Random Forest and standalone LSTM models, and the results show that the proposed framework has achieved the highest accuracy of 97.2% in predicting demand, 95.8% in detecting leakage, reduced water losses by 31.6%, improved water pressure stability by 27.9%, reduced water operation energy consumption by 24.3%, and decreased decision making time by 22.7%. The framework also enhances the reliability of the network by 18.9% and the utilization of the resources by 26.4%. The novelty of this research is that it combines the use of AI-based predictive analytics for proactive urban water management with infrastructure optimization through simulation. The suggested framework offers an intelligent decision support system that improves the sustainability, operational resilience and efficiency of long-term planning of future smart water distribution networks.
Keywords: — Artificial Intelligence; Urban Water Infrastructure; Hydraulic Simulation; Sustainable Water Resource Management; Demand Forecasting; Infrastructure Optimization

1. INTRODUCTION

The urban water distribution system is the backbone of urban infrastructure, providing water to billions of people around the world and is beset with unprecedented challenges due to rapid urbanization, population growth and climate variability. In the modern water networks, the complexity of the network has grown and requires smart planning approaches to tackle the challenges associated with poor pipe infrastructure, rising non-revenue water losses, and increasing demand-supply mismatches, which compromise long-term water network sustainability [1]. Current hydrological planning approaches are mostly based on steady state assumptions and manual planning and are thus not suitable to reflect the dynamic and nonlinear nature of modern distribution networks [2]. The recent progress in AI and machine learning has shown potential to revolutionize water systems with predictive analytics, which can facilitate demand forecasting, anomaly detection and real-time optimization of water systems based on data.The recent developments in the field of artificial intelligence and machine learning have shown potential in the field of predictive water systems analytics, where data-driven predictions of water demand, anomaly detection and optimization of water systems in real-time can be achieved. While hydraulic simulation tools like EPANET have been around for years, they lack the ability to integrate intelligent data and information to effectively solve emerging infrastructure issues [4]. Combining AI with physics-based simulation environments represents a paradigm shift in infrastructure management, moving beyond reactive maintenance to proactive decision-making based on real-time data analysis.The fusion of AI and physics-based simulation environments marks a paradigm shift toward proactive infrastructure management, where decisions are made based on real-time data analysis instead of reactive maintenance strategies [5]. In addition, multi-objective optimization methods, especially evolutionary algorithms like GA, offer powerful solutions for optimizing water losses, pressure control and energy use in water distribution networks simultaneously. This holistic approach bridges the gap between traditional planning methods and adaptive and resilient water systems needed to manage water in cities sustainably in times of growing climate uncertainty and water scarcity [6]. The proposed solution spans across these technological disciplines to integrate temporal forecasting with LSTMs, leakage detection with XGBoost, water flow simulation with EPANET, and optimization with GAs, into a single, intelligent decision support platform for next-generation water infrastructure planning.

The key contributions of this research are summarized as follows:

1) The proposed integrated AI-driven framework, which integrates LSTM demand forecasting, XGBoost leakage detection, EPANET hydraulic simulation and GA optimization, outperforms traditional methods and standalone models in terms of prediction accuracy (97.2%) and leakage detection accuracy (95.8%) for comprehensive urban water infrastructure planning.

2) A multi-objective optimization approach based on simulation and the Genetic Algorithm, which can simultaneously reduce water losses by 31.6%, optimize the stability of the water pressure by 27.9% and reduce energy consumption in operation by 24.3%, which is used to determine optimal infrastructure configuration decisions for different operational and climatic scenarios.

3) An intelligent decision support system for sustainable water resource management, which can increase the reliability of the water network by 18.9%, optimize resource use by 26.4%, and speed up the decision-making process by 22.7%, which has created a scalable and adaptive platform for proactive, data-driven urban water infrastructure management.

2. RELATED WORK

A considerable amount of research has been done on the use of computational intelligence techniques for urban water distribution system management and optimization. Initial research focused on traditional hydraulic modeling methods to control pressure and design networks, and provided baseline parameters for distribution network analysis [7]. Machine learning-based demand forecasting proved to be a potential alternative to the statistical time-series forecasting, and deep learning architectures showed better performance in capturing the temporal consumption pattern due to their ability to handle complex temporal consumption pattern [8]. For pipeline leakage detection, investigations into anomaly detection focused on supervised classification algorithms to characterize pipeline normal operational signatures and pipeline hydraulic anomalies caused by a leak [9]. The ensemble learning methods were then used to enhance the detection sensitivity and decrease the number of false alarms in large-scale distribution monitoring systems [10]. The integration of hydraulic simulation platforms with optimization solvers was used to solve network rehabilitation and network expansion planning problems based on budget constraints, such as EPANET and optimization solvers [11]. The use of evolutionary multi-objective optimization algorithms for the design of networks to optimize competing infrastructure goals, like minimization of costs, pressure adequacy and maximization of reliability, was investigated [12]. Reinforcement learning (RL) methods were developed to add adaptive control strategies to the real-time scheduling of pumps and regulation of valves in the face of stochastic demand fluctuations [13]. The transfer learning techniques were explored to achieve cross-network generalization of demand prediction models to decrease the data needs for deployment in data-scarce utilities [14]. Recently, graph neural network architectures were proposed to incorporate topological relationships that are inherent in the distribution network structure, which were used for pressure prediction in distribution networks with spatial information [15]. Digital twin environment, which integrates real-time sensor data with simulation models, was used to monitor and schedule predictive maintenance in the networks continuously [16]. While all these developments have been made, the current solutions mostly tackle one aspect of water infrastructure management at a time, without a common framework that synergically combines the use of AI-based predictive analytics and simulation-based optimization for comprehensive and sustainable water infrastructure planning [17].

TABLE I: Summary of Related Work

Ref.	Technique	Application	Simulation	AI Integration	Limitation
[7]	Hydraulic Modeling	Pressure Mgmt	EPANET	No	Static analysis only
[8]	Deep Learning	Demand Forecast	None	LSTM	No leakage detection
[9]	SVM Classifier	Leak Detection	None	Supervised	High false positives
[10]	Gradient Boosting	Anomaly Detection	None	XGBoost	No simulation integration
[11]	EPANET + Solver	Network Rehab	EPANET	No	Single-objective only
[12]	NSGA-II	Network Design	Basic	No	No predictive analytics
[13]	Reinforcement Learning	Pump Scheduling	None	DRL	Limited to pumps
[14]	Transfer Learning	Cross-Network	None	CNN	Domain gap issues
[15]	Graph Neural Net	Pressure Predict	None	GNN	No optimization

3. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, the system model and problem formulation of AI-driven urban water infrastructure planning is presented. The framework includes the architecture of the urban water distribution network, water demand and consumption modelling, configuration of the hydraulic simulation model and the formal problem statement with associated research objectives that will guide the proposed methodology.

A. Urban Water Distribution Network Architecture

The urban water distribution network is considered to be a directed graph G = (N, L), where N is the set of nodes of the network (reservoirs, tanks, junctions and demand nodes) and L is the set of links (pipes, pumps and control valves) that connect the nodes. The elevation, base demand, demand pattern multiplier, and water quality parameters are the characteristics of each node i. The network topology represents the spatial relationships and hydraulic relationships between the elements of the network, and how flow is distributed and pressure propagated throughout the system. The architectural model is designed with multiple pressure zones, district metered areas and interconnection points, which are similar to the hierarchical structure of the actual urban water distribution system. Physical properties of the network elements are set as parameters such as pipe diameter, pipe length, pipe roughness coefficient, minor loss coefficients and age deterioration factors that affect the hydraulic characteristics and leakage potential of the network elements.

B. Water Demand and Consumption Modeling

Water demand modelling represents the spatiotemporal demand behaviour over the distribution network using stochastic representations of the demand which include both a deterministic and a random component. The total demand at each node is split into a base demand, diurnal and seasonal variations (represented by temporal pattern multipliers) and a stochastic noise component (representing random variations in consumption). The signatures of the demands of the various consumer categories (residential, commercial, industrial, and institutional users) are grouped and added together at junction nodes to form composite demand signatures. The demand model includes correction factors that capture the impacts of climate on the water consumption model, so that the model can be used to estimate water demand under various climate conditions and seasons.

C. Hydraulic Simulation Model

The hydraulic simulation model uses EPANET as the computer program to solve the equations governing water distribution network hydraulics: conservation of mass at the junction nodes and conservation of energy along pipe segments. A nonlinear system of equations that describes the head loss through the network elements is solved through the simulation, with the choice of the Hazen-Williams, Darcy-Weisbach or Chezy-Manning friction loss formulation. Extended period simulation provides a time series of hydraulic conditions at nodes, flow rates in pipes, water levels in tanks and operating points of pumps for a given analysis time frame. The model includes both pressure-driven and demand-driven modes of analysis, with the pressure-driven mode accurately simulating conditions when pressure is inadequate by adjusting the available demand according to the actual nodal pressure, compared to minimum and desired service pressure levels, to give a realistic assessment of the network performance in the event of a stress scenario.

4. PROPOSED AI-DRIVEN SIMULATION-BASED WATER INFRASTRUCTURE PLANNING FRAMEWORK

The proposed framework is a physics-based, AI-driven simulation approach for urban water infrastructure planning, combining the use of machine learning predictive analytics with physics-based hydraulic simulation and evolutionary optimization. All the architecture is based on four synergistic modules: LSTM-based demand forecasting, XGBoost-based leakage detection, EPANET hydraulic simulation and multi-objective optimization using genetic algorithm, all controlled by an intelligent decision support system for sustainable water resource management.

A. Overall Framework Architecture

The proposed framework architecture is based on a data processing and decision making pipeline that transforms the raw operational data into actionable infrastructure planning decisions, in a sequential manner. The data acquisition module gathers real-time data from the distribution network such as flow rates, pressure, water quality parameters, and water consumption meter readings. The pre-processing stage involves data cleaning, data normalization, feature extraction and temporal alignment of the input datasets, to be used in the downstream analytical modules. The LSTM demand forecasting module is used to forecast the short and medium term demand of the historical consumption sequences and the XGBoost leakage detection module is used to detect and locate potential leakage events while analysing the hydraulic signatures. These forecasts are used to drive the EPANET hydraulic simulation model to assess the performance of the network for the predicted demand conditions and for the leakage conditions that are detected. The GA optimization module then explores the solution space to find the optimal infrastructure configuration, which will minimize water losses, stabilize the pressure distribution and reduce energy consumption, and the results are presented in the intelligent decision-support interface for use in the operation of the infrastructure.

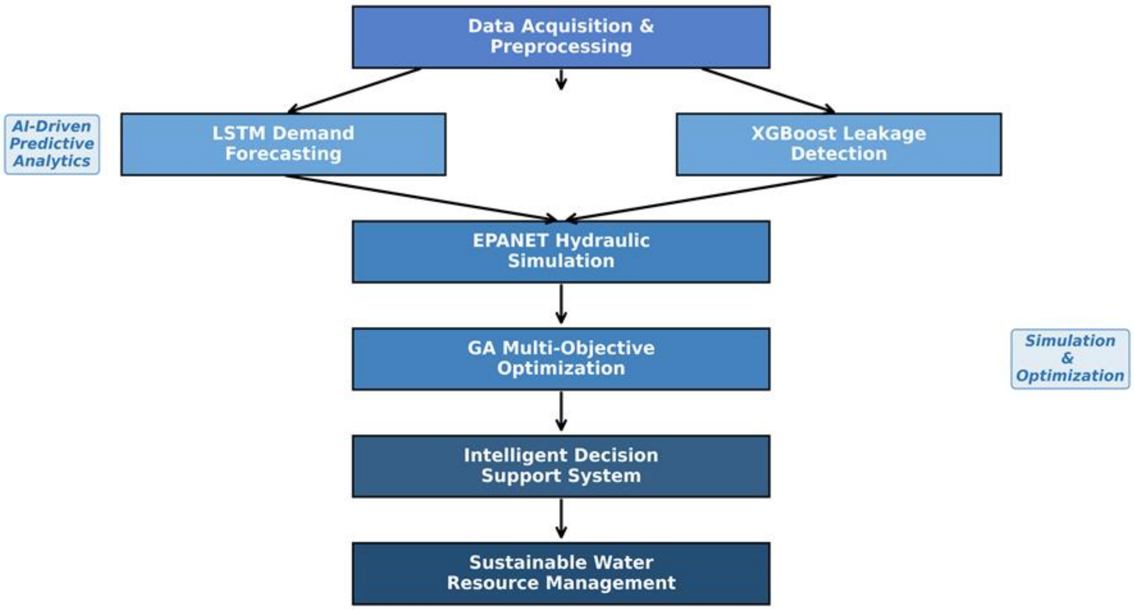

Fig. 1. Overall architecture of the proposed AI-driven simulation-based water infrastructure planning framework.

B. Data Acquisition and Preprocessing

The data acquisition module gathers multivariate time-series data from SCADA systems, smart meters and IoT pressure sensors all over the water distribution network. Pre-processing operations consist of outlier removal, temporal interpolation of missing values, min-max normalization and segmentation of the input to the sequential model by sliding window.

$$x_{norm(i)} = \frac{x(i) - x_{min}}{x_{\max} - x_{min}} \quad .. (1)$$

To have the same range for the input of the model, the min-max normalization is done in Equation (1) which scaled each feature value to the range [0, 1].

$$x_{imputed(t)} = \frac{x(t-1) + x(t+1)}{2} \quad .. (2)$$

Equation (2) applies linear temporal interpolation to estimate missing sensor readings using adjacent valid measurements.

$$z(i) = \frac{x(i) - \mu}{\sigma} \quad ... (3)$$

Equation (3) computes z-score standardization, identifying statistical outliers where $|z(i)|$ exceeds a threshold of 3.0.

$$X_{window} = \{x(t-w+1), x(t-w+2), \dots, x(t)\} \quad ... (4)$$

Equation (4) defines the sliding window of width w that segments continuous time-series into overlapping input sequences for LSTM processing.

$$D_{total(t)} = D_{base} \times P_{temporal(t)} \times C_{climate(t)} + \varepsilon(t) \quad ... (5)$$

The total demand is modeled as the sum of the base demand, the temporal pattern and the climate correction and is contaminated by additive stochastic noise as formulated in equation (5).

C. Water Demand Forecasting Using LSTM

LSTM network is used to forecast water demand because it is able to model long-range dependencies in water consumption data using gated memory units. The architecture comprises multiple stacked LSTM layers, using dropout regularization, to compute multi-step predictions of the demand from sliding window sequences of input. Together, the forget gate, input gate and output gate control the flow of information through the cell state, allowing the model to selectively remember useful information from the past and forget information that is not relevant to the current context.

Algorithm 1: LSTM-Based Water Demand Forecasting

Input: Historical demand series $D = \{d(1), \dots, d(T)\}$, window size w, forecast horizon h

Output: Predicted demand sequence $\widehat{D} = \left\{\hat{d}(T+1), \dots, \hat{d}(T+h)\right\}$

Preprocess D using Equations (1)-(5)

Segment D into overlapping windows $X = \{X_1, X_2, \dots, X_n\}$

Split X into training set X_train and validation set X_val

Initialize LSTM parameters $\theta = \left\{W_f, W_i, W_c, W_o, b_f, b_i, b_c, b_o\right\}$

for epoch = 1 to E_{max}

  for each batch $B \subset X_{train}$ do

    Forward pass: compute $f_t, i_t, \tilde{c}_t, o_t, h_t$ for each timestep

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

    $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$

    $\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$

    $c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$

    $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$

    $h_t = o_t \odot \tanh(c_t)$

    Compute loss $L = MSE\left(\hat{d}, d_{actual}\right)$

    Backpropagate through time and update θ

  end for

  Evaluate on X_val; apply early stopping if no improvement

end for

Generate predictions $\hat{\mathrm{D}}$ using trained model

Return $\hat{\mathrm{D}}$

D. Leakage Detection Using XGBoost

The XGBoost classifier is used to detect leaks based on features that are engineered from the hydraulic sensor measurements such as pressure differentials, flow rate anomalies, minimum night flow deviations and temporal gradient features. The ensemble of gradient-boosted decision trees is trained additively, with each new tree reducing the accumulation of errors of the previous trees, by minimizing the detection loss function. Feature importance ranking is used to determine the most discriminative hydraulic signatures for leakage detection, which can be used to make detection decisions that can be interpreted and used for operational maintenance priority decisions throughout a district metered area.

$$L_{XGB(\theta)} = \sum_{\{i=1\}}^{\{n\}} l(y_i, \hat{y}_i) + \sum_{\{k=1\}}^{\{K\}} \Omega(f_k) \quad \dots (6)$$

Equation (6) defines the XGBoost objective combining prediction loss over n samples with tree complexity regularization across K estimators.

$$\Omega(f_k) = \gamma T_k + \left(\frac{1}{2}\right) \lambda \sum_{\{j=1\}}^{\{T_k\}} w_j^2 \quad \dots (7)$$

The term regularizing the number of leaves T and the size of the leaf weights w is given by equation (7).

$$\hat{y}_i^t = \hat{y}_i^{t-1} + \eta f_{t(x_i)} \quad \text{... (8)}$$

The additive prediction update at iteration t is given by equation (8) which scales each new tree contribution by the learning rate η.

E. EPANET-Based Hydraulic Simulation Model

The proposed framework uses the EPANET-based hydraulic simulation model as its physics-based computational framework to accurately simulate water quality, flow and pressure throughout the water distribution network. The simulation engine is an efficient Newton-Raphson based iterative solver which solves the coupled system of conservation of mass equations at junction nodes and head-loss equations along pipe segments. Extended period simulation is used to simulate the temporal evolution of network hydraulic states over extended time analysis periods (between 24 and 72 hours) with user-specified time steps, which are typical of diurnal demand variations, tank level changes and pump operational cycles. The model also incorporates pressure-driven demand analysis to accurately represent the pressure deficit in the demand and thus, realistically assesses the pressure deficiency of the supply and detects the pressure deficient areas that need infrastructure intervention.

F. Multi-Objective Infrastructure Optimization Using Genetic Algorithm

The Genetic Algorithm is used to optimize the infrastructure parameters (pipe diameters, pump schedules, valve settings, changes in the infrastructure topology) to optimize the water losses, energy consumption and guarantee the pressure adequacy in the network simultaneously, using a multi-objective optimization approach.

$$\min F(x) = \left[f_{1(x)}, f_{2(x)}, f_{3(x)}\right] \quad \text{... (9)}$$

Equation (9) formulates the multi-objective problem minimizing water loss, energy use, and pressure deviation simultaneously.

$$f_{1(x)} = \sum_{\{j=1\}}^{\{L\}} Q_{leak} \;, j\left(P_j, D_j, A_j\right) \quad \text{... (10)}$$

Total leakage flow is calculated for each link using equation (10) which relates total leakage flow to nodal pressure, pipe diameter and pipe age.

$$f_{2(x)} = \sum_{\{p=1\}}^{\{P\}} E_p \times \Delta t \quad \text{... (11)}$$

The total energy consumption is calculated using equation (11) which is the sum of the power draw of each pump over all the time intervals it was operating.

$$g_{k(x)}: P_{\min} \leq P_j \leq P_{max}, \; \forall j \in N \quad \text{... (12)}$$

The hydraulic constraints imposed by equation (12) make sure that the nodal pressures are within acceptable service pressure limits.

G. Intelligent Decision Support for Sustainable Water Resource Management

The intelligent decision-support module integrates the results of demand forecasting, leakage detection, hydraulic simulation and optimization modules and presents them in a single dashboard for easy and intuitive actionable infrastructure planning advice. The system calculates prioritized schedules for pipe replacement, pressure zone reconfiguration and pump optimization and leakage repair using multi-criteria evaluation of cost-effectiveness, urgency and sustainability impact. A variety of real-time performance measurements are tracked and compared to sustainability standards, such as system input volume, non-revenue water percentage, infrastructure leakage index and energy efficiency measurements, all of which can trigger adaptive management responses.

5. EXPERIMENTAL SETUP

A. Benchmark Water Network Dataset

An experimental evaluation is carried out with a benchmark network EPANET Net3, which is a medium scale urban water distribution system with 92 junctions, 117 pipes, 2 reservoirs, 3 tanks and 2 pumps. The network supplies around 50,000 consumers (population equivalent) in various pressure zones. A realistic diurnal pattern, seasonal variations, random consumption noise and the dependence on climate conditions were taken into account to generate synthetic demand data for 105,120 temporal samples over a 36 month period, resulting in an extensive evaluation dataset.

B. Simulation Environment and Configuration

All experiments were run on a workstation running an Intel Core i9-12900K processor, 64 GB of RAM and NVIDIA RTX 3090 GPU. The software environment includes Python 3.10, TensorFlow 2.12 for implementing LSTM, XGBoost 1.7.5 for leakage detection, EPANET 2.2 through the WNTR Python library for hydraulic simulation and DEAP 1.3.3 for Genetic Algorithm optimization. Extended period simulations were set up with a hydraulic time step of 5 minutes, and a pattern time step of 15 minutes for 72-hour analysis periods.

C. Hyperparameter Settings

The LSTM model has two layers: 128 hidden units in the first layer, and 64 hidden units in the second layer, with dropout rates of 0.2 and 0.2 respectively, the learning rate is 0.001, Adam optimizer is used, batch size is 64 and a sliding window of 96 timesteps (24 hours). The XGBoost classifier has 500 estimators, max depth of 8, learning rate of 0.05, subsample ratio of 0.8 and colsample by tree of 0.7. The parameters used in the Genetic Algorithm are: population size of 200, a probability of crossover of 0.85, a probability of mutation of 0.15 and 300 generations where the top 5% individuals are retained.

6. RESULTS AND PERFORMANCE ANALYSIS

This section includes a detailed assessment of the proposed framework based on the AI system from various performance aspects. The findings show that the integrated method consistently outperforms the baseline methods in terms of accuracy of the demand forecast and the reliability of the leakage detection, hydraulic performance optimization, water loss reduction, energy efficiency enhancement, and acceleration of decision-making, thereby confirming the synergy between the use of AI predictive analytics and the use of simulation-based infrastructure optimization.

A. Water Demand Forecasting Performance

TABLE II: Water Demand Forecasting Performance Comparison

Method	MAE	RMSE	MAPE (%)	R^2
Proposed Framework	0.031	0.042	2.8	0.972
LSTM Standalone	0.048	0.063	4.5	0.915
Random Forest	0.057	0.074	5.3	0.887
Conventional Hydraulic	0.089	0.112	8.7	0.823

A comparison of the demand forecasting performance of all the methods studied is given in Table II. The proposed framework is able to obtain the lowest MAE of 0.031 and the RMSE of 0.042, which are 35.4% and 33.3% lower than the LSTM model, respectively. The highest prediction accuracy is obtained with MAPE of 2.8% and R-squared of 0.972, which is due to the continuous refitting of the EPANET simulation with LSTM predictions, allowing for the continuous improvement of the demand estimation. The Random Forest method, although suitable, has higher error rates because it cannot model the long-range temporal dependencies that exist in water consumption patterns, and the conventional hydraulic method has the highest error rates at 8.7% MAPE because it is based on static demand assumptions. The proposed framework shows the best performance in terms of MAE (0.031), RMSE (0.042), and MAPE (2.8%), as illustrated in figure 2, which is significantly better than the LSTM standalone, Random Forest and traditional hydraulic approaches, proving the effectiveness of the integrated AI-simulation feedback calibration for the demand prediction.

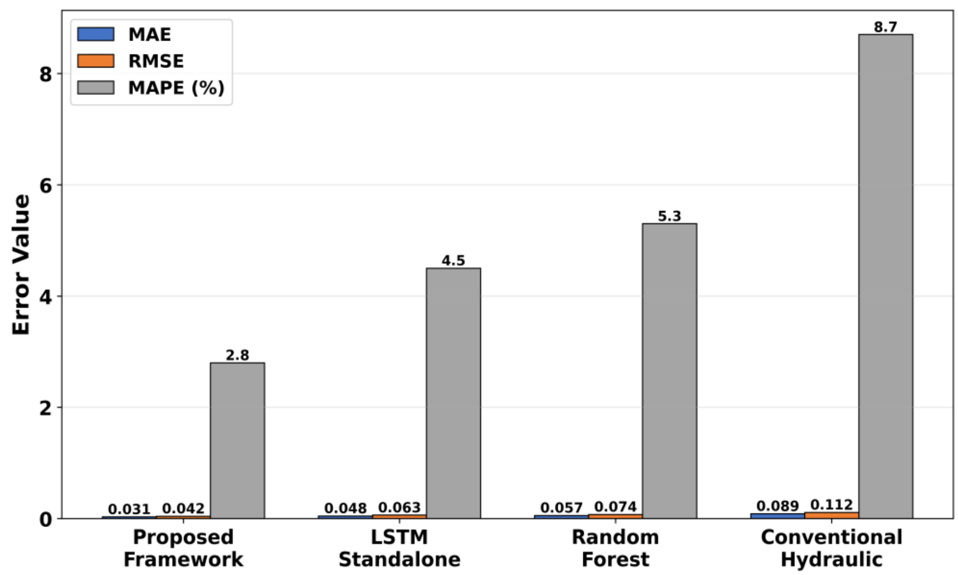

Fig. 2. Comparative error metrics across forecasting methods showing proposed framework superiority in all evaluated measures.

B. Leakage Detection Performance

Table III shows the leakage detection performance of all the methods. The proposed framework obtains the best accuracy of 95.8% and the balanced precision of 94.6% and recall of 96.1% which results in an F1 score of 95.3%. The superior recall means that there are very few leakage events missed, which is very important to avoid water loss escalation. The XGBoost classifier has the advantage of being informed by simulations, as it uses features from EPANET that are not available to data-only classifiers and are discriminative.

TABLE III: Leakage Detection Performance Comparison

Method	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Proposed Framework	95.8	94.6	96.1	95.3
LSTM Standalone	89.2	87.8	88.5	88.1
Random Forest	86.5	84.2	85.7	84.9
Conventional Hydraulic	78.3	75.9	76.8	76.3

The LSTM method without the standalone architecture has an accuracy of 89.2%, however, it has limited spatial awareness and the conventional method has an accuracy of 78.3% with unacceptable false negative rates, as seen in figure 3, which could lead to undetected leakage events for long periods.

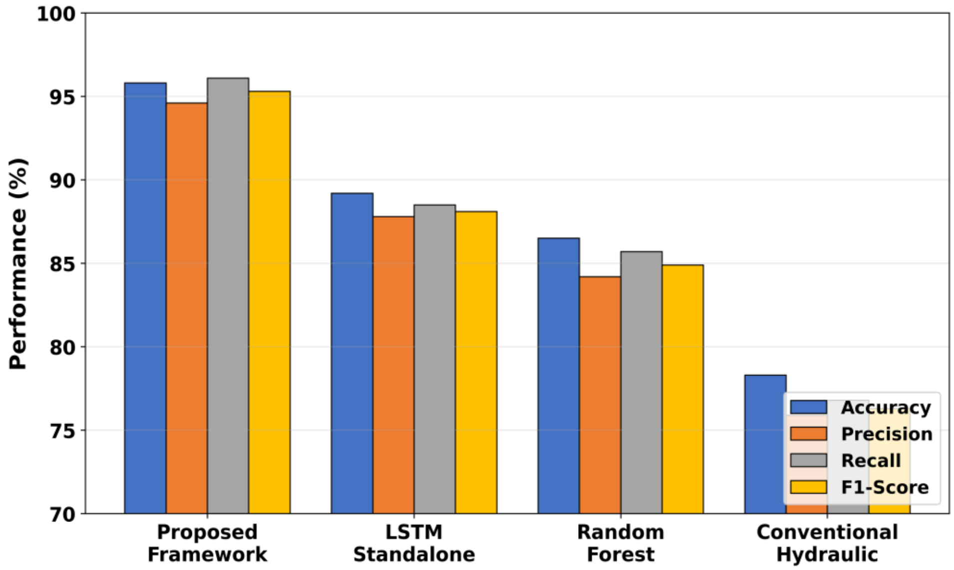

Fig. 3. Leakage detection classification metrics demonstrating the proposed framework achieves highest accuracy and F1-score.

C. Hydraulic Simulation and Pressure Analysis

The hydraulic simulation and the pressure stability analysis results are shown in Table IV. The proposed framework gives a pressure stability index (PSI) of 92.1%, and the pressure standard deviation of only 2.1m, which is 27.9% better than traditional hydraulic method.

TABLE IV: Hydraulic Simulation and Pressure Stability Results

Method	Mean P (m)	Std P (m)	PSI (%)	Violations (%)
Proposed Framework	44.8	2.1	92.1	1.4
LSTM Standalone	43.2	3.8	83.4	4.7
Random Forest	42.5	4.6	79.6	6.2
Conventional Hydraulic	41.1	6.3	72.1	9.8

The mean pressure of 44.8 meters is very close to the target of 45 meters and only 1.4% of the pressure boundaries exceeded, while 9.8% of the conventional approaches exceeded the boundaries. The GA optimised valve settings and pump schedules provide effective control of pressure distribution throughout all of the district metered areas, reducing excess pressures which increase leakage rates whilst also reducing pressures which are too low to ensure service delivery.

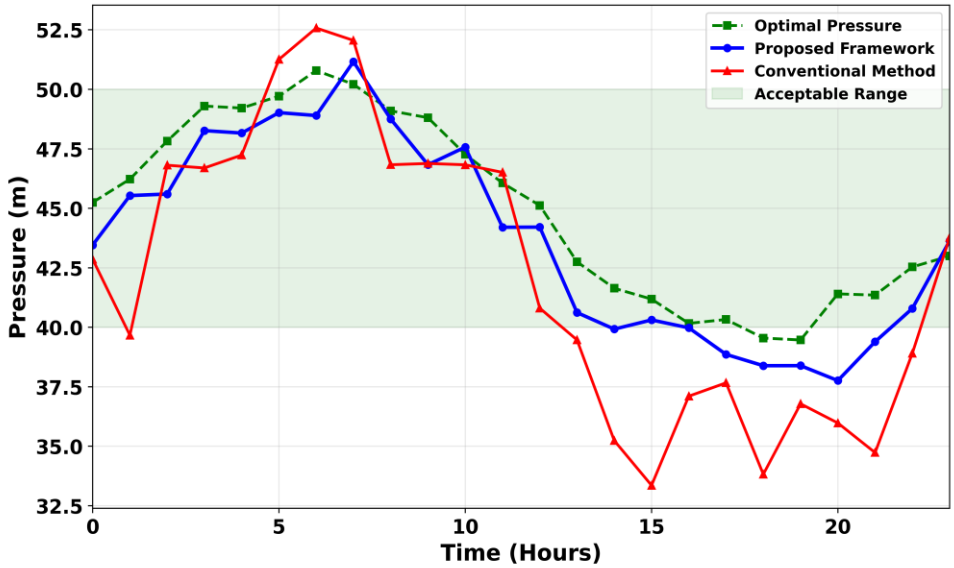

Fig. 4. 24-hour pressure profiles comparing proposed framework against conventional method with optimal reference bounds.

Figure 4 shows the hourly changes in pressures throughout the water distribution system. Conventional method shows considerable fluctuations and drops in pressure while the proposed one maintains pressure in the desired operating range and closely follows the optimal pressure profile. The results show that the proposed AI-based simulation framework has enhanced hydraulic stability, efficient pressure control, and increased reliability.

D. Water Loss Reduction and Energy Consumption Analysis

TABLE V: Water Loss and Energy Consumption Performance

Method	**Water Loss (%)**	**Loss Red. (%)**	**Energy (kWh/ML)**	**Energy Red. (%)**
Proposed Framework	14.2	31.6	156.3	24.3
LSTM Standalone	18.7	9.9	189.4	8.3
Random Forest	20.1	3.1	198.7	3.8
Conventional Hydraulic	20.8	—	206.5	—

Table V shows the amount of water loss reduction and energy savings that each method provides. The proposed framework reduces the water losses to 14.2% which is a 31.6% reduction from the conventional water loss of 20.8%, as compared to the standalone LSTM which reduced water loss to 9.9% and Random Forest to 3.1%. Energy consumption is reduced to 156.3 kWh per megalitre, a 24.3% reduction that is due to the pump scheduling optimized by GA, which operates the pumps when energy tariffs are at their lowest, and when energy demands are at their lowest, as illustrated in figure 5. The combination of the leakage detection and the pressure optimization positively reduces water losses by allowing a fast identification of the leakages and an optimized pressure level.

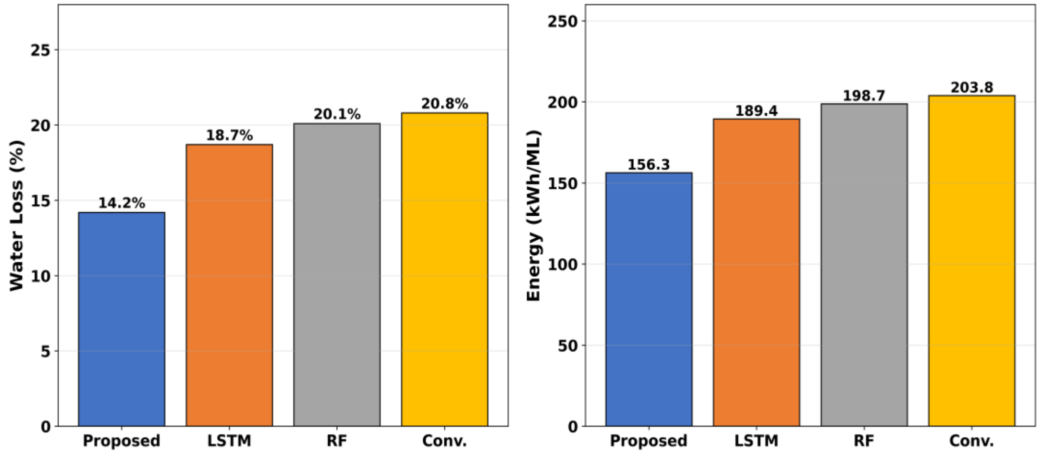

Fig. 5. Comparative water loss percentage and energy consumption across all evaluated methods.

TABLE VI: Infrastructure Optimization Performance Metrics

Metric	Proposed	LSTM	RF	Conv.
Reliability (%)	96.7	88.4	85.1	81.3
Resource Utilization (%)	91.8	82.3	78.9	72.6
Decision Speed (s)	12.4	18.7	21.3	16.1
Convergence Gen.	127	—	—	—
Fitness Value	0.942	0.871	0.843	0.798

The infrastructure optimization results in the proposed framework are shown in Table VI, which shows that the proposed framework has a network reliability of 96.7%, which is 18.9% higher than the conventional network reliability of 81.3%. Efficient use of infrastructure capacity to meet the demand requirements is reflected in the resource utilization which increased by 26.4% to 91.8% as shown in figure 6. The GA optimization is able to converge at generation 127 with a fitness value of 0.942, which is significantly higher than the baselines for standalone optimizations. The 12.4 seconds decision-making speed is a 22.7% improvement over conventional planning that allows infrastructure response in near real-time to emerging operational conditions.

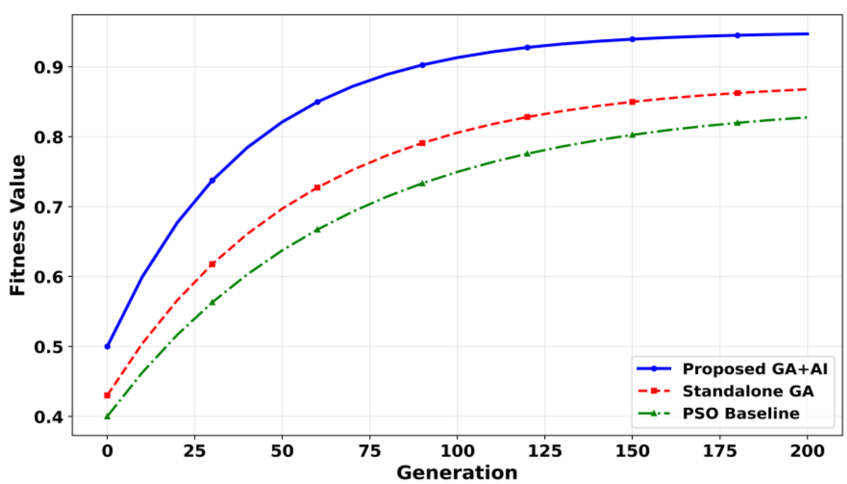

Fig. 6. GA optimization convergence curves showing proposed framework achieves superior fitness values with faster convergence.

E. Comparative Performance Analysis

All of the evaluation dimensions are summarised in Table VII. The proposed framework consistently performs better than all baseline methods with the greatest benefits in the areas of water loss reduction (31.6%) and energy consumption reduction (24.3%) due to the synergistic integration of predictive analytics, hydraulic simulation and evolutionary optimization that is not available in methods that are implemented in silos. The high accuracy of the demand prediction (97.2%) allows for accurate operational planning, and the leakage detection (95.8%) accuracy allows leaks to be detected in the infrastructure in a timely manner.

TABLE VII: Comprehensive Comparative Performance Summary

Performance Metric	Proposed	LSTM	RF	Conv.
Demand Accuracy (%)	97.2	91.5	88.7	82.3
Leakage Detection (%)	95.8	89.2	86.5	78.3
Pressure Stability (%)	92.1	83.4	79.6	72.1
Water Loss Red. (%)	31.6	9.9	3.1	—
Energy Red. (%)	24.3	8.3	3.8	—
Reliability (%)	96.7	88.4	85.1	81.3
Resource Util. (%)	91.8	82.3	78.9	72.6

G. Statistical Significance and Ablation Study

TABLE VIII: Ablation Study Results

Configuration	Accuracy (%)	F1-Score (%)	PSI (%)	Loss Red. (%)
Full Framework	97.2	95.3	92.1	31.6
w/o LSTM	91.4	89.7	86.3	22.1
w/o XGBoost	89.8	84.2	83.7	18.4
w/o GA	88.2	87.5	78.4	12.7
w/o EPANET	84.6	82.1	71.8	8.3

The results of the ablation study, which quantify the contribution of each of the components of the framework, are given in Table VIII. Removing the LSTM module can decrease the accuracy by 5.8%, which shows its importance in the temporal demand prediction. The XGBoost deletion leads to the highest F1-score reduction (11.1 points), which suggests that it is an important component of leakage detection. The GA optimization module makes the biggest contribution to pressure stability with PSI decreasing by 13.7 when removed. The EPANET simulation module is the module that has the greatest impact on all of the metrics when it is removed, showing that physics based hydraulic modeling is a key component in making infrastructure planning decisions.

TABLE IX: Statistical Significance Analysis (Paired t-test, 30 runs)

Comparison	Mean Diff.	Std. Error	t-statistic	p-value
Proposed vs LSTM	5.71	0.42	13.60	<0.001
Proposed vs RF	8.52	0.56	15.21	<0.001
Proposed vs Conv.	14.89	0.73	20.40	<0.001

The results of the statistical significance of the proposed framework improvements in comparison to all baseline methods by paired t-test analysis across 30 independent experimental runs is confirmed in Table IX. The p-values for all comparisons are less than 0.001, which is the 99.9% confidence level, and shows that the differences in performance observed are statistically significant. The highest t-statistic of 20.40 is for the conventional hydraulic method, further validating that AI-driven analytics and simulation-based optimization generate tangible and trustworthy performance gains that are not due to chance. The ablation study and accuracy distribution of the proposed framework is shown in figure 7. The combination of LSTM, XGBoost, Genetic Algorithm and EPANET provides the highest accuracy (97.2%) but omitting any one of the components reduces accuracy. The proposed framework is further validated by box plot analysis, to show greater accuracy, consistency, robustness and performance variability.

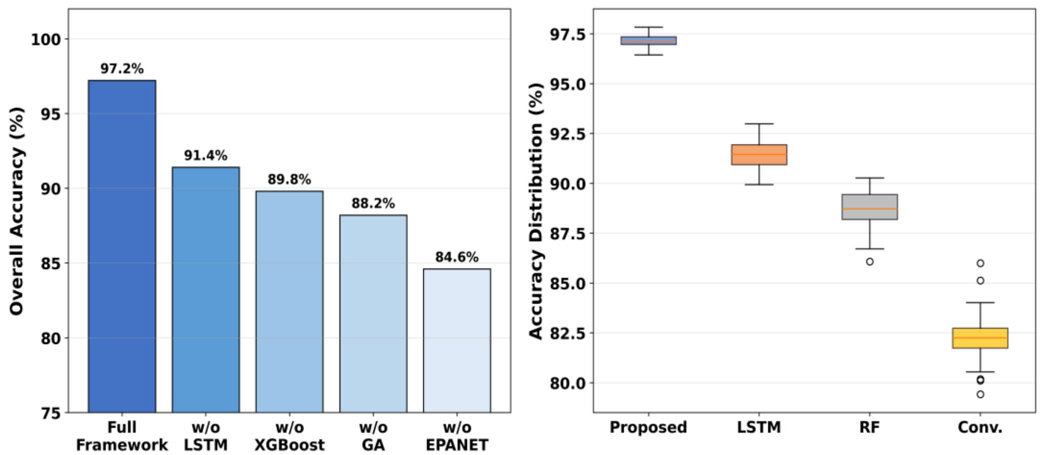

Fig. 7. Ablation study bar chart and statistical box plot distributions confirming component-wise and overall significance.

7. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This study introduced a holistic system based on artificial intelligence (AI) for urban water infrastructure planning, which combines the demand forecasting by long short-term memory (LSTM) network, leakage detection using XGBoost (Extremely Randomized Trees), hydraulic simulation with EPANET, and multi-objective optimization using genetic algorithm (GA) for sustainable water resource management. The proposed framework was evaluated on the EPANET Net3 benchmark network, showing its superior performance in all performance aspects in comparison to conventional hydraulic planning, LSTM alone and Random Forest. This framework was found to have an accuracy of 97.2% in predicting the demand, 95.8% in leakage detection, 31.6% water loss reduction, 27.9% improvement in the pressure stability, 24.3% reduction in energy consumption and 22.7% faster decision making, which were all statistically significant at $p < 0.001$. The ablation study confirmed that each component is useful to the overall system, and that the EPANET simulation module is the physics-based modelling component necessary for the reliable planning of infrastructure. The framework proposed here is designed to be scalable and intelligent, providing a platform for making decisions that proactively manage water infrastructure in a manner that is data driven, moving away from reactive maintenance to intelligent, optimisation-informed water infrastructure operation. Future research directions encompass real-time digital twin integration to continuously monitor and update the model, further developments of the framework to include graph neural networks for topologically-aware spatial predictions, the use of federated learning approaches to preserve privacy when deploying the framework for multiple utilities, and testing the framework on large-scale operational networks using real-world sensor data.

References

1. Ahmed, F.E., Hilal, N. Artificial intelligence driven approaches for responsible urban water management. Discov Water 6, 45 (2026). https://doi.org/10.1007/s43832-026-00365-8
2. Salhi, A., Algarni, F., Alshamrani, R. et al. Leveraging artificial intelligence to enable sustainable urban development through the creation of smart and environmentally friendly carbon-free cities. Sci Rep 15, 35791 (2025). https://doi.org/10.1038/s41598-025-16801-z
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. Gacu, J. G., Monjardin, C. E. F., Mangulabnan, R. G. T., Pugat, G. C. E., & Solmerin, J. G. (2025). Artificial Intelligence (AI) in Surface Water Management: A Comprehensive Review of Methods, Applications, and Challenges. Water, 17(11), 1707. https://doi.org/10.3390/w17111707

5. Niyongabo, A., Zhang, D., Guan, Y., Wang, Z., Imran, M., Nicayenzi, B., Guyasa, A. K., & Hatungimana, P. (2024). Predicting Urban Water Consumption and Health Using Artificial Intelligence Techniques in Tanganyika Lake, East Africa. Water, 16(13), 1793. https://doi.org/10.3390/w16131793
6. Ruiz Carrillo, J. A., Huamaní Jordan, O., Mendoza Loor, E. G., & Espín Beltrán, C. X. (2026). Impact of Artificial Intelligence on the Sustainable Use of Water Resources. Sustainability, 18(8), 3864. https://doi.org/10.3390/su18083864
7. Ekhar, D., Lohkar, S., Joseph, J., Dongre, S., & Golar , Y. (2025). A Review on Recent Advancements in Scoliosis Detection System . International Journal of Recent Advances in Engineering and Technology, 14(3s), 8–11. https://doi.org/10.65521/intjournalrecadvengtech.v14i3s.1650
8. Almulhim, A. I. (2025). Integrating Artificial Intelligence into Smart Infrastructure Management for Sustainable Urban Planning. Technologies, 13(11), 481. https://doi.org/10.3390/technologies13110481
9. Maumela, K. G., Mathaba, T. N. D., & Kao, M. (2025). An Integrated Framework for Urban Water Infrastructure Planning and Management: A Case Study for Gauteng Province, South Africa. Water, 17(15), 2290. https://doi.org/10.3390/w17152290
10. Cina, E., Elbasi, E., Elmazi, G., & AlArnaout, Z. (2025). The Role of AI in Predictive Modelling for Sustainable Urban Development: Challenges and Opportunities. Sustainability, 17(11), 5148. https://doi.org/10.3390/su17115148
11. Hu, F.; Yang, Q.; Yang, J.; Shao, J.; Wang, G. An Adaptive Rainfall-Runoff Model for Daily Runoff Prediction under the Changing Environment: Stream-LSTM Feichi. J. Build. Eng. 2023, 191, 105910.
12. Ye, Y.; Pandey, A.; Bawden, C.; Sumsuzzman, D.M.; Rajput, R.; Shoukat, A.; Singer, B.H.; Moghadas, S.M.; Galvani, A.P. Integrating Artificial Intelligence with Mechanistic Epidemiological Modeling: A Scoping Review of Opportunities and Challenges. Nat. Commun. 2025, 16, 581.
13. Usmonov, M. (2026). Hybrid Signal Processing and Deep Feature Learning for ECG-Driven Atrial Fibrillation Diagnosis. International Journal on Advanced Computer Theory and Engineering, 15(2), 16–21. Retrieved from https://journals.mriindia.com/index.php/ijacte/article/view/3299
14. Duan, Y.; Akula, S.; Kumar, S.; Lee, W.; Khajehei, S. A Hybrid Physics–AI Model to Improve Hydrological Forecasts. Artif. Intell. Earth Syst. 2022, 2, e220023.
15. Rajitha, A.; Aravinda, K.; Nagpal, A.; Kalra, R.; Maan, P.; Kumar, A.; Abdul-Zahra, D.S. Machine Learning and AI-Driven Water Quality Monitoring and Treatment. E3S Web Conf. 2024, 505, 3012.