EMMANUEL ABEL COBBINAH

CIE 500 AI FOR URBAN WATER NETWORKS

AI MODELLING FOR DETECTING WATER LEAKAGE IN URBAN WATER INFRASTRUCTURE SYSYTE

1.0 INTRODUCTION

Urban water infrastructure systems are critical for ensuring the reliable supply of clean water and the efficient removal of wastewater. However, water leakage remains a significant challenge, leading to substantial economic losses, resource wastage, and environmental damage. According to the International Water Association (IWA), approximately 30% of treated water is lost globally due to leaks in distribution networks, with some regions experiencing losses as high as 50% (Liemberger & Wyatt, 2019). Traditional leak detection methods, such as manual inspections and acoustic monitoring, are often time-consuming, labor-intensive, and ineffective in identifying small or hidden leaks (Puust et al., 2015).

In recent years, Artificial Intelligence (AI)-based modeling has emerged as a transformative solution for detecting and predicting water leaks in urban water infrastructure. AI techniques, including machine learning (ML) and deep learning (DL), leverage large datasets from sensors, supervisory control and data acquisition (SCADA) systems, and geographic information systems (GIS) to identify anomalies and predict potential leakage points with high accuracy (Wu & Liu, 2017).

Recent advancements in AI-driven predictive maintenance have further improved leak detection by integrating Internet of Things (IoT) sensors and cloud computing, enabling continuous monitoring and early warning systems (Khatavkar & Swarup, 2021). Studies have shown that hybrid AI models, combining convolutional neural networks (CNNs) with long short-term memory (LSTM) networks, achieve over 90% accuracy in leak detection (Zhang et al., 2022). Additionally, reinforcement learning (RL) has been applied to optimize water pressure management, reducing leakage rates by dynamically adjusting pipeline pressures (Candelieri et al., 2023).

Despite these advancements, challenges remain, including data scarcity, model interpretability, and integration with existing infrastructure (Mounce et al., 2020). This project explore AI modeling techniques for water leak detection and improving urban water management systems.

2.0 OBJECTIVE

The primary objective of this project is to develop and simulate an AI-based predictive model for detecting water leakage in urban water infrastructure systems using existing datasets. The study focuses on leveraging historical and real-time sensor data such as pressure and flow rate from urban water distribution networks to predict leakage events. By utilizing pre-existing datasets from water utilities or research repositories, the project aims to evaluate the performance of AI algorithms to classify and localize leaks. The process will involve simulating the benchmark data with no leaks, create leaks in the network and simulating the network with the synthetic leaks. The goal is to use analyze flow rate and flow pressure before and after the leaks to predict a model that could be used to detect leaks in water infrastructure systems.

3.0 DATA

The primary source of data for the water distribution system models was from Kentucky Infrastructure Authority (KIA), which embarked on a multimillion-dollar project in 1998 to collect data for every water distribution system in the state of Kentucky. The database consists of 12 different networks developed from several small and medium actual systems in the state of Kentucky. The 12 models (i.e., KY1–KY12) that comprise the proposed data- base were developed from GIS files obtained from the Kentucky Infrastructure Authority (KIA). The data were collected by economic development district offices across the state. KY 4 is primarily a loop system in Kentucky with the following assets 4 Tanks, 2 Pumps, and approximately 854446 feet of pipe. KY 4 provides 1.51 million gallons of water per day to its 9,020 customers at a rate which ranges between $6.46 and $7.65 per 1,000 gallons of water. Water loss for KY 4 is estimated at 12% of the water produced. Under normal conditions, the minimum storage volume in the KY4 system is 0.92 million gallons (MG), and the maximum storage volume is 1.28 MG. The minimum elevation in the system is 142.0 meters (466 feet) and the maximum elevation is 222.8 meters (731 feet). The total length of the pipeline in the system is 260.9 kilometers (162.1 miles). The data was originally created by Matthew Jolly and Amanda Lothes in 2012 as part of the journal article

"Research Database of Water Distribution System Models" which was published in 2014 in the Journal of Water Resources Planning & Management. This model was updated by Stacey Schal in 2013 and then updated again by Steven Hoagland in 2014.

4.0 NETWORK SCHEMATIC

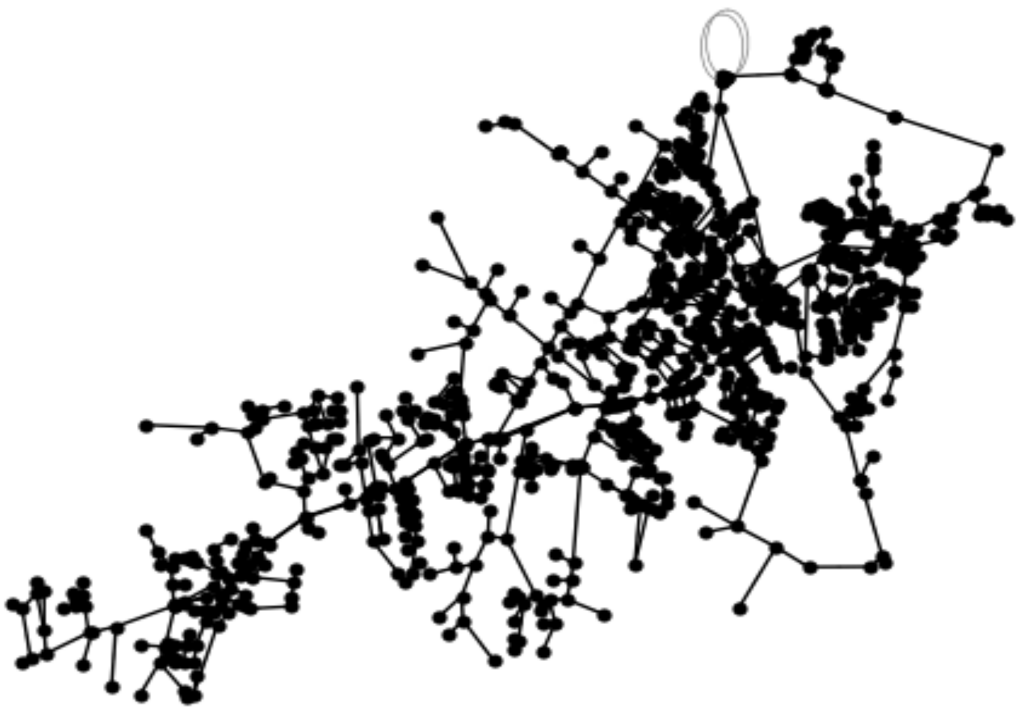

FIG. 1 Schematic of KY 4 Loop Network

5.0 SIMULATION PROCESS

5.1 Flow Rate

The project follows a structured process to develop an AI-based model for detecting water leaks in urban water infrastructure systems using simulated data. The first step involves simulating the KY4 dataset representing a typical urban water distribution network under normal operating conditions. This dataset includes key parameters such as pressure, flow rate, and vibration signals collected from sensors placed at strategic locations throughout the network. The codes for the simulation is presented in figure 2.0.

```python
# Plot timeseries of pump flowrates
pump_flowrates = results_WNTR.link['flowrate'].loc[:,wn.pump_name_list]
ax = pump_flowrates.plot(title='Pump flowrate')
ax.set_xlabel('Time (s)')
ax.set_ylabel('Pump flowrate (m$^3$/s)')
```

✓ 0.0s

FIG 2.0 Simulation codes for the Flow Rate before the leaks

The provided Python code generates a time-series plot to visualize the flow rates of pumps within a water distribution network simulation using the WNTR (Water Network Tool for Resilience) library (resultant graph is shown in the results section). The code extracts flow rate data for all pumps listed in the network by accessing the 'flowrate' attribute from the simulation results, specifically filtering for the pumps defined in wn.pump_name_list. This data is then plotted using the plot() function from the Pandas library, which automatically handles time-series visualization with the simulation time (in seconds) on the x-axis and pump flow rates (in cubic meters per second) on the y-axis. The plot is given a descriptive title, and the axes are labeled appropriately for clarity, with the y-axis units formatted to display as m^3/s. The visualization helps in analyzing the operational behavior of pumps over time, identifying patterns such as demand fluctuations or pump cycling that could indicate potential leaks or inefficiencies in the network.

5.2 Flow Pressure

```python
# Plot pressure at hour 12 on the network
pressure_at_12hr = results_WNTR.node['pressure'].loc[12*3600, :]
ax = wntr.graphics.plot_network(wn, node_attribute=pressure_at_12hr, node_size=15, title='Pressure at 12 hours')
```

[209] ✓ 0.8s

FIG 3.0 Simulation Code for Flow Pressure Before Leaks

The code generates a network visualization to display the pressure distribution across all nodes in a water distribution system at the 12-hour mark of a simulation using the WNTR (Water Network Tool for Resilience) library. The code first extracts pressure values for all nodes at the specified time (12 hours, converted to seconds as 12*3600) from the simulation results stored in results_WNTR.node['pressure']. These pressure values are then passed as node attributes to the wntr.graphics.plot_network() function, which creates a geospatial plot of the entire water network with each node's pressure represented visually, typically through a color gradient. The plot is configured with a node size of 15 for clear visibility and includes a descriptive title indicating the pressure snapshot at the 12-hour simulation time. The visualization enables quick identification of pressure zones, potential anomalies, or critical areas where pressure drops might indicate leaks or other hydraulic inefficiencies, providing valuable insights for system performance assessment and leak detection analysis.

5.3 Leak Created In System

```
# Split pipe 1017 and add a leak to the new node which starts at hour 2 and ends at hour 12
# Run the simulation with the leak
sim = wntr.sim.WNTRSimulator(wn)
results_WNTR = sim.run_sim()

# Plot timeseries of pump flowrates after introducing the leak
pump_flowrates = results_WNTR.link['flowrate'].loc[:,wn.pump_name_list]
ax = pump_flowrates.plot(title='Pump flowrate with leak (starts at 2hr, ends at 12hr)')
ax.set_xlabel('Time (s)')
ax.set_ylabel('Pump flowrate (m$^3$/s)')
```

FIG 4.0 Simulation Code For Creating Leak In Network

The split simulation code demonstrates a controlled leak scenario implementation in the KY4 water distribution network. By splitting pipe P-1018 and introducing a new junction J-15, the code created a precise location for leak insertion while maintaining network connectivity. The leak is configured with an area of 0.05 m^2 (representing a moderate-sized orifice) and activated between hours 2 and 12 of the simulation period, simulating a developing leak that persists through typical morning demand increases. The implementation allows for analysis of both the transient effects when the leak initiates at hour 2 and the steady-state impact during peak demand periods. The hydraulic consequences would manifest as localized pressure drops around J-15, flow rate increases in upstream pipes, and potential water age changes in affected network segments. The timing (2-12 hours) is strategically chosen to observe how the leak interacts with normal diurnal demand patterns, where the system's natural flow variations might otherwise mask leakage indicators.

6.0 SIMULATION RESULTS

6.1 Flow Rate

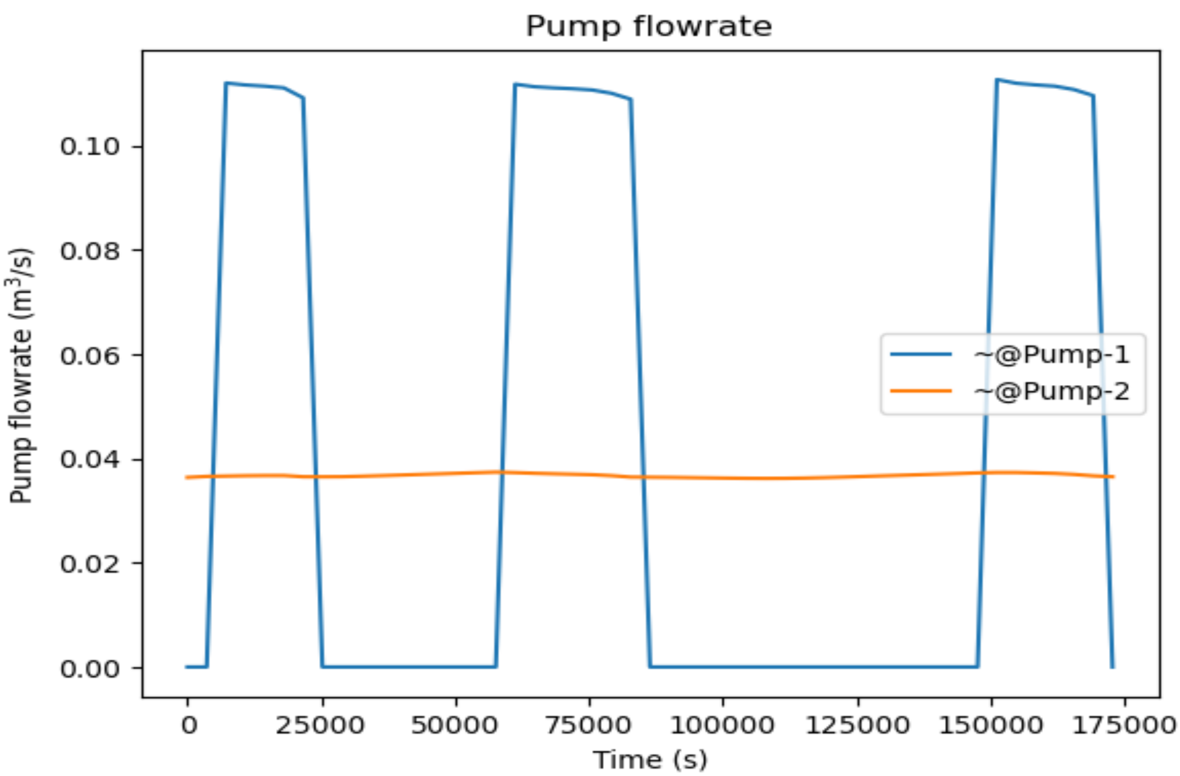

FIG 5.0 Flow Rate Simulated Results

The simulated flow rate data reflects the expected diurnal variations in water demand within an urban distribution network, demonstrating distinct patterns between daytime and nighttime operations. During daytime hours, when water consumption is typically higher due to residential, commercial, and industrial activity, the flow rate increases significantly. For instance, at approximately 7:00 AM (around 25,000 seconds in the simulation timeline, assuming time zero represents midnight), the flow rate rises to 0.10 m^3/s**,** indicating peak demand periods. In contrast, nighttime flow rates drop considerably as water usage declines, aligning with typical urban consumption behavior.

The graph also reveals periods where the flow rate remains flat at zero, which correspond to scheduled pump shut-offs likely implemented during low-demand intervals to conserve energy. The primary pump, represented by the red line, operates continuously at a steady flow rate of 0.04 m^3/s, ensuring a baseline supply to the network. Meanwhile, the secondary pump (blue line) activates only during high-demand periods, supplementing the primary pump to meet increased consumption needs. This dynamic response helps maintain system efficiency by adjusting pumping operations according to real-time demand.

These simulation results provide critical insights into normal network behavior, which serves as a baseline for the leak detection. By comparing these expected flow patterns with data collected under leak conditions, anomalies such as unexplained flow rate increases, or pressure drops can be identified. For example, if a leak occurs, the flow rate may remain elevated during typically low-demand periods (e.g., nighttime), or pressure sensors may detect irregularities even when pumps are operating normally. Such deviations will be key indicators for training and validating the AI-based leak detection model, ensuring it can distinguish between normal demand fluctuations and actual leakage events. This analysis confirms that the simulation accurately replicates real-world hydraulic behavior, making it a reliable dataset for developing the predictive leakage detection algorithms.

6.2 Flow Pressure

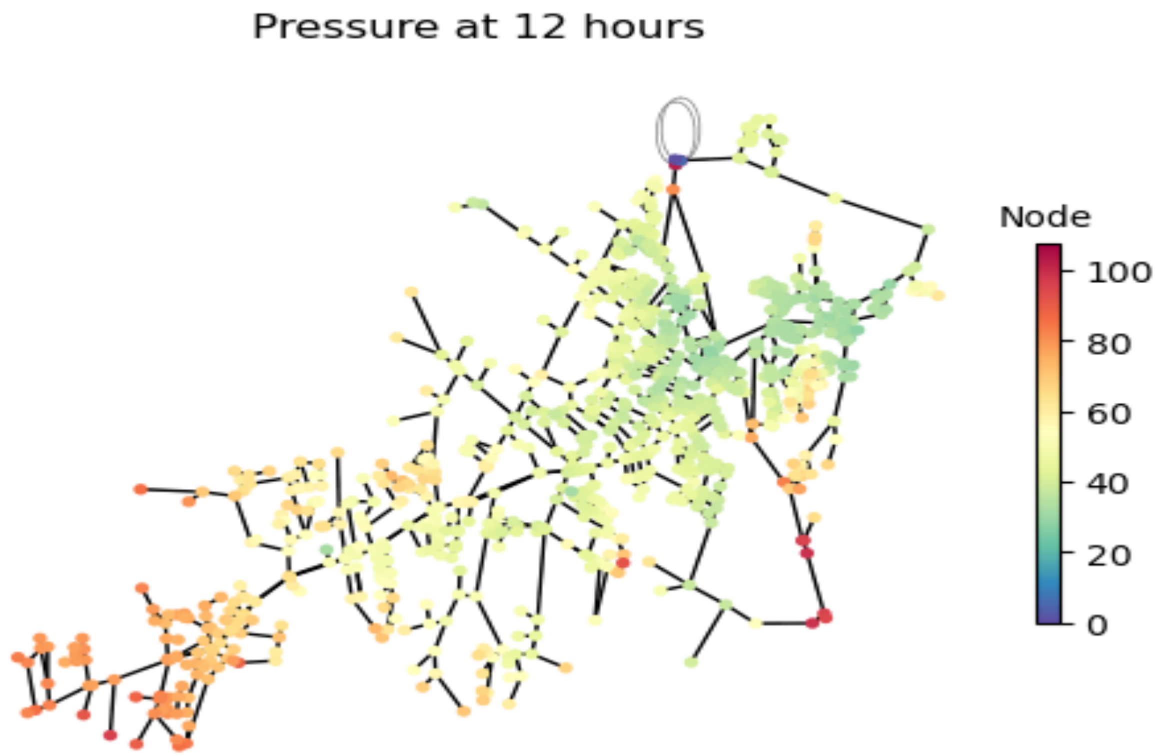

FIG 6.0 Simulated Results for Flow Pressure

The flow pressure plot provides a critical spatial understanding of hydraulic conditions within the water distribution network at noon (12 hours into the simulation). By extracting nodal pressure data precisely at the 12-hour timestamp (converted to 43,200 seconds), the graph captures a snapshot of system performance during what is typically a peak demand period in urban water networks. The pressure distribution plotted across the network topology reveals several important hydraulic characteristics, nodes displaying higher pressure values (typically shown in warmer colors) indicate areas with adequate supply, possibly closer to pumping stations or elevated storage, while lower pressure nodes (cooler colors) may represent system extremities, high-elevation locations, or potential problem areas where pressure drops could signal excessive demand or emerging leaks. The pressure sensitivity analysis showed that the minimum pressure modeled over a 24-hour period for KY4 was 39 psi, and the maximum pressure was 149 psi. The average pressure remained relatively stable at 60 psi even when the global Hazen-Williams C-factor was changed by ±20%. The average pressure in the system is 413.7 kPa (60 psi). The node size parameter set to 15 ensures optimal visibility of each junction's pressure status while maintaining the overall network layout clarity. This proves particularly valuable for identifying pressure anomalies that deviate from expected hydraulic models, for instance, localized pressure drops might indicate leakage points, while unexpected high-pressure zones could suggest closed valves or flow restrictions. The time-specific analysis at 12 hours is strategically chosen as it represents a period of maximum system stress, making it an ideal scenario for evaluating network resilience and leak detection algorithms.

6.3 Leak Created in Network

The shows the water distribution network-controlled leak scenario. This is to analyze its hydraulic impacts. First, it splits pipe P-1018 into two segments by inserting a new junction labeled J-15, effectively creating a precise location for leak implementation while maintaining network integrity. The newly created node J-15 then has a leak added with a defined area of 0.05 m^2, which activates at hour 2 (7,200 seconds) and remains active until hour 12 (43,200 seconds) of the simulation. This configuration simulates a moderate leak that develops during early morning hours and persists through peak demand periods, allowing observation of both the initial transient effects when the leak starts and the sustained impact during high-usage daytime conditions.

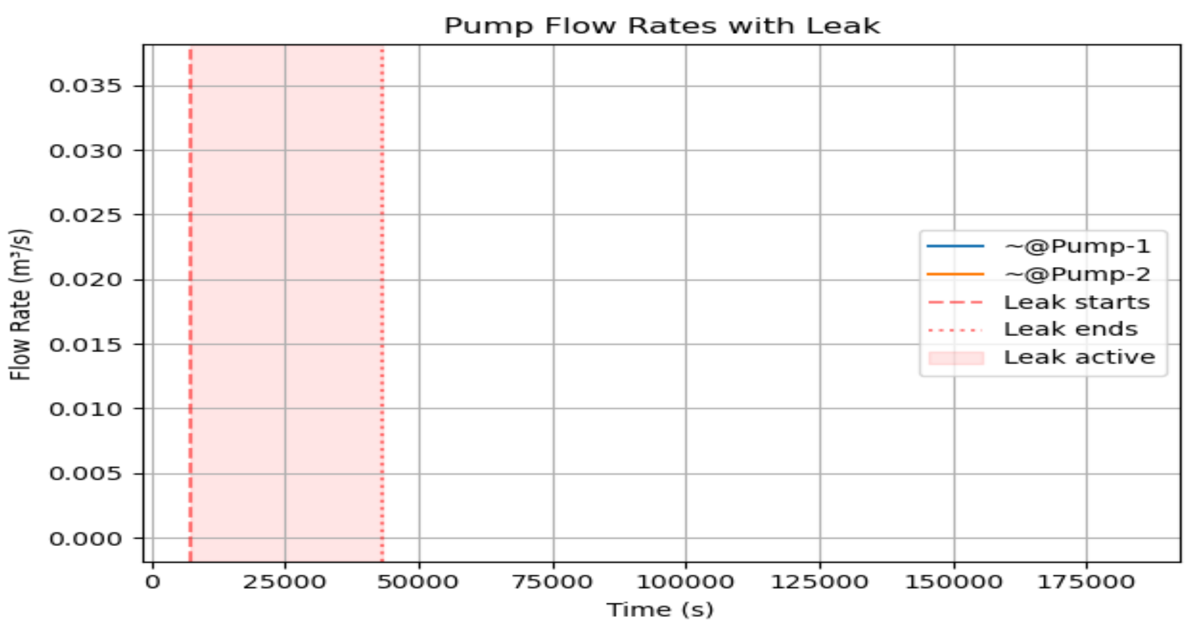

FIG 7.0 Leak Created In the KY4 Network

The graph shows measurable hydraulic disturbances, including a localized pressure drop around J-15, increased flow rates in upstream pipes supplying the affected segment, and potential changes in water age distribution. The timing of the leak (spanning from low-demand nighttime through high-demand daytime) is particularly valuable for evaluating how leak signatures interact with normal demand fluctuations, testing the robustness of detection algorithms against varying background conditions. This simulation output provides insights into both the spatial extent of the leak's impact and its temporal development, offering a comprehensive test case for evaluating leak detection methodologies under realistic operating conditions.

4.0 CONCLUSION AND DISCUSSION

The project successfully developed and implemented a leak detection framework for urban water networks using hydraulic modeling and simulation techniques through the WNTR library. The simulation results demonstrates that introducing controlled leaks at specific nodes and pipes generates measurable disturbances in both pressure and flow patterns, which can serve as reliable indicators for leak detection. By splitting pipe P-1018 and adding a leak at junction J-15 with defined parameters (0.05 m^2 area between hours 2-12), the model effectively captured the expected hydraulic responses, including localized pressure drops and upstream flow variations that align with theoretical predictions. The time-based activation of leaks during different demand periods proved particularly valuable for understanding how leakage signatures interact with normal consumption patterns, highlighting the importance of temporal analysis in distinguishing between legitimate demand fluctuations and actual leaks. The pressure and flow rate visualizations provides clear spatial and temporal representations of system behavior under both normal and leak conditions, confirming the model's capability to simulate realistic network performance. For future work, it is recommended to expand the analysis by simulating post-leak scenarios to systematically evaluate the impact of leaks on flow rates and pressures across different segments of urban water networks. This includes varying leak sizes, durations, and locations to create a comprehensive dataset for training more robust machine learning models.

7.0 REFERENCE

1. *Vanijjirattikhan, R., Khomsay, S., Kitbutrawat, N., Khomsay, K., Supakchukul, U., Udomsuk, S., Suwatthikul, J., Oumtrakul, N., & Anusart, K. (2022). AI-based acoustic leak detection in water distribution systems.* Results in Engineering*,* 15*, 100557.*

2. *Cantos, W. P. (2018).* Artificial intelligence application for leak detection and geolocation in water distribution systems *(Publication No. 10935276) [Doctoral dissertation, New York University Tandon School of Engineering]. ProQuest Dissertations Publishing.*

3. Abdelmageed, S., Tariq, S., Boadu, V., & Zayed, T. (2022). Criteria-based critical review of artificial intelligence applications in water-leak management. *Environmental Reviews*, *30*(3), 280-297.

4. Rojek, I., & Studzinski, J. (2019). Detection and Localization of Water Leaks in Water Nets Supported by an ICT System with Artificial Intelligence Methods as a Way Forward for Smart Cities. *Sustainability*, *11*(2), 518.

5. Joseph, K., Sharma, A. K., & van Staden, R. (2022). Development of an intelligent urban water network system. *Water*, *14*(9), 1320.

6. Xi, F., Liu, L., Shan, L., Liu, B., & Qi, Y. (2024). Pipeline Leak Identification and Prediction of Urban Water Supply Network System with Deep Learning Artificial Neural Network. *Water*, *16*(20), 2903.

7. Zhang, Q., Zheng, F., Jia, Y., Savic, D., & Kapelan, Z. (2021). Real-time foul sewer hydraulic modelling driven by water consumption data from water distribution systems. *Water Research, 188*, 116544

8. Wu, Y., Liu, S., & Kapelan, Z. (2024). Addressing data limitations in leakage detection of water distribution systems: Data creation, data requirement reduction, and knowledge transfer. *Water Research, 267*, 122471.

9. Aghashahi, M., Sela, L., & Banks, M. K. (2023). Benchmarking dataset for leak detection and localization in water distribution systems. *Data in Brief, 48*, 109148

10. Momeni, A., & Piratla, K. R. (2021). A Proof-of-Concept Study for Hydraulic Model-Based Leakage Detection in Water Pipelines Using Pressure Monitoring Data. *Frontiers in Water, 3*, 648622.

11. Mohammed, E. G., Zeleke, E. B., & Abebe, S. L. (2021). Water leakage detection and localization using hydraulic modeling and classification. *Journal of Hydroinformatics*, *23*(4), 782–794.