Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 57-1 (2024) 280–285

Digital Twin Framework for Leakages Detection in Large-scale Water Distribution Systems: A Case Study of IIT-Jodhpur Campus

Anushka Singh*. Abhilasha Maheshwari. Shobhana Singh*****

** Indian Institute of Technology, Jodhpur, N.H. 62, Nagaur Road, Karwar Jodhpur 342030, Rajasthan, India (email: m22ch003@iitj.ac.in)*

*** Indian Institute of Technology, Jodhpur, N.H. 62, Nagaur Road, Karwar Jodhpur 342030, Rajasthan, India (email: abhilasham@iitj.ac.in)*

**** Indian Institute of Technology, Jodhpur, N.H. 62, Nagaur Road, Karwar Jodhpur 342030, Rajasthan, India (email: shobhana@iitj.ac.in)*

Abstract: Sustainable development goals and industry 4.0 push for a holistic plan of action for smart water infrastructure enabling advance digital technologies such as Digital Twins for water networks through an integrated use of machine and physical counterparts. This paper proposes a Digital Twin framework for leakage detection applications in large scale water distribution systems. The framework elucidates digital map generation of the network, hydraulics modelling, calibration and leakage detection model in an integrated manner using python interface. The hydraulic model accounting for spatial and temporal variations of network hydraulics and an optimization formulation for calibration and graph neural networks for leakage identification has been developed. The framework is applied, and results have been demonstrated on a real-life case study of IIT Jodhpur campus water distribution system.

Keywords: optimization, smart water infrastructure, EPANET, python, calibration, neural networks, sustainable development goals

INTRODUCTION

1.1 Background and Motivation

Digital Twin (DT) refers to a virtual representation that works synchronously with a physical system to predict, maintain, and remote control over the physical system, thereby improving quality of service. Grieves introduced the concept of DT in 2002, in manufacturing as part of Product life cycle management (PLM), later Vickers popularized it through NASA (Grieves and Vickers 2016). In recent trends, DT finds use in diverse fields such as design, diagnosis, optimization, manufacturing, PLM, health, smart water (Pedersen et al. 2021; Wu et al. 2023), water Distribution networks (WDN) (Brahmbhatt, Maheshwari, and Gudi 2023; Hall et al. 2000; Ramos et al. 2022) and also to plan any further improvement to product or service, considering real-time data to facilitate better and smooth operation by the user (Rosen et al. 2015). The convenience and adequacy of DTs is conspicuous to both corporate and academia as several companies such as Deloitte, Siemens and Gartner speak volumes favoring the use of DTs for system efficiency under the umbrella of industry 4.0 (Enric Escorsa 2018). To this end, the agenda 2030 of Sustainable Development Goals set by United Nations includes in it, SDG 6 which is focused on water wherein the target 6.4 of SDG 6 aims achieving increased water efficiency in supply management. Thus, there is a need to increase operational efficiency of WDN through improved management and performance enhancement. Recently (Brahmbhatt et al. 2023) have simulated a large scale benchmark model of C-Town WDN for demonstrating higher operational efficiency in water systems. However, the authors did not validate on a real-world case study with actual sensor data from the field.

The water demand exerted on limited water resources can be reduced by optimizing water supply through large scale WDN This is facilitated by the simulation of water hydraulics and water quality in distribution system. (Sankar et al. 2015) demonstrated that by use of a predictive controller model, up to 40% of the daily deficit can be reduced in a water scant region. Furthermore, WDN must cope with challenges such as population growth, infrastructure wear, resource dearth and pure quality of water for everyone and to achieve such goals of smart water infrastructure (SWI), water utilities look for the lucrative possibilities provided by DTs. Gathering of real time data, preprocessing the huge data, when complemented with the IoT, AI/ML, appropriate hydraulic models for simulation of the network behaviour can address challenges such as: (i) early leak detection, (ii) optimum pressure and flowrate control (iii) water quality control (Conejos et al. 2020). Owing to the above-described compelling issues, the whole value chain of water is looking for digital technology-based solutions that could ensure a sustainable water future, plausible through smart water infrastructure. Thus, as shown in Fig. 1, smart water infrastructure design consists of three main

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2024.05.048

systems layers (namely, physical system, decision support system and management and control systems). The sensors and other data gathering units in physical systems layer transfers the data to decision support layer where optimization methods are developed using data analytics tools and platforms are developed for modelling, simulating and control followed be the management and real-time control of the physical system. In this direction, the aim of this work is to propose a decision support framework that can facilitate decision making for leakages detection through smart water infrastructure.

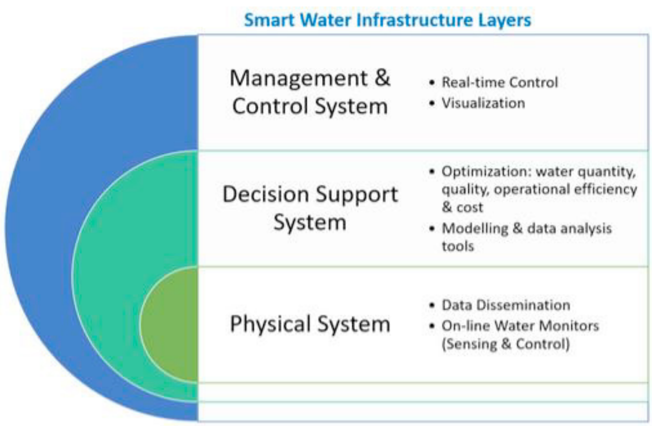


Fig.1: Systems layers of Smart Water Infrastructure

1.2 Components of a Digital Twin for Smart water Infrastructure

A DT has several components integrated in a holistic framework. The main components of a DT include:

1.Data Collection: It is crucial that accurate, regularly updated, and real time data from the field is fed to the DT of the physical system to perform effectively and provide beneficial analysis and take correct measures. Online sensing, GIS, SCADA, smart metering etc. can provide useful data for this step.

2.Data processing: The collected data, when cleaned, sorted and free of redundancy, can form a useful database. Various data analytics tools and techniques are essential to rid of inconsistencies of missing time stamps/data and or irregular information to draw useful conclusions, accurate predictions, and informed decisions Therefore, pre-processing of raw data is important for the model to act upon.

3.Models: The processed data is then fed to appropriate models to simulate the physical system and be able to depict all plausible scenarios for reference. In the case of WDN, hydraulic models should be calibrated for demand variables and other time dependent parameters which when coupled with advanced algorithms, machine learning, Artificial intelligence, Deep learning, Augmented and Virtual Reality counterparts can simulate, analyze, predict, and then suggest appropriate action to optimize the physical system.

4.Actuators: The virtual equivalent of the physical system gives smart analysis and run diagnosis to aid in planning, management, and generate optimum operation rules, which is relayed to the physical system through actuators such as valves, pumps to implement proper action.

2. METHODOLOGY AND PROCEDURE

2.1 Problem statement

Of the many challenges water utilities encounter in the water distribution network operational management, the most prominent is the leakage in pipeline or at junctions that may occur over time due to material degradation, water and tear over time, pipe bursts due to overpressure or even some natural calamity. This leakage problem further uproots the issue of sustainability and inadequate water supply to the consumers. Advanced information technologies that are used in the smart water management systems ensure the following benefits: better understanding of the water system, early detection of leaks and efficient control of water losses, economic benefits to water and energy conservation, reduction of financial losses (up to 30% reductions in water bills), improvement of system efficiency and customer service quality, among others) Therefore early detection of leaks in the pipeline is a significant concern and need of the hour for achieving operational excellence in WDN management and several authors have contributed to this end (Capelo et al. 2021; Manzi et al. 2019; Mounce and Machell 2006; Sadeghioon et al. 2018). A digital twin of a water network can solve the crisis and help in the early detection of leakages to ensure timely maintenance and planning. The challenges, however, for developing a DT of the same include:

(i) Developing a DT using a hydraulic model for large scale water distribution networks

(ii) Calibrating the hydraulic model.

(iii) Developing algorithms (AI or ML based) to predict leakage in the system accurately

(iv) Testing and validating the AI/ML models

Taken together there is a research gap addressing holistic framework consisting of model development of large-scale water network, hydraulic simulation, optimization, and visualization. In this paper preliminary aspects of this challenge have been addressed.

2.2 Case Study description

This paper is focused on IIT Jodhpur campus water distribution network to detect potential leakage in pipeline through a DT. Indian Institute of Technology (IIT), Jodhpur at Karwar, is a vast campus spanning 852 acres on National Highway 65 toward Nagaur. network. This study considers the following characteristics of the water network:

(b)

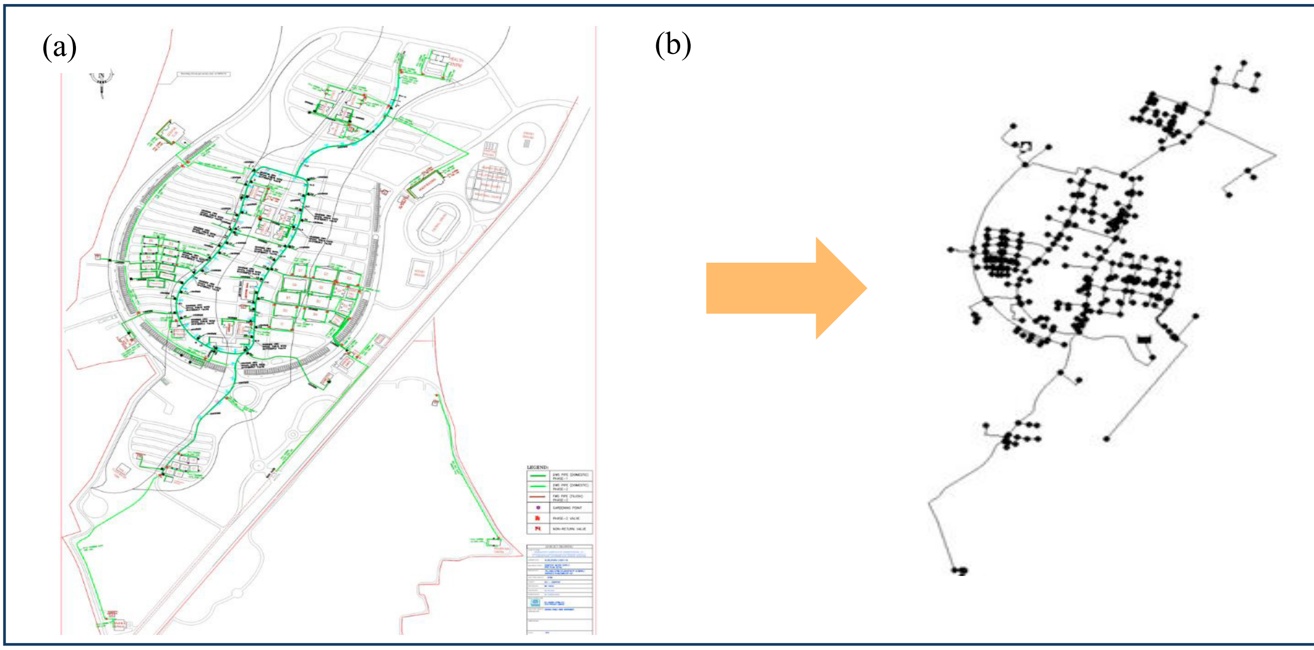


Fig.2: (a) Schematic diagram of campus WDN and (b) its skeletonized model

Table 1: Water Distribution network characteristics of IITJ campus

Nodes	366
Pipes	375
DMAs	5
Material of Construction of pipes	HDPE pipes (size ε [⌀90mm, ⌀250mm]
Roughness coefficient	120-150
Pumps	2
Reservoir	1
Study area Location	Pocket A of IIT Jodhpur, Near Karwar, Jodhpur
Length of network	13 Km
Water Source	From Indira Gandhi Canal via Manaklao pumping station
Serving population	5000 people (floating)

The campus has three parcels of land, A, B, and C, of which study zone, A has an area of about 659 acres that houses (i) hostels, (ii) staff and faculty residential quarters, (iii) educational premises including lecture hall complex and dept buildings, (iv) Sports complex. Indira Canal is the primary water supplier for the campus through the Manaklao pumping station. However, the WDN of the campus is an independently managed system. The overall future water management plan to become NET-ZERO water, as well as the expansion of student intake, calls for efficient management of the water distribution.

2.3 DT Framework

As discussed in the previous section, a holistic framework for developing a DT includes hydraulic model of WDN, dynamic simulation of hydraulics and calibration of the hydraulic model parameters followed by Neural network model for leakage identifications. EPANET is a well-known open-source simulator used for simulating hydraulics and water quality (Klise et al. 2020; Rossman 2000). In this paper, the assumption follows that, data generated from EPANET simulator is close to the field data. The next paragraphs elucidates the step-by step procedure for the developed DT framework in detail.

1.Digital Model development; In this framework, firstly, the available CAD file of the campus water network is converted to a universal exchange format (i.e. dxf) using ODA FILE CONVERTER (Version 23.2.0) since the CAD file do not extend support to other applications. The file is then converted to EPANET compatible, (.inp) file using EpaCAD (Version 1.0). In the EPANET file, the whole network is divided into 5 District Metered Areas (DMAs), and base demands as well as demand multipliers of each DMA, including all nodes, is randomly generated for 1 week, and generic value of base head as well as pump curve is assigned to the reservoir and both pumps respectively. The roughness coefficient of the pipes is assigned standard values based on the material of construction.

Fig. 3 describes the steps followed in extracting the WDN network file using above mentioned methodology. Fig. 2(b) also shows skeletonized (extracted) model of the WDN.

2.Hydraulics Simulation: Secondly, extended period simulation of network hydraulics using Python interface on WNTR simulation (which is based on EPANET engine) is being carried out for a period of one week. And the data for pressure and flow values of all the nodes are simulated, these represent the predicted values of pressure and flow rate.

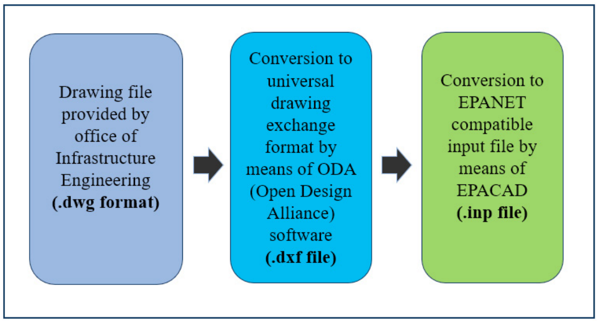


Fig.3: Schematic of steps for generating digital network file from CAD drawing files

3.Calibration: Thirdly, calibration of hydraulic model is done for optimized time dependent parameters i.e., Roughness Coefficients (C) values and Demand multipliers (D) values for a period of 24 h. The actual and predicted values of pressure and flow rates using standard and optimized parameters are compared using an objective function which is to be minimized.

4.Neural Network based Leak Detection Model: A neural network-based approach has been proposed in this paper to develop an estimation model for leakages. The leakage scenarios are functions of leak size in the pipe. Leakage scenarios data is generated, stored, and then fed to a Graph Convolution Neural Network (GCN) model to estimate the leakage location.

5.Model Validation: Neural Network validation: After the model is made, it is trained and tested on a subset of data, which in this case was the pressure from sensor nodes. Of this dataset 80% is used for training the neural network and the rest for testing the model for leak detection. Finally, an estimation can be made to indicate the leak location. And using Neural network prediction results the accuracy has been determined.

The entire framework with above steps is schematically shown in Fig 4.

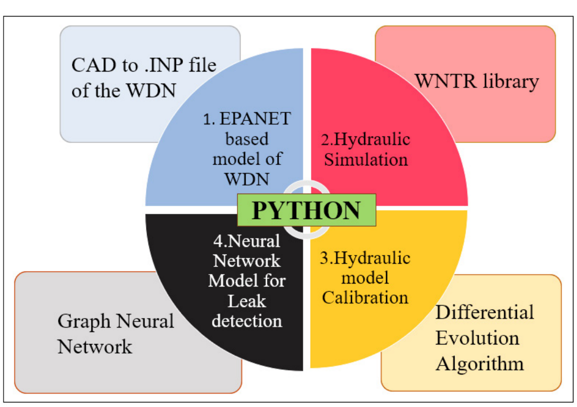


Fig.4: Digital twin framework for leakage detection application in WDN

3. OPTIMIZATION FORMULATION

This section explains mathematical formulation of the optimization problem for calibration and Neural network steps of the DT framework for leakage detection.

Calibration Model: To represent real-time sensor data, noise is added to the data of time-dependent variables in the hydraulic model (D & C values). Calibration compares the simulated data (generated during the data input/generation step) and the real-time data from sensors (depicted by adding noise). If the simulated and the actual data match, a conclusion can be drawn that the DT agrees with the system, and the model is calibrated. For this purpose, 50 nodes are randomly selected to work as sensor locations. Pressure and flow rates generated using the D & C values with noise from these locations is equivalent to the real-time actual data. The objective function as described in (1) consisted of mean square error in pressure and flow rates, which is minimized using a differential evolution algorithm, generating the optimized values of the time-dependent parameters, giving the estimated pressure and flow readings from the calibrated model. Finally, a comparison in graphs is drawn for the actual and estimated values. The Objective function and constraints for calibration as described in (1)-(3), is:

$$MSE = \left(\frac{1}{n}\right)\sum_{i=1}^{n}((Pi^{obs} - Pi^{sim})^2 + (Fi^{obs} - Fi^{sim})^2) \quad (1)$$

Where Pi^{obs}, Fi^{obs} *and* Pi^{sim}, Fi^{sim} *are the* observed and predicted values of Pressure and flowrates in the node *'i'*.

Subject to: $0.1 < D < 0.9$ (2)

$85 < C < 150$ (3)

Where D and C are demand multipliers and roughness coefficients. The calibration model was tested for 24 h to obtain a set of optimized D and C values using a differential evolution algorithm to optimize the objective function by trying to improve the candidate solution model. Finally, a comparison in graphs is drawn for the estimated values. The differential evolution algorithm is used to minimize the error function (objective) subject to constraints on the time varying coefficients, for this, appropriate search space is provided to store the optimized values obtained after 30 iterations.

Leakage Detection: Leakage scenarios are created for every pipe to train the algorithm for detecting any potential leakage which is a function of the leak size in the pipe. The logic follows that in case of leak, pipes in vicinity to the leak pipe would experience a pressure drop, which can be detected by the trained model to predict the leak location.

Leakage scenarios are a function of the leak size in a pipe. WNTR simulator provides the ability to add leaks to the network using leak model. Leak can be added to junctions and tanks. WNTR includes methods to add leaks to any location along a pipe by splitting the pipe into two sections. WNTR model leaks, where mass flow rate of fluid through the hole is expressed in (4), (Klise et al. 2020)

$$d_{leak} = Cd \times A \times (P^{\alpha}) \times \sqrt{\frac{2}{\rho}} \qquad ,\alpha=0.5 \qquad (4)$$

and, $\frac{P}{\rho} = g \times h$ (5)

Where, d_{leak} is the leak demand (m^3/s), C_d=discharge coefficient (unitless), A is the area (m^2), and α is an exponent related to characteristic of leak (unitless). P is gauge pressure (Pa), h is gauge head (m), g is acceleration due to gravity (m/s^2) and ρ is density of the fluid (kg/m^3)

21 leak sizes were simulated in the range of 0.05 to 0.3 m diameter in all the pipes. A total of 15,540 leak scenarios were simulated for 24 hour period, generating 372960 data samples. This leakage scenarios dataset is used for training of the NN model to estimate the leak pipe Hydraulic parameters which pose as the main variability in the model are (i) the demand multipliers, (D), and the (ii) roughness coefficients (C). The pressure and flowrate readings for the nodes are the function of these two parameters. Graph rendered using data of sensor node connections is used as input, known as adjacency matrix, and fed to a GCN model to test for accuracy on a sufficiently large test dataset.

Neural Network Modeling and Estimation: A Graph Convolution Neural network (GCN) algorithm has the capabilities to effectively use the neighbourhood information by operating on graph data structures (Şahin and Yüce 2023) resulting in better relationship formulations and pattern recognition to make even more accurate predictions than techniques such as convolution neural networks. Owing to these advantages, a GCN model is used for the prediction of leakage in pipes for which scaled pressure data of sensor nodes is used as input after going through a binary labeler. GCN aims to encode graph structures and features of nodes into low dimensional representations and to morph and modify these representations such that they fit the node labels: G-(V, E), (Toulouse, Dai, and Le 2022) where G is the graph data, V is the set of vertices and E is a set of edges and to classify vertices (Yao, Mao, and Luo 2019) according to some node label. The input for the GCN model is essentially a graph that is sequentially processed through several hidden layers. A Chebyshev convolutional layer (ChebConv) as the initial convolutional layer. For this paper, embedding size 64 and 256 respectively, yielded the optimum results to train 498551 number of parameters to find the optimized values of weights and biases. The train-test set ratio of data is taken as 0.8:0.2. A generic equation for GCN is described by (6) (Şahin and Yüce 2023; Toulouse et al. 2022)

$$Z = \sigma\left(\left(\widetilde{D}^{-\frac{1}{2}}\right) * \tilde{A} * \left(\widetilde{D}^{-\frac{1}{2}}\right) * X * W\right) = \sigma\left(\hat{A}XW\right) \qquad (6)$$

$$\tilde{A} = A + I \qquad (7)$$

$$\widetilde{D}ii = \sum_j \tilde{A}\,ij \qquad (8)$$

Wherein, "σ" in (5), represents the activation function, which for this paper is ReLU and softmax. 'X' denotes the input features i.e., pressure values from sensors. 'W' are the weights that needs to be trained and 'Z' denotes the output of the GCN that holds the k (which is 2 for this paper) degree information of neighbourhood information of neurons. Ã as given, (7) is the normalized adjacency matrix with self-loops. The degree of the normalized adjacency matrix is denoted by D̃ii in (8).

The specifications of the GCN model parameters and a pictorial architecture are described in Table 2 and Fig. 5.

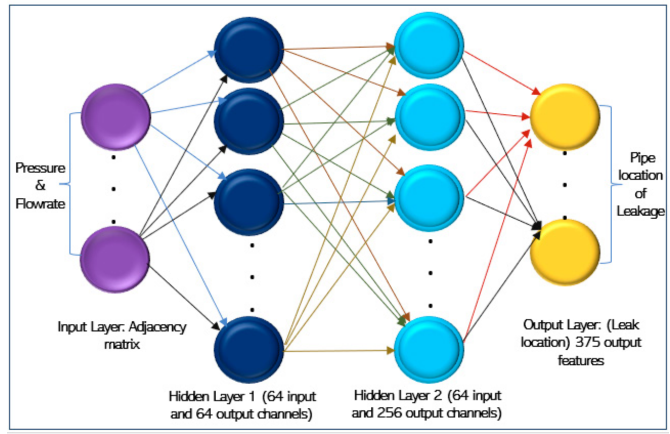


Fig.5 GCN Architecture

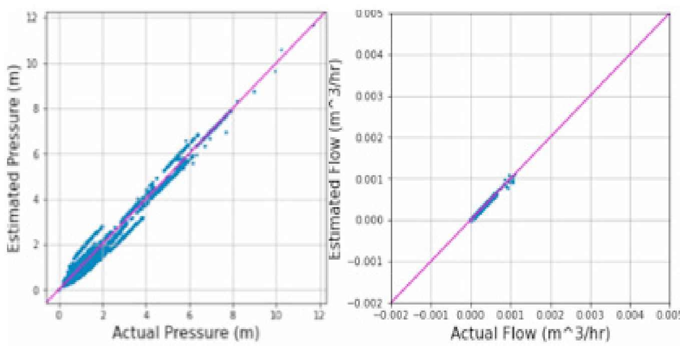
Table 2: Parameters of the GCN Model

Epochs	2700
Activation function	ReLU,softmax
Loss Function	Cross entropy
Optimization method	Adam (lr=0.01)
Model Train-test ratio	0.8:0.2

4. RESULT AND DISCUSSION

This section describes hydraulic calibration model and leakage detection results for the case study network. The results are generated on a 64-bit operating system Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz, 20 GB RAM, x64- intel based processor Dell machine. Python 3.9.12 and Python 3.9.17, WNTR version 1.0.0 and EPANET 2.2 were used for simulations.

Fig. 6 shows the calibration results i.e., the comparison of actual and predicted values of pressure and flow rate of all nodes for 24-hour data. The results show a good overlap between estimated and actual values for all nodes which were also quantitatively assessed with the criteria of R^2 value which is equal to 0.972 for pressure and 0.995 for flow rate data. Thus, the WNTR simulation model holds good with calibrated parameters. Furthermore, accuracy results can also be drawn by comparison of actual and predicted leak locations in test data. As shown in Fig. 7, the first and second row indicates predicted and actual locations of leak respectively in the WDN. The green box indicates correct estimate and red box shows erroneous estimates of leak location. Thus the model demonstrated an accuracy of 81.74%.



(a) (b)

Fig.6: Observed and estimated values comparison of all the nodes for (a) Pressures and (b) flow

```
Cls = pd.DataFrame([test3,pred3])
Cls
```

	0	1	2	3	4	5	6	7	8	9	...	3140	3141	3142	3143	3144	3145	3146	3147	3148	3149
0	P-119	P-76	P-141	p1083	P-216	p927	p499	P-214	p497	P-152	...	P-39	P-268	P-146	P-284	P-17	P-167	P-97	P-268	P-238	P-66
1	P-119	p531	P-141	p1083	P-216	p927	P-130	P-214	p497	P-152	...	P-39	P-227	P-146	P-262	P-15	P-167	P-97	P-227	P-238	P-66

2 rows × 3150 columns

Fig.7: Accuracy on test data, the predicted and actual leak nodes

5. SUMMARY & FUTURE RESEARCH DIRECTIONS

A python-based DT framework for leakage identification tasks in large-scale water distribution systems has been developed using EPANET based WNTR Simulator and Graph Convolution Neural network method for a real case study of IIT Jodhpur campus WDN. It is observed that such decision support frameworks in smart water infrastructure design have practical implications to assist in leakage detection and prevent water losses due to pipe ruptures and high pressures, increasing water use efficiency. The accuracy of presented leakage detection model can be further increased by fine tuning of GCN architecture and hyperparameters. As a subject matter of future research, the authors are also working upon to test this model with the real time data from the field sensors to be deployed in the network.

6. REFERENCES

Brahmbhatt, Parth, Abhilasha Maheshwari, and Ravindra D. Gudi. 2023. "Digital Twin Assisted Decision Support System for Quality Regulation and Leak Localization Task in Large-Scale Water Distribution Networks." *Digital Chemical Engineering* 9(September)

Capelo, Miguel, Bruno Brentan, Laura Monteiro, and Dídia Covas. 2021. "Near–Real Time Burst Location and Sizing in Water Distribution Systems Using Artificial Neural Networks." *Water (Switzerland)* 13(13)

Conejos, Pilar, Fernando Martínez Alzamora, Marta Hervás Carot, and Joan Alonso Campos. 2020. "Building and Exploiting a Digital Twin for the Management of Drinking Water Distribution Networks." *Urban Water Journal* 17:1–10.

Enric Escorsa. 2018. *Digital Twins A Glimpse at the Main Patented Developments*.

Grieves, Michael, and John Vickers. 2016. "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems." *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches* (August):85–113.

Hall, R., B. Bowerman, Joseph Braverman, J. Taylor, Helen Todosow, and U. Wimmersperg. 2000. "The Vision of a Smart City." *2nd Int. Life* .

Klise, K. A., D. B. Hart, Bynum, M., Hogge, J., Haxton, T., Murray, R., Burkhardt, and J. 2020. *Water Network Tool for Resilience (WNTR) User Manual*.

Manzi, Daniel, Bruno Brentan, Gustavo Meirelles, Joaquín Izquierdo, and Edevar Luvizotto. 2019. "Pattern Recognition and Clustering of Transient Pressure Signals for Burst Location." *Water (Switzerland)* 11(11):1–13.

Mounce, Stephen R., and John Machell. 2006. "Burst Detection Using Hydraulic Data from Water Distribution Systems with Artificial Neural Networks." *Urban Water Journal* 3(1):21–31.

Pedersen, Agnethe N., Morten Borup, Annette Brink-Kjær, Lasse E. Christiansen, and Peter S. Mikkelsen. 2021. "Living and Prototyping Digital Twins for Urban Water Systems: Towards Multi-Purpose Value Creation Using Models and Sensors." *Water (Switzerland)* 13(5).

Ramos, Helena M., Maria Cristina Morani, Armando Carravetta, Oreste Fecarrotta, Kemi Adeyeye, P. Amparo López-Jiménez, and Modesto Pérez-Sánchez. 2022. "New Challenges towards Smart Systems' Efficiency by Digital Twin in Water Distribution Networks." *Water (Switzerland)* 14(8).

Rosen, Roland, Georg von Wichert, George Lo, and Kurt D. Bettenhausen. 2015. "About The Importance of Autonomy and Digital Twins for the Future of Manufacturing." *IFAC-PapersOnLine* 48(3):567–72.

Rossman. 2000. *EPANET 2 USERS MANUAL*.

Sadeghioon, Ali M., Nicole Metje, David Chapman, and Carl Anthony. 2018. "Water Pipeline Failure Detection Using Distributed Relative Pressure and Temperature Measurements and Anomaly Detection Algorithms." *Urban Water Journal* 15(4):287–95.

Şahin, Ersin, and Hüseyin Yüce. 2023. "Prediction of Water Leakage in Pipeline Networks Using Graph Convolutional Network Method." *Applied Sciences (Switzerland)* 13(13).

Sankar, Gokul Siva, S. Mohan Kumar, Sridharakumar Narasimhan, Shankar Narasimhan, and S. Murty Bhallamudi. 2015. "Optimal Control of Water Distribution Networks with Storage Facilities." *Journal of Process Control* 32:127–37.

Toulouse, Michel, H. K. Dai, and Truong Giang Le. 2022. "Distributed Load-Balancing for Account-Based Sharded Blockchains." *International Journal of Web Information Systems* 18(2/3):100–116.

Wu, Zheng Yi, Alvin Chew, Xue Meng, Jianping Cai, Jocelyn Pok, Rony Kalfarisi, Kah Cheong Lai, Sock Fang Hew, and Jia Jie Wong. 2023. "High Fidelity Digital Twin-Based Anomaly Detection and Localization for Smart Water Grid Operation Management." *Sustainable Cities and Society* 91(February).

Yao, Liang, Chengsheng Mao, and Yuan Luo. 2019. "Graph Convolutional Networks for Text Classification."In *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33.