Deep Learning-Based Digital Twin Framework for Real-Time Water Network Anomaly Detection

Hussain Ahmad Awad[1,*], Mohammed-Issa Riad Jaradat[2]

[1]Amman Arab University, Jordan
[2]Al Al-Bayt University, Jordan

ABSTRACT

Urban water distribution systems are critical infrastructures that require continuous monitoring to prevent water loss caused by leaks and operational anomalies. Conventional monitoring approaches often struggle to detect early-stage anomalies due to the complexity and temporal dynamics of multivariate sensor data. This study proposes a Deep Learning based Digital Twin framework for real time anomaly detection in urban water networks. The framework integrates a Deep LSTM Autoencoder with a Digital Twin architecture to model normal operational behavior and identify deviations through reconstruction error analysis. The model was trained exclusively on normal operational data and optimized using a Receiver Operating Characteristic based threshold selection strategy. Experimental evaluation on time series sensor data demonstrated that the proposed approach achieved a recall of 0.74 for leak detection and an Area Under the Curve value of 0.750, indicating reliable discrimination between normal and anomalous conditions. The results show that temporal sequence learning significantly enhances anomaly detection capability in complex water distribution systems. By synchronizing real time anomaly detection with a Digital Twin environment, the proposed framework enables proactive monitoring, improved situational awareness, and predictive maintenance support for smart city water management applications.

Keywords Digital Twin, Deep Learning, LSTM Autoencoder, Anomaly Detection, Smart Water Network

INTRODUCTION

Urban water distribution systems play a crucial role in supporting residential, industrial, and public activities within modern cities. As urban populations continue to grow, the demand for reliable and efficient water supply systems increases significantly. However, water leakage remains one of the most persistent challenges in water infrastructure management [1]. Undetected leaks contribute to substantial water loss, financial inefficiencies, increased energy consumption, and long-term infrastructure deterioration. In many regions, non-revenue water caused by leakage represents a significant percentage of total water production, highlighting the need for advanced monitoring and early detection mechanisms [2]. Traditional leak detection methods often rely on manual inspections, acoustic sensing, or simple threshold-based monitoring, which may not adequately capture the dynamic and complex behavior of large-scale water networks.

The rapid development of Internet of Things technologies has enabled the deployment of distributed sensors across water distribution systems, generating large volumes of multivariate time series data [3]. These sensors monitor parameters such as pressure, flow rate, pump status, and water levels in real time. While this data rich environment creates opportunities for intelligent monitoring, it also introduces challenges related to data complexity, temporal

Submitted 15 December 2025
Accepted 28 February 2026
Published 22 August 2026

Corresponding author
Hussain Ahmad Awad,
haawad@aau.edu.jo

Additional Information and Declarations can be found on page 190

DOI: 10.47738/ijrm.v3i3.56

How to cite this article: H. A. Awad, M.-I. R. Jaradat, "Deep Learning-Based Digital Twin Framework for Real-Time Water Network Anomaly Detection," *Int. J. Res. Metav.*, vol. 3, no. 3, pp. 180-192, 2026.

dependencies, and anomaly identification. Conventional statistical techniques and rule-based approaches are often insufficient to model nonlinear interactions among multiple variables over time. As a result, advanced artificial intelligence methods have gained attention for their ability to learn complex patterns directly from data without requiring explicit physical modeling assumptions.

Recent state of the art research has explored the application of machine learning and deep learning techniques for leak detection and anomaly identification in water networks [4]. Supervised learning algorithms such as Support Vector Machines and Random Forest have been employed for classification-based detection tasks. More recently, deep learning architectures including Convolutional Neural Networks and Long Short-Term Memory networks have demonstrated strong performance in modeling temporal dependencies within time series data. In particular, LSTM based Autoencoders have shown promising capability in learning normal operational behavior and detecting anomalies through reconstruction error analysis. At the same time, Digital Twin technology has emerged as an advanced framework for synchronizing physical infrastructure systems with virtual representations, enabling simulation, monitoring, and predictive analytics within smart city environments [5].

Despite these advancements, several research gaps remain in the integration of deep learning-based anomaly detection within Digital Twin architectures for urban water systems. Many existing studies focus either on predictive modeling or on Digital Twin simulation without fully combining both components into a unified real time monitoring framework. Furthermore, threshold selection strategies for anomaly classification are often determined heuristically rather than through systematic optimization methods, which can limit operational reliability. There is also limited research that investigates the use of sequence based Autoencoder models trained exclusively on normal operational data for anomaly detection in water networks, particularly within a synchronized Digital Twin environment that supports real time visualization and decision support.

To address these limitations, this study proposes a Deep Learning based Digital Twin framework for real time anomaly detection in urban water distribution systems. The proposed approach integrates a Deep LSTM Autoencoder trained on normal operational sequences with a Receiver Operating Characteristic based threshold optimization method to enhance detection performance. By embedding anomaly detection outputs into a Digital Twin monitoring environment, the framework enables synchronized visualization, temporal analysis of system behavior, and proactive maintenance decision making. This research contributes to the advancement of intelligent infrastructure management by combining deep learning sequence modeling with Digital Twin technology to improve reliability, efficiency, and sustainability in smart city water networks.

Literature Review

Water leakage detection in urban distribution systems has attracted significant research attention due to its economic and environmental consequences. Early approaches primarily relied on hydraulic modeling and pressure monitoring techniques to identify abnormal system behavior [6], [7]. These methods commonly employed threshold-based analysis and mass balance principles to detect deviations in flow and pressure patterns. Although effective in controlled environments, such techniques often struggle to handle large scale networks with dynamic operational variability [8]. Statistical time series modeling methods were later introduced to improve detection sensitivity, but their ability to model nonlinear multivariate interactions remained limited [9].

With the advancement of machine learning technologies, data driven approaches became increasingly prominent in leak detection applications. Supervised classification models such as Support Vector Machines, Decision Trees, and Random Forests have been widely adopted to distinguish between normal and abnormal operating conditions [10], [11]. These approaches demonstrated improved detection accuracy compared to rule-based systems; however, their performance heavily depends on labeled datasets and effective feature engineering strategies. Ensemble learning techniques were later proposed to enhance classification robustness and reduce variance in prediction outcomes across complex water network datasets [12].

More recently, deep learning architectures have been introduced to address limitations in conventional machine learning methods. Convolutional Neural Networks have been applied to extract spatial features from sensor signals, while Recurrent Neural Networks and Long Short-Term Memory models have shown superior capability in modeling temporal dependencies within multivariate time series data [13], [14]. LSTM based Autoencoders, in particular, have gained attention for unsupervised anomaly detection tasks, where the model learns normal operational behavior and identifies anomalies through reconstruction error analysis [15], [16]. These methods reduce reliance on labeled anomaly data and are well suited for real world infrastructure systems characterized by rare fault events [17].

In parallel with artificial intelligence advancements, Digital Twin technology has emerged as a transformative framework for infrastructure monitoring and smart city applications. A Digital Twin provides a synchronized virtual representation of physical assets, enabling real time monitoring, simulation, and predictive analysis [18]. In the context of water distribution networks, Digital Twins have been utilized for hydraulic simulation, operational optimization, and system resilience assessment [19]. However, many implementations focus primarily on visualization and simulation rather than embedding advanced anomaly detection mechanisms within the Digital Twin environment [20].

Despite substantial progress in both deep learning-based anomaly detection and Digital Twin technologies, limited research has fully integrated temporal deep learning models into a unified Digital Twin framework for real time water leakage detection. This gap highlights the need for comprehensive approaches that combine sequence-based learning, systematic threshold optimization, and synchronized virtual monitoring to enhance intelligent water infrastructure management.

Methodology

This study proposes a Deep Learning based Digital Twin framework for real time anomaly detection in urban water distribution systems by integrating multivariate time series modeling with a synchronized virtual monitoring environment. The overall research workflow is illustrated in figure 1, which outlines the sequential stages of the proposed framework, including data preprocessing, temporal sequence construction, deep learning model training, anomaly scoring and threshold optimization, and Digital Twin integration. These stages form a comprehensive pipeline that transforms raw sensor data into actionable anomaly detection outputs within a virtual monitoring system.

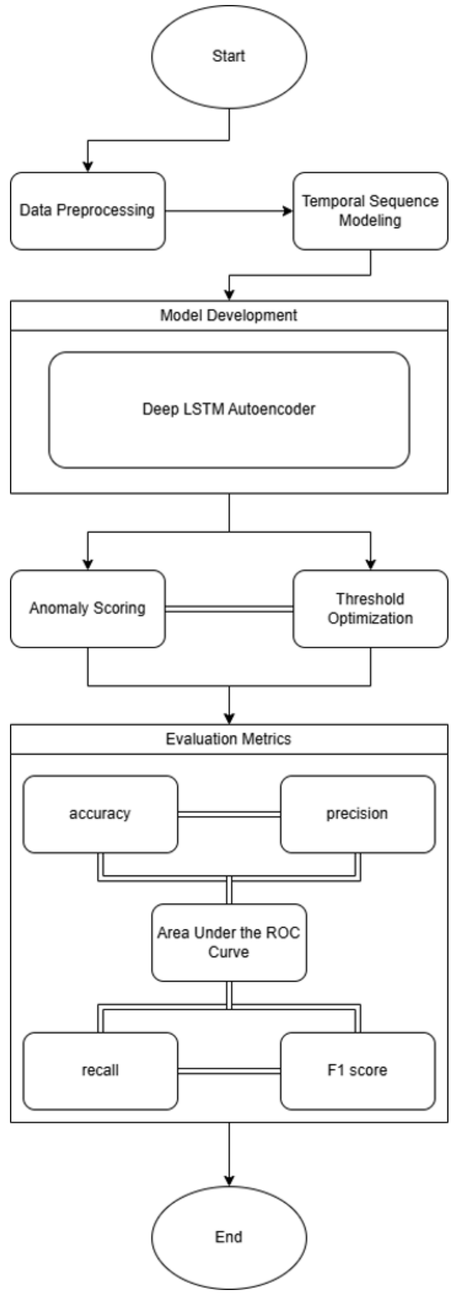

Figure 1 **Research Step**

The dataset consists of sensor measurements such as pressure levels, flow rates, pump status indicators, and other hydraulic variables, each associated with a timestamp and a binary anomaly label. The preprocessing stage begins with merging multiple training datasets into a unified dataset to increase variability and robustness. All timestamps are converted into standardized datetime format and sorted chronologically to preserve temporal continuity. Feature scaling is applied using Min Max normalization to ensure numerical stability during neural network training. Each feature is transformed according to

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

which rescales the data into the range between 0 and 1 and prevents dominant influence from features with larger numeric magnitudes.

To capture temporal dependencies within the water distribution system, a sliding window mechanism is employed to construct sequential input samples. Given a predefined time step length of $T = 20$, each input sequence is defined as

$$S_t = \{X_{t-19}, X_{t-18}, \dots, X_t\} \quad (2)$$

$X_t$represents the multivariate feature vector at time step t. The corresponding label is assigned based on the operational state at time $t + 1$. This sequence-based representation allows the model to learn dynamic system behavior over time rather than analyzing isolated observations.

A Deep Long Short-Term Memory Autoencoder is developed to model normal operational patterns. The encoder consists of stacked LSTM layers that compress the sequential input into a lower dimensional latent representation. The decoder reconstructs the original sequence from this latent vector. The model is trained using the Mean Squared Error loss function defined as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{X}_i)^2 \quad (3)$$

$X_i$denotes the original input sequence and $\hat{X}_i$denotes the reconstructed sequence. The Adam optimizer with a learning rate of 0.0005 is employed to minimize reconstruction loss, and early stopping is implemented to prevent overfitting. Importantly, the model is trained exclusively on sequences labeled as normal, enabling it to learn stable operational dynamics while producing higher reconstruction errors for abnormal patterns.

During inference, anomaly scores are computed using the mean absolute reconstruction error across all time steps and features as

$$Error_t = \frac{1}{T \times F}\sum_{i=1}^{T}\sum_{j=1}^{F} | X_{ij} - \hat{X}_{ij} | \quad (4)$$

Trepresents the number of time steps and Frepresents the number of features. Higher reconstruction error values indicate greater deviation from learned normal behavior. To determine the optimal anomaly classification threshold, a Receiver Operating Characteristic based Youden Index method is applied. The Youden Index is calculated as

$$J = TPR - FPR \quad (5)$$

and the threshold that maximizes J is selected as the optimal decision boundary.

This systematic threshold optimization enhances detection reliability compared to heuristic threshold selection methods.

Finally, the anomaly detection outputs are integrated into a Digital Twin environment that synchronizes real time sensor inputs with the trained deep learning model. Reconstruction error values and detected anomalies are visualized within the virtual representation of the water network, enabling continuous monitoring, early warning generation, and proactive maintenance decision support. Through the integration of temporal deep learning modeling, optimized thresholding, and Digital Twin synchronization, the proposed methodology establishes a comprehensive framework for intelligent real time water infrastructure monitoring in smart city applications.

Algorithm 1 Deep Learning Based Digital Twin Anomaly Detection

Input: Training dataset D_{train}, Test dataset D_{test}, Window length T, Learning rate α
Output:
Predicted anomaly labels $\hat{y}_t$, Optimal threshold τ^*
Process:
Start
Normalize dataset

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Construct sequential data using sliding window
For $t = T$ to N:

$$S_t = \{X_{t-T+1}, X_{t-T+2}, \dots, X_t\}$$

Train LSTM Autoencoder using normal sequences
Define encoder

$$h_t = Encoder(S_t)$$

Define decoder

$$\hat{S}_t = Decoder(h_t)$$

Compute reconstruction loss

$$L = \frac{1}{n} \sum || S_t - \hat{S}_t ||^2$$

Update model parameters

$$\theta \leftarrow \theta - \alpha \nabla_\theta L$$

End training
For each test sequence S_t:
Compute reconstruction error

$$E_t = \frac{1}{T \cdot F} \sum_{i=1}^{T} \sum_{j=1}^{F} | S_{ij} - \hat{S}_{ij} |$$

Compute ROC curve

$$J(\tau) = TPR(\tau) - FPR(\tau)$$

Select optimal threshold

$$\tau^* = arg\,max\, J(\tau)$$

Classify anomaly
If $E_t > \tau^*$
$\hat{y}_t = 1$
Else
$\hat{y}_t = 0$
Update Digital Twin state

End

Result

The optimized Deep LSTM Autoencoder was evaluated using 2,069 time-sequence samples derived from the test dataset to assess its ability to distinguish between normal operational conditions and leak events. To enhance anomaly sensitivity, the model was trained exclusively on normal operational data so that it could learn the underlying temporal patterns of stable system behavior without being influenced by anomalous signals. This training strategy allows the autoencoder to reconstruct normal sequences with low error while producing higher reconstruction errors for abnormal patterns. The anomaly classification threshold was determined using a Receiver Operating Characteristic based Youden Index approach, which identifies the point that maximizes the difference between the true positive rate and the false positive rate. The optimal reconstruction error threshold was calculated as 0.05856, indicating the value that provides the best balance between detection sensitivity and false alarm rate.

The classification performance of the proposed model is illustrated in figure 2, which presents the confusion matrix of the optimized Deep LSTM Autoencoder. The confusion matrix provides a detailed breakdown of true positives, true negatives, false positives, and false negatives, enabling a clear evaluation of detection effectiveness. By analyzing this matrix, it is possible to assess how well the model identifies leak events while maintaining correct classification of normal conditions. This visualization supports a comprehensive understanding of the model's behavior and highlights the trade-off between sensitivity and false alarm generation within the Digital Twin based anomaly detection framework.

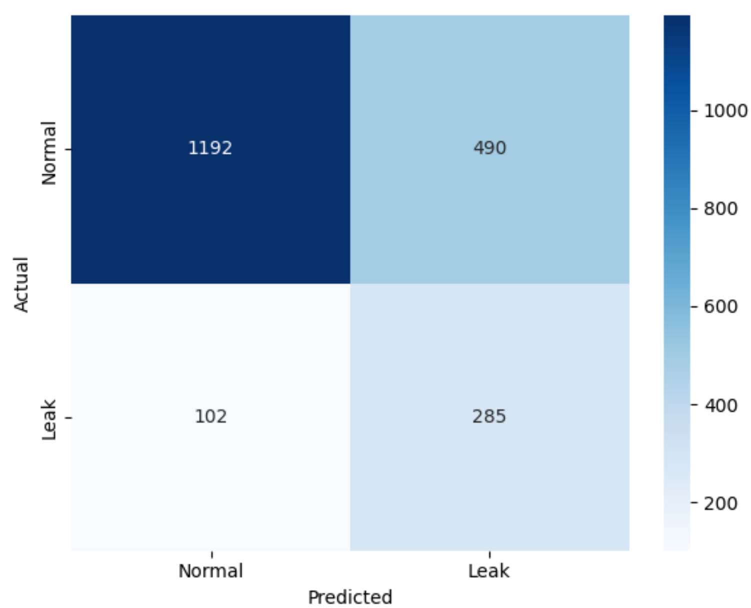

Figure 2 Confusion Matrix of the Optimized Deep LSTM Autoencoder

From the confusion matrix, the model correctly classified 1,192 normal operational instances as normal and successfully detected 285 leak events. These correctly identified samples represent the true negatives and true positives of the system, indicating that the model is capable of recognizing stable operational behavior while also identifying a substantial portion of anomalous conditions. However, 490 normal samples were incorrectly classified as leaks, which are categorized as false positives. In addition, 102 actual leak events were not detected by the model and were classified as normal conditions, representing false negatives. These outcomes highlight the inherent trade-off between sensitivity and specificity in anomaly detection systems.

Overall, the model achieved an accuracy of 71 percent across the test dataset. While accuracy provides a general indication of classification performance, it does not fully capture the model's effectiveness in identifying minority leak events. Therefore, additional evaluation metrics are necessary to provide a more comprehensive assessment of detection capability. A detailed summary of precision, recall, F1 score, and ROC AUC values is presented in table 1, offering deeper insight into the balance between leak detection sensitivity and false alarm generation within the proposed Digital Twin based monitoring framework.

Table 1 **Performance Metrics of the Optimized Model**

Metric	Value
Accuracy	0.71
Precision (Leak)	0.37
Recall (Leak)	0.74
F1-score (Leak)	0.49
ROC-AUC	0.750

As shown in Table 1, the model achieved a recall value of 0.74 for the leak class, meaning that 74 percent of actual leak events were correctly identified by the system. This result indicates that the proposed Deep LSTM Autoencoder demonstrates strong sensitivity to anomalous conditions within the water distribution network. In practical smart city water management scenarios, high recall is particularly important because undetected leaks may result in substantial water loss, increased operational costs, and potential infrastructure damage. Therefore, the ability of the model to detect the majority of leak events reflects its suitability for safety critical monitoring applications integrated within a Digital Twin environment.

The precision value for the leak class was 0.37, indicating that a portion of detected leak events corresponded to false alarms. While this moderate precision reflects the presence of false positives, such behavior is common in anomaly detection systems that are optimized for higher sensitivity. In infrastructure monitoring contexts, prioritizing recall over precision is often a deliberate design choice, since missing a true leak event can have more severe consequences than generating additional inspection alerts. The trade-off observed in the results suggests that the model is calibrated toward proactive detection, which aligns with preventive maintenance strategies in intelligent urban water networks.

The discriminative capability of the proposed framework is further demonstrated by the ROC curve shown in figure 3. The ROC curve evaluates the relationship between true positive rate and false positive rate across various threshold settings, providing a comprehensive assessment of model separability. The achieved AUC value of 0.750 indicates good classification performance and confirms that reconstruction error effectively differentiates between normal and leak conditions. This result demonstrates that the Deep LSTM Autoencoder successfully captures temporal dependencies in multivariate sensor data, reinforcing its effectiveness within the proposed Digital Twin based anomaly detection framework.

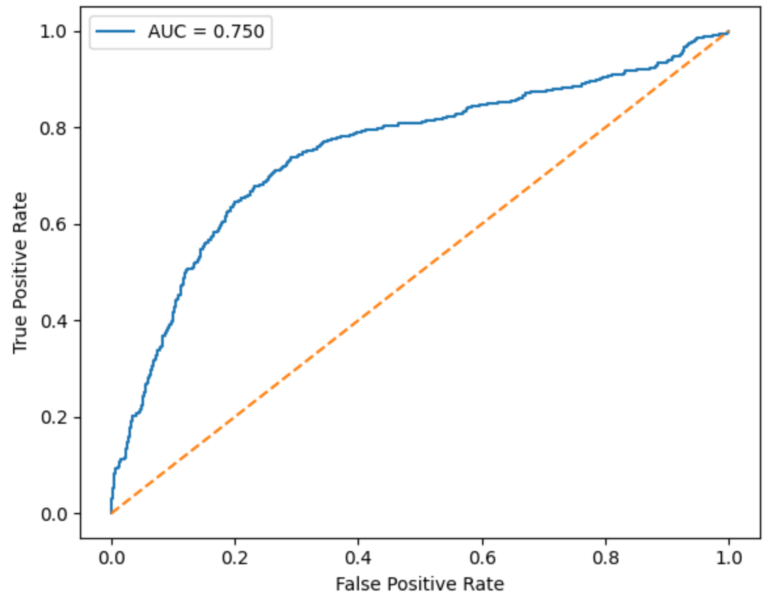

Figure 3 ROC Curve of the Optimized Deep LSTM Autoencoder

The model achieved an Area Under the Curve value of 0.750, which indicates a good level of separability between normal operational conditions and leak events. An AUC value above 0.70 generally reflects reliable classification capability, suggesting that the reconstruction error generated by the Deep LSTM Autoencoder provides meaningful distinction between stable and abnormal system states. The ROC curve remains consistently above the diagonal reference line, which represents random classification performance. This behavior confirms that the proposed model performs substantially better than chance and that reconstruction error serves as an effective anomaly scoring mechanism within the Digital Twin based monitoring framework.

In addition to conventional classification metrics, the temporal detection behavior of the model is illustrated in figure 4, which presents the real time reconstruction error along with the detected leak events. The figure demonstrates how anomaly scores evolve over time and how leak events correspond to significant increases in reconstruction error beyond the optimal threshold. Distinct temporal clusters of elevated reconstruction error can be observed, indicating that the model captures sustained abnormal operational patterns rather than isolated fluctuations. This temporal consistency is essential for real world deployment in smart water distribution systems, where early identification of persistent anomalies enables proactive maintenance and resource optimization within the Digital Twin environment.

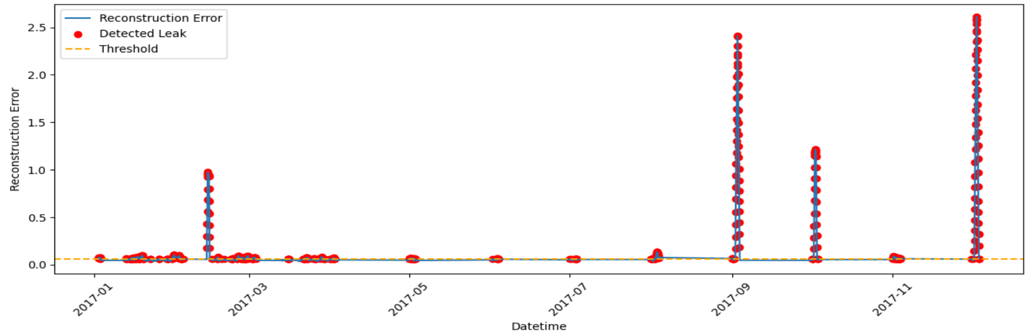

Figure 4 **Real-Time Reconstruction Error and Detected Leak Events**

Several distinct anomaly clusters were observed throughout the monitoring period, particularly during February, September, October, and December, where reconstruction error values exceeded the optimal threshold. These results indicate that the proposed Deep Learning-Based Digital Twin framework successfully captures sustained abnormal operational patterns rather than isolated outliers. The ability to detect temporal anomaly clusters demonstrates the suitability of the framework for real-time deployment in smart city water management systems integrated within a digital twin environment.

Discussion

The experimental results indicate that the proposed Deep Learning-based Digital Twin framework demonstrates reliable capability in detecting leak events within an urban water distribution system. The achieved recall value of 0.74 shows that the model successfully identifies the majority of anomalous conditions, which is a critical requirement in infrastructure monitoring applications [21], [22]. In real-world water management systems, undetected leaks may result in significant water loss, increased operational expenses, and gradual infrastructure deterioration [23]. Therefore, the strong sensitivity achieved by the Deep LSTM Autoencoder suggests that the model effectively captures abnormal temporal patterns in multivariate sensor data [24]. The ROC AUC value of 0.750 further supports this finding by confirming that the reconstruction error provides meaningful separation between normal and leak conditions. These results validate that temporal sequence learning through LSTM architecture enhances anomaly detection performance compared to simple threshold-based monitoring approaches [25].

Although the model shows strong sensitivity, the presence of false positives indicates that there is still room for optimization in terms of precision. In safety-critical environments such as urban water networks, prioritizing higher recall is often justified because missing a true leak event can cause more severe consequences than generating additional inspection alerts [21], [23]. However, reducing false alarm rates would improve operational efficiency and reduce unnecessary maintenance actions [22]. The temporal analysis of reconstruction error also demonstrates that the model detects sustained abnormal patterns rather than isolated fluctuations, which is important for practical deployment in real-time monitoring systems [24]. When integrated into a Digital Twin environment, these anomaly scores can be visualized and interpreted in a synchronized virtual representation of the physical network, enabling proactive decision-making and predictive maintenance strategies [26]. Overall, the findings highlight the potential of combining deep learning and Digital Twin technology to enhance intelligent water infrastructure management in smart city applications [25], [26].

Conclusion

This study presented a Deep Learning based Digital Twin framework designed for real time anomaly detection in urban water distribution systems by integrating a Deep LSTM Autoencoder with continuous sensor data monitoring. The model was trained exclusively on normal operational conditions to learn stable temporal patterns and applied an ROC based threshold optimization strategy to enhance anomaly discrimination capability. Experimental results demonstrated that the proposed framework achieved a recall of 0.74 for leak detection and an AUC value of 0.750, indicating reliable capability in distinguishing between normal and abnormal operational states. These findings confirm that temporal sequence modeling through LSTM networks effectively captures multivariate dependencies within complex water network sensor data. Although the system produced a moderate number of false positives, the prioritization of higher recall aligns with the operational requirements of safety critical infrastructure, where early leak detection is essential to prevent water loss, economic impact, and structural degradation. The integration of anomaly detection outputs within a Digital Twin environment further strengthens the practical contribution of this research by enabling synchronized visualization, real time monitoring, and proactive maintenance decision support. Overall, the proposed framework provides a scalable and intelligent solution for smart city water management and establishes a foundation for future enhancements through spatial modeling, cost sensitive optimization, and explainable artificial intelligence integration.

Declarations

Author Contributions

Conceptualization: H.A.A., M.-I.R.J.; Methodology: H.A.A., M.-I.R.J.; Software: H.A.A.; Validation: M.-I.R.J.; Formal Analysis: H.A.A.; Investigation: M.-I.R.J.; Resources: H.A.A., M.-I.R.J.; Data Curation: H.A.A.; Writing – Original Draft Preparation: H.A.A.; Writing – Review and Editing: M.-I.R.J.; Visualization: H.A.A.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Puust, Z. Kapelan, D. A. Savić, and T. Koppel, "A review of methods for leakage management in pipe networks," *Urban Water J.*, vol. 7, no. 1, pp. 25–45, Feb. 2010, doi: 10.1080/15730621003610878.

[2] A. O. Lambert and T. Hirner, "Losses from water supply systems: Standard terminology and recommended performance measures," *IWA Publishing*, vol. 2000, no. May, pp. 1–40, May 2000.

[3] M. Singh and S. Ahmed, "IoT based smart water management systems: A systematic review," *Mater. Today Proc.*, vol. 46, no. 11, pp. 5211–5218, Oct. 2020, doi: 10.1016/j.matpr.2020.08.588.

[4] M. Romano, Z. Kapelan, and D. Savić, "Automated detection of pipe bursts and other events in water distribution systems," *J. Water Resour. Plan. Manage.*, vol. 140, no. 4, pp. 457–467, Apr. 2014, doi: 10.1061/(ASCE)WR.1943-5452.0000339.

[5] F. Tao, Q. Qi, A. Liu, and A. Kusiak, "Digital twins and cyber–physical systems toward smart manufacturing and Industry 4.0," *Engineering*, vol. 5, no. 4, pp. 653–661, Aug. 2019, doi: 10.1016/j.eng.2019.01.014.

[6] S. M. Mounce, J. B. Boxall, and R. M. Machell, "Development and verification of an online artificial intelligence system for detecting bursts and other abnormal flows," *J. Water Resour. Plan. Manage.*, vol. 136, no. 3, pp. 309–318, May 2010, doi: 10.1061/(ASCE)WR.1943-5452.0000030.

[7] R. S. Pudar and J. A. Liggett, "Leaks in pipe networks," *J. Hydraul. Eng.*, vol. 118, no. 7, pp. 1031–1046, Jul. 1992, doi: 10.1061/(ASCE)0733-9429(1992)118:7(1031).

[8] A. F. Colombo, P. Lee, and B. W. Karney, "A selective literature review of transient-based leak detection methods," *J. Hydro-Environ. Res.*, vol. 2, no. 4, pp. 212–227, Mar. 2009, doi: 10.1016/j.jher.2009.02.003.

[9] J. Mashford, D. De Silva, D. Marney, and S. Burn, "An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine," *NSS*, vol. 2009, no. Oct., pp. 1–8, Oct. 2009, doi: 10.1109/NSS.2009.38.

[10] I. Lučin, B. Lučin, Z. Carija, and A. Sikirica, "Data-driven leak localization in urban water distribution networks using big data for random forest classifier," *Mathematics*, vol. 9, no. 6, pp. 1–15, Mar. 2021, doi: 10.3390/math9060672.

[11] T. Ma, "Data-driven leak detection and identification in water distribution networks using transductive long short-term memory," *ICDSIS*, vol. 2024, no. May, pp. 1–10, May 2024, doi: 10.1109/ICDSIS61070.2024.10594295.

[12] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *ACM SIGKDD*, vol. 2016, no. Aug., pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[15] Y. Chen, J. Liu, and L. Peng, "Auto-encoding variational Bayes," *Cambridge Explor. Arts Sci.*, vol. 2, no. 1, pp. 1–10, Feb. 2024, doi: 10.61603/ceas.v2i1.33.

[16] P. Malhotra, L. Vig, P. Shroff, and G. Agarwal, "Long short-term memory networks for anomaly detection in time series," *ESANN*, vol. 2015, no. Apr., pp. 89–94, Apr. 2015.

[17] A. I. Provotar, Y. Linder, and M. M. Veres, "Unsupervised anomaly detection in time series using LSTM-based autoencoders," *ATIT*, vol. 2019, no. Dec., pp. 1–6, Dec. 2019, doi: 10.1109/ATIT49449.2019.9030505.

[18] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," *Transdisciplinary Perspectives on Complex Systems*, vol. 2017, no. May, pp. 85–113, May 2017, doi: 10.1007/978-3-319-38756-7_4.

[19] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital twin: Enabling technologies, challenges and open research," *IEEE Access*, vol. 8, no. 2020, pp. 108952–108971, Jun. 2020, doi: 10.1109/ACCESS.2020.2998358.

[20] Y. Liu, L. Chen, D. Shangguan, and C. Yu, "A data-driven fault diagnosis method for marine steam turbine condensate system based on deep transfer learning," *Machines*, vol. 13, no. 8, pp. 1–15, Aug. 2025, doi: 10.3390/machines13080708.

[21] H. Wen, X. Gu, J. Han, X. Huo, Z. Ye, and W. Lu, "A review of anomaly detection methods for smart water distribution networks," *AIAAS*, vol. 1, no. Mar., pp. 1–8, Mar. 2026, doi: 10.1145/3813723.3813732.

[22] P. M. Doss, M. Rokstad, and F. Tscheikner-Gratl, "The performance of encoder–decoder neural networks for leak detection in water distribution networks," *Water Supply*, vol. 2024, no. May, pp. 1–12, May 2024, doi: 10.2166/ws.2024.174.

[23] G. Rajan and S. Li, "A systematic literature review on flow data-based techniques for automated leak management in water distribution systems," *Smart Cities*, vol. 8, no. 3, pp. 78, Jun. 2025, doi: 10.3390/smartcities8030078.

[24] P. M. Doss, M. Rokstad, and F. Tscheikner-Gratl, "The performance of encoder–decoder neural networks for leak detection in water distribution networks," *Water Supply*, vol. 2024, no. May, pp. 1–12, May 2024, doi: 10.2166/ws.2024.174.

[25] R. J. L. Taloma, F. Cuomo, D. Comminiello, and P. Pisani, "Machine learning for smart water distribution systems: Exploring applications, challenges and future perspectives," *Artif. Intell. Rev.*, vol. 58, no. 2025, pp. 1–25, Jan. 2025, doi: 10.1007/s10462-024-11093-7.

[26] M. H. Homaei, V. Morales, O. Mogollon Gutierrez, R. Molano Gómez, and A. Caro, "Smart water security with AI and blockchain-enhanced digital twins," *arXiv*, vol. 2025, no. Apr., pp. 1–15, Apr. 2025, doi: 10.48550/arXiv.2504.20275.