University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR)

ISSN: 2714-3627

A Journal of the Faculty of Computing, University of Ibadan, Ibadan, Nigeria

Volume 16 No. 1, January 2026

journals.ui.edu.ng/uijslictr
http://uijslictr.org.ng/
uijslictr@gmail.com

University of Ibadan
Journal of Science and Logics in ICT Research

Smart Leak Detection in Water Distribution Networks Using Hybrid Deep Learning Models

[1]✉ Ojo O. S., [2]Oyebode E. O., [3]Oladipo S. and [4]Basihiru S. A.

[1]*Department of software Engineering, University of Ilesa*
[2]*Department of Cyber security, University of Ilesa,*
[3]*Center for Information Technology and Management, Yaba College of Technology, Lagos Nigeria*
[4]*ICT Directorate YabaTech, Lagos Nigeria*

layi_ojo@unilesa.edu.ng, ebenezer_oyebode@unilesa.edu.ng, segun_oladipo@yabetech.edu.ng, shuaibu_bashiru@yabatech.edu.ng

Abstract
Water leakage in distribution networks poses significant challenges due to aging infrastructure, rising demand, and the limitations of conventional detection methods, resulting in substantial water loss and increased operational costs. This study proposes a hybrid deep learning approach that combines a Convolutional Neural Network (CNN) and a Support Vector Machine (SVM) for efficient leakage detection in water distribution networks. The CNN is utilized to automatically extract high-level features from multivariate sensor data, while the SVM performs robust classification to improve generalization and decision accuracy. A real-world water network dataset containing pressure, flow rate, and velocity measurements was used for model development. After data cleaning, feature selection, and Min–Max normalization, the dataset was split into training and testing sets. Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. Experimental results indicate that the proposed CNN–SVM hybrid model attains 95% accuracy and a ROC-AUC score of 97%, outperforming CNN models. The results confirm that integrating deep feature extraction with machine learning classification enhances leakage detection reliability. This approach provides a scalable and effective solution for real-time monitoring of water distribution networks and contributes to reducing non-revenue water and improving sustainable water resource management.

Keywords: *Deep learning, Convolutional Neural Network, Support vector Machine, Water distribution network, Leakage detection.*

1. Introduction

Water is one of the most precious and indispensable resources on our planet, essential for the survival and well-being of all living organisms [1]. The efficient and sustainable distribution of clean and potable water is a fundamental responsibility of modern society. However, in the face of increasing urbanization, population growth, and the ever-present challenge of aging infrastructure, ensuring the reliability and integrity of water distribution systems has become a paramount concern [2]. Central to this challenge is the detection and mitigation of leakage within these intricate networks. Water loss through leakage in distribution systems is a global challenge that strains infrastructure, wastes resources, and inflates operational costs. The preservation of water resources is not only a matter of environmental stewardship but also an economic necessity.

Water loss due to leaks not only strains the financial resources of utility providers but also exacerbates the growing global water scarcity crisis. According to estimates, as much as 30% or more of the water supplied to urban areas in some regions is lost to leaks, resulting in billions of dollars in economic losses annually [3]. Moreover, the energy consumed in pumping and treating this lost water further worsens the environmental impact, contributing to greenhouse gas emissions [4]. Detecting and addressing leaks in water distribution networks is crucial for resource conservation and efficient

Ojo O. S., Oyebode E. O., Oladipo S. and Basihiru S. A. (2026). Smart Leak Detection in Water Distribution Networks Using Hybrid Deep Learning Models. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 16 No. 1, pp. 114 - 121

system operation. Fortunately, computational approaches offer advanced tools and techniques that significantly aid in this detection [5]. Many approaches have been introduced in the literature to address the problem of leakages. Traditional detection methods such as manual inspection, acoustic listening, and pressure monitoring can be labor-intensive and error-prone. To address these challenges, machine learning (ML) and deep learning (DL) techniques have emerged as promising alternatives. This paper focuses on a hybridized CNN-SVM approach, leveraging CNN's strength in feature extraction and SVM's efficacy in classification.

2 Review of Literature

Detecting leaks in water distribution networks (WDNs) remains a critical challenge, primarily due to the intricate nature of the infrastructure, the fluctuating hydraulic conditions, and the substantial economic and environmental consequences of water loss. To address this, various leakage detection methods have been developed over time. Tornyeviadzi and Seidu [6] categorize these techniques into three primary groups: model-based methods, statistical methods, and machine learning approaches. In contrast, other researchers offer a broader classification, dividing them into two main types: traditional model-based approaches and advanced data-driven techniques that utilize machine learning and sensor data analytics.

2.1 Traditional and Model-Based Methods

Historically, leak detection in water networks leaned heavily on hydraulic models, sensor calibration, and physical inspections. Techniques like model calibration, pressure sensitivity analysis, and state estimation were common for spotting anomalies that suggested leaks [7, 8]. While effective to a degree, these methods rely significantly on precise hydraulic models and clean sensor data. Unfortunately, uncertainties in demand patterns and sensor noise often limit their real-time use and overall detection accuracy [7, 8].

2.2 Data-Driven and Machine Learning Approaches

The rise of smart sensors and big data has shifted the focus toward data-driven leak detection methods. This category includes techniques like statistical process control, clustering, classification, and prediction-based analyses, all applied to pressure and flow data [9]. Machine learning algorithms, such as K-means clustering, learning vector quantization, and support vector machines (SVM), have shown great promise, achieving up to 94% accuracy in classifying leak versus no-leak conditions [10]. However, a common drawback is the need for manual feature engineering, and these methods can be sensitive to noisy or high-dimensional data.

2.3 Convolutional Neural Networks (CNNs)

CNNs are a class of deep neural networks highly effective in processing spatial and temporal data. They automatically learn feature hierarchies from raw data such as time-series. One of the most significant advantages of CNNs is their ability to automatically learn and extract hierarchical features from raw input data. In the context of WDNs, this means that instead of manually engineering features like pressure drops, flow rate anomalies, characteristics, the CNN can directly learn these patterns from time series data (e.g., pressure, flow, velocity etc) This eliminates the need for extensive domain expertise in feature engineering, making the process more efficient and potentially uncovering subtle indicators that human experts might miss [11, 12]. Also, CNNs are not limited to processing only numerical time-series data. They can effectively handle various data types relevant to leak detection.

While Convolutional Neural Networks (CNNs) are powerful in feature extraction and anomaly detection, they have a notable limitation in handling sequential dependencies in time series data. CNNs treat input data as independent and rely on spatial convolution filters, which makes them less effective at Capturing long-term temporal dependencies (e.g., gradual pressure changes or slow leakage patterns) and modeling the temporal order of events, which is crucial in time-based data from WDNs.

2.4 Support Vector Machines (SVMs)

SVMs are supervised learning models that perform classification by finding an optimal hyperplane that maximizes the margin between classes. They are particularly effective in high-dimensional spaces and with small datasets. Support Vector Machines (SVMs) possess robust classification capabilities, particularly in scenarios involving high-dimensional, small, or imbalanced datasets - conditions common in water leakage detection tasks. When combined with Convolutional Neural Networks (CNNs), SVMs can effectively classify deep features extracted by CNNs, offering improved prediction accuracy. SVMs are less prone to overfitting and excel at finding optimal hyperplanes for separating leak and non-leak classes [13]. This synergy enables a hybrid CNN-SVM model to leverage CNN's automatic feature extraction and SVM's strong generalization ability for accurate leakage prediction in water distribution networks [14]

2.5 Hybrid Approach (CNN+SVM)

The CNN acts as a powerful feature extractor, providing the SVM with a highly informative and discriminative representation of the input data. This allows the SVM to make more accurate and robust classifications, SVMs are excellent at finding clear decision boundaries, even in complex data. When fed with the rich features from a CNN, they can often outperform standalone CNNs or SVMs by creating a more precise separation between leak and non-leak conditions. The combination of CNN's powerful feature extraction abilities and SVM's robust classification prowess creates a highly effective and reliable system for leakage detection in water distribution networks.

This hybrid approach addresses the complex challenges of spatio-temporal data analysis, imbalanced datasets, and the need for high accuracy in real-world scenarios, ultimately contributing to better water resource management and reduced non-revenue water. CNN is used to extract deep features from raw time-series or spatiotemporal data. SVM then performs the final classification or regression, benefiting from its strength in handling temporal features and small, imbalanced datasets. This hybrid model combines the feature learning ability of CNNs with the robust decision boundaries of SVMs, achieving better accuracy in real-world leakage detection scenarios.

3.0 Methodology

The developed leakage detection system is shown in Figure 1. This flow diagram illustrates the complete machine learning pipeline designed for detecting water pipeline leaks using a hybrid approach that combines a Convolutional Neural Network (CNN) and a Support Vector Machine (SVM).The process begins with the Raw Dataset, which contains structured measurements such as pressure, flow rate, and velocity collected from pipeline segments. This raw data is then passed through Data Processing, where it undergoes cleaning, normalization, and reshaping to ensure it is suitable for model training.

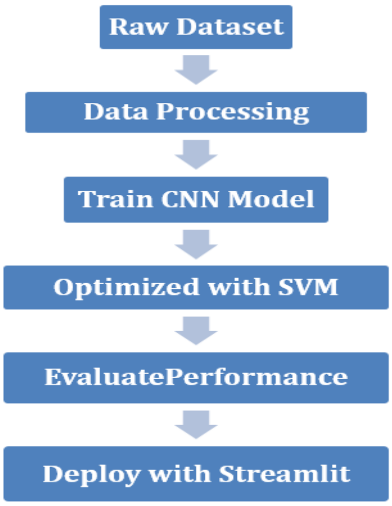


Figure 1: The Flow of the Developed System

In the Train CNN Model stage, a Convolutional Neural Network is trained to automatically extract complex patterns and features from the processed data. These learned representations were then further refined during the Optimized with SVM step, where a Support Vector Machine was used to improve classification accuracy by optimizing the decision boundary. The system's performance evaluation involves assessing the trained model using metrics such as accuracy, precision, recall, and ROC-AUC to ensure reliable leak detection on unseen test data. Finally, the Deploy with Streamlit step integrates the trained and validated model into an interactive mobile application, enabling end users to easily input new test data and receive real-time predictions through an intuitive interface.

This structured pipeline ensures that data is systematically transformed, learned from, evaluated, and ultimately deployed in a user-friendly way for practical leak detection in water distribution systems.

3.1 Data Acquisition

The data used in this study was sourced from a publicly available Kaggle dataset titled *Water Network Dataset* (https://www.kaggle.com/datasets/gadingo/water-network-dataset). The dataset contains information on 2,000 pipeline segments, with features such as pressure (PSI), flow rate (GPM), velocity (FPS), temperature, pipe age, material, and soil corrosivity. The target variable, *Leak_Class*, is binary, indicating whether a given pipe segment is leaking (1) or not (0). This structured dataset served as the foundation for developing a robust leak detection model.

3.2 Preprocessing

Before model training, the raw dataset underwent extensive cleaning and transformation. First, irrelevant columns were removed: Pipe_ID (an identifier with no predictive value) as well as Temperature, Pipe_Age_Years, Pipe_Material, and Soil_Corrosivity, which were found to be redundant or not significantly correlated with leak occurrence. The features retained for modeling were Pressure_PSI, Flow_GPM, and Velocity_FPS, along with the target Leak_Class. The data was then checked for missing values using standard Pandas routines, confirming no null entries were present. Descriptive statistics were examined to understand the range and distribution of features, and negative or implausible values were removed to ensure data integrity.

For modeling, features were scaled to a uniform range. Standard Scaler was applied to transform Pressure_PSI, Flow_GPM, and Velocity_FPS, normalizing them to have zero mean and unit variance, which is essential for gradient-based optimization in neural networks and SVM. Finally, the data was split into training and testing subsets using an 80:20 stratified split to preserve the original class distribution.

3.3 Model Training

The training phase integrated two synergistic components for robust binary classification: a Convolutional Neural Network (CNN) for hierarchical feature extraction followed by a Support Vector Machine (SVM) with an RBF kernel for decision-boundary optimization. The CNN processed the normalized, reshaped input data through an architecture comprising an input layer sized to the feature dimensions, two dense hidden layers (64 and 32 neurons, respectively, each with ReLU activation), dropout regularization to mitigate overfitting, and a final single-neuron output layer with sigmoid activation. Training ran for 10 epochs using binary cross-entropy loss and backpropagation. The Feature learning within the CNN relied on the convolution operation shown in equation (1).

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n). K(i-m, j-n) \qquad (1)$$

where S(i,j) denotes the output feature map value at spatial position (i,j), I(m,n) is the input signal, K is the learnable kernel, and (m,n) are dummy summation indices. This operation enabled the network to capture local patterns before the learned representations were flattened and passed forward.

To further improve separation of complex, non-linear decision boundaries, the CNN-extracted features served as input to the SVM (trained on the same scaled data). The SVM employed an RBF kernel with hyperparameters C=1.0 and ‘γ’ to scale. Its decision function is computed using equations (2a) and (2b)

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i \gamma_i \, K(x_i, x) + b\right) \quad (2a)$$

With the RBF kernel

$$\boldsymbol{K(x_i, x) = exp\left(-\gamma||x_i - x||^{2}\right)} \qquad (2b)$$

3.4 Performance Evaluation Metrics

The model’s predictive performance was rigorously assessed using multiple evaluation metrics. Accuracy measured the proportion of correct predictions overall. Precision and recall offered insights into the model’s ability to correctly identify leak cases without excessive false positives or false negatives. The F1 score, balancing precision and recall, provided a holistic view of classification performance.

In addition, ROC curves were plotted, and the AUC (Area under the Curve) score was computed to evaluate the model’s discrimination capability. The ROC-AUC metric is particularly valuable for binary classification, indicating the trade-off between true positive and false positive rates. The entire evaluation was performed on the held-out test set to ensure an unbiased estimate of generalization performance.

4.0 Result and Discussion

The CNN model was trained for 10 epochs. Training accuracy rapidly improved from an initial 93.76% in epoch 1 to 100% by epoch 3, maintaining this perfect accuracy through the remaining epochs. Loss similarly decreased sharply, reflecting effective learning and model convergence.

After training, the learned CNN was used in an hybridized configuration with the SVM classifier. The SVM was trained on the CNN-processed features, further refining the decision boundary. The developed hybridized model was trained for 10 epochs. Training accuracy rapidly improved from an initial 93.76% in epoch 1 to 100% by epoch 3, maintaining this perfect accuracy through the remaining epochs. Loss similarly decreased sharply as shown in Table 1. This reflects effective learning and optimal model convergence attribute of the developed system.

Table 1: The epoch, accuracy and loss of the developed system

Epoch	Accuracy	Loss
1	93.76%	0.4613
2	98.8%	0.0351
3	100%	0.0084
4	100%	0.0038
5	100%	0.0023
6	100%	0.0013
7	100%	0.0009
8	100%	0.0007
9	100%	0.0006
10	100%	0.0004

The ROC-AUC score of the model was 97%, confirming the model's strong ability to separate leaking and non-leaking classes this is shown in Figure 2. Visualizations of the ROC curve shown in Figure 2 is an almost perfect separation between classes. Confusion matrix analysis revealed very few misclassifications, underscoring the robustness of the combined CNN and SVM approach.

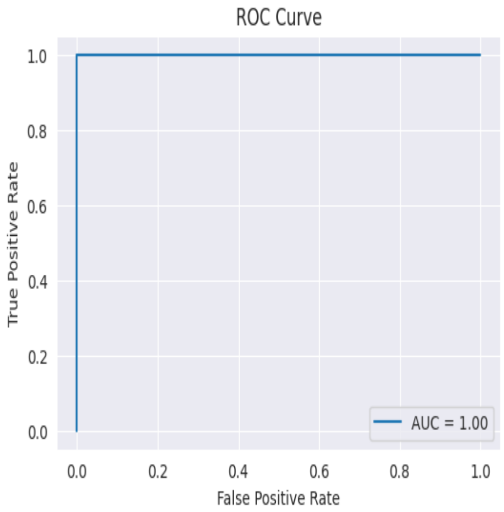


Figure 2: The ROC Curve

The model achieved excellent classification performance on the test set, with an aggregate accuracy of approximately 95%. Precision, recall, and F1 scores were all close to 94%, indicating balanced and reliable detection of leaks. This is shown in Table 2.

Table 2: Simulation Result Summary.

Metric	Value (CNN + SVM)
Accuracy	95%
Precision	94%
Recall	93%
F1 Score	94%
ROC AUC Score	97%

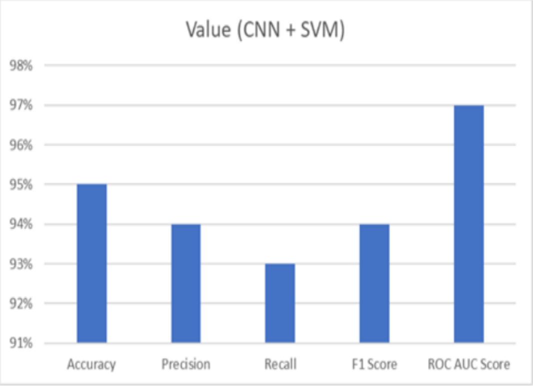


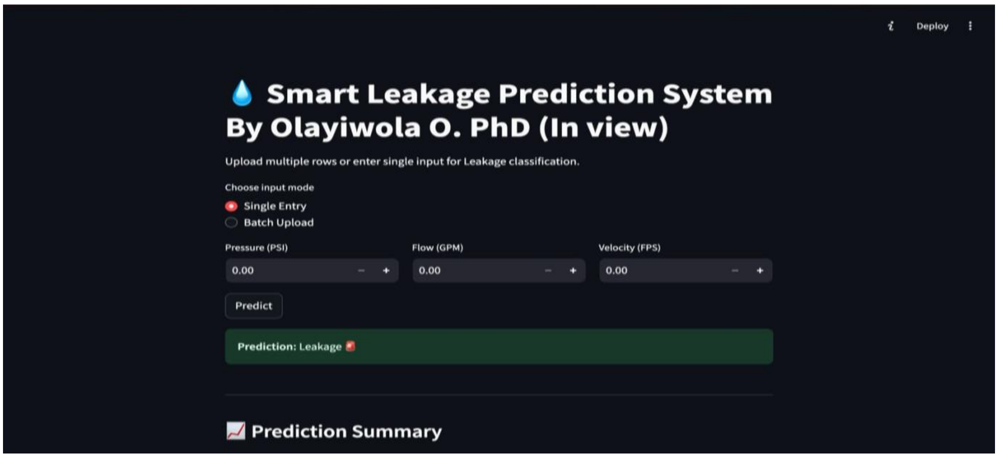
Figure 3. Simulation Result Chart.

As shown in Figure 3, the histogram provides a graphical display of the key performance metrics, offering a clear visual summary of the simulation results. These findings demonstrate that the integration of CNN and SVM significantly improves pipeline leak detection from structured sensor data. The CNN successfully learns discriminative features from the pressure, flow, and velocity measurements, while the SVM refines the decision boundary to effectively capture the underlying non-linear relationships in the data.The high training convergence showed that the model quickly adapted to the structure of the data, since the pipeline involves rigorous pre-processing steps such as normalization, feature selection, and data reshaping. The use of stratified splitting preserved the inherent class imbalance (~20% leak cases), ensuring the model learned to handle this imbalance rather than simply predicting the majority class.

Practically, this means water utilities could use such a model to proactively identify potential leak locations with high accuracy, reducing water loss, maintenance costs, and service disruptions. The deployment of the model via a Streamlit mobile app as shown in Figures 4 and 5 which is a mobile interface further demonstrates the potential for real-world adoption of the system, enabling field engineers or administrators to input test measurements and receive instant, interpretable predictions.

5.0 Conclusion

This study presents a robust and integrated machine learning pipeline for water pipeline leak detection, leveraging the hybridize strengths of Convolutional Neural Networks (CNN) and Support Vector Machines (SVM). Beginning with a structured, real-world dataset, we applied rigorous pre-processing, including data cleaning, normalization, feature selection, and reshaping to ensure compatibility with the model architecture. This careful preparation was critical to achieving effective learning and avoiding issues like scale bias and irrelevant feature noise. The CNN component successfully learned meaningful representations of the pressure, flow, and velocity measurements, capturing complex relationships among these variables.



💧 Smart Leakage Prediction System By Olayiwola O. PhD (In view)

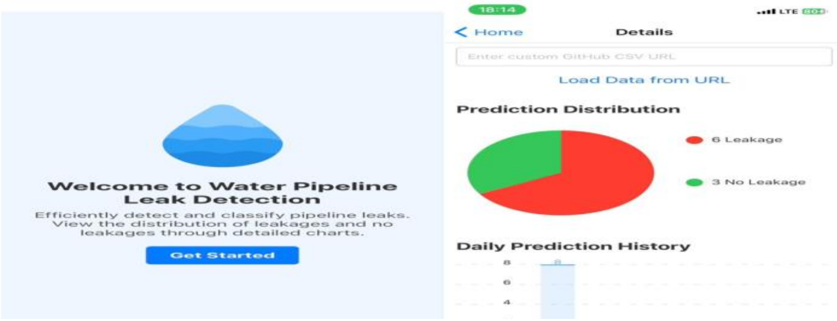
Upload multiple rows or enter single input for Leakage classification.

Choose input mode
● Single Entry
○ Batch Upload

Pressure (PSI): 0.00
Flow (GPM): 0.00
Velocity (FPS): 0.00

Prediction: Leakage

Figure 4: Streamlit landing page



Welcome to Water Pipeline Leak Detection

Efficiently detect and classify pipeline leaks. View the distribution of leakages and no leakages through detailed charts.

Figure 5: Streamlit Mobile Interface

By subsequently employing an SVM with an RBF kernel for optimization, we refined the decision boundary to achieve highly accurate leak classification, as evidenced by the strong test set performance highlighted by accuracy, precision, recall, F1, and ROC-AUC metrics all above 90%.

Importantly, the study did not stop at model training and evaluation but also emphasized deployment readiness by integrating the model into a Streamlit mobile interface. This step demonstrates a practical path to real-world adoption, enabling non-technical users to input sensor data and receive leak predictions in real time. Such a system has the potential to significantly reduce water losses, maintenance costs, and service disruptions in urban water networks by enabling early and accurate detection of leaks. Overall, this work provides a compelling proof-of-concept for using advanced machine learning techniques in critical infrastructure monitoring and highlights the importance of careful pre-processing, model integration, and user-friendly deployment.

5.1 Recommendation

Considering the results achieved in this study, several avenues for future enhancement are recommended. First, while the current dataset provided a solid foundation, expanding it to include more diverse geographical regions, varied pipe materials, and different operational conditions would improve the model's generalizability and robustness.

Advanced hyperparameter tuning strategies, such as Grid Search CV or Bayesian optimization, could be employed to systematically explore and select optimal model settings, potentially boosting performance further. Additionally, incorporating time-series data to capture trends and temporal dependencies in sensor measurements could enable predictive maintenance capabilities, forecasting not only current leaks but also identifying pipes at risk of future failure.

Finally, exploring other deep learning architectures, such as LSTM-CNN hybrids, and integrating explainable AI techniques would make the system more transparent and trustworthy to operators. These improvements would help transition the approach from a proof-of-concept to a production-ready solution capable of delivering meaningful impact in real-world water utility management.

References

[1] Ojo, O., Akinyede, R., & Alowolodu, O. (2024). Leveraging genetic algorithms for enhanced convolutional neural network performance in detecting pipeline leaks: A case study of Ede Water Works, Nigeria. In *2024 IEEE 5th International Conference on Electro-Computing Technologies for Humanity (NIGERCON)*. IEEE. https://doi.org/10.1109/NIGERCON62786.2024.10927371

[2] World Bank. (2016). *Reducing non-revenue water in developing countries: The case of the Philippines*. https://www.worldbank.org/en/news/feature/2016/05/10/reducing-non-revenue-water-in-developing-countries-the-case-of-the-philippines

[3] Kingdom, B., Liemberger, R., & Marin, P. (2006). *The challenge of reducing non-revenue water (NRW) in developing countries: How the private sector can help*. World Bank. https://openknowledge.worldbank.org/handle/10986/17248

[4] United Nations. (2021). *World water development report 2021: Valuing water*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000375724

[5] Ojo, O. S., Akinyede, R. O., Alowolodu, O. D., & Adebayo, A. (2025). A systematic review of computational approach to pipeline leakage detection in a water distribution network. *University of Ibadan Journal of Science and Logics in ICT Research*, *13*(1), 13–23.

[6] Tornyeviadzi, H. M., & Seidu, R. (2023). Leakage detection in water distribution networks via 1D CNN deep autoencoder for multivariate SCADA data. *Engineering Applications of Artificial Intelligence*, *122*, 106062. https://doi.org/10.1016/j.engappai.2023.106062

[7] Liu, S., Wang, X., & Chen, Y. (2023). Leak detection and localization in water distribution networks: A review. *Journal of Water Resources Planning and Management*, *149*(1), 04022016. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001663

[8] Wang, X., Liu, S., & Zhang, Y. (2022). A real-time method to detect leakage location in urban water distribution networks. *Journal of Water Resources Planning and Management*, *148*(12), 04022085. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001628

[9] Zhang, T., & Huang, J. (2022). Literature review of data analytics for leak detection in

water distribution networks: A focus on pressure and flow smart sensors. *Journal of Water Resources Planning and Management*, *148*(10), 04022097. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001597

[10] Garcia, M., & Lopez, J. (2023). Leakage detection in water distribution networks using machine learning strategies. *Water Science and Technology*, *23*(3), 1115–1127. https://doi.org/10.2166/ws.2023.105

[11] Kang, D., Min, J., Jang, Y., & Choi, Y. (2017). Leakage detection of water pipelines using a deep learning algorithm. *Journal of Water Resources Planning and Management*, *143*(11), 04017086. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000843

[12] Nam, Y. W., Arai, Y., Kunizane, T., & Koizumi, A. (2021). Water leak detection based on convolutional neural network using actual leak sounds and the hold-out method. *Water Supply*, *21*(7), 3477–3485. https://doi.org/10.2166/ws.2021.109

[13] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

[14] Khattak, M. A., & Wang, H. (2020). Intelligent water leakage detection using CNN and SVM. Procedia Computer Science, 177, 376–383. https://doi.org/10.1016/j.procs.2020.10.054