

A Hybrid Machine Learning Algorithm for Pipeline Leak Detection and Localisation in Water Distribution Networks

Giresse M. Komba¹, Topside E. Mathonsi², Pius A. Owolawi^{3*}

Department of Computer Systems-Engineering, Information and Communication Technology,
Tshwane University of Technology, Pretoria, South Africa^{1, 3}
Information Technology-Information and Communication Technology,
Tshwane University of Technology, Pretoria, South Africa²

Abstract—Water Distribution Networks (WDNs) frequently experience significant water losses due to pipeline leakages. These losses not only create economic challenges for water utilities but also intensify global concerns regarding water scarcity. This study aims to enhance the accuracy and reliability of leak detection and localisation within WDN infrastructures. Traditional leak detection techniques often exhibit limitations such as high operational costs, inefficient detection processes, and susceptibility to false alarms, particularly when sensors are deployed randomly across the network. Furthermore, detecting concealed or low-intensity leaks remains a difficult task. To address these challenges, this study introduces a hybrid supervised machine learning framework that combines Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Graph Theory (GT). The integration of these techniques enables the proposed model to analyse multiple parameters influencing leak behaviour and improve the reliability of detection outcomes. The hybrid model, referred to as the SVM-ANN-GT algorithm, is evaluated using the EPANET hydraulic simulation environment and compared with conventional machine learning approaches. Experimental results indicate that the proposed hybrid model significantly improves leak detection performance. The model achieves an average detection accuracy of approximately 96%, outperforming standalone SVM and ANN models, which achieved accuracies of 85% and 80%, respectively. The improved performance is primarily attributed to the integration of graph-theoretic optimisation for sensor placement, which enhances monitoring coverage and reduces redundancy within the network.

Keywords—SVM-ANN-GT; leak detection and localisation; EPANET; WDNs; ML

I. INTRODUCTION

Water Distribution Networks (WDNs) represent vital municipal infrastructure responsible for supplying safe and clean water to consumers, ensuring that communities have reliable access to potable water for everyday use [1–3]. These systems typically consist of interconnected components such as reservoirs, storage tanks, pipelines, pumps, and valves, as illustrated in Fig. 1. Proper planning, operation, and maintenance of these elements are necessary to guarantee efficient and dependable water distribution throughout the network [4, 5].

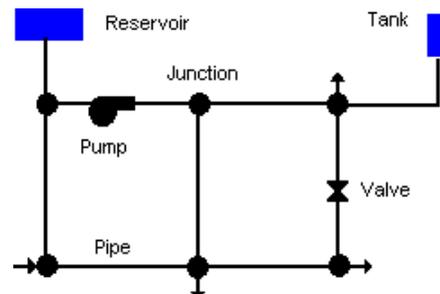


Fig. 1. Typical water distribution networks.

WDNs provide several important benefits to society, including the protection of public health, the maintenance of sanitation and hygiene standards, and the reliable provision of water for essential activities such as drinking, cooking, sanitation, and fire fighting [6]. In addition, these systems play a key role in supporting economic development by supplying water to industries and commercial sectors that rely on a consistent and stable water source [7, 8]. WDNs also contribute to water conservation by minimising losses and promoting fair and efficient water distribution. By maintaining appropriate pressure levels and stable flow conditions within the network, WDNs ultimately improve service reliability and enhance the overall quality of life for communities [9, 10].

Pipelines represent a fundamental component of WDNs because they function as the primary pathways through which water is conveyed to consumers [11, 12]. Their proper functioning is essential for maintaining a reliable and continuous water supply that satisfies the demands of growing populations [13]. However, pipelines are frequently exposed to several environmental and operational risks. These include soil movement, structural deterioration due to ageing, geological disturbances, excessive water pressure, poor connections, and installation faults. Such factors can compromise pipeline integrity and lead to leaks, ultimately causing significant water losses [14].

Water leakage within WDNs is generally classified into three categories: reported leakage, unreported leakage, and unavoidable or background leakage, as illustrated in Fig. 2 [15].

*Corresponding author.

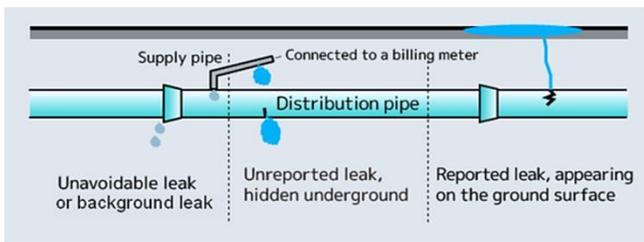


Fig. 2. Categories of water leaks.

Reported leaks are typically visible and are detected and communicated either by consumers or by water utility personnel during routine monitoring [16]. In contrast, unreported leaks remain unnoticed for extended periods because they often occur in buried pipelines or concealed locations, making them difficult to identify without specialised monitoring technologies [17, 18]. Background leakage, on the other hand, occurs naturally within distribution systems as infrastructure ages and materials gradually deteriorate. These leaks are usually small and difficult to detect because they do not produce obvious signs of water loss [19].

The magnitude of water loss resulting from pipeline leaks is substantial. Globally, billions of cubic meters of water are lost every year due to leakage in WDNs. These losses can account for approximately 5% to 50% of the total water supplied, representing a major challenge for sustainable water management and conservation initiatives [20]. In economic terms, water utilities worldwide are estimated to lose approximately 14 billion US dollars annually due to leak-related water wastage [21, 22]. In South Africa, leakage is responsible for about 37% of the country's unexplained water losses, costing the nation nearly 10 billion rands each year [23].

Although several leak detection techniques have been developed, many existing systems still rely heavily on manual inspection methods, which limit their ability to provide continuous or real-time monitoring and often lead to operational inefficiencies [24, 25]. Furthermore, inaccurate detection and frequent false alarms often occur when sensors are installed without a strategic placement strategy, leading to unnecessary maintenance actions and increased operational costs [26]. Detecting and locating small or background leaks remains particularly difficult because these leaks typically occur underground and produce minimal observable changes in system parameters.

To address these challenges, this study proposes the SVM-ANN-GT algorithm, a hybrid Machine Learning (ML) approach that combines Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Graph Theory (GT) to enhance leak detection and localisation in WDNs. Unlike traditional approaches, the proposed framework enables more accurate real-time leak detection while reducing false alarms through improved data classification and decision fusion mechanisms. In addition, the use of graph-theoretic concepts facilitates optimised sensor placement, which improves monitoring coverage, reduces energy consumption, and lowers both deployment and maintenance costs.

In practical water distribution systems, random or poorly planned sensor placement often results in undetected leaks,

increased false alarms, and unnecessary energy consumption caused by redundant or inefficient monitoring nodes. Strategic sensor deployment helps minimise the number of sensors required while still ensuring sufficient monitoring coverage. This approach reduces installation and maintenance expenses while simplifying system management, particularly in large or geographically complex networks. Since this research focuses on pipeline leak detection, efficient sensor placement plays a critical role in enabling rapid and precise leak localisation, which is essential for timely repair interventions and improved water conservation.

The remainder of this study is organised as follows: Section II reviews existing studies on water leak detection techniques. Section III describes the architecture and theoretical foundations of the proposed algorithm. Section IV presents the experimental setup, dataset description, simulation results, and performance evaluation. Finally, Section V summarises the major findings and outlines directions for future research.

II. RELATED WORK

Numerous studies have investigated different approaches for detecting and locating leaks in WDNs.

Porwal et al. [27] introduced a weighted-sample Support Vector Machine (SVM) model designed to enhance leak detection in WDNs. Their method improved classification performance by effectively addressing noisy data and outliers, outperforming conventional techniques such as acoustic signal analysis, transient signal analysis, and temperature variation analysis. Building on these findings, the proposed SVM-ANN-GT algorithm extends the concept by incorporating an optimisation mechanism for sensor placement, which helps reduce operational costs and energy consumption. By combining these elements, the proposed approach offers a more efficient and dependable framework for leak identification in WDNs.

Zhou et al. [28] developed a leak detection framework that integrates SVM with Kernel Principal Component Analysis (KPCA). In their model, KPCA was used to reduce the dimensionality of flow measurement data while extracting meaningful features, after which SVM performed the classification task. The approach was evaluated using Flowmaster simulation software and compared with alternative techniques such as Support Vector Data Description (SVDD) and k-means clustering. Experimental results indicated that the KPCA-SVM model achieved superior detection accuracy compared with the competing approaches. However, the study did not address the effect of sensor placement along the pipeline network, which is an important factor for improving detection performance while reducing installation and operational costs.

Lang et al. [29] proposed a leak detection and localisation technique that combines Least Squares Support Vector Machines (LS-SVM) with Local Maxima Decomposition (LMD). The proposed framework was evaluated using Flowmaster software on both simulated and real-world datasets. Results showed improved accuracy in identifying and locating leaks compared with existing methods such as wavelet transform and particle swarm optimisation. Despite its promising performance, the LS-SVM-LMD approach lacks a

mechanism for strategic sensor deployment. As a result, the system required a large number of sensors, which increased both operational costs and energy consumption. Furthermore, poorly positioned sensors introduced additional noise in the data, which negatively affected overall detection accuracy.

Amini et al. [30] developed an Artificial Neural Network (ANN)-based model for real-time leak detection in pipelines using pressure measurements. The algorithm achieved high precision and recall values, demonstrating strong performance. Nevertheless, the study relied exclusively on pressure data and did not incorporate other relevant parameters such as flow rate or water quality indicators. Additionally, the evaluation was performed using a single dataset, which limits the generalisability of the results. Although this approach shares similarities with the SVM-ANN-GT framework, particularly in the use of pressure-based data, it suffers from limitations related to random sensor placement, which leads to increased operational costs and higher energy consumption. In contrast, the proposed SVM-ANN-GT model integrates graph theory to optimise sensor positioning, thereby improving detection accuracy while minimising energy usage.

Mashhadi et al. [31] presented a machine learning-based method for leak detection and localisation in WDNs by combining logistic regression with spectral clustering. Their objective was to reduce false alarm rates while improving localisation accuracy. The method was implemented on real WDN systems and evaluated using the EPANET hydraulic simulator, with comparisons made against other ML models such as decision trees and ANN. The results showed that their approach achieved a detection and localisation accuracy of approximately 94.3%, outperforming the alternative algorithms. However, the model required a large number of sensors, which increased both operational expenses and energy consumption. Similar to the SVM-ANN-GT algorithm, their approach utilised machine learning techniques to detect and locate leaks, although the proposed method reported a maximum false alarm rate of 1.9%.

Rayaroth [32] introduced the Random Decision Tree Bagging Classifier with Shuffled Frog Leaping Optimisation (RDTBC-SFLO) algorithm for detecting leaks in WDNs. The model combines a random ID3 decision forest classifier with the SFLO optimisation algorithm to improve sensor placement and detection accuracy. The Iterative Dichotomiser 3 (ID3) algorithm constructs decision trees by selecting attributes with the highest information gain, enabling effective classification of pressure measurements as either normal or abnormal. The SFLO optimisation component identifies optimal sensor locations within the WDN, thereby reducing classification errors and improving detection performance. Simulation experiments conducted using EPANET showed that the RDTBC-SFLO approach significantly improved leak detection while reducing classification time compared with existing methods. However, the approach was only tested on small-scale WDN systems, which raises concerns regarding its scalability for larger networks. In contrast, the SVM-ANN-GT algorithm proposed in this study was evaluated on large-scale WDN environments, demonstrating its potential applicability in real-world scenarios.

Wu et al. [33] proposed a two-stage burst detection method based on a data-driven clustering algorithm. Their approach identified anomalies in flow meter data collected from district metering areas and analysed statistical patterns to detect pipeline bursts. One notable advantage of the method was its ability to adapt to non-stationary conditions in historical datasets, eliminating the need for manual data selection. Evaluation using simulated flushing-induced bursts demonstrated that the method could effectively identify relatively large and short-duration bursts while maintaining a low false positive rate. Additionally, the system was able to detect abnormal water usage patterns associated with weather variations, suggesting its applicability in real operational environments. Nevertheless, further investigation is needed to assess the method's performance under noisy conditions or highly dynamic network environments. Moreover, the study did not compare its performance with other burst detection algorithms, making it difficult to assess its relative effectiveness.

Lučin et al. [34] proposed a machine learning-based leak detection framework using a Random Forest (RF) classifier, particularly targeting WDN environments with limited sensor availability. The study used Monte Carlo simulations to generate leak scenarios with varying demand patterns and leak characteristics. Pressure data collected over a 24-hour period were analysed to identify and locate leaks within EPANET-simulated networks. The model was implemented using the Scikit-learn library and supported by high-performance computing resources, achieving reliable detection results even with sparse sensor deployment. However, the study had several limitations. It did not incorporate additional hydraulic or environmental variables, lacked validation using real WDN data, and did not account for prediction uncertainty. Furthermore, the study explored only a single machine learning model, which may limit its robustness. While the proposed SVM-ANN-GT algorithm shares similarities with this approach in terms of ML implementation and sensor placement considerations within a Python framework, the RF-based method struggled to detect small-scale leaks, potentially leading to undetected water losses and operational inefficiencies. The hybrid SVM-ANN-GT algorithm addresses these challenges by integrating SVM, ANN, and GT to optimise network monitoring and improve detection of small leaks while reducing energy consumption and false alarm rates.

Radha and Ramamoorthy [35] proposed a leak detection framework based on association rule learning combined with particle swarm optimisation. Their study highlighted the importance of leak detection for reducing water loss, improving distribution efficiency, and preventing contamination within WDNs. The Apriori algorithm was applied to simulated WDN datasets to identify patterns associated with water leakage by analysing frequent item sets. Particle swarm optimisation was then used to optimise rule thresholds, improving detection accuracy. The authors reported a leak detection rate exceeding 90% with a false alarm rate below 1%. They also compared their method with other association rule mining techniques such as FP-Growth and Eclat, demonstrating superior performance. However, the approach has not yet been validated using real-

world WDN data or tested on large-scale network infrastructures.

Fan et al. [36] proposed a hybrid machine learning approach that integrates Artificial Neural Networks (ANN) with Autoencoder (AE) models for leak detection in WDNs. Using the EPANET simulator, they generated pressure datasets under various conditions, including different user demands, noise levels, and leak severities. While the supervised ANN model achieved high classification accuracy, the study identified challenges related to imbalanced datasets in Wireless Sensor Networks (WSNs). To address this limitation, the authors introduced an unsupervised AE model capable of detecting leaks in imbalanced datasets. Although the AE approach performed well in monitored sensor regions, its effectiveness declined in other network areas. To overcome this limitation, the authors proposed a strategy involving multiple independent detection attempts using the AE model, which significantly reduced false alarms. The method was evaluated using a WSN testbed and demonstrated promising results. Similar to the proposed SVM-ANN-GT algorithm, their approach utilised an ANN and focused on reducing false alarms. However, their framework did not address small background leaks or include sensor placement optimisation.

Overall, many previous studies have applied machine learning techniques such as SVM, ANN, decision tree models, and clustering algorithms to detect leaks in WDNs. Nevertheless, most of these approaches do not integrate real-time leak detection with optimised sensor placement strategies. In contrast, the proposed SVM-ANN-GT algorithm combines accurate ML-based classification with graph theory-based sensor placement, resulting in a more robust, energy-efficient, and cost-effective leak detection framework.

III. SYSTEM DESIGN AND MODELING

A. Proposed Leak Detection and Architecture

The Internet of Things (IoT)-based Wireless Sensor Network (WSN) architecture designed for WDNs, illustrated in Fig. 3, provides an advanced framework for detecting and locating pipeline leaks. This architecture utilises IoT-enabled sensing devices to support continuous monitoring of network conditions, improve the precision of leak detection, and optimise the use of resources within WDN infrastructure. Through the strategic placement of pressure, vibration, and flow sensors, the system can detect and localise leaks effectively while minimising operational costs, energy consumption, and false alarm rates.

Pressure sensors are deployed to continuously measure pressure variations within the pipeline system. Any abnormal pressure fluctuations caused by leaks are immediately detected, and the collected data are transmitted in real time to support accurate leak localisation. Vibration sensors capture characteristic vibration signals produced when water escapes through cracks or faults in pipelines, allowing the system to identify leak locations with high precision. In addition, flow sensors measure water flow rates across the network and detect sudden reductions that may indicate the presence of a leak.

The integration of these sensing technologies with IoT communication infrastructure significantly improves the leak

detection and localisation process within WDNs. Real-time transmission of sensor data to a central monitoring platform allows operators to quickly identify anomalies and initiate corrective actions, thereby reducing water loss. This architecture enhances the efficiency, reliability, and cost-effectiveness of water distribution monitoring systems by enabling proactive detection and localisation of pipeline faults.

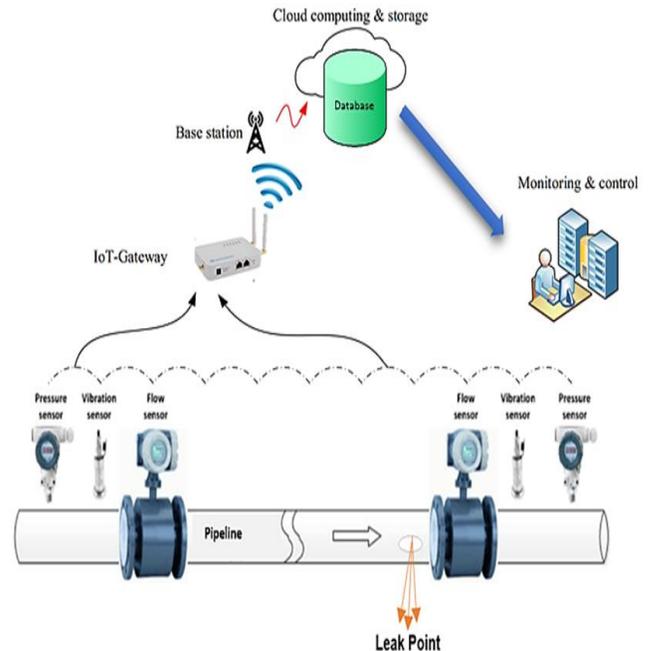


Fig. 3. Pipeline leak detection and location architecture.

Furthermore, IoT connectivity combined with cloud-based platforms supports advanced data processing and management capabilities. Sensor measurements are transmitted through an IoT gateway, where the information can be analysed in real time to detect potential leak events. Historical data stored in cloud systems can also be used to evaluate pipeline performance trends, identify recurring issues, and support predictive maintenance strategies that improve the long-term reliability of leak detection systems.

B. EPANET

EPANET is a widely used hydraulic simulation tool for analysing Water Distribution Networks and is recognised for its specialised capabilities in modelling water infrastructure systems. The software is specifically designed for water network analysis and includes features for hydraulic modelling, water quality simulation, and pump scheduling, making it highly suitable for studying the behaviour of WDNs.

One of the major advantages of EPANET is that it is an open-source, which enables researchers and practitioners to freely access the platform, customise simulations, and collaborate on water system studies. Over time, the software has gained a strong reputation due to its validation using real-world water distribution data, demonstrating reliable performance in practical applications. In addition, EPANET provides a user-friendly graphical interface, allowing users with different levels of technical expertise to easily design, simulate, and analyse network models.

Compared with network simulation tools commonly used in computer networking research, such as Network Simulator-2 (NS-2), Objective Modular Network Testbed in C++ (OMNET++), and Optimised Network Engineering Tools (OPNET), EPANET offers several advantages for water engineering studies. Its dedicated focus on hydraulic processes, combined with its modelling accuracy and intuitive interface, makes it particularly well suited for analysing water distribution networks and evaluating leak detection strategies.

C. Graph Theory

Graph Theory (GT) is a branch of mathematics that deals with the study of graphs, which are mathematical structures used to represent relationships between objects [37, 38]. In GT, a graph is composed of vertices (also known as nodes) and edges (also known as arcs or links) that connect pairs of vertices [39, 40].

This study applies GT for strategic sensor placement in pipelines, minimising costs while ensuring efficient monitoring. GT employs the Minimum Dominating Set Algorithm (MDS) to identify optimal node subsets, reducing sensors and simplifying maintenance for cost-effective, network-wide coverage.

In GT, an MDS in a graph G is a subset D of vertices where every vertex in G is either in D or adjacent to a vertex in D . The domination number $\gamma(G)$ is the smallest size of a dominating set in G . The dominating set problem involves determining whether $\gamma(G) \leq K$ for a given graph G and input K .

The WDN's structure was represented by the connected graph $G (V, E)$, with nodes as junctions and edges as pipes. Given an undirected graph $G (V, E)$, a subset of vertices $D \subseteq V$ is called an MDS if, for every vertex $u \in V \setminus D$, there is a vertex $v \in D$ such that $(u, v) \in E$.

In every graph, there exists a minimum of one MDS: if $D=V$ is the set of all vertices, then by definition, D is an MDS since there is no vertex $u \in V \setminus D$. Finding small MDS presents an intriguing challenge: The domination number of G is defined as $\gamma(G) = \min \{|D|: D \text{ is an MDS of } G\}$. Fig. 4 illustrates the MDS of a given graph, represented as $\{4, 7\}$, where each vertex in the graph is either in the set or adjacent to a vertex in the set.

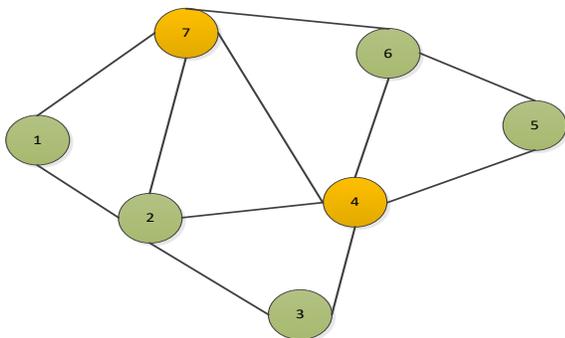


Fig. 4. A typical graph-dominating set.

The rest nodes in the graph are connected to either node 4 or node 7 in the graph. However, other dominating sets, such as $\{2, 4\}$, etc., can also be defined for the same connected graph.

D. Support Vector Machines

SVM stands as a highly regarded supervised ML algorithm, demonstrating exceptional proficiency in classification and prediction tasks by capitalising on its widespread recognition in both regression and classification domains [41, 42]. Fig. 5 visually illustrates the extensive application of this technique across different tasks, including pattern recognition, data classification, and outlier detection [43].

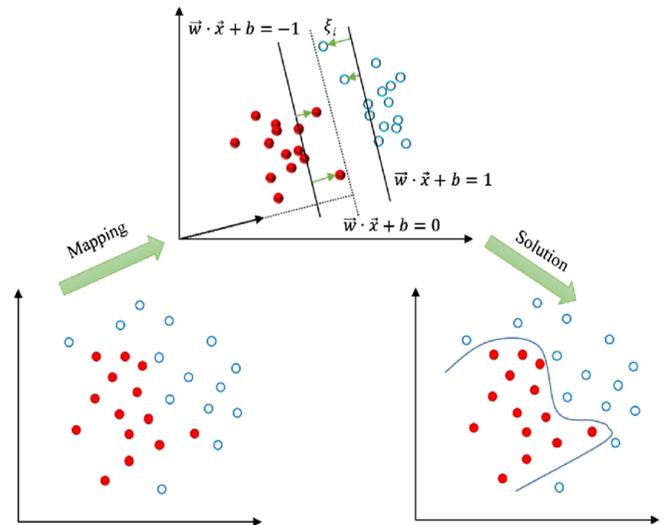


Fig. 5. Classification process of SVM.

This description demonstrates how the algorithm utilises a hyperplane to distinguish between two classes within a dataset, making it an effective technique for classification problems. SVMs are well known for their capability to accurately categorise data by identifying the optimal separating hyperplane that maximises the margin between classes. Due to this capability, SVMs have been successfully applied in a wide range of fields, including image recognition, text classification, and bioinformatics.

Within the context of WDNs, SVM models have been employed to identify and predict water losses associated with pipeline leaks. By analysing key parameters such as water pressure, flow rate, temperature, and leakage indicators, the algorithm can determine the location of leaks and assess their severity with a high degree of accuracy. The application of SVM in this domain improves leak detection performance and supports better management of water distribution systems, ultimately enhancing operational efficiency and enabling more effective utilisation of resources within WDNs.

Training Phase:

Given a labelled training dataset with input features x_i and corresponding leak information y_i :

x_i : Input features of the i -th instance, including water pressure, flow rates, and temperature.

y_i : Leak information of the i -th instance, such as the presence or absence of a leak and its severity.

The goal is to identify the optimal hyperplane that efficiently distinguishes instances of leaks from non-leaks. This is accomplished by minimising Eq. (1) while adhering to the constraints outlined in Eq. (2), as presented below:

$$\text{Minimize: } \frac{1}{2} \|w\|^2 \quad (1)$$

$$\text{Subject to: } y_i(w \cdot x_i + b) \geq 1 \text{ for all } i \quad (2)$$

where, w represents the vector that is perpendicular to the hyperplane, b is the bias term, and $\|w\|$ is the Euclidean norm of the weight vector.

1) *Dual optimisation problem:* The optimisation problem can be reformulated using Lagrange multipliers α_i to solve the dual problem. This involves maximising Eq. (3) and subjecting it to Eq. (4), as shown below:

$$\text{Maximize: } \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3)$$

$$\text{Subject to: } \sum_i \alpha_i y_i = 0 \text{ and } 0 \leq \alpha_i \leq C \text{ for all } i \quad (4)$$

where, α_i is the Lagrange multiplier associated with each training instance, y_i is the leak information, $K(x_i, x_j)$ is the kernel function that measures the similarity between two instances, and C is the regularisation parameter.

2) *Decision function and predictions:* Once the SVM model is trained, the decision function for predicting new instances can be represented as follows, utilising Eq. (5):

$$f(x) = \sum_i \alpha_i y_i K(x, x_i) + b \quad (5)$$

To predict the presence and severity of leaks in a new instance x , the sign and magnitude $f(x)$ can be represented by Eq. (6):

$$\text{If } f(x) > 0, \quad (6)$$

The instance is predicted to leak with a certain severity, which can be represented by Eq. (7):

$$\text{If } f(x) < 0, \quad (7)$$

The instance is predicted to be leak-free.

The magnitude $f(x)$ can indicate the severity of the leak.

These equations provide a mathematical representation of the SVM framework used to predict the location and severity of leaks within WDNs. By examining operational variables such as water pressure, flow rate, temperature, and leakage-related data, the SVM model learns underlying patterns and relationships that indicate the presence of leaks. The selection of appropriate kernel functions and parameter settings is adapted to the characteristics of the specific dataset and problem context. This flexibility allows the model to produce accurate and reliable predictions, ensuring that the leak

detection process is well-suited to the dynamic conditions of WDN systems.

E. Artificial Neural Network

ANNs are computational models designed based on the structure and operation of biological neural systems, particularly the human brain [44, 45]. They are extensively applied in ML and AI for tasks such as pattern recognition, classification, regression analysis, and decision-making processes [46, 47]. Fig. 6 illustrates the architecture of a three-layer ANN consisting of four input nodes, five neurons in the hidden layer, and two output nodes, demonstrating how information flows through the network during processing.

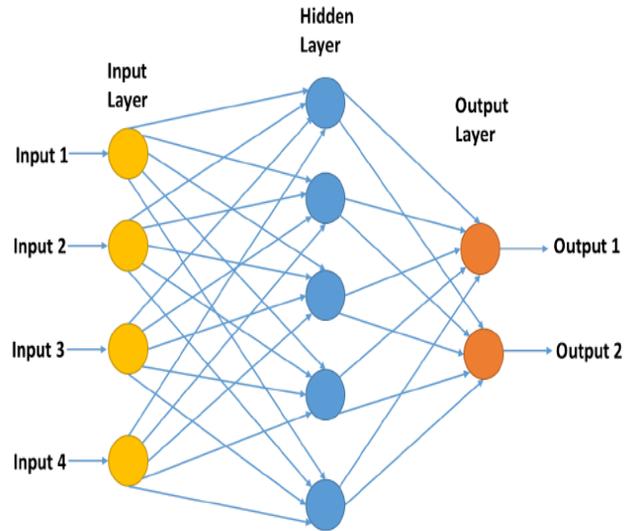


Fig. 6. Three-layer ANN architecture with input, hidden, and output layers.

In this study, ANNs were employed to detect and localise leaks within the pipeline infrastructure of WDNs. By leveraging the computational capabilities of ANNs, the proposed model was able to capture complex non-linear relationships present in the sensor data and extract relevant features required for accurate leak detection. Several key operational variables, including water pressure, flow rate, temperature, and leakage measurements, were analysed to identify patterns associated with potential pipeline faults.

The ability of ANNs to support continuous learning, real-time data processing, and resilience to noisy or incomplete data makes them particularly suitable for monitoring complex WDN environments. The primary purpose of the ANN-based model was to analyse the input data obtained from sensors and predict the occurrence of leaks while simultaneously estimating their locations within the distribution network.

The mathematical formulations used in the prediction and detection process are presented as follows:

1) Weighted Sum Calculation (Single Neuron) represented by Eq. (8):

$$z = \sum_{i=1}^n (w_i \cdot x_i) + b \quad (8)$$

where, z denotes the weighted sum, w_i and x_i are the weight and input corresponding to the i -th element, n is the number of inputs, and b is the bias term. The weighted sum is obtained by multiplying each input with its corresponding weight, summing these products, and adding the bias.

2) Activation Function (Non-linearity Introduction) can be represented by Eq. (9):

$$a = f(z) \quad (9)$$

where, a is the neuron output, z is the weighted sum, and f is the activation function applied to z . This function introduces non-linearities into the network, allowing it to model complex relationships in the data.

3) Forward Propagation (Layer-wise Output Computation) can be represented by Eq. (10):

$$a^l = f(W^{(l)} \cdot a^{(l-1)} + b^{(l)}) \quad (10)$$

where, $a^{(l)}$ represents the output of layer l , $a^{(l-1)}$ is the output of the preceding layer, $W^{(l)}$ and $b^{(l)}$ are the weights and bias vectors for layer l and f is applied element-wise.

4) Output Layer Calculation can be represented by Eq. (11):

$$h_\theta(x) = a^{(L)} \quad (11)$$

where, $h_\theta(x)$ represents the predicted output or hypothesis function for a given input x . $a^{(L)}$ represents the output of the final layer (output layer) of the ANN. This output is used to make predictions or compute the final result of the ANN model.

5) Thresholding (for leak detection) can be represented by Eq. (12).

$$\text{If } h_\theta(x) \geq \text{threshold}, \quad (12)$$

where, $h_\theta(x)$ represents the output of the leak detection model (such as an ANN) for input x , and the threshold is a predefined value. If the output is greater than or equal to the threshold, it indicates a detected leak.

The threshold value can be selected according to the required sensitivity and specificity of the leak detection system. In this study, a threshold of 0.5 was adopted for binary classification within the leak detection framework. This value was selected to maintain a balance between sensitivity and specificity and aligns with common practice in many machine learning classification models.

The equations presented describe the key mathematical operations used by the ANN to process sensor inputs and predict potential leaks. During the training phase, the network learns optimal values for the weights (W) and biases (b) by adjusting these parameters iteratively to minimise prediction errors. Through this optimisation process, the ANN improves

its ability to accurately detect and predict leaks within the Water Distribution Network.

F. SVM-ANN-GT Algorithm

The proposed SVM-ANN-GT algorithm combines the capabilities of SVM, ANN, and GT to improve the efficiency of sensor deployment along pipelines in WDNs. The main goal of this integrated approach is to optimise leak detection while simultaneously minimising energy consumption within the network. The operational steps of the proposed method are outlined in Algorithm 1.

Algorithm 1: Proposed SVM-ANN-GT algorithm

- 1) Graph Theory:
 1. Construct a connected graph $G(V, E)$ to model the Water Distribution Network (WDN), where V represents junctions and E represents pipes.
 2. Apply the dominating set technique to identify a subset $D \subseteq V$ that can monitor water flow and detect leaks.
 3. Perform optimal sensor placement by selecting nodes in D as optimal locations, minimising installation and maintenance costs while ensuring an effective monitoring system.
- 2) SVM Training Phase:
 4. Initialise weight vector w and bias term b .
 5. **For each** instance (x, y) in the training dataset:
 6. Set input features $X = x$ and leak information $Y = y$.
 7. Perform optimisation:
 8. **For each** iteration:
 - Update w and b to minimise the objective function using Eq. (1) and Eq. (2). Store the optimised weight vector w and bias term b .
 9. SVM Decision Function and Predictions:
 10. Function predictLeak(x):
 11. Calculate the decision function $f(x) = w^T x + b$ using Eq. (5).
 12. **If** $f(x) \geq 0$:
 13. **Return** "Leak detected with severity " + $|f(x)|$
 14. **Else:**
 - a. **Return** "No leak detected"
- 3) ANN for Leak Detection:
 - Set up an Artificial Neural Network (ANN) architecture with the input layer, hidden layers, and output layer.
 17. Initialise weights and biases for each neuron.
 18. **For each** epoch in training:
 - For each** instance (x, y) in the training dataset:
 19. Set input features $X = x$ and leak information $Y = y$.
 20. **For each** layer in the network:
 - For each** neuron in the layer:
 - Calculate the weighted sum z using Eq. (8).
 - Apply the activation function ϕ to z using Eq. (9).
 - Calculate the output of the final layer using Eq. (10).
 - Return** the predicted output or hypothesis function using Eq. (11).
 21. Thresholding (for leak detection):

22. If the output of the leak detection model (e.g., SVM or ANN) for input x is greater than or equal to a predefined threshold:
23. **Return** "Leak detected"
Else:
Return "No leak detected"

IV. SIMULATION RESULTS

In this study, the proposed SVM-ANN-GT algorithm was implemented using the EPANET hydraulic simulation tool (version 2.2). The simulations were conducted on a computer running the Windows 11 operating system with 16 GB of RAM. EPANET 2.2 was configured with appropriate parameter settings to accurately model the behaviour of WDNs. The simulated network consists of 36 nodes, representing different junction points within the distribution system. The network configuration also includes one storage tank and one pump station. The storage tank is located at an elevation of 100 m, with a minimum water level of 20 m and a maximum level of 30 m, as illustrated in Fig. 7.

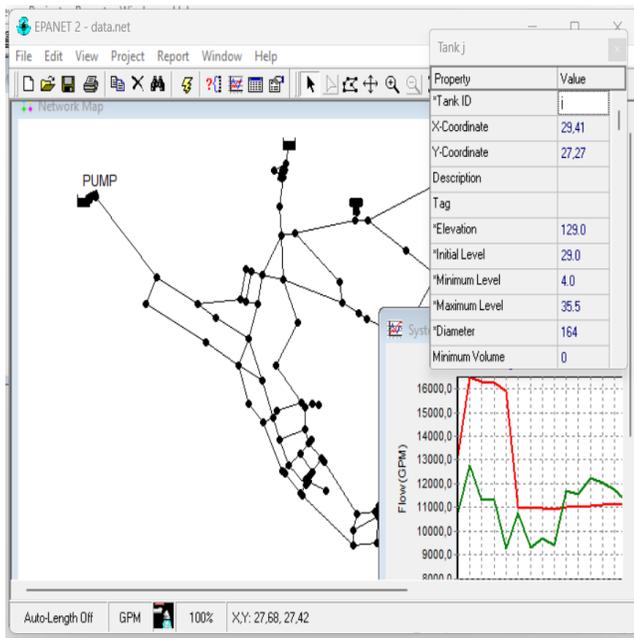


Fig. 7. EPANET simulation scenario.

The simulation environment was designed to represent a typical urban and peri-urban water distribution system in South Africa. The network incorporates pipelines constructed from ductile iron and PVC materials, with pipe diameters ranging from 100 mm to 300 mm. These pipes are assumed to be installed in heterogeneous soil conditions, including clay, sandy loam, and gravel, each of which influences the likelihood and characteristics of leak formation differently. In addition, environmental factors such as rainfall patterns, temperature fluctuations, and soil moisture levels were considered, as they can increase pipeline stress and the probability of leakage.

More challenging conditions, including high groundwater levels and expansive soil types, were also incorporated into the simulation environment to reflect realistic operational conditions. These variables were modelled within EPANET to

ensure that the performance evaluation of the SVM-ANN-GT algorithm reflects real-world pipeline behaviour and provides a realistic assessment of its leak detection capabilities.

EPANET is a widely recognised hydraulic modelling tool used for analysing WDNs. Despite its strong capabilities in hydraulic simulation, the software has certain limitations when applied to large-scale systems. EPANET operates using a single-threaded architecture, which means it does not utilise multi-core processors or GPU acceleration. As a result, simulations involving large or complex networks may require longer processing times and higher memory usage, particularly on systems with limited computational resources. Additionally, EPANET does not natively support real-time sensor data integration, parallel computing, or concurrent simulations, which can limit its applicability for dynamic monitoring environments or machine learning-driven analysis. To overcome these limitations, researchers frequently integrate EPANET with external platforms such as MATLAB or Python, although this integration increases system complexity and often requires advanced programming expertise.

Within the simulated network, the pump operates according to a pump curve labelled "Pump 1 Curve". The pump functions with an efficiency of 80% and a power rating of 50 kW. It is initially configured in the "On" state and operates at 100% of its nominal speed at the start of the simulation. The pipelines within the network each have a length of 500 m and a diameter of 300 mm. A roughness coefficient of 0.01 is assigned to represent the internal surface condition of the pipes. At the beginning of the simulation, all pipes are configured to remain fully open, allowing unrestricted water flow, and the initial flow rate is set to 0 L/s.

The simulation period spans 24 hours, with a time step of one hour for system calculations. Performance reports are generated at hourly intervals, allowing the monitoring of system behaviour throughout the simulation period. The flow change tolerance is set to 0.001 L/s, representing the allowable variation in flow rate during calculations, while the accuracy tolerance is also defined as 0.001 to maintain high precision in the simulation results.

System pressure is expressed in meters of head (m), and reporting intervals occur every 3600 seconds (one hour). The simulation applies the Hazen-Williams equation to estimate flow behaviour and head losses within the pipes. This empirical formula calculates flow rates and pressure losses based on hydraulic characteristics such as pipe diameter, roughness, and flow conditions.

Table I summarises the key simulation parameters used during the implementation of the proposed model.

TABLE I. SIMULATION PARAMETERS

Parameters	Values
Nodes	36
Tanks	1
Pump Stations	1
Tank Elevation	100 meters
Minimum Water Level	20 meters

Maximum Water Level	30 meters
Pump Curve	Pump1Curve
Efficiency	80%
Power	50kw
Initial Status	On
Initial Speed	100%
Pipe Length	500 meters
Pipe Diameter	300 mm
Roughness Coefficient	0.01
Pipe Status	Open
Initial Flow Rate	0 L/s
Simulation Time	24 hours
Simulation Timestep	1 hour
Report Duration	1 hour
Flow Change Tolerance	0.001 L/s
Accuracy Tolerance	0.001
Pressure Units	meters of the head (m)
Reporting Steps (seconds)	3600 seconds
Flow Change Formula	Hazen-Williams

TABLE II. DATASET

Timestamp	Flow rate	Pressure	Temperature	Leak
2023/03/05 09:28	10.8	35.9	50.2	1
2023/03/05 09:29	11.1	36.4	49.5	0
2023/03/05 09:30	11.7	35.7	50.0	0
2023/03/05 09:31	12.0	36.0	50.5	0
2023/03/05 09:32	11.3	36.2	50.8	1
2023/03/05 09:33	10.5	35.8	49.8	1
2023/03/05 09:34	10.2	36.3	49.5	0
2023/03/05 09:35	9.9	36.1	50.0	0
2023/03/05 09:36	10.3	35.9	50.5	1
2023/03/05 09:37	11.0	36.4	50.2	1
2023/03/05 09:38	11.5	35.7	50.0	0
2023/03/05 09:39	12.2	36.1	50.8	0
2023/03/05 09:40	11.5	36.3	51.0	1
2023/03/05 09:41	10.8	35.9	50.2	1
2023/03/05 09:42	11.1	36.4	49.5	0
2023/03/05 09:43	11.7	35.7	50.0	0
2023/03/05 09:44	12.0	36.0	50.5	0
2023/03/05 09:45	11.3	36.2	50.8	1
2023/03/05 09:46	10.5	35.8	49.8	1

Table II presents a summary of the main attributes associated with the pipes and nodes within the network. The parameters included in the table consist of timestamps, flow rate measurements, pressure values, temperature readings, and leak status indicators. These variables provide valuable information about the hydraulic behaviour of the system and support the identification of abnormal conditions that may indicate the presence of leaks or other operational issues within the network.

A. Identified Leakage Zone with Sensor Placement

Fig. 8 shows the leakage zones detected through the strategic deployment of sensors within the network. By applying the proposed SVM-ANN-GT algorithm to the selected WDN, a total of 14 distinct leakage locations were identified. These findings demonstrate the algorithm's effectiveness in accurately detecting and mapping multiple leaks across the network, highlighting its potential as a reliable solution for addressing leak detection challenges in WDN environments.

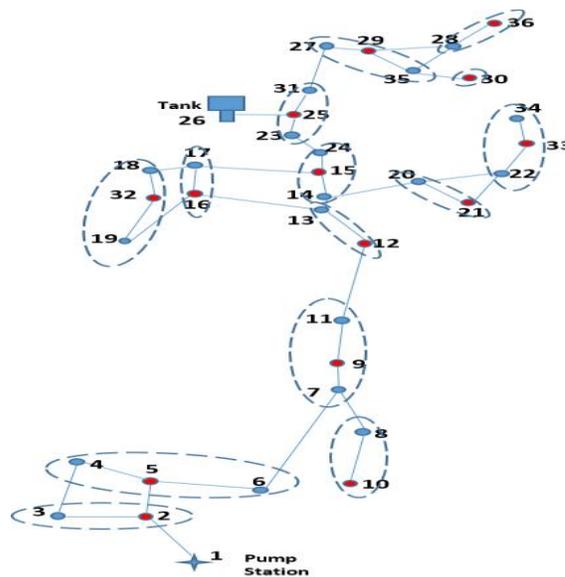


Fig. 8. Identification of leakage zone through sensor placement in the EPANET case study network.

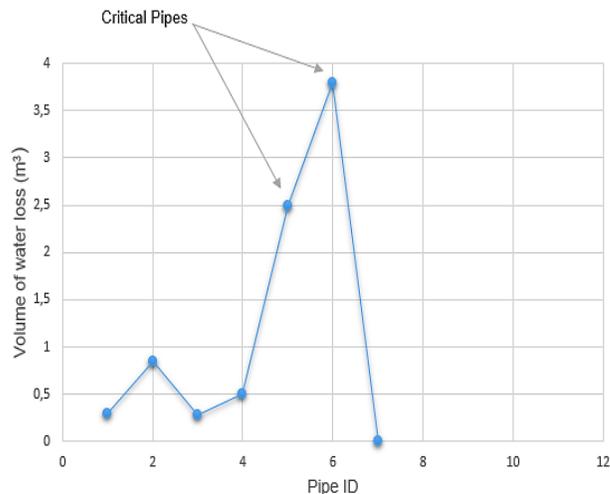


Fig. 9. Water loss volumes per pipe in the case study network.

Fig. 9 presents the discharge behaviour and corresponding leak flow rates observed across all pipes in the analysed WDN. The results indicate that pipes 5 and 6 experience the highest levels of water loss, with the leakage in pipe 6 approaching approximately 4 m³. This significant loss is largely attributed to pressure variations within the network, which underscores the importance of the nodes connected to these pipes. Adjusting

pressure levels at these critical nodes can therefore be an effective strategy for mitigating water losses.

Furthermore, when compared with random sensor deployment, the proposed algorithm achieved better detection performance, accurately identifying high-risk leakage zones and enhancing the overall efficiency of leak monitoring and management within the network.

B. Correlation Matrix of System Variables

Fig. 10 displays a correlation matrix as a heatmap, showing the relationships between flow rate, pressure, temperature, and leak.

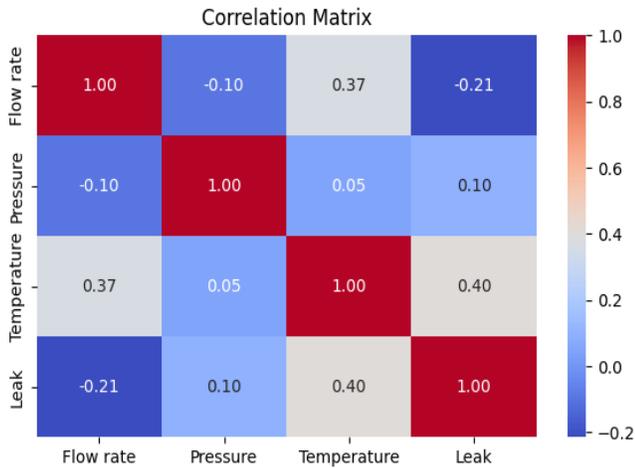


Fig. 10. Correlation matrix heatmap of flow rate, pressure, temperature, and leak variables.

Each cell in the matrix corresponds to the Pearson correlation coefficient calculated between a pair of variables, with values spanning from -0.2 to 1.0 . Warmer colour tones represent stronger positive relationships, whereas cooler tones indicate negative associations. Lighter colours suggest weak or negligible correlations between the variables. The diagonal entries show perfect correlations (1.00) since each variable is correlated with itself.

The results reveal several noteworthy relationships. For example, temperature shows a moderate positive association with leak (0.40) as well as with flow rate (0.37). Conversely, there is a weak negative relationship between flow rate and leak (-0.21). Meanwhile, pressure exhibits very little correlation with the other variables, indicating a limited direct relationship with the parameters considered in this analysis..

C. Accuracy of Leak Detection

Leak detection accuracy reflects the ability of a water leak detection system to correctly identify the presence of leaks and determine their exact locations within a WDN. Table III presents a comparative analysis of the proposed SVM-ANN-GT algorithm and a baseline approach based on random sensor deployment. The comparison focuses on important performance metrics, including the average detection distance and the total number of network hops, which are used to evaluate the efficiency and effectiveness of the detection process.

TABLE III. PERFORMANCE COMPARISON BETWEEN THE SVM-ANN-GT ALGORITHM AND RANDOM SENSOR DEPLOYMENT.

Parameters	SVM-ANN-GT algorithm	Random Deployment
Average distance (m)	190	340
Total Hops	18	28

The SVM-ANN-GT algorithm demonstrated improved performance compared with random sensor deployment for water leak detection when evaluated using two main metrics: average distance and total network hops. The average distance between sensors and detected leakage points was considerably smaller for the proposed method, measuring approximately 190 metres, whereas the random deployment approach produced an average distance of about 340 metres. This reduction indicates that the sensors selected by the SVM-ANN-GT algorithm are positioned closer to potential leak locations, leading to more effective monitoring and faster leak identification.

Similarly, the total number of hops, which represents the number of intermediate nodes required for data transmission, was also reduced using the proposed method. The SVM-ANN-GT algorithm required 18 hops, while the random deployment approach required 28 hops. Fewer hops facilitate faster communication between sensors and monitoring systems, thereby reducing delays in the detection process and improving overall system responsiveness.

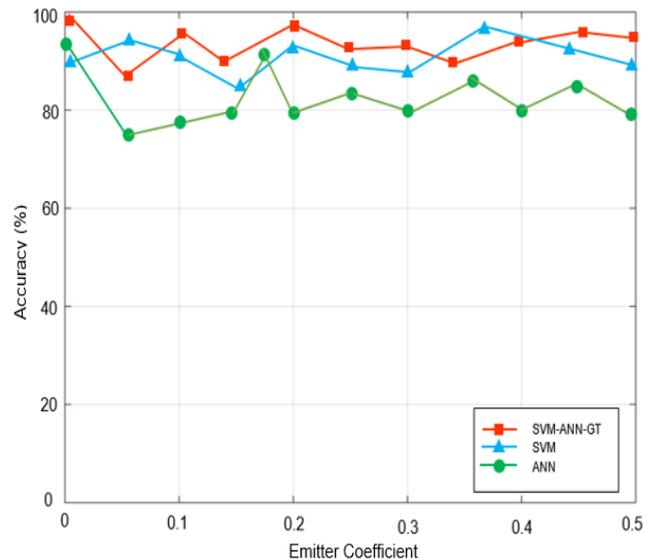


Fig. 11. Accuracy vs. emitter coefficient.

Fig. 11 further illustrates that the SVM-ANN-GT model consistently achieves higher accuracy compared with individual SVM and ANN models across an emitter coefficient range from 0 to 0.5. The hybrid model attains accuracy values between 93% and 98%, with an average accuracy of approximately 96%. In comparison, the standalone SVM model achieves accuracy levels ranging from 89% to 95%, with an average of around 92%, while the ANN model records lower accuracy values between 82% and 89%, averaging about 85%. Although variations in the emitter coefficient do not significantly degrade performance, the hybrid SVM-ANN-GT

approach manages these variations more effectively than the individual machine learning models.

The improved performance of the proposed algorithm can largely be attributed to its strategic sensor placement mechanism, which is based on GT. In particular, the Minimum Dominating Set (MDS) algorithm is applied to identify optimal sensor locations within the network. This approach ensures comprehensive monitoring coverage while minimising the number of sensors required. By reducing redundancy and prioritising structurally important nodes within the WDN, the algorithm improves leak localisation accuracy, shortens detection time, and decreases the likelihood of false alarms. In contrast, the standalone SVM and ANN models lack this optimisation component, resulting in lower detection coverage, a higher number of sensors, and less efficient identification of leakage areas.

WDNs can vary considerably in scale, ranging from small municipal systems to extensive metropolitan infrastructures. When applied to larger networks, leak detection systems often require additional sensors, which can increase computational complexity and data processing requirements. The proposed SVM-ANN-GT framework addresses this challenge by using the MDS algorithm from graph theory to minimise the number of required sensors while still maintaining full network coverage. As a result, the system can scale to larger networks without proportional increases in cost or system complexity.

Furthermore, the hybrid design of the algorithm contributes to its scalability. Within the framework, graph theory manages network topology and sensor placement, SVM performs classification of leak-related patterns, and ANN captures complex non-linear relationships in the sensor data. This modular structure supports parallel processing and localised updates, allowing the system to adapt to network changes without requiring complete retraining of the model. Consequently, the algorithm remains both efficient and accurate as network size increases.

The computational cost associated with the SVM-ANN-GT algorithm increases sublinearly with the number of network nodes due to the optimisation mechanisms embedded in the model. In traditional leak detection systems, computational demands often increase linearly or even exponentially with network size, as each additional sensor contributes new data streams and decision paths. However, by applying the MDS algorithm to select a minimal set of monitoring nodes, the proposed approach reduces the number of sensors and the overall data volume required for analysis. This ensures that computational requirements remain manageable even as the network expands.

Simulation experiments conducted using EPANET further confirm the algorithm's capability to operate efficiently in larger WDNs without causing proportional increases in computational or hardware resource requirements.

WDNs are inherently dynamic systems, influenced by fluctuating water demand, seasonal variations, and ongoing infrastructure expansion. Leak detection algorithms must therefore be able to adapt to changing operating conditions while maintaining high levels of accuracy. The integration of

SVM and ANN models allows the proposed framework to incorporate multiple input variables, including pressure, flow rate, and temperature. In particular, the ANN component is effective at modelling non-linear relationships, enabling the system to adapt to varying network conditions. Additionally, the algorithm's real-time processing capabilities allow it to continuously learn from new data, which supports sustained performance over time.

Despite these advantages, the initial deployment of sensors across a large WDN may still require considerable resources. Although the proposed method optimises sensor placement, installing sensors throughout an extensive network can involve significant costs, particularly in areas with challenging terrain or densely built urban environments. The application of graph-theoretic optimisation reduces this burden by identifying only the most critical monitoring locations, thereby limiting unnecessary expenditure while maintaining effective leak detection coverage.

Another practical challenge encountered in real-world deployments involves noise and faults in sensor data, which can affect the accuracy of leak detection. The ANN component of the proposed algorithm is capable of handling noisy datasets, and the hybrid combination of SVM and ANN increases the system's resilience even when certain sensors malfunction.

In addition, sensor calibration drift may occur over time. Sensors that operate for extended periods without recalibration may produce inaccurate pressure, temperature, or flow measurements, which can degrade detection accuracy. Environmental factors such as heavy rainfall, frost, or extreme heat can also influence ground conditions, potentially affecting sensor stability and the reliability of collected data. These challenges highlight the need for regular sensor maintenance and adaptive algorithms to ensure the long-term reliability of the monitoring system.

Reducing false positive detections is also critical for maintaining operational efficiency, particularly in large networks. Systems based on random sensor placement often generate high numbers of false alarms, leading to unnecessary inspections and maintenance interventions. By strategically positioning sensors, the SVM-ANN-GT algorithm significantly reduces false positive rates and improves the reliability of leak detection, thereby supporting more efficient management of WDNs.

D. Recall

As shown in Fig. 12, recall, also referred to as sensitivity or the true positive rate, evaluates how well a model can correctly identify all relevant positive instances within a dataset. It is computed as the proportion of correctly predicted positive cases relative to the total number of actual positive samples. In this study, the SVM-ANN-GT model achieved the highest recall value of 90%, demonstrating a strong ability to detect the majority of true positive cases. The SVM model obtained a recall of 70%, while the ANN model recorded the lowest performance with a recall value of 60%. These findings emphasise the importance of recall in applications where it is crucial to identify as many true positive instances as possible, such as in leak detection, medical diagnosis, and fraud

detection. The high recall achieved by the SVM-ANN-GT model indicates that it is particularly suitable for scenarios that require strong detection sensitivity. In contrast, the lower recall observed in the ANN model suggests that additional improvements may be necessary to enhance its capability to correctly identify positive cases.

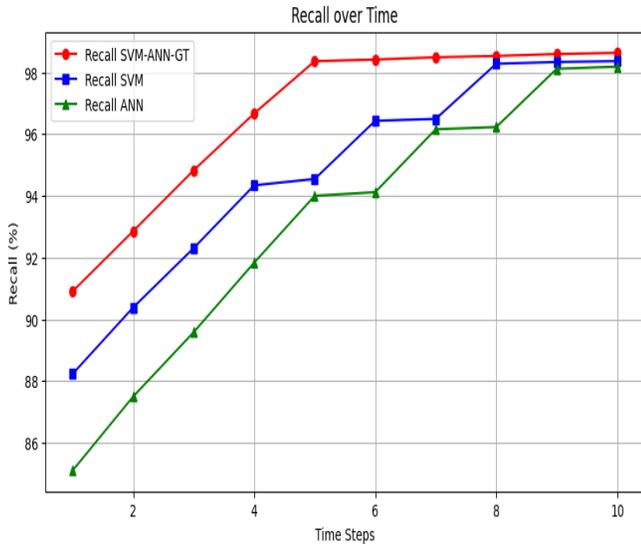


Fig. 12. Recall vs. time steps.

E. F1-Score

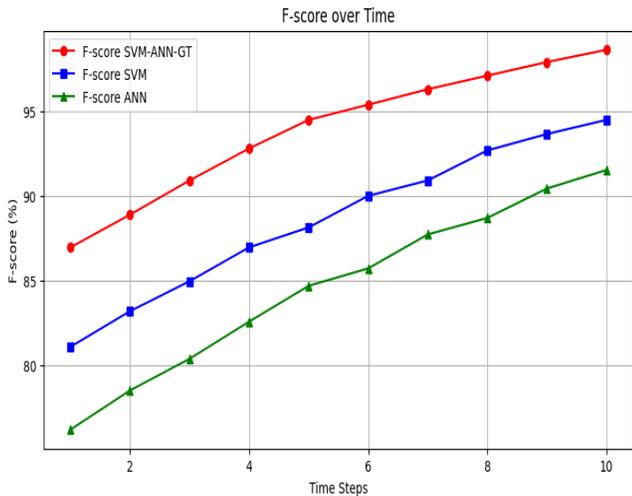


Fig. 13. F-score vs. time steps.

As illustrated in Fig. 13, the F1-score provides a comprehensive evaluation of model performance by integrating precision and recall into a single metric. This measure is particularly useful when analysing datasets with imbalanced class distributions, as it considers both the accuracy of positive predictions and the model's ability to detect actual positives. The F1-score is calculated as the harmonic mean of precision and recall, thereby assigning equal importance to both measures. In this evaluation, the SVM-ANN-GT hybrid model achieved the highest F1-score of 87.5%, indicating strong and well-balanced predictive performance. The SVM model

produced a moderate F1-score of 72.5%, whereas the ANN model demonstrated comparatively lower performance with a score of 62.5%. These results highlight the significance of the F1-score in evaluating classification models, particularly in scenarios where both precision and recall are essential, such as in leak detection and other critical monitoring systems.

The superior F1-score achieved by the SVM-ANN-GT model demonstrates its effectiveness as a reliable predictive framework across different application domains. Conversely, the relatively lower F1-score of the ANN model suggests opportunities for improvement, particularly in enhancing both its precision and recall capabilities to achieve better overall performance.

F. False Positive and False Negative Rates

The SVM-ANN-GT algorithm demonstrates strong accuracy in leak detection; however, its overall performance can also be evaluated through its false positive (FP) and false negative (FN) rates. Elevated FP rates may result in unnecessary maintenance activities and increased operational costs, while high FN rates can allow leaks to remain undetected, potentially leading to significant water loss and damage to infrastructure.

Simulation experiments conducted using EPANET produced relatively low error rates, with FP at 3% and FN at 4%, indicating that the proposed approach achieves a good balance between sensitivity and specificity. The incorporation of GT supports optimal sensor placement by ensuring that key areas of the network are effectively monitored, which helps reduce the occurrence of false negatives. At the same time, the hybrid SVM-ANN model improves classification performance, thereby lowering the number of false positives. These combined characteristics make the algorithm well-suited for applications where high detection accuracy and reliability are essential.

G. Balancing Sensitivity and Specificity

The SVM-ANN-GT algorithm is designed to effectively detect genuine leaks while minimising false alarms by integrating accurate classification, sensitivity to relevant features, and optimised sensor placement. First, the SVM component establishes optimal decision boundaries that distinguish between leak and non-leak conditions. By analysing historical patterns in variables such as pressure, flow rate, and temperature, the SVM model helps reduce false positives by clearly separating the two classes.

Second, the ANN component is capable of modelling complex non-linear relationships within sensor data, enabling the system to detect subtle anomalies that might otherwise remain unnoticed. This capability contributes to reducing false negatives by identifying leak-related patterns that simpler models may fail to capture.

Third, the GT component determines optimal sensor locations within the network by applying the Minimum Dominating Set (MDS) algorithm. This ensures that sensors are positioned at structurally important nodes, improving leak localisation accuracy and reducing misclassification that could arise from incomplete monitoring coverage.

By combining these complementary mechanisms, the proposed framework achieves a balanced detection strategy that is both highly sensitive to actual leak events and resistant to false alarms. This balance between precision and recall is reflected in the model's strong performance, as demonstrated by its F1-score of 87.5%.

V. CONCLUSION AND FUTURE WORK

This study introduced the SVM-ANN-GT algorithm, a hybrid machine learning approach that combines SVM, ANN, and GT to enhance leak detection and localisation in WDNs. The method improves monitoring performance by applying the MDS algorithm to optimise sensor placement, ensuring that sensors are positioned at critical nodes within the network. This optimisation strategy contributes to higher detection accuracy, fewer false alarms, and more efficient monitoring coverage.

Experimental results indicate that the proposed framework achieved a detection accuracy of 96%, a recall of 90%, and an F1-score of 87.5%, while maintaining false positive and false negative rates below 5%. These results outperform the performance of standalone SVM and ANN models, demonstrating the advantages of integrating machine learning techniques with graph-theoretic optimisation. The findings suggest that the proposed algorithm has strong potential to support efficient water resource management, reduce operational costs, and minimise energy consumption in WDN monitoring systems.

Despite these promising outcomes, the study has certain limitations. The evaluation was conducted primarily using controlled simulations in EPANET, which may not fully capture the complexities of real-world water distribution systems. Practical challenges such as real-time system operation, large-scale network processing, sensor failures, noisy measurements, and environmental influences were not fully represented in the simulation environment. Consequently, further investigation is required to assess the algorithm's robustness under realistic operating conditions.

Future research will therefore focus on evaluating the proposed approach in dynamic and non-ideal environments to better understand its reliability and adaptability. In practical implementations, the algorithm could be integrated with SCADA systems or IoT-based monitoring platforms, enabling scalable and cost-effective leak detection while supporting broader water conservation strategies. Additional research directions include incorporating advanced deep learning models, deploying the system in large-scale real-world WDNs, and examining its scalability, adaptability, and operational performance within complex infrastructure environments.

ACKNOWLEDGMENT

The authors extend their gratitude to the Tshwane University of Technology for providing financial support essential for the completion of this research. Additionally, the authors declare that no conflicts of interest are associated with the publication of this study.

REFERENCES

- [1] Shao, Y., et al., Global energy and leakage optimization in water distribution systems from water treatment plants to customer taps. *Resources, Conservation and Recycling*, 2023. 194: p. 107003.
- [2] Adedeji, K.B., et al., Towards Digitalization of Water Supply Systems for Sustainable Smart City Development—Water 4.0. *Applied Sciences*, 2022. 12(18): p. 9174.
- [3] Wu, Y., et al., Defender–attacker–operator: Tri-level game-theoretic interdiction analysis of urban water distribution networks. *Reliability Engineering & System Safety*, 2021. 214: p. 107703.
- [4] Wang, S., A.F. Taha, and A.A. Abokifa, How Effective is Model Predictive Control in Real-Time Water Quality Regulation? State-Space Modeling and Scalable Control. *Water Resources Research*, 2021. 57(5): p. e2020WR027771.
- [5] Ramos, H.M., et al., Integrated Smart Management in WDN: Methodology and Application. *Water*, 2023. 15(6): p. 1217.
- [6] Mosetlhe, T.C., et al., Appraising the impact of pressure control on leakage flow in water distribution networks. *Water*, 2021. 13(19): p. 2617.
- [7] Taiwo, R., I.A. Shaban, and T. Zayed, Development of sustainable water infrastructure: A proper understanding of water pipe failure. *Journal of Cleaner Production*, 2023: p. 136653.
- [8] Ramos, H.M., et al., New challenges towards smart systems' efficiency by digital twin in water distribution networks. *Water*, 2022. 14(8): p. 1304.
- [9] Mehryar, M., et al., Dynamic zoning of the network using cooperative transmission and maintenance planning: A solution for sustainability of water distribution networks. *Reliability Engineering & System Safety*, 2023. 235: p. 109260.
- [10] Bello, O., et al., Solving management problems in water distribution networks: A survey of approaches and mathematical models. *Water*, 2019. 11(3): p. 562.
- [11] Ghobadi, F., G. Jeong, and D. Kang, Water pipe replacement scheduling based on life cycle cost assessment and optimization algorithm. *Water*, 2021. 13(5): p. 605.
- [12] Phan, H.C., et al., Managing water main breaks in distribution networks—A risk-based decision making. *Reliability Engineering & System Safety*, 2019. 191: p. 106581.
- [13] MSIBI, N.A., Management of a Water Distribution Network by Coupling GIS and Hydraulic Modeling: A case of Ghazaouet, Algeria. 2020, PAUWES.
- [14] Ávila, C.A.M., et al., Leakage management and pipe system efficiency. Its influence in the improvement of the efficiency indexes. *Water*, 2021. 13(14): p. 1909.
- [15] Adedeji, K.B., et al., Pressure management strategies for water loss reduction in large-scale water piping networks: A review. *Advances in Hydroinformatics: SimHydro 2017—Choosing The Right Model in Applied Hydraulics*, 2018: p. 465–480.
- [16] Cody, R., Acoustic Monitoring for Leaks in Water Distribution Networks. 2020.
- [17] Wan, X., R. Farmani, and E. Keedwell, Gradual Leak Detection in Water Distribution Networks Based on Multistep Forecasting Strategy. *Journal of Water Resources Planning and Management*, 2023. 149(8): p. 04023035.
- [18] Moni, N.A., Vibration interpretation technique for water leak detection in small diameter pipelines. 2020, Botswana International University of Science & Technology (Botswana).
- [19] Adedeji, K.B., et al., Leakage detection and estimation algorithm for loss reduction in water piping networks. *Water*, 2017. 9(10): p. 773.
- [20] Tomyeviadzi, H.M. and R. Seidu, Leakage detection in water distribution networks via 1D CNN deep autoencoder for multivariate SCADA data. *Engineering Applications of Artificial Intelligence*, 2023. 122: p. 106062.
- [21] Tshimbalanga, M.A.E., Real water loss within water distribution networks of the Tshwane Metropolitan City in South Africa. University of Johannesburg.

- [22] Yang, G. and H. Wang, Optimal pressure sensor deployment for leak identification in water distribution networks. *Sensors*, 2023. 23(12): p. 5691.
- [23] Ringas, N., et al., Dry, the Beloved Country: Space and Water: The Cape Town Water Crisis. *Space Fostering African Societies: Developing the African Continent through Space, Part 1*, 2020: p. 273-332.
- [24] Farah, E. and I. Shahrour, Water Leak Detection: A Comprehensive Review of Methods, Challenges, and Future Directions. *Water*, 2024. 16(20): p. 2975.
- [25] Fares, A., et al., Leak detection in real water distribution networks based on acoustic emission and machine learning. *Environmental Technology*, 2023. 44(25): p. 3850-3866.
- [26] Komba, G.M., T.E. Mathonsi, and P.A. Owolawi. Optimizing leak detection and location in water distribution networks using SVM-RF algorithm. in *2024 15th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT)*. 2024. IEEE.
- [27] Porwal, S., S. Akbar, and S. Jain. Leakage detection and prediction of location in a smart water grid using SVM classification. in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*. 2017. IEEE.
- [28] Zhou, M., et al., An integration method using kernel principal component analysis and cascade support vector data description for pipeline leak detection with multiple operating modes. *Processes*, 2019. 7(10): p. 648.
- [29] Lang, X., et al., Pipeline leak aperture recognition based on wavelet packet analysis and a deep belief network with ICR. *Wireless Communications and Mobile Computing*, 2018. 2018: p. 1-8.
- [30] Amini, I., *Data-Driven Methods in Pipeline Leakage Detection*. 2021.
- [31] Soldevila, A., et al., Data-driven approach for leak localization in water distribution networks using pressure sensors and spatial interpolation. *Water*, 2019. 11(7): p. 1500.
- [32] Rayaroth, R., Random bagging classifier and shuffled frog leaping based optimal sensor placement for leakage detection in WDS. *Water Resources Management*, 2019. 33: p. 3111-3125.
- [33] Wu, Y., et al., Burst detection in district metering areas using a data driven clustering algorithm. *Water research*, 2016. 100: p. 28-37.
- [34] Lučin, I., et al., Data-driven leak localisation in urban water distribution networks using big data for random forest classifier. *Mathematics*, 2021. 9(6): p. 672.
- [35] Suresh, P., et al., Model-Based Leak Detection in Pipe Networks Using Optimisation Techniques.
- [36] Fan, X., X. Zhang, and X.B. Yu, Machine learning model and strategy for fast and accurate detection of leaks in water supply network. *Journal of Infrastructure Preservation and Resilience*, 2021. 2(1): p. 1-21.
- [37] Bollobás, B., *Modern graph theory*. Vol. 184. 1998: Springer Science & Business Media.
- [38] Komba, G.M. and O.P. Kogeda, Gateway Location Algorithm for Wireless Mesh Networks. *Journal of Engineering and Applied Sciences*, 2018. 13(17): p. 7288 - 7297.
- [39] Knauer, U. and K. Knauer, *Algebraic graph theory*, in *Algebraic Graph Theory*. 2019, de Gruyter.
- [40] Marcus, D.A., *Graph theory*. Vol. 53. 2020: American Mathematical Soc.
- [41] Burkov, A., *The hundred-page machine learning book*. Vol. 1. 2019: Andriy Burkov Quebec City, QC, Canada.
- [42] Kurani, A., et al., A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Annals of Data Science*, 2023. 10(1): p. 183-208.
- [43] Khairandish, M.O., et al., A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images. *Irbm*, 2022. 43(4): p. 290-299.
- [44] Jo, T., *Machine Learning Foundations. Supervised, Unsupervised, and Advanced Learning*. Cham: Springer International Publishing, 2021.
- [45] Fradkov, A.L., Early history of machine learning. *IFAC-PapersOnLine*, 2020. 53(2): p. 1385-1390.
- [46] Khaleel, M., A.A. Ahmed, and A. Alsharif, *Artificial Intelligence in Engineering. Brilliance: Research of Artificial Intelligence*, 2023. 3(1): p. 32-42.
- [47] Zhang, J., et al., Applications of artificial neural networks in microorganism image analysis: a comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer. *Artificial Intelligence Review*, 2023. 56(2): p. 1013-1070.