AQUA
Water Infrastructure, Ecosystems and Society

IWA PUBLISHING

 AQUA — Water Infrastructure, Ecosystems and Society Vol 73 No 3, 504 doi: 10.2166/aqua.2024.264

Identifying failure types in cyber-physical water distribution networks using machine learning models

Utsav Parajuli and Sangmin Shin *

School of Civil, Environmental and Infrastructure Engineering, Southern Illinois University, Carbondale, IL 62901, USA
*Corresponding author. E-mail: sangmin.shin@siu.edu

UP, 0000-0003-4647-9689; SS, 0000-0002-0391-6319

ABSTRACT

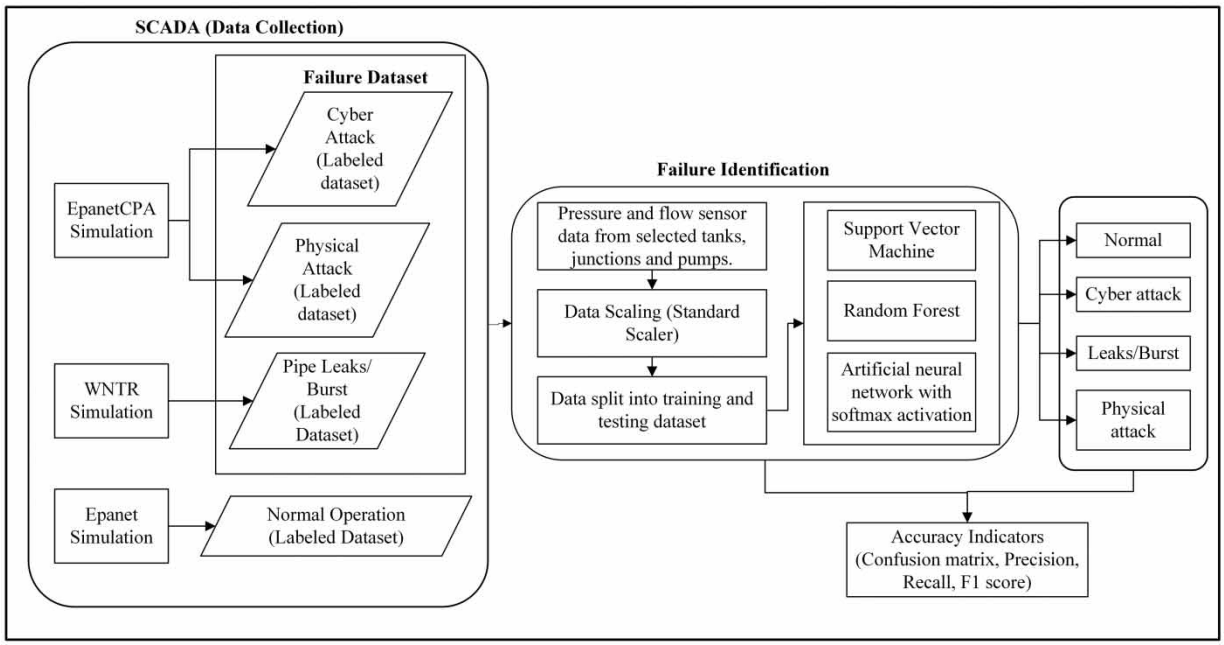
Water cyber-physical systems (CPSs) have experienced anomalies from cyber-physical attacks as well as conventional physical and operational failures (e.g., pipe leaks/bursts). In this regard, rapidly distinguishing and identifying a facing failure event from other possible failure events is necessary to take rapid emergency and recovery actions and, in turn, strengthen system's resilience. This paper investigated the performance of machine learning classification models – support vector machine (SVM), random forest (RF), and artificial neural networks (ANNs) – to differentiate and identify failure events that can occur in a water distribution network (WDN). Datasets for model features related to tank water levels, nodal pressure, and water flow of pumps and valves were produced using hydraulic model simulation (WNTR and epanetCPA tools) for C-Town WDN under pipe leaks/bursts, cyber-attacks, and physical attacks. The evaluation of accuracy, precision, recall, and F1-score for the three models in failure type identification showed the variation of their performances depending on the specific failure types and data noise levels. Based on the findings, this study discussed insights into building a framework consisting of multiple classification models, rather than relying on a single best-performing model, for the reliable classification and identification of failure types in WDNs.

Key words: anomaly detection, cyber-physical system, cyber-physical attacks, resilience, smart water networks, water distribution network

HIGHLIGHTS

- Further investigation for anomaly detection machine learning models in identifying a specific failure type is needed for WDN resilience.
- Machine learning models showed reliable failure identification performance.
- The models' performance varied with the failure types and data noise levels.
- The models produced misclassification between different failure events that produced similar hydraulic responses.
- Insights into a framework with multiple classification models were discussed to improve the reliable failure identification of WDNs.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Numerous smart meters, sensors, and data acquisition systems are being used to monitor and autonomously control WDNs (water distribution networks) owing to the ongoing technological advances over a couple of decades, particularly the development of affordable sensors and universal internet access (Wu *et al.* 2022). The smart system deploys Information and Communication Technologies (ICTs), which have received attention to achieve efficiency, sustainability, livability, and resilience goals in urban water management (Mutchek & Williams 2014). ICT employs complex architecture including sensors, communication, programable logic controllers (PLCs), actuators, remote terminal units (RTUs), and data and control servers – called Supervisory Control and data acquisition system (SCADA) (Nazir *et al.* 2021). Transforming conventional WDNs to water CPSs (Cyber-Physical Systems) with SCADA has supported real-time monitoring and data collection and remote control to improve the system's operational efficiency and capabilities related to rapid, accurate failure detection and timely recovery actions, which, in turn, strengthens system resilience (Walsby 2013; Mutchek & Williams 2014; Shin *et al.* 2018).

However, the water CPSs have become more susceptible to cyber and physical attacks. Water infrastructure, such as wastewater treatment facilities and WDNs, is one of the most targeted by cybercriminals since it is essential to the sustainable growth of modern society. In 2000, a former employee of the wastewater treatment facility in Maroochy, Australia, changed the pumps' operation by maliciously sending the incorrect command, causing the wastewater to overflow and produce an unpleasant odor (Ramotsoela *et al.* 2018). Similarly, in Georgia, USA, a drinking water system was physically attacked in 2013; the attacker gained physical access to the system and changed the fluorine and chlorination settings (Do *et al.* 2017). It is also reported that the water sector had the third largest number of cyber-physical incidents among critical infrastructures (Clark *et al.* 2017).

Numerous studies have introduced detection models and algorithms for operational anomalies from cyber-physical attacks in WDNs. For example, Housh & Ohar (2018) used a simulation-model-based approach for cyber-physical attack detection for WDNs. Abokifa *et al.* (2019) used a combination of ANN (artificial neural network) and principal component analysis (PCA) for the real-time detection of attacks in WDNs. Taormina & Galelli (2018) employed deep learning autoencoders (AEs) with a threshold for reconstruction error that can detect cyber-attacks. Tsiami & Makropoulos (2021) introduced the algorithm of graph convolutional neural network considering the temporal and spatial relationships of SCADA data to improve the detection of cyber-physical attack events. Housh *et al.* (2022) proposed a semi-supervised detection algorithm with dimensionality reduction followed by support vector data description (SVDD), which does not require labeled attack

datasets in its training to consider real-world applications. Brentan *et al.* (2021) introduced a two-step process for cyber-attack detection, which includes a fast Independent Component Analysis algorithm to separate multiple sensors data into the individual component (flow, pressure, and tank water level) and a statistical control algorithm (abrupt change point detection algorithm) to detect changes in control variables due to the attacks. A more detailed description of the approaches can be found in the Battle of the Attack Detection Algorithm (BATADAL) (Taormina *et al.* 2018).

However, the operational anomalies in the WDNs can be caused by not only cyber-physical failures due to attacks or malfunctions but also conventional failures/disruptions such as significant pipe leaks or bursts. In this regard, the first step to rapidly take emergency or recovery actions against the WDN disruptions – i.e., strengthen the resilience of WDNs – is the rapid identification of failure events that caused operational anomalies (Shin *et al.* 2018). However, while previous studies have tested their models and algorithms for a specific failure type, they have made few efforts to distinguish and identify a failure type during a WDN disruption from other failure types – which can occur in a WDN or water CPS. For example, the approaches introduced in BATADAL were evaluated for the detection of cyber-attack events only.

For conventional physical failures and disruptions, a common way to detect pipe leaks or bursts is monitoring minimum night flow in district metered areas (DMAs) (Amoatey *et al.* 2021). DMAs are hydraulically independent sectors of a WDN, which typically have inlet flow meters and pressure sensors to monitor pipe leakage. The analysis of minimum night flow for a DMA using probabilistic approaches or machine learning models, considering minimal human activities during the night, has suggested the effective detection of background pipe leakage or bursts. For example, Głomb *et al.* (2023) investigated the performance of multiple machine learning anomaly detectors in the rapid and accurate detection of pipe leaks using the data of DMAs' water consumption, inflow, and pressure. In addition, the analysis of acoustic sensor data from a DMA is also used to detect and localize pipe leaks (Xue *et al.* 2020). Siddique *et al.* (2023) used an acoustic emission scalogram combined with a deep learning algorithm (convolution neural network) to diagnose pipe conditions.

Similarly, Nam *et al.* (2019) proposed hybrid PCA and exponentially weighted moving average (EWMA) for the detection and isolation monitoring of the pipe burst. Mashhadi *et al.* (2021) discussed the use of machine learning algorithms for leak detection and localization in WDNs. Fan *et al.* (2021) used ANN (supervised) and AE (unsupervised) algorithms for leak detection. Ahmad *et al.* (2023) used a novel vulnerability index and 1-D convolutional neural network for pipe leak and size detection. Here, the acoustic emission hit feature was used for pipe leak detection. Asghari *et al.* (2023) employed machine learning-based transient analysis for leak detection, which substitutes complex inefficient optimization algorithms with machine learning models. In this context, further investigation is needed into how well the data-driven algorithms and models perform in identifying a specific failure type from multiple failure types that can occur in water CPSs. The rapid identification of the failure type will help the system manager quickly implement the response and recovery actions to return to the normal operating conditions of WDNs (Shin *et al.* 2020).

Other infrastructure sectors have investigated the classification of failure events in their systems using data-driven models. Anwar *et al.* (2015) used different machine learning models to differentiate cyber-attacks from physical faults in a smart electrical grid. Patil *et al.* (2019) used and compared RF (random forest), SVM (support vector machine), K-nearest neighbor (KNN) and Bagging Tree to classify sensor faults and cyber-attacks in smart buildings. Hashim *et al.* (2020) used PCA and multiclass SVM for detecting and identifying faults (leakage and equipment malfunction) in nonresidential building water pipes. Nazir *et al.* (2021) used KNN and SVM for multiclass classification as supervised learning and unsupervised AE for detecting anomalies in IT operations in WDNs. However, to the best of our knowledge, less attention has been paid to distinguishing and identifying failure types for WDNs with CPSs.

Also, the data-driven models in the previous studies are trained, validated, and tested using clean datasets and the assumption of faultless sensor monitoring. The real-world sensors consist of faults and noise in their measurements (El-Zahab & Zayed 2019). These alterations are either uniformly or unevenly reported in the dataset. The uniform noises in the dataset make it challenging to identify anomalies (Abokifa *et al.* 2019), which leads to misinterpretations of WDN failures. Therefore, it is crucial to assess the model's performance to outliers brought on by measurement noise, which may not always signify failure and has not yet been fully investigated for WDNs under cyber-physical failures.

Thus, this study evaluates machine learning classification models to differentiate and identify the failure types among cyber-physical attacks and conventional disruptions (pipe leaks/bursts) using datasets including noise, with the following question: can the machine learning classification models that have been used to detect WDN's anomalies from a specific type of failure also differentiate and identify a failure event from other possible failure events? The contributions of this study include providing insights into advancing data-driven models for identifying different failure types in WDNs. This will help rapid

emergency and recovery actions to WDN disruptions from cyber-physical attacks and conventional physical failures and, in turn, enhance the WDN's resilience.

2. METHODS

Figure 1 summarizes the process of failure identification with machine learning classification models in this study.

2.1. Datasets for WDN failures

2.1.1. Study of the WDN

This study selected the C-town WDN to generate datasets for multiple failure types, which was also used in Taormina *et al.* (2018) and Fan *et al.* (2021) for the dataset generation to test their anomaly detection models. The C-town WDN has 432 pipes, 388 nodes, 11 pumps, one actuated valve, and seven tanks (Figure 2). The WDN is divided into five DMAs. All the actuators, pumps, and tanks are connected to the SCADA and operated through nine PLCs, which allows rapid detection of pipe leaks/bursts and adaptive control of the WDN components within the DMAs. The status of each pump and valve is controlled by the PLC and is reported to the SCADA. The SCADA system of the C-town WDN gathers and monitors the data for 43 operational variables, including water level at tanks (seven variables), flow status of pumps and valves (24 variables), and pressure at the nodes near pumps and valves (12 variables). The data for the sensing variables are continuous, except for the binary state of the pumps and valves – which indicates the pumps and valve turned on and off. To evaluate the performance of machine learning models in distinguishing and identifying different failure types, a total of 29 variables among the 43 variables for system status were selected as input (feature) datasets for training the machine learning models – water levels at seven tanks, pressure at the nodes near 11 pumps and one valve, and flow status of nine pumps and one valve.

2.1.2. Characterization of C-town WDN failures

This study considered three types of WDN disruptions – i.e., conventional disruptions (pipe leaks/bursts), cyber-attacks, and physical attacks. Collecting real-world data balanced between normal operational and failure states is a challenge. This is because the occurrence of cyber-physical attacks in WDNs is rare, despite its growing risk, and the lack of a dataset (unbalanced dataset) can affect the performance of failure identification models (Dogo *et al.* 2020). Thus, as also considered in Taormina *et al.* (2018), the datasets for 29 variables for failure type identification were created through the simulation of hydraulic

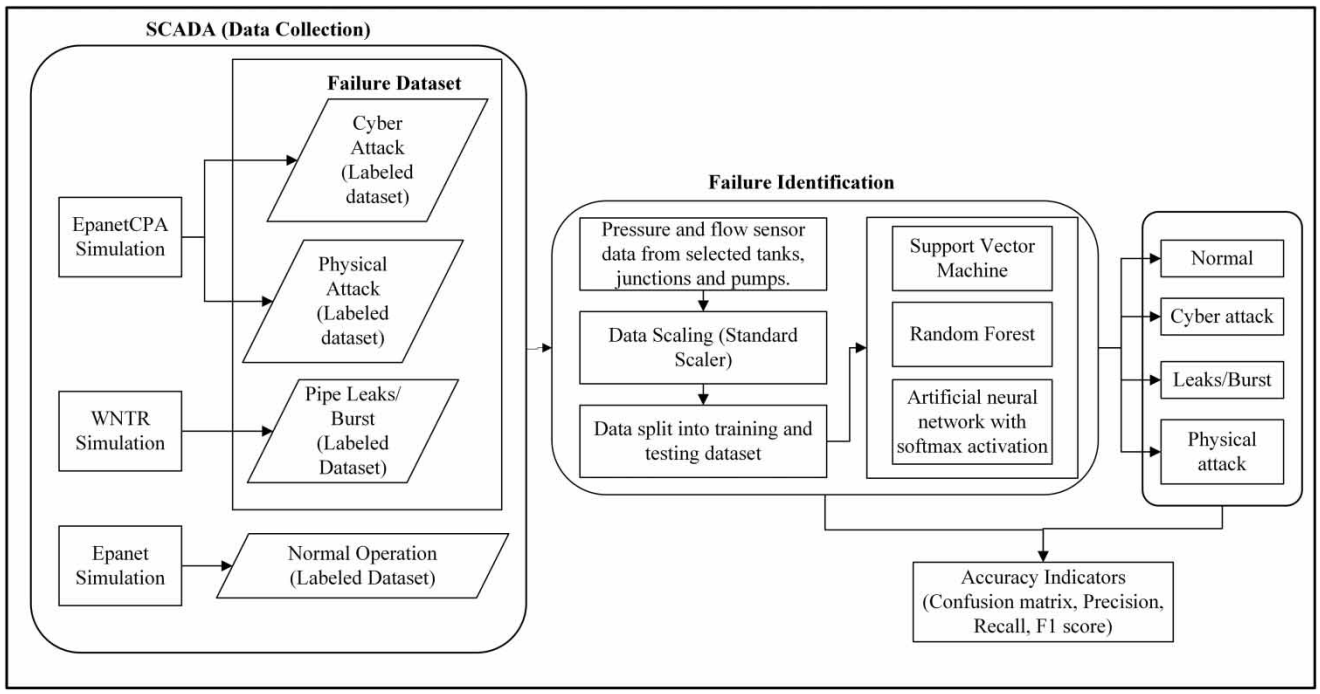


Figure 1 | The process of testing the performance of classification models in WDN failure type identification.

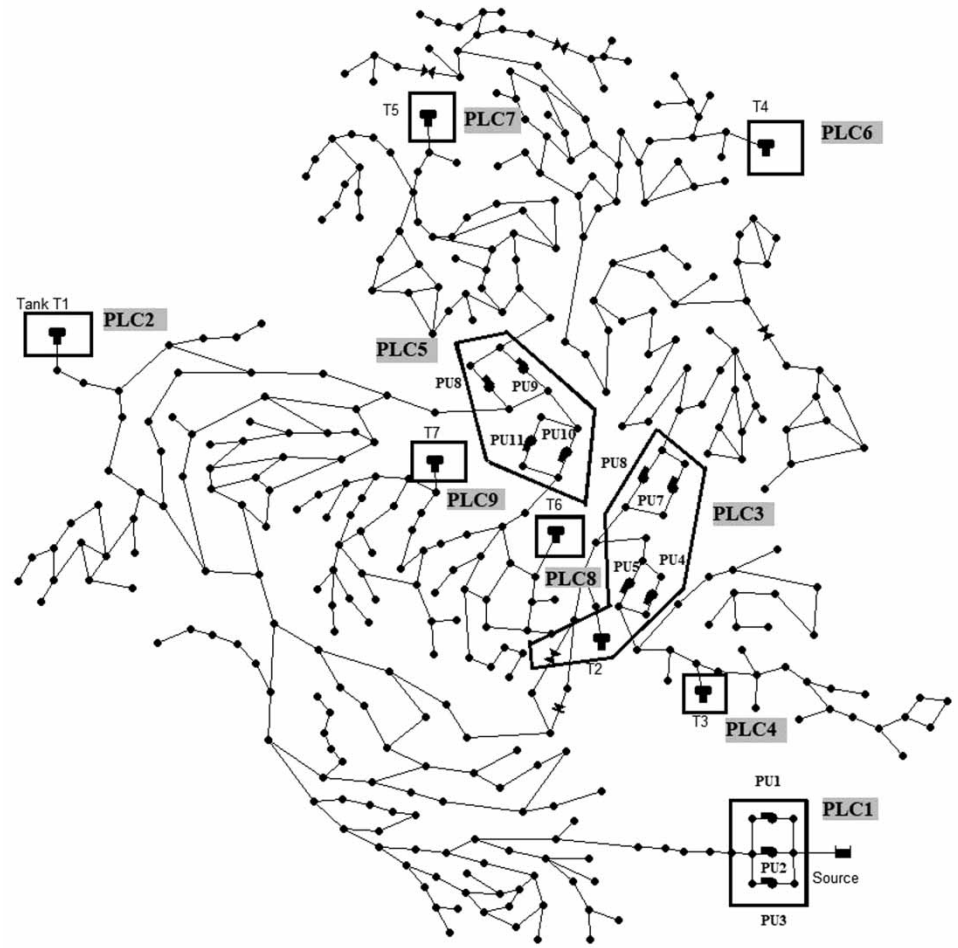


Figure 2 | Illustration of C-town WDN (adapted from Taormina *et al.* (2017)).

models – i.e., Water Network Tool for Resilience (WNTR) and epanetCPA (Klise *et al.* 2018; Taormina *et al.* 2019). WNTR is an open-source Python-based model that runs on the EPANET engine. This study used WNTR to generate datasets for conventional disruption scenarios (pipe leaks/bursts) by pressure-driven analysis. WNTR was iteratively simulated to obtain the operational data (e.g., nodal pressure, tank water level, water flow, and status of the pump), changing the parameters and leak nodes. The epanetCPA, an object-oriented MATLAB toolbox based on EPANET, was used for simulating both cyber and physical attacks in the C-town WDN. The epanetCPA tool can simulate the interactions between WDN's physical components (e.g., tanks, pumps, and valves) and cyber components (PLC and SCADA), which provides the flexibility to design the cyber and physical attack scenarios that can occur in WDNs. The datasets consisting of the events of 388 pipe leak, 128 cyber-attack, and 72 physical attack records, respectively, were created for training the failure identification models.

Conventional disruption scenarios

The datasets for the physical failure due to pipe leaks/bursts were generated using WNTR, which can simulate leaks in a WDN through a leak model (Klise *et al.* 2018). The leak scenarios assumed the leakage at a single node during 96, 108, 120, and 132 h, respectively, which was applied for all nodes. Then, the dataset for the scenarios of pipe leaks/bursts was generated by randomly selecting the leakage nodes, considering the pipe leaks/bursts in different locations. The leakage is modeled in WNTR through the orifice Equations (1) and (2):

$$Q_{\text{leak}} = C_d A p^{\alpha} \sqrt{\frac{2}{\rho}} \tag{1}$$

$$Q_{\text{leak}} = C_d A \sqrt{2gh}, \text{ when } \alpha = 0.5 \tag{2}$$

where Q_{leak} is the leak demand, C_d is the unitless discharge coefficient (0.75 assuming turbulent flow), A is the leak area (assuming a leak diameter of 0.05 m), p is the internal water pressure, α is an exponent related to the leaks and assumed to be 0.5, ρ is the water density, g is the gravitational acceleration, and h is the gauge head. A leak can be modelled at junctions or pipes under the user requirement. WNTR simulates the leak at the location by splitting the pipe at that point and considering it as a junction.

Cyber-attack scenarios

Cyber-attack scenarios for water CPSs are presented by Taormina *et al.* (2017, 2019). In this study, 6 days (144 h) duration of each scenario is created for the cyber-attack datasets. A total of 128 datasets for cyber-attack scenarios described in Table 1 are generated for training and testing machine learning models for failure identification in this study. Cyber-attack scenario 1 in Table 1 signifies the attack on the communication link between the sensors for the tank water level and PLCs. For example, the reading on the water level in a tank is manipulated as being constantly higher than a threshold level, regardless of its actual condition, which directs the PLC to close pumps and valves. Similarly, the same attack is simulated for other components (T2:PLC3, T3:PLC4, T4:PLC5, T5:PLC7, and T7:PLC9) for 96, 108, 120, and 132 h.

In scenario 2, the control logic of PLCs is manipulated, resulting in intermittent switching on/off of pumps. This attack scenario was also carried out for different components and duration, as shown in Table 1.

Scenario 3 is a Denial-of-Service attack (DOS) that was designed for the PLCs. In this scenario, PLC fails to receive the data for updated water levels for the tank and keeps the pump on. Scenarios 4, 5, and 6 were designed with a replay attack in scenarios 1, 2, and 3, respectively, but hide the attacks as if the WDN is in normal conditions by deliberately replaying data of the WDN status under a normal state. Here, the cyber-attack scenarios presented above are implemented by the

Table 1 | Cyber-attack scenario specification

No	Scenario	Attacked components	Duration (h)
1	Communication between tank water level and PLC	T1 and PLC2; T2 and PLC3; T3 and PLC4; T4 and PLC6; T5 and PLC7; T7 and PLC9	96,108,120,132
2	Modification of control logic of PLC (Switches the pump intermittently)	PLC1 and Pump 1 and 2; PLC3 and Pump 4 and 5; PLC3 and Pump 6 and 7; PLC5 and Pump 8; PLC5 and Pump 10 and 11	96,108,120,132
3	Denial of Service (DOS): Connection link between PLCs	PLC2 and PLC1; PLC4 and PLC3; PLC9 and PLC5; PLC6 and PLC3; PLC7 and PLC5	96,108,120,132
4	Replay attacks in Scenario 1	T1 and PLC2, and SCADA T2 and PLC3, and SCADA T3 and PLC4, and SCADA T4 and PLC6, and SCADA T5 and PLC7, and SCADA T7 and PLC9, and SCADA	96,108,120,132
5	Replay attacks in Scenario 2	PLC1 and Pump 1 and 2, and SCADA PLC3 and Pump 4 and 5, and SCADA PLC3 and Pump 6 and 7, and SCADA PLC5 and Pump 8, and SCADA PLC5 and Pump 10 and 11, and SCADA	96,108,120,132
6	Replay attacks in scenario 3	PLC2 and PLC1, and SCADA PLC4 and PLC3, and SCADA PLC9 and PLC5, and SCADA PLC6 and PLC3, and SCADA PLC7 and PLC5, and SCADA	96,108,120,132

attack on the cyber assets of WDNs. It is also noted that the attack scenarios (including physical attacks in the following section) can similarly occur from malfunctions or failures in sensors or actuators.

Physical attack scenarios

This study considered a physical attack as physically breaching the system's control and directly altering the system's operation – e.g., changing the pump status (on/off) being hidden against the control rule. Table 2 summarizes the specification of physical attack scenarios. Altogether 72 datasets were created every 6 days (144 h) with hourly interval data using epanetCPA in the C-town WDN.

Normal operation scenario

WNTR simulation was carried out with a pressure-driven analysis model for a 6-month period to produce the normal condition dataset (4,320 h). Data of 1-h intervals were used for training the machine learning models for failure identification. The simulation used the C-town WDN with the default base demand and demand patterns.

2.1.3. Generation of data noise

In this study, the training datasets for normal and failure conditions were obtained through the hydraulic models (WNTR and epanetCPA) simulation, which considered no noise in the sensor data. However, in practice, the sensor data of WDN contains errors and noise in their measurement. In this regard, this study additionally tested the machine learning model's performance in failure type identification against different noise levels. The noise was added to the datasets for continuous features, which follows Gaussian distribution. For every observation of a clear signal, a randomly generated noise value based on the Gaussian distribution was produced and added to the dataset that was obtained through hydraulic models' simulation (Abokifa *et al.* 2019). The noise value had varied standard deviation values from zero mean. The noisy datasets were produced using a standard deviation range of 0–0.3 with an interval of 0.05, with zero denoting data that is entirely clear and 3.0 denoting an increasing order of noise.

2.2. Failure identification models

Identifying and differentiating the type of a failure event among the ones that can occur in a WDN is a multiclass classification problem. In this regard, this study adopted three supervised machine learning models – ANN, SVM, and RF, which have been widely used for WDN anomaly detection (Jain *et al.* 1996; Breiman 2001; Widodo & Yang 2007).

2.2.1. ANN

The ANN is a supervised machine learning model with a network of multiple layers fully connected consisting of neurons. The basic frameworks are the input, hidden, and output layers (Fan *et al.* 2021). Links connect the nodes of a layer, and the network becomes increasingly deep as hidden layers are added. The ANN model can explain complicated nonlinear relations by increasing the number of neurons and hidden layers, which also produces good accuracy despite the expense of high computation requirements and the risk of overfitting (Fan *et al.* 2021).

In this study, the ANN architecture has one input layer with 29 features that comprise hourly interval pressure data from nodes and tanks, two hidden layers with 76 neurons each, and an output layer providing the probabilistic classification of the normal states and three disrupted states from pipe leaks, cyber-attacks, and physical attacks in the C-town WDN. Rectified linear unit (RELU) was selected as the activation function in hidden layers (Agarap 2018). The values in the output layer are scaled using the SoftMax function to reflect probabilities of normal, cyber-attack, physical attack, and pipe leak/burst events, which are added up to 1. Instead of simply dividing each probability by the total, it employs the exponential function, which helps highlight higher values and suppress lower ones. In contrast to linear regression, the SoftMax function allows for the presence of many classes that assist in multiclass classification (Qi *et al.* 2017).

Table 2 | Physical attack scenarios

No	Scenario	Components	Duration (h)
1	Turn on the pump	Pump 1, Pump 2, Pump 4, Pump 5, Pump 6, Pump 7, Pump 8, Pump 10, Pump 11	96,108,120,132
2	Turn off the pump	Pump 1, Pump 2, Pump 4, Pump 5, Pump 6, Pump 7, Pump 8, Pump 10, Pump 11	96,108,120,132

The ANN model was built using Keras (Kim *et al.* 2022) dense function, with the weights initialized automatically as biases. Keras dense function was selected because of its simplicity and fast iteration, even in complex models. A sequential layer activates feed-forward neural networks, and layers are added sequentially. Sequential models construct deep neural networks by adding layers on top of each other. The model was compiled with stochastic gradient descent (SGD) optimizer to minimize the loss function, which was set to 0.9. The learning rate was set to 0.01. Training data is normalized with a standard scaler. One hot encoding was applied that transfers categorical value to the multiple class columns and assigns a binary value of 0 and 1 to the respective class. Seventy-five per cent of data was used for training, 25% for testing with stratified sampling, and a batch size of 48 and 100 epochs was used while training and testing the dataset.

2.2.2. SVM

SVM is also a supervised machine learning model that has been mainly used for classification, regression, and outlier detection (Pedregosa *et al.* 2011). SVM can handle very large features, making it efficient in complex classification tasks (Widodo & Yang 2007). When using the SVM model, each data point is represented as a point in n number of dimensional spaces (features), with each feature's value being the value of a specific coordinate. Next, classification is performed by identifying the hyper-plane that effectively distinguishes the two classes. Higher-dimensional spaces are mapped using kernel functions to convert the original dataset, which includes both linear and nonlinear data, into a linear dataset. The most used are three types of kernels: linear, polynomial, and radial basis function (RBF). All SVM kernels include two main parameters – i.e., regularization parameter C and kernel coefficient, gamma (γ) (Sunkad & Soujanya 2016). Parameter C balances the misclassification of training samples versus decision surface simplicity. A small C soothes the decision surface, whereas a greater C attempts to categorize every training sample accurately. The gamma parameter defines the influence of a particular training data. Kernel RBF was selected with hyperparameters C and gamma equal to 10 and 0.07, respectively. A combined dataset for the normal and disrupted states from pipe leaks, cyber-attacks, and physical attacks was created and divided into a 75% training and a 25% test set to train and test the model for failure identification. Stratified sampling was conducted to uniformly distribute each failure class's samples into training and test datasets.

2.2.3. RF

RF is a supervised machine learning model that has been used as both a classifier and regressor (Breiman 2001). RF generates the decision tree by random data sampling and obtains the prediction from each tree, selecting the most appropriate solution by voting (Breiman 2001). It also provides the importance of each feature for classification and regression, which assists in the feature selection (Hasan *et al.* 2016). RF is one of the popular classification models because of its fast execution, minimal tuning parameters, ability to produce generalization error, and its applicability in high dimensional datasets (Cutler *et al.* 2012). In this study, the number of trees in the forest (n_estimators) was set to 10, and the criteria to measure the quality of a split were selected as Entropy. The combined failure dataset was split into 75% training, and 25% test. Samples were normalized with a standard scaler before training and testing.

2.3. Evaluation indicators

In general, the performance of data-driven models can be represented using four types: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) (Sokolova & Lapalme 2009). A desirable performance of the models is to detect the failure types without missing the failure conditions and reduce the false alarms. In this regard, the performance of the machine learning models in failure type identification was evaluated using the indicators of accuracy, precision, recall, and F1-score derived from the confusion matrix (Sokolova & Lapalme 2009). These indicators provide the values bounded between 0 and 1. The value 0 indicates poor performance, while the value 1 implies the best performance. The mathematical representation of the indicators is provided in Equations (3)–(6).

- Accuracy is defined as the ratio of the number of failure types that are correctly classified to the total number of failure events in the dataset, which is represented as:

$$\text{Accuracy} = \frac{\text{True}_{\text{Positive}} + \text{True}_{\text{Negative}}}{\text{True}_{\text{Positive}} + \text{True}_{\text{Negative}} + \text{False}_{\text{Positive}} + \text{False}_{\text{Negative}}} \tag{3}$$

- Precision shows what fraction of positive classification (failure identification) is actually correct, which is calculated as:

$$\text{Precision} = \frac{\text{True}_{\text{Positive}}}{\text{True}_{\text{Positive}} + \text{False}_{\text{Positive}}} \tag{4}$$

- Recall shows how many times the model provides true classification among all actual failure events, of which type should be identified as true. It is calculated as:

$$\text{Recall} = \frac{\text{True}_{\text{Positive}}}{\text{True}_{\text{Positive}} + \text{False}_{\text{Negative}}} \tag{5}$$

- F1-score is the harmonic average of precision and recall values. This indicator shows the balance (tradeoff) between the indicators of precision and recall, as represented:

$$F1_{\text{score}} = 2\frac{\text{Recall}*\text{Precision}}{\text{Recall} + \text{Precision}} \tag{6}$$

3. RESULTS AND DISCUSSION

3.1. Hydraulic response to failure events

Figure 3 shows the time variation of the water levels in Tank 2 (Figure 2) under normal operating conditions, pipe leaks/bursts (Junction 211), cyber-attacks (attack on the communication link between the Tank 2 water level sensor and PLC 3 by manipulating the tank water level readings), and physical attacks (turning off pump 1). It can be observed that the different types of disruptive events produced different hydraulic responses compared to the normal conditions. In the scenario of pipe leaks/bursts, the water level in Tank 2 dropped due to an increase in water discharge with pipe leaks. However, it is also observed that Tank 2 was partially filled intermittently due to the normal operation of the pumps feeding water into the WDN. On the other hand, in the scenario of the cyber-attack, the manipulated readings of tank water level were sent to PLC 3, which controls pumps 4, 5, 6, and 7 and valve 2. This resulted in the closure of valve 2 and the deactivation of the pumps, which, in turn, led to a significant and rapid drop in the tank water level. In the scenario of the physical attack, the attacker closed pump 1, which is a main pump feeding water into the WDN from a water source. This has also caused a significant drop in the tank water level.

It is also noted from Figure 3 that the cyber and physical attack events produced similar hydraulic responses (emptying tank water) with time, even though they were different attack scenarios. The rapid and significant drop of the tank water level could be consequently produced due to the cyber-attack manipulating tank water level readings, leading to the closure of

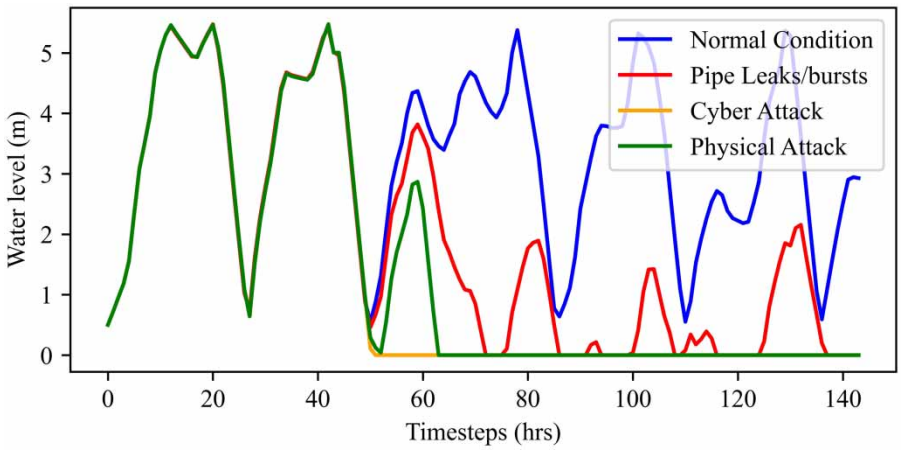


Figure 3 | Time variation of Tank 2 water levels during normal and disrupted conditions.

valve 2 and the deactivation of the pumps, and the physical attack turning off the main pump directly. Based on similar results from this study, Taormina *et al.* (2017) also noted the importance of identifying the cause or approach of cyber-physical attacks as well as detecting anomalous conditions of the system due to the attacks. In this context, it is considered that the simulation of hydraulic models (WNTR and epanetCPA) produced appropriate datasets containing both different and similar hydraulic responses of C-town WDN under disruptive events, which are used to test the failure identification performance of the target machine learning models in the following section.

3.2. Failure type identification performance

Three supervised machine learning algorithms – SVM, RF, and ANN – were used to test their capability to identify the types of failures, i.e., pipe leaks/bursts, cyber-attacks, and physical attacks, using the dataset for 29 selected features under the failures. SVM and RF produced labels for classification failures, while ANN with SoftMax activation produced probabilistic values for each failure occurrence. The class with the highest probability was considered as the anticipated failure class to evaluate the model and compare it with SVM and RF. Figure 4 shows the confusion matrix for the constructed SVM, RF, and ANN models. It can be noted that all three machine learning models provide overall reliable performance in differentiating and identifying the types of failure events. A review of the confusion matrix for the SVM model in Figure 4(a) revealed that 98.87% of pipe leak events, 87.89% of physical attacks, 86.18% of normal conditions, and 81.62% of cyber-attacks were correctly identified. On the other hand, 16.17% of physical attack events were classified as cyber-attacks and 7.48% of cyber-attack events were classified as physical attacks. Similarly, when the confusion matrix of the RF in Figure 4(b) is analyzed,

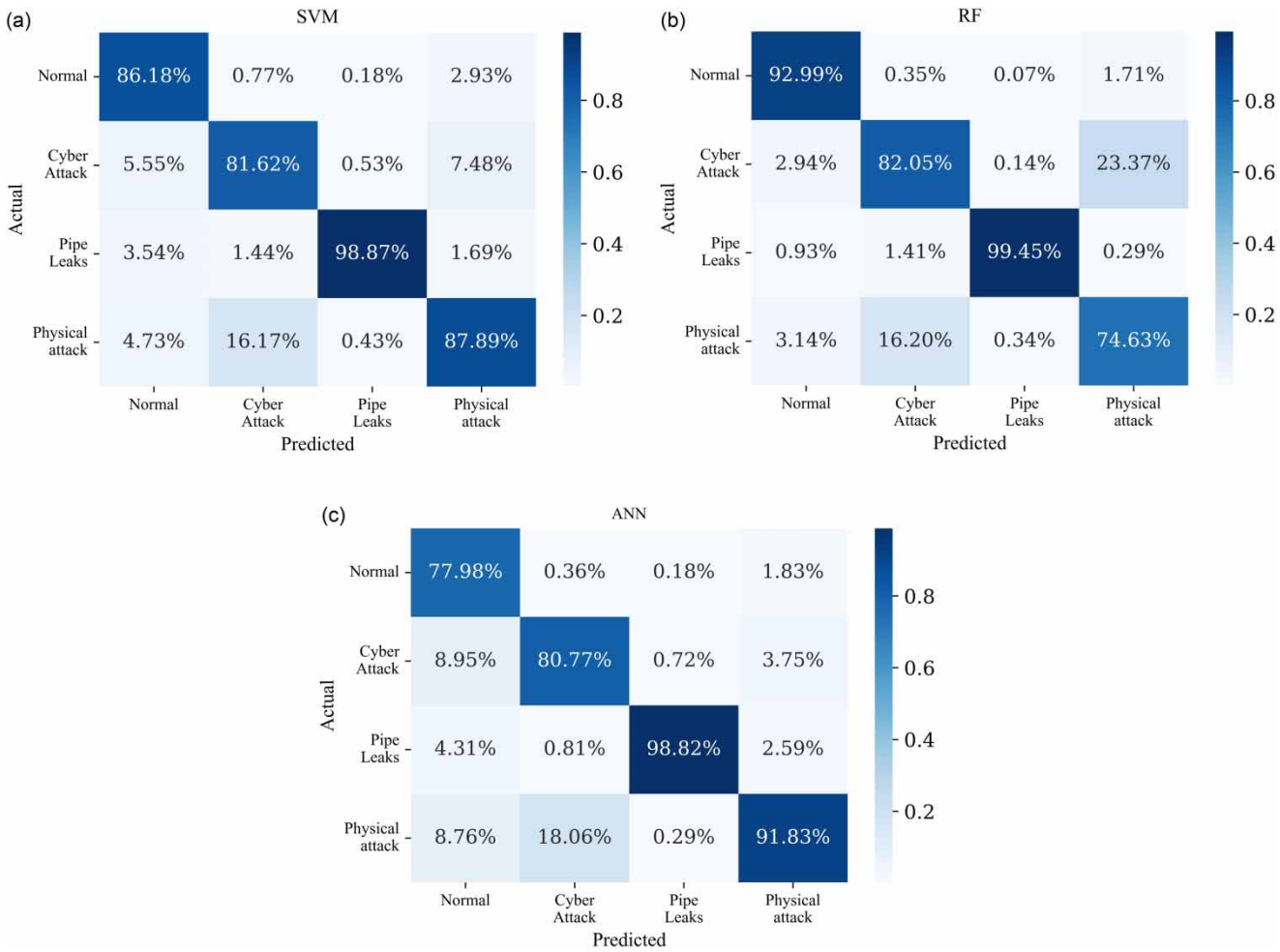


Figure 4 | The confusion matrix for the accuracy of (a) SVM, (b) RF, and (c) ANN in failure identification.

99.45% of pipe leak events, 74.63% of physical attacks, 92.99% of normal conditions, and 82.05% of cyber-attack events were accurately classified, which shows slightly better performance in differentiating physical and cyber-attacks from pipe leak/burst events than SVM. However, the RF model produced more false identification of the attacks – e.g., 23.37% of cyber-attack events were classified as physical attacks. For the ANN confusion matrix in Figure 4(c), 98.82% of pipe leak events, 91.83% of physical attacks, 77.98% of normal conditions, and 80.74% of cyber-attacks were accurately classified.

The higher rate of misclassification between cyber-attack and physical attack events is attributed to the similarity in hydraulic response of the WDN from different failure processes between the two types of attacks. In a physical attack scenario designed in this study, the attacker physically gains access to the pumps, runs them continuously, or shuts them down when they are active. Similar results can also be achieved during a cyber-attack via an attack on communication, denial of service, and attack on PLCs. For instance, in a denial-of-service attack, the PLC is unable to receive updated water levels from the tank, causing the pump to run or stop for an extended period. These different scenarios can produce similar hydraulic performance; however, their resolution may require different options to recover the disrupted system. In this regard, the SVM, RF, and ANN models need to be trained and improved further to distinguish and identify the failure types that have similar hydraulic responses from different failure events, especially which require different approaches for emergency and recovery actions to improve WDN resilience.

Table 3 presents the precision, recall, and F1-score of each model based on the confusion matrix (Figure 4). It is seen that all failure types except physical attacks were identified with higher performance scores. The three models showed higher performance in identifying conventional disruptions with pipe leaks, based on the F1 scores, compared to the disruptions from cyber and physical attacks. As described above, this occurred due to the misclassification of the models between cyber and physical attack events, some of which produced similar hydraulic responses in the C-town WDN (Figure 3). The F1-score for cyber-attack identification was 0.87/0.84/0.83 for SVM/RF/ANN models, whereas the F1-score for physical attack identification was 0.72/0.68/0.57, which showed a higher rate of misclassification with physical attack events. Similarly, SVM and RF had accuracy of 0.88 and ANN had accuracy of 0.84 in failure type identification.

Overall, the evaluation of the model's performance with accuracy, precision, recall, and F1-score suggested that SVM and RF models had superior performance in distinguishing and identifying overall failure types, compared to ANN. However, it should be noted that the three machine learning models had different performances depending on the failure types. It is considered that the different performances are attributed to the impacts of various factors such as models' algorithms (the degree of linearity/nonlinearity between input and output variables), selected features and their scaling, failure event specifications, and training data size (Ahsan *et al.* 2021; Zhang *et al.* 2021; Umoh *et al.* 2022). For example, SVM is more sensitive to feature scaling than RF, while ANN can benefit from scaled features to accelerate convergence (Ahsan *et al.* 2021). In addition, ANN shows good performance with large training datasets in capturing complex relationships of the features, while RF and SVM are more effective in training limited or smaller datasets (Zhang *et al.* 2021). In this regard, it would be suggested to couple multiple machine learning models in a single framework to differentiate different failure types, rather than relying on a single best-performing model.

Given the superiority of supervised learning in multiclass classification, the three machine learning models, especially SVM and RF, had reliable performance in identifying the failure types. However, the machine learning model's performance can

Table 3 | Performance value obtained from SVM, RF, and ANN

Failure class	Models	Precision	Recall	F1-score
Normal	SVM	0.90	0.98	0.94
	RF	0.93	0.99	0.96
	ANN	0.70	1	0.82
Pipe leaks/bursts	SVM	0.99	0.93	0.96
	RF	0.99	0.97	0.98
	ANN	1	0.90	0.95
Cyber-attack	SVM	0.82	0.92	0.87
	RF	0.82	0.85	0.84
	ANN	0.83	0.84	0.83
Physical attack	SVM	0.86	0.62	0.72
	RF	0.74	0.63	0.68
	ANN	0.98	0.41	0.57

vary significantly based on the placement of the sensors or selection of the features. The monitoring sensors need to be installed strategically based on feature selection or optimization based on model performance, and the model's parameters must be optimized for global monitoring and detection of diverse failures in WDN.

In addition, a real-world challenge in training and testing the machine learning models or data-driven models is to find the cyber-physical failure datasets in balance to the datasets of normal operating conditions, which can affect the failure identification performance of the models in this study (Fan *et al.* 2021). Failure events due to cyber-physical attacks in WDNs rarely occur, compared to the period of normal operating conditions and other conventional disruptive events (e.g., pipe leaks). Thus, the unbalanced datasets consisting of system's performance under normal and disrupted conditions can impair the classification performance of the machine learning models in a real WDN. In this regard, incorporating synthetic data that are produced using hydraulic simulation models (e.g., WNTR and epanetCPA) into unbalanced datasets can be a way to improve the performance of the machine learning models in failure type identification.

3.3. Failure identification under data noise

The machine learning models – SVM, RF, and ANN – were trained using the noisy datasets of normal and failure conditions of the C-town WDN and their performances were tested depending on the noise levels of the datasets. Figure 5 shows the performance of SVM, RF, and ANN in failure type identification with the noisy datasets. As expected, it is noted that the overall performance of the three models in failure type identification decreased as the signal noise increased. This was because the models misinterpreted the noise as extended failure conditions. For relatively small data noise, SVM and ANN nearly maintained the level of accuracy in the cases of training them with noise-free datasets, while RF showed a rapid decline in its performance. However, it can be observed from the slope of accuracy curves in Figure 5, that the decreasing rate of accuracy of RF was less, compared to SVM and ANN as the data noise levels increased.

As seen in Figure 5, SVM demonstrated relatively higher accuracy for overall data noise levels, compared to RF and ANN. However, it also showed a consistent drop in the performance with increasing data noise. In turn, its accuracy decreased lower than the accuracy of ANN, which was the least sensitive among the three models to the data noise levels. Considering the trend in the RF performance from Figure 5, RF was expected to be less sensitive to the data noise at high noise levels compared to SVN and ANN. This implies that the three different models can demonstrate different performances in failure type identification depending on the noise levels of training data. In practice, sensing data has noise at various levels from various noise sources such as sensor ageing, malfunctions, and miscalibration, communication disruptions, traffic, and human errors (Rousso *et al.* 2023). As observed in Figure 5, the best-performing models can vary with the levels of data noise. Therefore, it is suggested to not rely on a single best-performing model for distinguishing and identifying a failure event but rather integrate the results from multiple models for more reliable failure type identification with confidence across different levels of data noise.

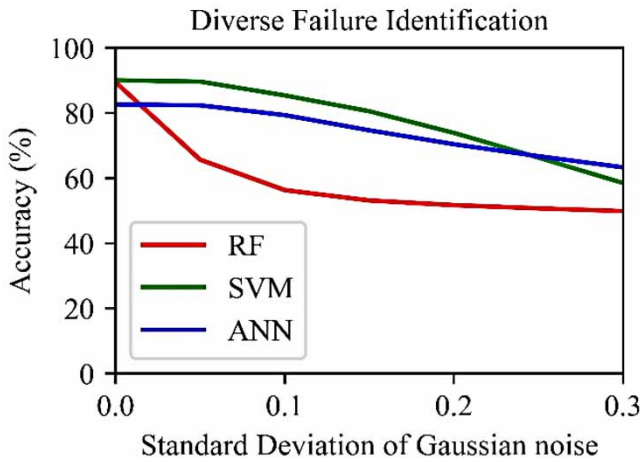


Figure 5 | Effect of data noise on models' performance in failure type identification.

4. CONCLUSIONS

Resilience-based strategies aiming at the minimization of WDN's disruptions from a failure event and rapid recovery have received great attention in recent years to design, operate, and manage critical infrastructure including WDNs. In this regard, a smart systems approach with ICT-based sensors and controllers has been considered and employed in the infrastructure to effectively and resiliently manage disruptive events. However, despite the potential effects of the smart systems approach, they are more exposed to subtle cyber threats. Consequently, this may lead to an increase in the complexity of detecting and identifying failure types and further exacerbating the vulnerability to both cyber and conventional failure events, giving rise to concerns in infrastructure services. Thus, the first step of reactive actions to secure the resilience of smart infrastructure systems will be rapid detection and identification of a failure event, which will be followed by emergency and recovery actions. Anomaly detection and localization are also a critical step in responding to a disruptive event with emergency actions. However, when a failure event occurs, the challenge is how to distinguish the actual failure event from the potential failure events that could occur in the infrastructure systems.

In this regard, this study investigated the performance of three supervised machine learning models – SVM, RF, and ANN – in identifying failure types among cyber-physical attacks and conventional physical disruptions (pipe leaks/bursts). They were trained and tested using the datasets including 29 features related to tank water levels, nodal pressure, and flow status of pumps and valves under the three types of WDN failures – i.e., pipe leaks/bursts, cyber-attacks, and physical attacks. Overall, three models showed reliable performance in identifying the failure types. However, their performances varied depending on the specific failure types, and no single model with consistently superior performance for all failure types was identified. In addition, testing the three models with data noise showed a decrease in their performance in failure type identification. However, the variation of their performance was also different depending on the classification models and levels of data noise. Thus, the use of multiple classification models, rather than relying on a single best-performing model, is recommended to improve the capability of WDNs to distinguish and identify a failure event from different potential failure events.

In addition, the classification models produced a higher rate of misclassification between cyber-attack and physical attack events, due to the similarity in hydraulic response of the C-town WDN from the different failure events. Thus, a failure type identification framework with multiple classification models needs to be designed to distinguish the failure events that can produce similar hydraulic responses, which require different emergency and recovery options during system disruptions. These results suggest insights into building a data-driven analytics framework for the reliable classification and identification of failure types in real-world WDNs.

The findings of this study will contribute to improving the capability of WDNs to rapidly and reliably differentiate and identify failure types and, in turn, find adequate emergency and recovery options depending on the failure events. It is also considered that the findings and insights in this study can be further discussed in distinguishing and identifying contamination events (e.g., malicious contaminant injection or accidental contamination intrusion). However, the application of machine learning classification models for reliable failure identification suggests the following challenges as future work. The SVM, RF, and ANN produced misclassification between cyber-attack and physical attack events, due to the similar hydraulic responses of the C-town WDN from different failure types, which can require different approaches to emergency and recovery options. Thus, further studies on data-driven models to characterize and differentiate failure types that can produce similar hydraulic responses are needed.

Second, this study considered three supervised classification models (SVM, RF, and ANN) and three failure types (conventional pipe leaks/bursts, cyber-attacks, and physical attacks). In this regard, more diverse data-driven (machine learning and deep learning) models can be tested to identify various failure types including attacks that maliciously open a fire hydrant, contamination and mechanical failures with various specific failure scenarios. In addition, the failure identification performance of the models can be further discussed with the different sizes/locations (e.g., proximity to critical storage tanks or reservoirs), severity, and timing and the various types of conventional disruptive events and the operational failures due to not only cyber-physical attacks but also malfunctions/errors in cyber and physical assets of WDNs.

Third, the SVM, RF, and ANN models in this study showed a reliable performance in the presence of data noise. However, as the level of data noise increased, their performances varied with the noise levels. Thus, further investigation of their performance using real-world datasets (e.g., missing data, poor sensor data quality) is suggested, which can increase the chance of practical applicability.

Fourth, the incidents of cyber-physical attacks in the water sector are reported as the third most frequently targeted area among critical infrastructure systems. Considering the interconnected sectors such as energy/power systems (that have the first-largest incidents), the vulnerability of the water systems to cyber-physical attacks is relatively high. Therefore, future studies can be guided more toward understanding the cascading failures between interdependent infrastructure systems due to cyber-physical attacks.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abokifa, A. A., Haddad, K., Lo, C. & Biswas, P. 2019 Real-time identification of cyber-physical attacks on water distribution systems via machine learning–based anomaly detection techniques. *Journal of Water Resources Planning and Management* **145** (1), 04018089. doi:10.1061/(ASCE)WR.1943-5452.0001023.

Agarap, A. F. 2018 Deep Learning using Rectified Linear Units (ReLU). *arXiv preprint arXiv:*1803.08375. doi:10.48550/arXiv.1803.08375.

Ahmad, Z., Nguyen, T.-K. & Kim, J.-M. 2023 Leak detection and size identification in fluid pipelines using a novel vulnerability index and 1-D convolutional neural network. *Engineering Applications of Computational Fluid Mechanics* **17** (1). doi:10.1080/19942060.2023.2165159.

Ahsan, M., Mahmud, M., Saha, P., Gupta, K. & Siddique, Z. 2021 Effect of data scaling methods on machine learning algorithms and model performance. *Technologies* **9** (3), 52. doi:10.3390/technologies9030052.

Amoatey, P. K., Obiri-Yeboah, A. A. & Akosah-Kusi, M. 2021 Impact of active night population and leakage exponent on leakage estimation in developing countries. *Water Practice and Technology* **17** (1), 14–25. doi:10.2166/wpt.2021.124.

Anwar, A., Mahmood, A. N. & Shah, Z. 2015 A Data-driven approach to distinguish cyber-attacks from physical faults in a smart grid. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, Melbourne, Australia, pp. 1811–1814. doi:10.1145/2806416.2806648.

Asghari, V., Kazemi, M. H., Duan, H.-F., Hsu, S.-C. & Keramat, A. 2023 Machine learning modeling for spectral transient-based leak detection. *Automation in Construction* **146**, 104686. doi:10.1016/j.autcon.2022.104686.

Breiman, L. 2001 Random forests. *Machine Learning* **45** (1), 5–32. doi:10.1023/a:1010933404324.

Brentan, B., Rezende, P., Barros, D., Meirelles, G., Luvizotto, E. & Izquierdo, J. 2021 Cyber-attack detection in water distribution systems based on blind sources separation technique. *Water* **13** (6), 795. doi:10.3390/w13060795.

Clark, R. M., Panguluri, S., Nelson, T. D. & Wyman, R. P. 2017 Protecting drinking water utilities from cyberthreats. *American Water Works Association* **109**, 50–58.

Cutler, A., Cutler, D. R. & Stevens, J. R. 2012 Random forests. *Ensemble Machine Learning* 157–175. doi:10.1007/978-1-4419-9326-7_5.

Do, V. L., Fillatre, L., Nikiforov, I. & Willett, P. 2017 Security of SCADA systems against cyber–physical attacks. *IEEE Aerospace and Electronic Systems Magazine* **32** (5), 28–45. doi:10.1109/MAES.2017.160047.

Dogo, E. M., Nwulu, N. I., Twala, B. & Aigbavboa, C. O. 2020 Empirical comparison of approaches for mitigating effects of class imbalances in water quality anomaly detection. *IEEE Access* **8**, 218015–218036. doi:10.1109/access.2020.3038658.

El-Zahab, S. & Zayed, T. 2019 Leak detection in water distribution networks: An introductory overview. *Smart Water* **4** (1), 5. doi:10.1186/s40713-019-0017-x.

Fan, X., Zhang, X. & Yu, X. B. 2021 Machine learning model and strategy for fast and accurate detection of leaks in water supply network. *Journal of Infrastructure Preservation and Resilience* **2** (1), 10. doi:10.1186/s43065-021-00021-6.

Głomb, P., Cholewa, M., Koral, W., Madej, A. & Romaszewski, M. 2023 Detection of emergent leaks using machine learning approaches. *Water Supply* **23** (6), 2370–2386. doi:10.2166/ws.2023.118.

Hasan, Md. A. M., Nasser, M., Ahmad, S. & Molla, K. I. 2016 Feature selection for intrusion detection using random forest. *Journal of Information Security* **7** (3), 129–140. doi:10.4236/jis.2016.73009.

Hashim, H., Ryan, P. & Clifford, E. 2020 A statistically based fault detection and diagnosis approach for non-residential building water distribution systems. *Advanced Engineering Informatics* **46**, 101187. doi:10.1016/j.aei.2020.101187.

Housh, M. & Ohar, Z. 2018 Model-based approach for cyber-physical attack detection in water distribution systems. *Water Research* **139**, 132–143. doi:10.1016/j.watres.2018.03.039.

Housh, M., Kadosh, N. & Haddad, J. 2022 Detecting and localizing cyber-physical attacks in water distribution systems without records of labeled attacks. *Sensors* **22** (16), 6035. doi:10.3390/s22166035.

Jain, A. K., Mao, J. & Mohiuddin, K. M. 1996 Artificial neural networks: A tutorial. *Computer* **29** (3), 31–44. doi:10.1109/2.485891.

Kim, S., Wimmer, H. & Kim, J. 2022 Analysis of deep learning libraries: Keras, pytorch, and MXnet. In *2022 IEEE/ACIS 20th International Conference on Software Engineering Research, Management and Applications (SERA)*, pp. 54–62. doi:10.1109/SERA54885.2022.9806734.

Klise, K. A., Murray, R. & Haxton, T. 2018 An overview of the water network tool for resilience (WNTR). In *1st International WDSA/CCWI2018 Joint Conference*, Kingston, Ontario, Canada, pp. 1–8.

Mashhadi, N., Shahrour, I., Attoue, N., El Khattabi, J. & Aljer, A. 2021 Use of machine learning for leak detection and localization in water distribution systems. *Smart Cities* **4** (4), 1293–1315. doi:10.3390/smartcities4040069.

Mutchek, M. & Williams, E. 2014 Moving towards sustainable and resilient smart water grids. *Challenges* **5** (1), 123–137. doi:10.3390/challe5010123.

Nam, K., Ifaei, P., Heo, S., Rhee, G., Lee, S. & Yoo, C. 2019 An efficient burst detection and isolation monitoring system for water distribution networks using multivariate statistical techniques. *Sustainability* **11** (10), 2970. doi:10.3390/su11102970.

Nazir, S., Patel, S. & Patel, D. 2021 Autoencoder based anomaly detection for SCADA networks. *International Journal of Artificial Intelligence and Machine Learning* **11** (2), 83–99. doi:10.4018/IJAIML.20210701.oa6.

Patil, A., Kamuni, V., Sheikh, A., Wagh, S. & Singh, N. 2019 A machine learning approach to distinguish faults and cyberattacks in smart buildings. In: *2019 9th International Conference on Power and Energy Systems (ICPES)*. IEEE, Perth, Australia, pp. 1–6. doi:10.1109/ICPES47639.2019.9105507.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A. & Cournapeau, D. 2011 Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830.

Qi, X., Wang, T. & Liu, J. 2017 Comparison of support vector machine and softmax classifiers in computer vision. In *2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, pp. 151–155. doi:10.1109/ICMCCE.2017.49.

Ramotsoela, D., Abu-Mahfouz, A. & Hancke, G. 2018 A survey of anomaly detection in industrial wireless sensor networks with critical water system infrastructure as a case study. *Sensors* **18** (8), 2491. doi:10.3390/s18082491.

Rousso, B. Z., Lambert, M. & Gong, J. 2023 Smart water networks: A systematic review of applications using high-frequency pressure and acoustic sensors in real water distribution systems. *Journal of Cleaner Production* **410**, 137193.

Shin, S., Lee, S., Judi, D., Parvania, M., Goharian, E., McPherson, T. & Burian, S. 2018 A systematic review of quantitative resilience measures for water infrastructure systems. *Water* **10** (2), 164. doi:10.3390/w10020164.

Shin, S., Lee, S., Burian, S. J., Judi, D. R. & McPherson, T. 2020 Evaluating resilience of water distribution networks to operational failures from cyber-physical attacks. *Journal of Environmental Engineering* **146** (3), 04020003. doi:10.1061/(ASCE)EE.1943-7870.0001665.

Siddique, M. F., Ahmad, Z. & Kim, J.-M. 2023 Pipeline leak diagnosis based on leak-augmented scalograms and deep learning. *Engineering Applications of Computational Fluid Mechanics* **17** (1). doi:10.1080/19942060.2023.2225577.

Sokolova, M. & Lapalme, G. 2009 A systematic analysis of performance measures for classification tasks. *Information Processing & Management* **45** (4), 427–437. doi:10.1016/j.ipm.2009.03.002.

Sunkad, Z. A. & Soujanya 2016 Feature selection and hyperparameter optimization of SVM for human activity recognition. In *2016 3rd International Conference on Soft Computing & Machine Intelligence (ISCMI)*, pp. 104–109. doi:10.1109/ISCMI.2016.30.

Taormina, R. & Galelli, S. 2018 Deep-learning approach to the detection and localization of cyber-physical attacks on water distribution systems. *Journal of Water Resources Planning and Management* **144** (10), 04018065. doi:10.1061/(ASCE)WR.1943-5452.0000983.

Taormina, R., Galelli, S., Tippenhauer, N. O., Salomons, E. & Ostfeld, A. 2017 Characterizing cyber-physical attacks on water distribution systems. *Journal of Water Resources Planning and Management* **143** (5), 04017009. doi:10.1061/(ASCE)WR.1943-5452.0000749.

Taormina, R., Galelli, S., Tippenhauer, N. O., Salomons, E., Ostfeld, A., Eliades, D. G., Aghashahi, M., Sundararajan, R., Pourahmadi, M., Banks, M. K., Brentan, B. M., Campbell, E., Lima, G., Manzi, D., Ayala-Cabrera, D., Herrera, M., Montalvo, I., Izquierdo, J., Luvizotto, E., Chandy, S. E., Rasekh, A., Barker, Z. A., Campbell, B., Shafiee, M. E., Giacomoni, M., Gatsis, N., Taha, A., Abokifa, A. A., Haddad, K., Lo, C. S., Biswas, P., Pasha, M. F. K., Kc, B., Somasundaram, S. L., Housh, M. & Ohar, Z. 2018 Battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks. *Journal of Water Resources Planning and Management* **144** (8), 04018048. doi:10.1061/(ASCE)WR.1943-5452.0000969.

Taormina, R., Galelli, S., Douglas, H. C., Tippenhauer, N. O., Salomons, E. & Ostfeld, A. 2019 A toolbox for assessing the impacts of cyber-physical attacks on water distribution systems. *Environmental Modelling & Software* **112**, 46–51. doi:10.1016/j.envsoft.2018.11.008.

Tsiami, L. & Makropoulos, C. 2021 Cyber-physical attack detection in water distribution systems with temporal graph convolutional neural networks. *Water* **13** (9), 1247. doi:10.3390/w13091247.

Umoh, U. A., Eyoh, I. J., Murugesan, V. S. & Nyoho, E. E. 2022 Fuzzy-machine learning models for the prediction of fire outbreaks: A comparative analysis. *Artificial Intelligence and Machine Learning for EDGE Computing* 207–233. doi:10.1016/b978-0-12-824054-0.00025-3.

Walsby, C. 2013 The power of smart water networks. *American Water Works Association* **105** (3), 72–76.

Widodo, A. & Yang, B.-S. 2007 Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing* **21** (6), 2560–2574. doi:10.1016/j.ymssp.2006.12.007.

Wu, Z. Y., Chew, A., Meng, X., Cai, J., Pok, J., Kalfarisi, R., Lai, K. C., Hew, S. F. & Wong, J. J. 2022 Data-driven and model-based framework for smart water grid anomaly detection and localization. *Journal of Water Supply: Research and Technology-Aqua* **71** (1), 31–41. doi:10.2166/aqua.2021.091.

Xue, Z., Tao, L., Fuchun, J., Riehle, E., Xiang, H., Bowen, N. & Singh, R. P. 2020 Application of acoustic intelligent leak detection in an urban water supply pipe network. *Journal of Water Supply: Research and Technology-Aqua* **69** (5), 512–520. doi:10.2166/aqua.2020.022.

Zhang, C., Bengio, S., Hardt, M., Recht, B. & Vinyals, O. 2021 Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM* **64** (3), 107–115. doi:10.1145/3446776.

First received 17 October 2023; accepted in revised form 14 February 2024. Available online 28 February 2024