water

MDPI

Article

Combined Anomaly Detection Framework for Digital Twins of Water Treatment Facilities

Yuying Wei [1,2], Adrian Wing-Keung Law [1,3,*], Chun Yang [4] and Di Tang [3]

[1] School of Civil and Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore; yuying001@e.ntu.edu.sg
[2] Environmental Process Modelling Centre, Interdisciplinary Graduate Programme, Nanyang Environment and Water Research Institute (NEWRI), Nanyang Technological University, Singapore 637141, Singapore
[3] Environmental Process Modelling Centre, Nanyang Environment and Water Research Institute (NEWRI), Nanyang Technological University, 1 CleanTech Loop, CleanTech One, #06-08, Singapore 637141, Singapore; di.tang@ntu.edu.sg
[4] School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore; mcyang@ntu.edu.sg
* Correspondence: cwklaw@ntu.edu.sg

Abstract: Digital twins of cyber-physical systems with automated process control systems using programmable logic controllers (PLCs) are increasingly popular nowadays. At the same time, cyber-physical security is also a growing concern with system connectivity. This study develops a combined anomaly detection framework (CADF) against various types of security attacks on the digital twin of process control in water treatment facilities. CADF utilizes the PLC-based whitelist system to detect anomalies that target the actuators and the deep learning approach of natural gradient boosting (NGBoost) and probabilistic assessment to detect anomalies that target the sensors. The effectiveness of CADF is verified using a physical facility for water treatment with membrane processes called the Secure Water Treatment (SWaT) system in the Singapore University of Technology and Design. Various attack scenarios are tested in SWaT by falsifying the reported values of sensors and actuators in the digital twin process. These scenarios include both trivial attacks, which are commonly studied, as well as non-trivial (i.e., sophisticated) attacks, which are rarely reported. The results show that CADF performs very well with good detection accuracy in all scenarios, and particularly, it is able to detect all sophisticated attacks while ongoing before they can induce damage to the water treatment facility. CADF can be further extended to other cyber-physical systems in the future.

Keywords: anomaly detection; digital twin; NGBoost; probabilistic forecasting; programmable logic controller

Citation: Wei, Y.; Law, A.W.-K.; Yang, C.; Tang, D. Combined Anomaly Detection Framework for Digital Twins of Water Treatment Facilities. *Water* **2022**, *14*, 1001. https://doi.org/10.3390/w14071001

Academic Editor: Guido D'Urso

Received: 21 February 2022
Accepted: 20 March 2022
Published: 22 March 2022

1. Introduction

In recent years, the implementation of digital twins is beginning to play an important role in smart manufacturing in diverse fields, including metallurgical building blocks [1], automated flow-shop manufacturing [2], CNC machine tools [3], airframe health monitoring [4], hydraulic supports monitoring [5], electric power ecosystem [6], and numerous others. The digital twin mirrors the physical asset in real time through data communication across multiple dimensions, as shown in Figure 1. Subsequently, the digital twin's behavior can be established based on a combination of dynamic and data-driven methods for various objectives, including to safeguard the physical asset as well as to optimize the operations in order to minimize the consumption of energy and materials [7]. Overall, the implementation of a digital twin as a digital replica of the physical asset has matured significantly over the last decade with the advancement of digital technologies that enable fast data synchronization among sensors and actuators within the asset [8,9].

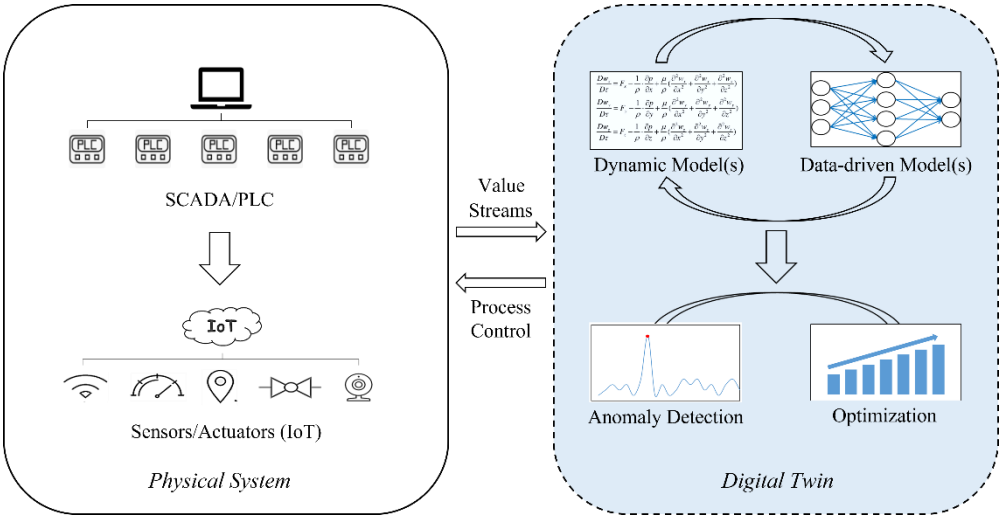

Figure 1. Schematic diagram of the digital twin concept (red point represents the detected anomaly).

The application of digital twins to critical infrastructures, such as water treatment facilities, can be susceptible to cyber-physical attacks that may adversely affect the functionality of the system, with potentially devastating consequences [10–12]. These security issues are of high concern and include, for example, the cyber-attack of Colonial Pipeline for ransoming, which occurred recently in May 2021 [13]. Attacks with malicious intent to damage the system can also be delivered when the system is operating in a local cloud environment [14]. This is because hackers can invade selected devices in the system and release cyber-physical attack commands after they breach the network security of the local cloud [10,15].

Programmable logic controllers (PLCs) have been the key common building block of industrial cyber-physical systems throughout the automation revolution [16]. Various methods of anomaly detection to detect attacks on PLCs have been proposed in the literature. They are usually classification-based and function by generating the whitelist for normal operations and then reporting unregistered operations as abnormal [17]. For example, Mochizuki et al. [18] summarized the normal operations for PLCs via the Petri net and applied a whitelist for anomaly detection. Similarly, Ghosh et al. [19] proposed a fault and behaviour monitoring tool for PLC, and Fujita et al. [17] developed a system toolset using the open-source software OpenPLC with extracted information from PLC programming. For sensors, existing anomaly detection methods typically adopt the same classification-type approach as far as we are aware. However, this is, in fact, undesirable because sensors with continuous values can face complicated attacks and therefore require more advanced regression methods for the detection. With the advancement of machine learning (ML), numerous data-driven models with both classification and regression capability [10,15,20–22] have now been established, and they can potentially be applied to improve the methods of anomaly detection involving continuous sensors going forward.

The attack scenarios on digital twins of cyber-physical infrastructure can be broadly classified into two categories: trivial and non-trivial [23]. Trivial attacks typically falsify discrete jumps to the reported values of the operational data in the digital twinning process with the intent to cripple the system instantly. These trivial attacks bear similarity to classifications, and thus, attacks on actuators can be classified into this category. Comparatively, non-trivial attacks are more sophisticated. They either introduce slow drifts to the system operation or falsify inconspicuous changes to the reported operational values within the specified threshold in the digital twinning process, and can therefore be considered advanced persistent threats to the system because the hackers are already able to compromise the system with some intelligence on how to avoid the alarms, for example making changes within the threshold levels [24–26]. Thereafter, hackers can build up a major disruption to the system through non-trivial attacks by modifying the system cumulatively over time,

remaining undetected until significant damages occur [27]. At present, existing anomaly detection algorithms are mostly not geared to identify these well-designed anomalies in the digital twin for real-time process control for critical cyber-physical infrastructures [24].

In this study, we develop a new combined anomaly detection framework (CADF) for digital twins of process control in water treatment facilities using the PLC-based whitelist synchrony system for actuators and the deep learning data-driven approach with probabilistic assessment for sensors. We note that the probabilistic anomaly detection approach has not been reported for digital twins of water treatment facilities as far as we are aware. A key reason for this is because the approach requires the quantification of both the transient changes of the system behavior as well as their uncertainties in real time. The level of technical details is thus substantially more demanding [28].

In the following, the details of CADF are first presented together with the assessment criteria for accuracy. Subsequently, the prototype facility of the Secure Water Treatment (SWaT) system is introduced. A total of 10 designed attacks in SWaT are then described, and the effectiveness of CADF for anomaly detection is summarized. Finally, a conclusion is drawn based on the results obtained.

2. Materials and Methods

2.1. Programmable Logic Controllers (PLC)-Based Whitelist System

In the IEC 61,131 standard, there are five main programming languages for PLCs: the two graphical programming languages of function block diagram (FBD) and ladder diagram (LD), the two textual programming languages of instruction list (IL) and structured text (ST), and the structuring tool and higher-level programming language sequential function chart (SFC) [29]. It should be noted that SFC is distinctly different from the other four languages because it is designed specifically for structurally complex applications. Transforming the PLC programs among the different languages or to other non-PLC programming languages is possible. For example, algorithms of transforming PLC programs from LD to ST were developed by Huang et al. [30]. Sadolewski [31] translated ST programs into ANSI C for verification purposes, and Darvas et al. [32] provided simple examples of the equivalency of PLC languages by generating temporal logic expressions from pattern-based requirements.

In our CADF for water treatment facilities, the PLC-based whitelist system applies the same strategy, but instead of using common methods such as Petri net [18] or introducing open-source software [17], we improve based on inspirations from the PLC transformations and summarize the logic expressions from the PLCs into conditions and if statements in Python. Our whitelist system consists of two kinds of logic expressions: sequence-based and value-based, as shown in Figure 2, which are complementary and not mutually exclusive. Sequence-based logic is derived from PLC routines that follow specific orders. For instance, in the ultrafiltration (UF) processes for water treatment, the UF drain valve must be open before the UF feed pump is activated, and then the on status should be maintained during the membrane filtration. Those orders are represented in the PLC state values (e.g., P3_STATE represents the PLC process sequence in PLC3 of the prototype testbed in Section 3.1 to be discussed later). Only when this state value changes are the corresponding actuators switched. Thus, we can detect the changes in the status of the actuators first and then judge whether they are due to normal sequence controls. On the other hand, the value-based logics rely on the reported values of sensors. A typical example in water treatment is to open the chemical dosing pump through PLC control to add sodium hypochlorite into the process water when the pH sensor sends a low value. In this case, the actuators should keep to their expected status with the specific conditions of pH sensor value; otherwise, anomalies can be considered to occur.

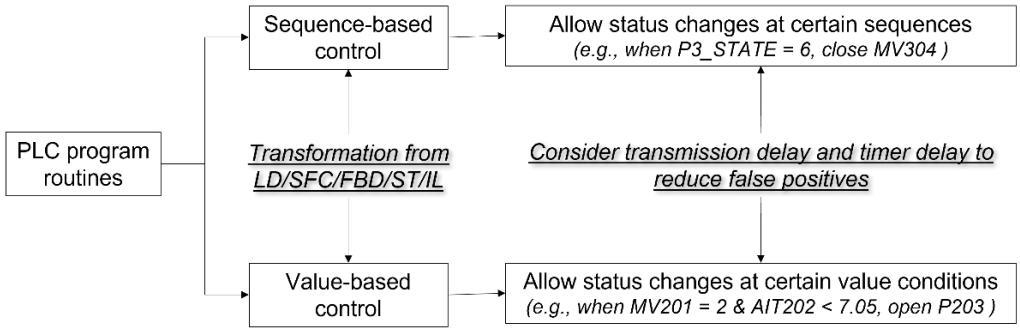

Figure 2. The two kinds of logic expressions from PLC.

In normal operations, time delays need to be considered in the whitelist to reduce the false positives for anomaly detection. The time delay can be either network-induced or due to designed timers as part of PLC instructions. Network-induced delays occur when the PLC transmits a large data packet [33]. The duration of the time delay due to the transmission varies with different water treatment facilities, although it is usually within a very short time. A suitable network-delay value can be quantified based on historical data and then included as additional tolerance when the actuators change their status. On the contrary, the designed timer delays can be longer depending on the judgement of the operators to buffer the input and output signals for safety reasons. Their magnitudes can usually be found directly from the PLC programs [34].

2.2. Probabilistic Machine Learning Model

Probabilistic assessment of the sensor data for real-time process evaluation is now common in many fields. For example, a probabilistic model has been established for hydrological drought prediction, which can integrate information from large datasets on persistent and prior meteorological droughts [35]. Recent probabilistic assessments of water systems include the multilevel probabilistic modelling of normal water usage established for the hourly data in northern Ethiopia's water network [36], as well as the resiliency modelling for flood control in Bangkok, Thailand [37]. We note that the input data to these models are typically uncertain, being influenced significantly by complex factors such as environmental noises and measurement durations.

In situations involving real-time control, such as digital twins of water treatment facilities, the computational requirement on hardware is much more demanding. Previously, Bayesian models are the common probabilistic approaches used in various fields [38]. However, they typically carry high computational costs and are difficult to update continuously during real-time system operations. The advancement of probabilistic machine learning algorithms in recent years overcomes this challenge. For example, a new long short-term memory (LSTM) neural network has been established for real-time soft sensoring of an underground drainage system in Copenhagen, Denmark with probabilistic assessment, and its field usage is now being tested in real applications [39]. In particular, dropout, which is a common tool to avoid the overfitting of deep neural networks, has been recently proven to approximate the Bayesian inference in data-driven modelling. The emerging idea of Monte Carlo (MC) dropout is being actively explored to be integrated with existing neural network structures for probabilistic assessment [40].

In the present study, we develop advanced probabilistic algorithms for real-time anomaly detection for water treatment facilities based on the deep learning approach of natural gradient boosting (NGBoost). NGBoost uses multi-parameter boosting and natural gradients to integrate the base learner, probability distribution and scoring rule into a modular algorithm [41], as shown in Equation (1).

$$\theta = \theta^{(0)} - \eta \sum_{m=1}^{M} \rho^{(m)} \cdot f^{(m)}(x) \quad (1)$$

In the equation, the parameter θ is optimized from its initial value $\theta^{(0)}$ by the M base learner outputs, which correspond to the different gradient boosting stages. Each base learner $f^{(m)}$ uses the inputs of x to calculate the best θ, where m represents the iteration. Then, the outputs are scaled with a stage-specific scaling factor $\rho^{(m)}$ and a common learning rate η. The learning algorithm starts with $\theta^{(0)}$ and continues to update θ by minimizing the scoring rule S over the response variables from all training examples. In each iteration m, the natural gradients of each sample x_i are represented as $g_i^{(m)}$ and obtained according to $I_S\left(\theta_i^{(m-1)}\right)$, which is the Riemannian metric of the statistical manifold at θ_i and $\nabla_\theta S\left(\theta_i^{(m-1)}, y_i\right)$. The gradient of the scoring rule S over the distribution and output y_i is shown in Equation (2) [41].

$$g_i^{(m)} = I_S\left(\theta_i^{(m-1)}\right)^{-1} \nabla_\theta S\left(\theta_i^{(m-1)}, y_i\right) \tag{2}$$

The base learners are fitted to predict the corresponding components of the natural gradients $g_i^{(m)}$ of each x_i as shown in Equation (3).

$$f^{(m)} \leftarrow fit\left(\left\{x_i, g_i^{(m)}\right\}_{i=1}^{n}\right) \tag{3}$$

Then, all parameters can be updated during the iterations. The scaling factor $\rho^{(m)}$ is obtained by minimizing the overall loss in the form of a line search far ρ as shown in Equation (4).

$$\rho^{(m)} \leftarrow argmin_\rho \sum_{i=1}^{n} S(\theta_i^{(m-1)} - \rho \cdot f^{(m)}(x_i), y_i) \tag{4}$$

More details that showed that NGBoost has similar or better performance than many existing methods for probabilistic regression can be found in [41]. In addition, different from other probabilistic methods which perform the uncertainty assessment by assuming homoscedasticity, NGBoost directly yields the parameters of the presumed distribution, which vary with the observed features. Since the base learner is a collection of weak learners using the boosting approach, the probability distribution is not restricted to the normal distribution.

Some recent studies have applied NGBoost as the probabilistic classifier, such as studies on brain tumors [42] and alcoholic EEG signals [43]. However, very few studies have utilized this approach for regression tasks so far, and none as a probabilistic forecasting tool. The present study aims to evaluate its suitability in real-time forecasting for water treatment facilities for the first time. However, we note that our NGBoost approach needs pre-processing to reframe the time series of measured data into pairs of input and output sequences before the training. The hyperparameters (including base learner, minibatch fraction, iterations as well as subsample fraction and the learning rate) will need to be tuned with grid search, as discussed in Section 3.3.

For the probability assessment, the uncertainty of machine learning model predictions consists of two categories: aleatoric uncertainty and epistemic uncertainty [44]. The aleatoric uncertainty captures the noise inherent in the input data and is a property that can be learned directly from the data by adopting the specific model and loss functions [45]. On the other hand, epistemic uncertainty refers to the uncertainty in the model predictions due to the lack of knowledge about the input features, i.e., insufficient training datasets. It is usually estimated by multiple sampling, which is essentially using multiple models and calculating the ensemble variance. In water treatment facilities, only infrequent historical data with limited sensors and actuators are typically kept, and the epistemic uncertainty can be high if the machine learning models are trained with limited datasets. Currently, CADF adopts the NGBoost method to calculate the aleatoric uncertainty only to meet the real-time process control requirement. Thus, additional consideration is needed for epistemic uncertainty to yield the total uncertainty.

2.3. Anomaly Detection Algorithms

A set of anomaly detection algorithms can be developed by considering the probabilistic assessment with the distributed forecasting outcomes. With water treatment facilities, we note that the posterior probability distribution can be very complicated due to the complexities of the physio-biochemical processes in the system. Here, we address the development of the anomaly detection algorithms based on normally distributed outputs as the first attempt.

Figure 3 shows the conceptual overview of CADF developed in this study. Generally, the procedures perform four types of real-time anomaly detection sequentially as follows:

1. Invalid inputs due to faulty devices or interrupted networks, which carry typical system signatures and can be typically identified without the need of model predictions;
2. Exceedance of confident intervals due to trivial (i.e., sudden) attacks;
3. Outliers distinguished from the normal value streams due to non-trivial (i.e., sophisticated) attacks;
4. Exceedance of whitelist due to trivial attacks or value drifts. We shall focus on the discussion of the second and third types in the following, which require data-driven modelling.

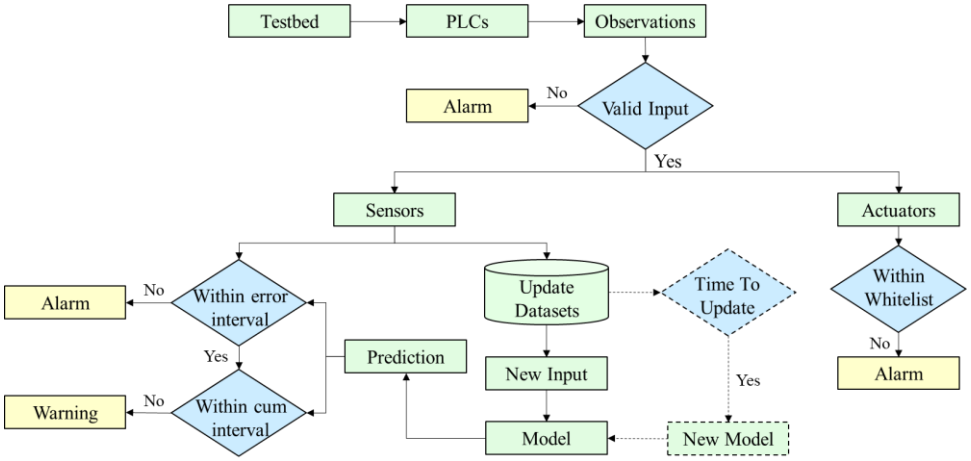

Figure 3. Conceptual overview of CADF. Note that model updating (dashed part) is currently not included in this study.

The third type uses the cumulative probability to detect sophisticated attacks in the form of falsified sensor values, whereby the algorithm cumulates the deviations between the sequential predictions and reported observations and raises warning or alarms based on the total difference as shown in Equation (5) and Figure 3.

$$\delta_c = \sum_{i=1}^{n} (o_i - \mu_i)/{\sigma_i}^2 \tag{5}$$

where δ_c is the cumulated error, o_i is the observation and μ_i and σ_i represent the mean and standard deviation of the output distribution of the sequential ith time sample ($i = 1, 2, \ldots n$) for the total n time samples considered.

With the simplified Gaussian distribution, the observations and predictions should have the same ensemble mean, and δ_c in Equation (5) should also be normally distributed, as shown in Equation (6):

$$\delta_c \sim N(0, \sum_{i=1}^{n} 1/{\sigma_i}^2) \tag{6}$$

However, when an anomaly occurs, the relationship for δ_c in Equation (6) will no longer be valid. Since the purpose of the sophisticated attack is to intentionally falsify the

reported observations in an inconspicuous manner, the cumulate deviation δ_c shall thus become increasingly biased during the attack and eventually exceed the threshold of $\varphi \cdot T_\delta$, where φ is a parameter of interval width and T_δ increases over time, as shown Equation (7):

$$T_\delta = \sqrt{\sum_{i=1}^{n} 1/{\sigma_i}^2} \tag{7}$$

The value of φ depends on the water treatment system and the sensing involved and thus needs to be established for the specific water treatment facility after extensive trial and error. However, we note that since the total uncertainty of the specific process is excluded, φ is likely to be more consistent as a non-dimensional parameter among the different facilities.

Anomaly detection in CADF is executed whenever the system receives newly reported observations through the digital twin. When an anomaly is identified, the corresponding alarm will be raised, and its format can be adjusted based on the requirements of the operator. Currently, the alarm message is designed to be 'detection time' + 'anomaly target' + 'sensor/actuator's tag'. Here, the detection time is the time that the CADF reports a positive, and the sensor/actuator's tag is the name of the attacked device in the system.

It is also important to point out that the above framework can be a flexible package, which can be integrated with different visualization platforms via suitable communication protocols. Furthermore, to advance the Industry 4.0 concept for the water industry, a standardized platform-independent communication architecture, such as the open platform communications unified architecture (OPC UA), can be highly beneficial to link up the numerous components of the water system into a comprehensive digital twin [46]. We note that for legacy systems, additional pre-processing and proper interfacing for the data acquisition may be needed.

2.4. Assessment of Accuracy

Different criteria are used to assess the performance of CADF. For PLCs, we compare the anomaly detection results between the support vector machine (SVM) and the PLC-based whitelist system using the two criteria of precision and recall [47], which are calculated based on the confusion matrix (shown in Table 1), which is a common tool to summarize the performance of two-class classification algorithms.

Table 1. Confusion matrix used in this study.

	Anomaly	**Normal**
Anomaly	True positive (TP)	False positive (FP)
Normal	False negative (FN)	True negative (TN)

The top row of Table 1 corresponds to samples predicted as an anomaly, and the second row contains the predicted normal events. The first column represents the truly anomalous samples, and the second column indicates other truly normal events. All predictions are classified into four categories: true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Precision is the ratio between TP and all the positives, while recall is the measure of correctly identifying TP, as follows:

$$\begin{cases} \text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \\ \text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \end{cases} \tag{8}$$

For sensors, we evaluate the accuracy of the probabilistic anomaly detection algorithms using three standard metrics, as follows: root-mean-square error (RMSE), negative log

likelihood (NLL) and continuous ranked probability score (CRPS). RMSE is the standard deviation of the prediction errors and indicates the spread for the residuals as follows:

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n}(\mu_i - o_i)^2/n} \tag{9}$$

NLL captures the fit between the predictions and observations as follows:

$$\text{NLL} = \sum_{i=1}^{n} -log[p(o_i|\mu_i)]/n \tag{10}$$

where $p(o_i|\mu_i)$ is the model's predicted probability density function evaluated with observation o_i. Smaller values of NLL indicate better model fit. We note that the log metric might have different expressions in the literature, but the basic idea is similar. CRPS can be expressed as follows:

$$\text{CRPS} = \sum_{i=1}^{n}(\int_{-\infty}^{\infty}(F(y) - \gamma)^2 dy)/n \tag{11}$$

where $F(y)$ is the predictive cumulative distribution function, and $\gamma = 1$ if $y \geq o_i$ and 0 otherwise. Here, CRPS requires a Gaussian distribution to be tractable analytically, while NLL does not. A prediction with low variance will receive a higher score in CRPS than NLL [48]. A perfect prediction with no variance yields a zero score for both.

Among the above three criteria, RMSE is typically used to assess the accuracy of point-based regression models, so this criterion might not represent well the distributed information from the model outcomes. Comparatively, NLL and CRPS are more commonly accepted as uncertainty metrics for probabilistic assessment [41,42,48].

3. Prototype Testing

3.1. Secure Water Treatment (SWaT) System

We perform the prototype testing of CADF in the SWaT system, which is a fully operational, scaled-down water treatment testbed facility in the Singapore University of Technology and Design [49]. The process flow diagram of the water treatment facility is shown in Figure 4.

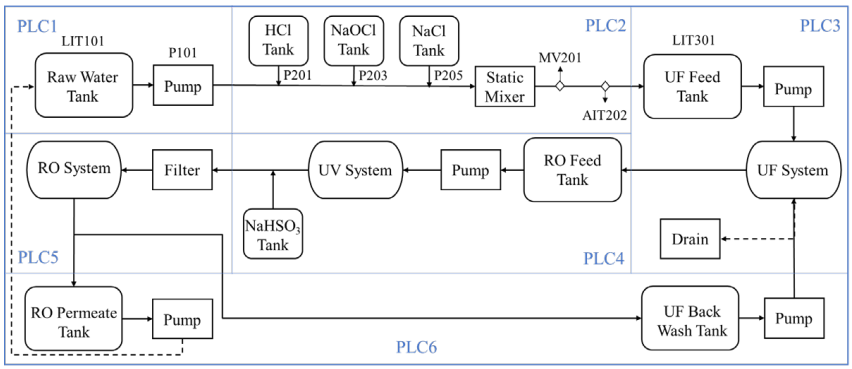

Figure 4. Diagram of SWaT testbed.

There are six treatment stages in SWaT, and each stage is controlled by an independent programmable logic controller (Allen-Bradley PLC) with control logic in RSLogix 5000. These PLCs are used with the EtherNet/IP and Common Industrial Protocol (CIP) stack for communication. The influent water is first conveyed from Stage 1 through the chemical dosing station to Stage 2 with the controlled additions of three types of pre-treatment chemicals, namely HCL, NOCL and NaCL. This is followed by the UF process in Stage 3. De-chlorination using ultraviolet (UV) lamps occurs in Stage 4, and then the treated water is fed to a reverse osmosis (RO) system in Stage 5 for the final processing. Stage

6 is a backwash process that cleans the UF membranes using the RO permeate. Here, it should be noted that Stage 6 is also connected to other units beyond the testbed facility; thus, all PLC6 components are not studied in this study. SWaT is designed to support the research on cyber-physical systems with 68 connected sensors and actuators in total to monitor the entire treatment process above. In SWaT, the sensor data can be continuous, while pumps only have 2 statuses (2 = open and 1 = closed), and valves have an additional status called travelling (=0) when the valves are changing their status. The physical and cyber setups of the facility are described in more detail in Mathur and Tippenhauer [50]. Additional information can also be found on the official SWaT website [51]. In this study, the transmission delay parameters in the PLC-based whitelist system are all set as 2 s according to the historical data, and the timer delay parameters vary with different actuators based on their PLC control logic.

3.2. Datasets and Attacks

In the present study, 10 offline datasets were used to test the performance CADF, as shown in Table 2. These datasets consisted of the data from normal operations as well as including data from designed attacks.

Table 2. General information of datasets used in this study.

	Content	Usage Purpose
Dataset 1	6 trivial attacks	Testing dataset
Dataset 2	7 trivial attacks	Testing dataset
Dataset 3	5 trivial attacks	Testing dataset
Dataset 4	No attack	Training & testing dataset
Dataset 5	2 sophisticated attacks	Testing dataset
Dataset 6	1 sophisticated attack	Testing dataset
Dataset 7	No attack	Training dataset
Dataset 8	1 sophisticated attack & 1 trivial attack	Training & testing dataset
Dataset 9	1 sophisticated attack	Testing dataset
Dataset 10	1 sophisticated attack	Training & testing dataset

Datasets 1 to 4 were used to test the detection accuracy of the PLC-based whitelist system. In datasets 1 to 3, trivial attacks were launched on different actuators, and their details can be found in Appendix A. In addition, the SVM analysis of MV201 (MV201 is the chemical dosing valve related to all 3 dosing tanks at stage 2) was used to benchmark the performance of the PLC-based whitelist system using Dataset 4 as the training dataset. MV201 was also attacked in Datasets 1 and 5. Datasets 5 and 6 were designed to test the intrinsic interactions between the actuators and sensors, so various sophisticated attacks were launched at different locations. Datasets 7 to 10 were used for the performance evaluation of the anomaly detection algorithms based on the probabilistic ML modelling of AIT202, which measured the pH value of the water after the chemical dosing from stage 1 to stage 2. Here, we note that the sensor values related to chemical concentrations in water treatment facilities are typically very difficult to predict or verify with scientific expressions. Thus, they are considered easy targets to attack from the cybersecurity point of view. In this case, the chemical dosing, including HCl, NaOCl and NaCl, can all affect the pH in a complex manner. Dataset 7 without attack was used to train the NGBoost model. However, it was found that dataset 7 alone was not sufficient to make accurate predictions during the testing phase, so part of the normal operation data from datasets 8 and 10 were also used as supplementary. The sensor values of AIT202 were falsified during the attack until the sensor status was reset to normal afterward.

The details of the sophisticated attacks included in datasets 5 to 10 are listed in Table 3, including the attack durations and descriptions with their locations marked in Figure 4. All attacks were launched by running corresponding Python scripts with the assistance of Pylogix API.

Table 3. Details of sophisticated attacks.

	Time	Descriptions
Dataset 5	Switch MV201 and P101 from close to open	
Dataset 5	From 12,582 s to 15,734 s	Decrease LIT301 0.05 every 1 s
Dataset 6	From 12,087 s to 12,357 s	Keep AIT202 around 7.06
Dataset 8	From 175 s to 455 s, normal at 482 s	Increase AIT202 0.05 every 10 s
Dataset 9	From 208 s to 349 s, normal at 482 s	Increase AIT202 0.005 every 1 s
Dataset 10	From 63 s to 1665 s, normal at 1682 s	Decrease AIT202 0.001 every 1 s

It is important to clarify that the attacks on sensors are only to falsify the values received by the PLCs, while the attacks on actuators can turn the physical devices on and off. The true sensor values cannot be changed directly, but they can be affected indirectly by the subsequent action of the PLCs of the related components. For example, P203 is the pump to control the HCl dosing, which influences the value of AIT202, and it will only be switched to its open status when the AIT202 value exceeds 7.05. Thus, if the attack aims to decrease the pH of water in SWaT, it can falsify an increase in the AIT202 value until the condition to open P203 is reached. Similarly, the water depth in the tank is controlled by PLCs within the safe operational limits to avoid the occurrence of damaging conditions such as physical overflows. A well-designed attack can falsify the tank water level to an arbitrary low value such that the related pumps are activated continuously, leading to the tank overflow. These sophisticated attacks require some intelligence of the system control so that they can hide within the trivial thresholds to avoid triggering the PLC alarms directly. Anomaly detection for these attacks is thus more challenging.

3.3. Hyperparameter Selection

This section presents the details in the tuning of the hyperparameters for the probabilistic data-driven models in CADF. For the probabilistic ML model, the selection of the main network hyperparameters was based on a grid search [52] with a three-fold cross validation. Table 4 shows the details of the grid search for the hyperparameters of the NGBoost model, with the best results displayed separately in the last column. The hyperparameters in the tuning range were first chosen around the default values and then optimized after trials to reduce the high computational cost.

Table 4. Hyperparameter tuning information of the NGBoost algorithm.

Hyperparameter	Tuning Range	Best Result
Base learner	DecisionTreeRegressor with max depth (2, 3, 4)	4
Minibatch fraction	(1, 0.5)	1
Iterations	(200, 300, 500)	500
Subsample fraction	(0.5, 0.8, 1.0)	1
Learning rate	(0.01, 0.05, 0.1)	0.1

3.4. Computational Performance of Probabilistic Algorithms

The basic requirements of real-time anomaly detection include high prediction accuracy and fast computational speed. In the present study, NGBoost with a 5-time-step is found to yield the best prediction performance among the different training datasets, as shown in Table 5. This can be attributed to the fact that fewer time steps than five would lead to insufficient information, while longer time steps introduce more noise to the training. From the table, the NLL with a 10-time-step is -0.16, which is much worse than a value of -2.89 for the 5-time-step and -2.01 for the 3-time-step.

Table 5. Probabilistic prediction performance of NGBoost on dataset 1.

	RMSE ($\times 10^{-2}$)	CRPS ($\times 10^{-2}$)	NLL
T10	0.91 ± 0.07	0.41 ± 0.01	-0.16 ± 0.09
T5	0.72 ± 0.07	0.33 ± 0.01	-2.89 ± 0.16
T3	0.76 ± 0.01	0.39 ± 0.03	-2.01 ± 0.21

Notation: T10 represents the model with the 10-time-step dataset, similarly for T5 and T3.

Figure 5 shows the sample probabilistic predictions with NGBoost on the pH sensor AIT202. In general, NGBoost always produces a narrow uncertainty band with two standard deviations for the operational dataset from this facility. As discussed above, the uncertainty from NGBoost primarily contains the aleatoric uncertainty of the data noises, and these noises are very small in the SWaT testbed due to its well-controlled environment. We note, however, that the actual noise in prototype water treatment facilities can be very large. For example, Cecconi and Rosso [53] analyzed the historical data of ammonium sensors in a real treatment facility. They found that the uncertainty is large and includes treatment process anomalies, calibration bias faults and fouling drifts.

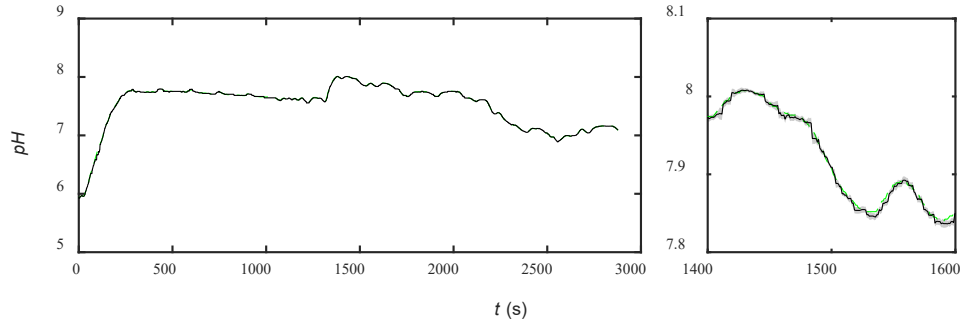

Figure 5. Sample results of NGBoost for the pH sensor AIT202 (bands delimited by 2σ).

The training and testing speeds of the probabilistic data-driven models are crucial in online field applications because the model predictions must be provided ahead of time for the process control. In this study, the NGBoost algorithm in CADF is built on Scikit-learn in Python, which cannot be sped up by a GPU. Hence, the algorithm is executed in a fast workstation with a $\times 64$ Platinum 8276 CPU @ 2.20 GHz, which requires ~66 s to train and ~0.4 s to test with a dataset of 2878 points. We find that the computation speed with NGBoost would be sufficiently fast and viable for the digital twin of the process control in SWaT in real time. The related parameters α and φ together decide whether the thresholds are deemed to be exceeded, and their values need to be fine-tuned for the specific water treatment facility (i.e., the SWaT system) to compensate for the epistemic uncertainty. In the computational speed evaluation on AIT202, the values of $\alpha = 6$ and $\varphi = 6$ are chosen for the best outcome for the subsequent verification based on the historical data of normal operations. We note that the thresholds are set to be wide intentionally to minimize the false positives. Another reason for the wide threshold is that sophisticated attacks typically do not usually trigger significant damage at the beginning.

4. Discussion

4.1. PLC-Based Whitelist System Performance

As discussed in Section 3.2, an SVM is used to benchmark the performance of the PLC-based whitelist system based on MV201 of dataset 1. SVM uses the values of the last timestep to predict the next timestep with dataset 4 as the training dataset. Comparatively, the conditions of MV201 in the PLC-based whitelist are all valued-based logics from the PLC control logic. MV201 was switched from closed to open for a duration of 70 s in dataset 1, including 8 s of travelling (MV201 = 0) and 62 s of being open (MV201 = 2). The detection results are shown in Table 6.

Table 6. Results from SVM and the PLC-based whitelist system of MV201 in dataset 1.

	TP	FP	FN	TN
SVM	70	94	0	14,235
Whitelist	58	0	12	14,341

SVM yields 94 false positives (FPs), which implies that the number of false alarms is even more than true attacks. Although its recall of 1.00 is good, the low precision of 0.43 is not acceptable because the operator may choose to ignore the true alarms when the number of FPs is excessive. In contrast, our PLC-based whitelist has a precision of 1.00 and a recall of 0.83, with no FP but with some false negatives (FNs) instead. All FNs occur at the beginning of the attack and are primarily due to the consideration of both the transmission delay and timer delay, as discussed in Section 2.1. Here, the number of FNs will not increase for longer period attacks; hence, the recall can still increase for more dangerous scenarios. Since only long-time attacks on actuators can induce true damage to the physical entities, we believe that the small percentage of FNs is acceptable. Although one can argue that the performance of SVM can be improved with more training datasets and fine-tuning of parameters, using the PLC-based whitelist system in CADF is definitely an easier and more efficient way to detect malicious anomalies.

Table 7 presents the anomaly detection results of datasets 1 to 4. It can be seen from the table that all trivial attacks are detected by CADF, and there is no false alarm. However, the number of detected anomalies is larger than the number for the launched attacks for datasets 1 and 2 because sometimes the physical values of the actuators may drift randomly. Here, we can confirm these anomalies are indeed random drifts since the normal value streams never show sudden spikes in the real operations. It should also be noted that although both precision and recall are used to compare the performance of SVM and our PLC-based whitelist system above, these two metrics are not the perfect fit for our detection results. For value-based logic, all conditions that exceed the normal operations are detected, while sequence-based logic only captures the anomalies when the actuators change their status. In other words, value-based logic yields continuous alarms when the anomalies are detected; however, sequence-based logic only sends alarms at the beginning and end of the anomalies.

Table 7. Results of PLC-based whitelist system of trivial attacks.

	Launched Attacks	Detected Anomalies
Dataset 1	6	7
Dataset 2	7	8
Dataset 3	5	5
Dataset 4	0	0

4.2. Combined Attacks

The combined attack refers to the modification of the sensor values in order to activate a change in the PLC conditions to damage the physical facility. In Section 4.1, we have demonstrated the ability of the PLC-based whitelist system in CADF to detect trivial attacks, which include direct attacks on actuators. However, since the PLC-based whitelist system can only handle issues with the actuators, we will need to rely on the probabilistic ML model in CADF for anomaly detection. In this section, the effectiveness of CADF is examined for these combined attacks.

In Table 3, dataset 5 includes one multi-point attack and one sophisticated attack. The multi-point attack targets two actuators (MV201 and P101) simultaneously, and it can be easily detected by the PLC-based whitelist system because it is equivalent to the combination of two trivial attacks. Figure 6 shows the alarm information of this attack. Both P201 and P203 are also reported by the anomaly detection algorithms because their value-

based logics are related to MV201. Thus, although our system can localize the anomaly, further judgements will be needed to identify the root cause among all the alarms.

```
13:23:31 anomaly target P101.Status
13:23:31 anomaly target P201.Status
13:23:31 anomaly target P203.Status
13:23:31 anomaly target MV201.Status
13:23:32 anomaly target P101.Status
13:23:32 anomaly target P201.Status
13:23:32 anomaly target P203.Status
13:23:32 anomaly target MV201.Status
```

Figure 6. Alarms information of the multi-point attack.

The sophisticated attack in dataset 5 did not trigger any direct alarms in the PLC-based whitelist system. This is due to the fact that the attack targeted the sensor LIT301 for the water level of the tank. MV201 has value-based logic with LIT301 in the following manner: (i) when LIT301 < 800, open MV201; (ii) when LIT301 > 1000, close MV201. From the left-hand side of Figure 7, it can be observed that LIT301 decreases at a slower speed under this attack. Thus, MV201 maintains its closed status for a longer period of time. The dashed line on the right-hand side of Figure 7 shows the real sensor value of LIT301 based on the declining trend, and the lowest water level is around 420 mm. Additionally, the true activation time of MV201 without the attack is also observed in the figure, which is much earlier than the attacked scenario. Therefore, without the inflow from Stage 2, it is possible to completely drain the tank and disrupt the treatment in Stage 3 by keeping LIT301 > 800 for a longer time.

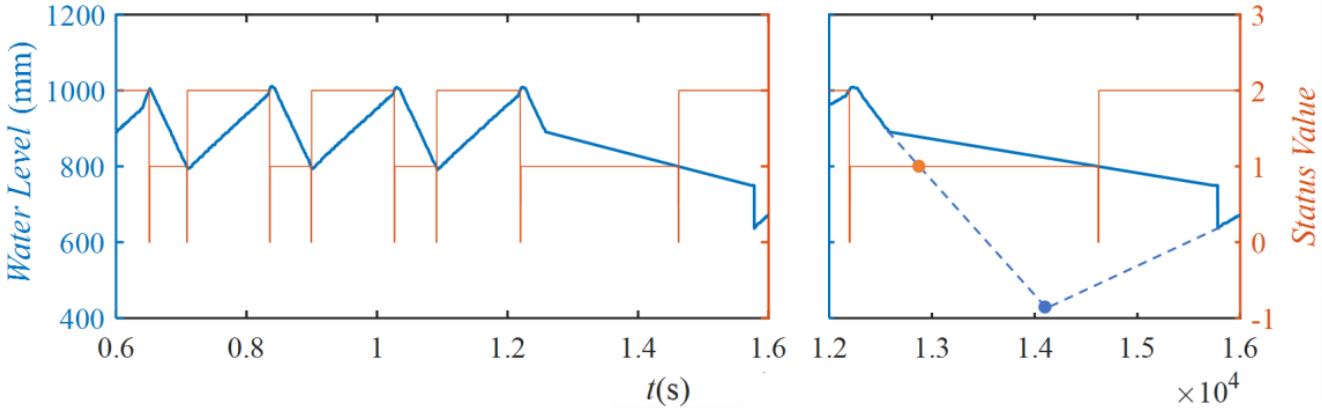

Figure 7. Sophisticated attack in dataset 5. Blue line = value of LIT301 and orange line = status value of MV201.

The sophisticated attack in dataset 6 has a similar outcome. It attacks AIT202 with consideration of the following logic: (i) when MV201 = 2 and AIT202 < 7.05, open P203, and (ii) when MV201 <> 2 or AIT202 < 6.95, close P203. The attack was started when MV201 was already open, and AIT202 was kept around 7.06. This led P203 to be open for a longer time and allowed the pH to reach a very low value, as shown in Figure 8. For water treatment facilities, which include bioreactors, the lowering of pH with this sophisticated attack can further lead to the destruction of bacteria that are sensitive to pH for the treatment processes, causing a major disruption that takes a long time to recover from.

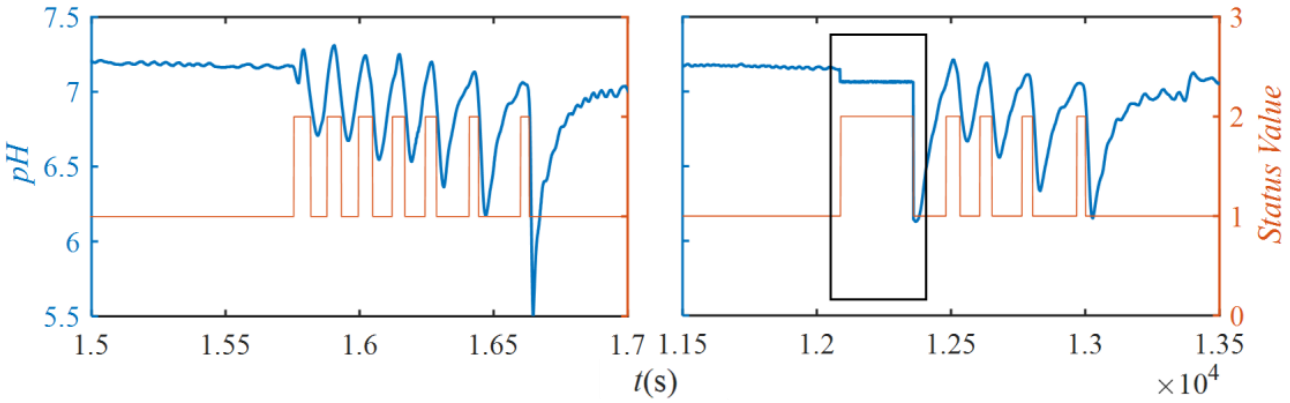

Figure 8. Sophisticated attack in dataset 6 (left: normal operation and right: attack period in boxed portion). Blue line = value of AIT202 and orange line = status value of P203.

Figure 9 further illustrates the anomaly detection performance with probabilistic assessment alone in Datasets 8, 9 and 10. The attacks in these three datasets are purely to falsify the sensor values but not to the extent of modifying the operational logic of the PLCs, and thus, the whitelist system is not needed. In Figure 9A1, the sensor value increases jaggedly under the sophisticated attack, which falsifies an increment of only 0.05 every 10 s. The first trivial anomaly detection is confirmed promptly when the sophisticated attack begins at 175 s. We stress that this does not imply that a trivial anomaly detection always precedes non-trivial anomaly detection for two reasons: (i) the aleatoric uncertainty can be much higher with the inherent data noises, such that the increment would be deemed to be within the normal fluctuation range; and (ii) the attacker can implement a more intelligent attack with exponential growth in the increment starting within the threshold. Nevertheless, the confirmation of the trivial anomaly detection in this designed sophisticated attack demonstrates that the algorithm is functioning as intended. More importantly, Figure 9A3 shows the accumulated error and corresponding interval, and the algorithm successfully detects the ongoing sophisticated attack at 308 s. Subsequently, 3 other anomalies are confirmed in Figure 9A2 by the algorithm, including at 482 s when the sensor data returns to the true value after the sophisticated attack, and at 642 and 663 s due to the start and end of another trivial attack, respectively.

For the sophisticated attacks in datasets 9 and 10 shown in Figure 9B2,C2, which start at 208 s and 63 s, respectively, a trivial anomaly is not triggered at the beginning due to the small increment within the threshold, as discussed above. Nonetheless, our anomaly detection algorithms successfully confirm the ongoing non-trivial attacks at 315 s in Figure 9B3 and at 1088 s in Figure 9C3. We emphasize that most existing anomaly detection algorithms cannot detect these sophisticated attacks well because their intelligence is primarily based on the exceedance of a specific/constant threshold based on the instantaneous difference between the predictions and observations. Trivial attacks are also confirmed at 482 s and 1682 s in the 2 figures, respectively, due to the end of the non-trivial attacks as discussed above.

Finally, although CADF is shown to be very effective in this study, its success cannot be fully guaranteed for all non-trivial attacks. CADF, therefore, needs to be further studied in real facilities. For example, if the attack in Figure 9C1 stops before 1088 s, our anomaly detection algorithms would not be able to identify this non-trivial attack. However, an inconspicuous attack with a short duration cannot usually incur damages in a water treatment facility. With the main objective to protect the testbed from physical damage, these 'mistakes' might be deemed to be tolerable. Otherwise, with small parameters in anomaly detection algorithms and strict thresholds, too many false alarms might be set off unnecessarily.

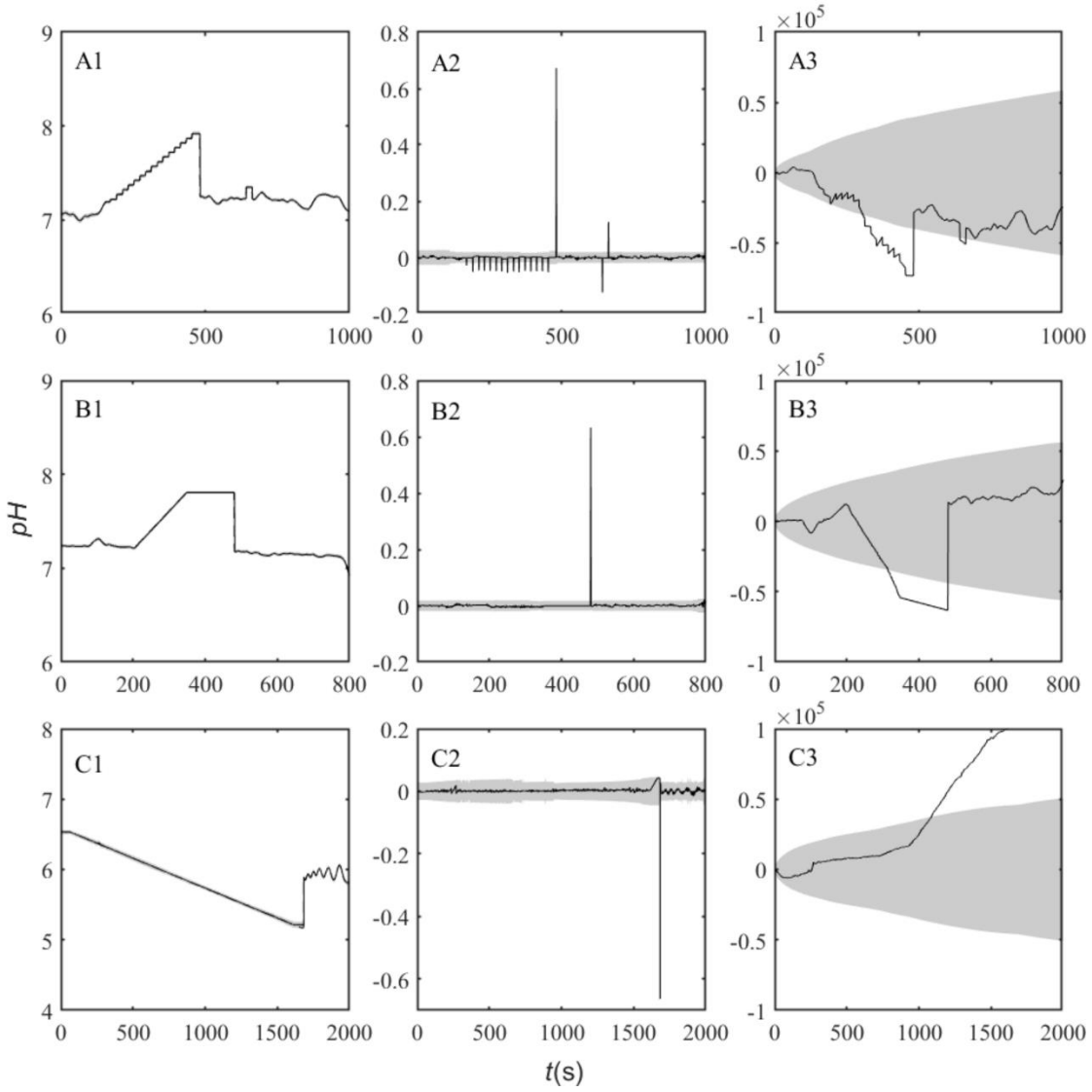

Figure 9. Anomaly detection results of AIT202 (α = 6, φ = 6). Title represents: (**A1**–**A3**) = dataset 2, (**B1**–**B3**) = dataset 3, (**C1**–**C3**) = dataset 4, 1 = observations and predictions, 2 = errors between predictions and observations and 3 = accumulated errors. Dashed-dotted line = observations, solid line = predictions.

5. Conclusions

A combined framework of anomaly detection called CADF is developed in this study for the digital twin of process control in water treatment facilities. CADF performs anomaly detection by utilizing the PLC-based whitelist system to generate normal operation logics from PLC programs and adopting the probabilistic ML model of NGBoost to obtain the uncertainty assessment. The implementation of CADF is comprehensively evaluated in the testbed facility of a SWaT based on the operational data with designed attacks. Overall, CADF is shown to be effective against both trivial and non-trivial attacks in the evaluation. The framework can be further extended to other cyber-physical systems that rely on PLCs in the future.

Author Contributions: Conceptualization, Y.W. and A.W.-K.L.; methodology and analysis, Y.W.; data curation, Y.W. and D.T.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W. and A.W.-K.L.; supervision, A.W.-K.L. and C.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Research Foundation (NRF), Prime Minister's Office, Singapore, under its National Cybersecurity R&D Programme and administered by the National Satellite of Excellence in Design Science and Technology for Secure Critical Infrastructure, Award No. NSoE_DeST-SCI2019-0011.

Acknowledgments: The authors would wish to thank all members in our project team for their valuable contributions during discussions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Details of Trivial Attacks in Datasets 1 to 3

	Time	Descriptions
Dataset 1	From 2928 s to 2998 s	Open MV101
Dataset 1	From 4871 s to 4940 s	Open MV201
Dataset 1	From 6763 s to 6832 s	Open MV301
Dataset 1	From 8554 s to 8623 s	Open MV302
Dataset 1	From 10,388 s to 10,448 s	Open MV303
Dataset 1	From 12,339 s to 12,482 s	Open MV304
Dataset 2	From 1645 s to 1674 s	Open P101
Dataset 2	From 3515 s to 3594 s	Close P301
Dataset 2	From 5390 s to 5450 s	Close P401
Dataset 2	From 7214 s to 7373 s	Close P403
Dataset 2	From 9125 s to 9189 s	Close P501
Dataset 2	From 11,055 s to 11,125 s	Close MV501
Dataset 2	From 12,807 s to 12,878 s	Close MV502
Dataset 3	From 983 s to 1054 s	Open MV504
Dataset 3	From 2746 s to 2807 s	Close UV401
Dataset 3	From 4790 s to 4851 s	Close P201
Dataset 3	From 6826 s to 6887 s	Open P203
Dataset 3	From 8568 s to 8629 s	Open P205

References

1. Knapp, G.L.; Mukherjee, T.; Zuback, J.S.; Wei, H.L.; Palmer, T.A.; De, A.; DebRoy, T. Building blocks for a digital twin of additive manufacturing. *Acta Mater.* **2017**, *135*, 390–399. [CrossRef]
2. Liu, Q.; Zhang, H.; Leng, J.; Chen, X. Digital twin-driven rapid individualised designing of automated flow-shop manufacturing system. *Int. J. Prod. Res.* **2019**, *57*, 3903–3919. [CrossRef]
3. Luo, W.; Hu, T.; Zhang, C.; Wei, Y. Digital twin for CNC machine tool: Modeling and using strategy. *J. Ambient. Intell. Humaniz. Comput.* **2019**, *10*, 1129–1140. [CrossRef]
4. Li, C.; Mahadevan, S.; Ling, Y.; Choze, S.; Wang, L. Dynamic Bayesian network for aircraft wing health monitoring digital twin. *Aiaa J.* **2017**, *55*, 930–941. [CrossRef]
5. Xie, J.; Wang, X.; Yang, Z.; Hao, S. Virtual monitoring method for hydraulic supports based on digital twin theory. *Min. Technol.* **2019**, *128*, 77–87. [CrossRef]
6. Salvi, A.; Spagnoletti, P.; Noori, N.S. Cyber-resilience of Critical Cyber Infrastructures: Integrating digital twins in the electric power ecosystem. *Comput. Secur.* **2022**, *112*, 102507. [CrossRef]
7. Francisco, A.; Mohammadi, N.; Taylor, J.E. Smart city digital twin–enabled energy management: Toward real-time urban building energy benchmarking. *J. Manag. Eng.* **2020**, *36*, 04019045. [CrossRef]
8. Kritzinger, W.; Karner, M.; Traar, G.; Henjes, J.; Sihn, W. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine* **2018**, *51*, 1016–1022. [CrossRef]
9. Conejos Fuertes, P.; Martínez Alzamora, F.; Hervás Carot, M.; Alonso Campos, J. Building and exploiting a Digital Twin for the management of drinking water distribution networks. *Urban Water J.* **2020**, *17*, 704–713. [CrossRef]
10. Junejo, K.N.; Goh, J. Behaviour-Based Attack Detection and Classification in Cyber Physical Systems Using Machine Learning. In Proceedings of the 2nd ACM International Workshop on Cyber-Physical System Security—CPSS '16, Xi'an, China, 30 May 2016; pp. 34–43.
11. Dunlap, S.; Butts, J.; Lopez, J.; Rice, M.; Mullins, B. Using timing-based side channels for anomaly detection in industrial control systems. *Int. J. Crit. Infrastruct. Prot.* **2016**, *15*, 12–26. [CrossRef]
12. Alcaraz, C.; Zeadally, S. Critical infrastructure protection: Requirements and challenges for the 21st century. *Int. J. Crit. Infrastruct. Prot.* **2015**, *8*, 53–66. [CrossRef]

13. Russon, M.A. US Fuel Pipeline Hackers 'Didn't Mean to Create Problems'. Available online: https://www.bbc.com/news/business-57050690 (accessed on 18 March 2022).
14. Adepu, S.; Mathur, A. An Investigation into the Response of a Water Treatment System to Cyber Attacks. In Proceedings of the 2016 IEEE 17th International Symposium on High Assurance Systems Engineering (HASE), Orlando, FL, USA, 7–9 January 2016; pp. 141–148.
15. Lin, C.T.; Wu, S.L.; Lee, M.L. Cyber attack and defense on industry control systems. In Proceedings of the 2017 IEEE Conference on Dependable and Secure Computing, Taipei, Taiwan, 7–10 August 2017; pp. 524–526.
16. Mellado, J.; Núñez, F. Design of an IoT-PLC: A containerized programmable logical controller for the industry 4.0. *J. Ind. Inf. Integr.* **2022**, *25*, 100250. [CrossRef]
17. Fujita, S.; Hata, K.; Mochizuki, A.; Sawada, K.; Shin, S.; Hosokawa, S. OpenPLC based control system testbed for PLC whitelisting system. *Artif. Life Robot.* **2021**, *26*, 149–154. [CrossRef]
18. Mochizuki, A.; Sawada, K.; Shin, S.; Hosokawa, S. On experimental verification of model based white list for PLC anomaly detection. In Proceedings of the 2017 11th Asian Control Conference (ASCC), Gold Coast, Australia, 17–20 December 2017; pp. 1766–1771.
19. Ghosh, A.; Qin, S.; Lee, J.; Wang, G.-N. FBMTP: An automated fault and behavioral anomaly detection and isolation tool for PLC-controlled manufacturing systems. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *47*, 3397–3417. [CrossRef]
20. Nicolaou, N.; Eliades, D.G.; Panayiotou, C.; Polycarpou, M.M. Reducing Vulnerability to Cyber-Physical Attacks in Water Distribution Networks. In Proceedings of the 2018 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater), Porto, Portugal, 10–13 April 2018; pp. 16–19.
21. Clotet, X.; Moyano, J.; León, G. A real-time anomaly-based IDS for cyber-attack detection at the industrial process level of Critical Infrastructures. *Int. J. Crit. Infrastruct. Prot.* **2018**, *23*, 11–20. [CrossRef]
22. Evangelou, M.; Adams, N.M. An anomaly detection framework for cyber-security data. *Comput. Secur.* **2020**, *97*, 101941. [CrossRef]
23. Dereszynski, E.W.; Dietterich, T.G. Probabilistic models for anomaly detection in remote sensor data streams. *arXiv* **2012**, arXiv:1206.5250.
24. Ghafir, I.; Hammoudeh, M.; Prenosil, V.; Han, L.; Hegarty, R.; Rabie, K.; Aparicio Navarro, F.J. Detection of advanced persistent threat using machine-learning correlation analysis. *Future Gener. Comput. Syst.* **2018**, *89*, 349–359. [CrossRef]
25. Milajerdi, S.M.; Gjomemo, R.; Eshete, B.; Sekar, R.; Venkatakrishnan, V. Holmes: Real-time apt detection through correlation of suspicious information flows. In Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 20–22 May 2019; pp. 1137–1152.
26. Stojanović, B.; Hofer Schmitz, K.; Kleb, U. APT datasets and attack modeling for automated detection methods: A review. *Comput. Secur.* **2020**, *92*, 101734. [CrossRef]
27. Toliupa, S.; Nakonechnyi, V.; Tereikovskyi, I.; Tereikovska, L.; Korystin, O. One-periodic template marks model of normal behavior of the safety parameters of information systems networking resources. In Proceedings of the 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), Kyiv, Ukraine, 8–11 November 2019; pp. 764–768.
28. Donkor, E.A.; Mazzuchi, T.A.; Soyer, R.; Alan Roberson, J. Urban water demand forecasting: Review of methods and models. *J. Water Resour. Plan. Manag.* **2014**, *140*, 146–159. [CrossRef]
29. Ayub, A.; Yoo, H.; Ahmed, I. Empirical study of PLC authentication protocols in industrial control systems. In Proceedings of the 2021 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 27 May 2021; pp. 383–397.
30. Huang, L.; Liu, W.; Liu, Z. Algorithm of transformation from PLC ladder diagram to structured text. In Proceedings of the 2009 9th International Conference on Electronic Measurement & Instruments, Beijing, China, 16–19 August 2009; p. 4-778-774-782.
31. Sadolewski, J. Conversion of ST control programs to ANSI C for verification purposes. *Inform. Softw. Eng. J.* **2011**, *5*, 65–76. [CrossRef]
32. Darvas, D.; Majzik, I.; Blanco Viñuela, E. Generic representation of PLC programming languages for formal verification. In Proceedings of the 23rd PhD Mini-Symposium, Budapest, Hungary, 8–9 February 2016; pp. 6–9.
33. Ju, C.; Yang, G.; Chen, Y.W.; Pan, C. Dynamic optimization of data packet-based communication for PLC visual monitoring. *Appl. Sci.* **2019**, *9*, 1721. [CrossRef]
34. Song, J.; Jee, E.; Bae, D.-H. Automated test sequence generation for function block diagram programs. In Proceedings of the 2016 23rd Asia-Pacific Software Engineering Conference (APSEC), Hamilton, New Zealand, 6–9 December 2016; pp. 305–312.
35. Hao, Z.; Hao, F.; Singh, V.P.; Sun, A.Y.; Xia, Y. Probabilistic prediction of hydrologic drought using a conditional probability approach based on the meta-Gaussian model. *J. Hydrol.* **2016**, *542*, 772–780. [CrossRef]
36. Tashman, Z.; Gorder, C.; Parthasarathy, S.; Nasr Azadani, M.M.; Webre, R. Anomaly Detection System for Water Networks in Northern Ethiopia Using Bayesian Inference. *Sustainability* **2020**, *12*, 2897. [CrossRef]
37. Law, A.W.K.; Zhu, F.; Yang, P.; Ho, H.L.; Sim, V.S.T.; Wu, X.; Lian, Y.; Loh, J.; Chan, H.; Chitwatkulsiri, D.; et al. *Development of 3D Visualization Platform for Compound Flooding and Transport Resiliency in Coastal Cities*; Singapore International Water Week (SIWW): Singapore, 2021.
38. Chen, S.H.; Pollino, C.A. Good practice in Bayesian network modelling. *Environ. Model. Softw.* **2012**, *37*, 134–145. [CrossRef]

39. Palmitessa, R.; Mikkelsen, P.S.; Borup, M.; Law, A.W.K. Soft sensing of water depth in combined sewers using LSTM neural networks with missing observations. *J. Hydro-Environ. Res.* **2021**, *38*, 106–116. [CrossRef]
40. Gal, Y.; Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1050–1059.
41. Duan, T.; Anand, A.; Ding, D.Y.; Thai, K.K.; Basu, S.; Ng, A.; Schuler, A. Ngboost: Natural gradient boosting for probabilistic prediction. In Proceedings of the International Conference on Machine Learning, Online. 13–18 July 2020; pp. 2690–2700.
42. Dutta, S.; Bandyopadhyay, S. Revealing Brain Tumor Using Cross-Validated NGBoost Classifier: NG Boost Classifier. *Int. J. Mach. Learn. Netw. Collab. Eng.* **2020**, *4*, 12–20.
43. Barus, D.T.; Masri, F.; Rizal, A. NGBoost Interpretation Using LIME for Alcoholic EEG Signal Based on GLDM Feature Extraction. In Proceedings of the Computational Methods in Systems and Software, Vsetin, Czechia, 14–17 October 2020; pp. 894–904.
44. Malinin, A.; Prokhorenkova, L.; Ustimenko, A. Uncertainty in gradient boosting via ensembles. *arXiv* **2020**, arXiv:2006.10562.
45. Scalia, G.; Grambow, C.A.; Pernici, B.; Li, Y.P.; Green, W.H. Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. *J. Chem. Inf. Modeling* **2020**, *60*, 2697–2717. [CrossRef]
46. Nicolae, A.; Korodi, A.; Silea, I. Complete Automation of an Energy Consumption Reduction Strategy from a Water Treatment and Distribution Facility, Inside an Industrial Internet of Things-Compliant Proactive Historian Application. *Sensors* **2021**, *21*, 2569. [CrossRef]
47. Buckland, M.; Gey, F. The relationship between recall and precision. *J. Am. Soc. Inf. Sci.* **1994**, *45*, 12–19. [CrossRef]
48. Teye, M.; Azizpour, H.; Smith, K. Bayesian uncertainty estimation for batch normalized deep networks. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 4907–4916.
49. Raman, M.G.; Dong, W.; Mathur, A. Deep autoencoders as anomaly detectors: Method and case study in a distributed water treatment plant. *Comput. Secur.* **2020**, *99*, 102055. [CrossRef]
50. Mathur, A.P.; Tippenhauer, N.O. SWaT: A water treatment testbed for research and training on ICS security. In Proceedings of the 2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater), Vienna, Austria, 11 April 2016; pp. 31–36.
51. iTrust. Secure Water Treatment—iTrust. Available online: https://itrust.sutd.edu.sg/testbeds/secure-water-treatment-swat/ (accessed on 18 March 2022).
52. Liashchynskyi, P.; Liashchynskyi, P. Grid search, random search, genetic algorithm: A big comparison for nas. *arXiv* **2019**, arXiv:1912.06059.
53. Cecconi, F.; Rosso, D. Soft Sensing for On-Line Fault Detection of Ammonium Sensors in Water Resource Recovery Facilities. *Environ. Sci. Technol.* **2021**, *55*, 10067–10076. [CrossRef]