Realtime Management of Wastewater Treatment Plants Using AI

Virginia Tech & DC Water

Feras A. Batarseh, PhD, JM, MSc
Mehmet Oguz Yardimci, MSc
Ryu Suzuki, MSc
Md Nazmul Kabir Sikder, MSc
Zhiwu Wang, PhD
Wan-Yi Mao

Table of Contents

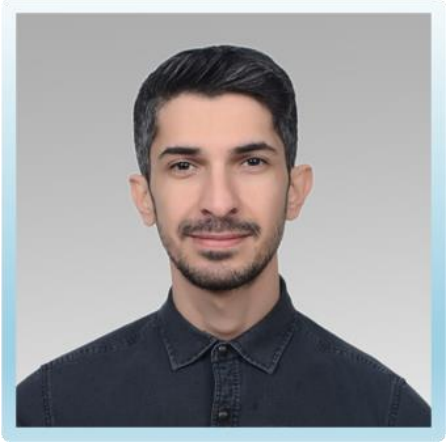
Our Team

Feras A. Batarseh – Team Leader

batarseh@vt.edu

Dr. Batarseh is an associate professor with the Department of Biological Systems Engineering at Virginia Tech. He is affiliated with the Center for Advanced Innovation in Agriculture (CAIA) at Virginia Tech and George Mason University's (GMU) School of Systems Biology. He obtained his Ph.D. and M.Sc. in computer engineering from the University of Central Florida in 2007 and 2011, a graduate certificate in project leadership from Cornell University in 2016, and a Juris Masters (JM) of Law from GMU in 2022.

Areas of expertise: AI for Agriculture, AI Assurance, Intelligent Water Systems, and Cyberbiosecurity



Mehmet Oguz Yardimci

oguzy@vt.edu

Yardimci is a Ph.D. student in Computer Science at Virginia Tech. He is a graduate research assistant affiliated with the Hume Center for National Security and Technology. He works in the AI Assurance and Applications Lab (A3) on AI Assurance for Cyberbiosecurity and Water Systems.

Areas of expertise: AI Assurance, Cyberbiosecurity, Intelligent Water Systems, Machine Learning, and Meta-Learning.

Ryu Suzuki

Ryu.Suzuki@dcwater.com

Suzuki serves as the manager of Process Engineering at DC Water's Blue Plains Advanced Wastewater Treatment Plant. He obtained his bachelor's degree in Civil Engineering from the University of Wisconsin-Madison and master's degree in Environmental Engineering from Michigan Technological University. Prior to DC Water, he served as a Peace Corps Volunteer in the Republic of Panama.

Areas of expertise: Wastewater Treatment Process, Process Control, and Automation

Md Nazmul Kabir Sikder

nazmulkabir@vt.edu

Sikder is a Ph.D. candidate in Electrical and Computer Engineering at Virginia Tech. He is affiliated with the Commonwealth Cyber Initiative (CCI) as a Graduate Research Assistant for developing Artificial Intelligence Assurance (AIA) methods for Cyber-Physical Systems. He has expertise in Data Analytics in Water Distribution Systems and developing MA and DL models for classification and prediction tasks for supervised and unsupervised problems. **Areas of expertise:** AI Assurance, Water Distribution Systems, Deep Learning, Data Mining, Anomaly Detection

Zhiwu (Drew) Wang - replaced Zhaohui An

wzw@vt.edu

Dr. Wang is the director of the Sustainable Environment Research Laboratory (SERL), which is a Virginia Tech Lab located in the Greater Washington D.C. area. The goal of SERL is to perform applied water quality research and apply it to solving the critical environmental problems facing the D.C. area, the Commonwealth of Virginia, and the nation. Dr. Wang's research focuses on the development of sustainable biotechnologies for the treatment of liquid and solid waste.

Areas of expertise: Water Treatment, Biotechnology, Solid Waste

Wan-Yi Mao

wanyi@vt.edu

Mao is a master's student in Computer Science at Virginia Tech. She is affiliated with the A3 lab and the Commonwealth Cyber Initiative (CCI) as a Graduate Research Assistant. She is working on achieving AI assurance and applying anomaly detection to critical cyber-physical systems using ML and DL models. She will receive her master's degree in 2023.

Areas of expertise: Data Analytics, Machine Learning, Water Security

Summary

This work introduces the Real-time AI-Driven Decision Support System (RADS). RADS is a real-time ***cybersecurity***, ***prediction***, and ***optimization*** framework for Wastewater Treatment Plants (WWTP). RADS addresses the need for intelligent systems (AI-driven) to manage immense data and water flow at facilities. We present a web-based, AI-driven framework that can be adapted and deployed to assist multiple tasks including but not limited to, data security, reservoir water level prediction, and cost-efficient effluent management for pumping goals and during extreme weather conditions. RADS provides 90% protection against cyberattacks, 87% accurate prediction of peak water levels in the WWTP reservoirs, and 27% saving on effluent release processes during extreme weather conditions. Our simulations also showed RADS was able to predict every single overflow incident from 2018 to 2022 at Blue Plains Advanced Wastewater Treatment Plant (DC Water).

About A3 Lab at Virginia Tech

We are a team of researchers focused on developing Artificial Intelligence (AI) applications and assurance algorithms. We deploy AI across different domains and provide data-driven solutions to persisting security and public policy problems, such as: optimizing water system operations, precision wastewater treatment, protecting national cyber-physical systems, evaluating technological trends, and understanding the effects of outlier events on society and the environment. However, serious show-stopper problems are persistent throughout those AI deployments, such as data poisoning, AI explainability, safety, trustworthiness, data bias, incompleteness, data democracy, security, and dark data. AI assurance methods are required to address such critical challenges; our research is at the intersection of these issues.

About DC Water

This project is a collaboration of Virginia Polytechnic Institute and State University (Virginia Tech) and DC Water. The work is also related to DC Water's Clean Rivers Project. The Clean Rivers Project is DC Water's ongoing program to reduce combined sewer overflows (CSOs) into the district's waterways - the Anacostia and Potomac Rivers and Rock Creek. The Project is a massive infrastructure and support program designed to capture and clean wastewater during rainfalls before it ever reaches our rivers. More details can be found in the link: www.dcwater.com/cleanrivers

As a part of our collaboration, we aim to improve the operational security and efficiency of Blue Plains Advanced Wastewater Treatment Plant (BPAWWTP) (Figure 1). BPAWWTP is located at 5000 Overlook Ave SW, Washington, DC 20032, and is the world's largest advanced wastewater treatment plant. The facility is operated by the District of Columbia Water and Sewer Authority (DC Water). The plant opened in 1937 as a primary treatment facility, and advanced treatment capacity was added in the 1970s and 1980s. The effluent that leaves Blue Plains is discharged to the Potomac River and meets some of the most stringent permit limits in the United States (Figure 2).

Figure 1 Blue Plains Advanced WWTP (BPAWWTP).

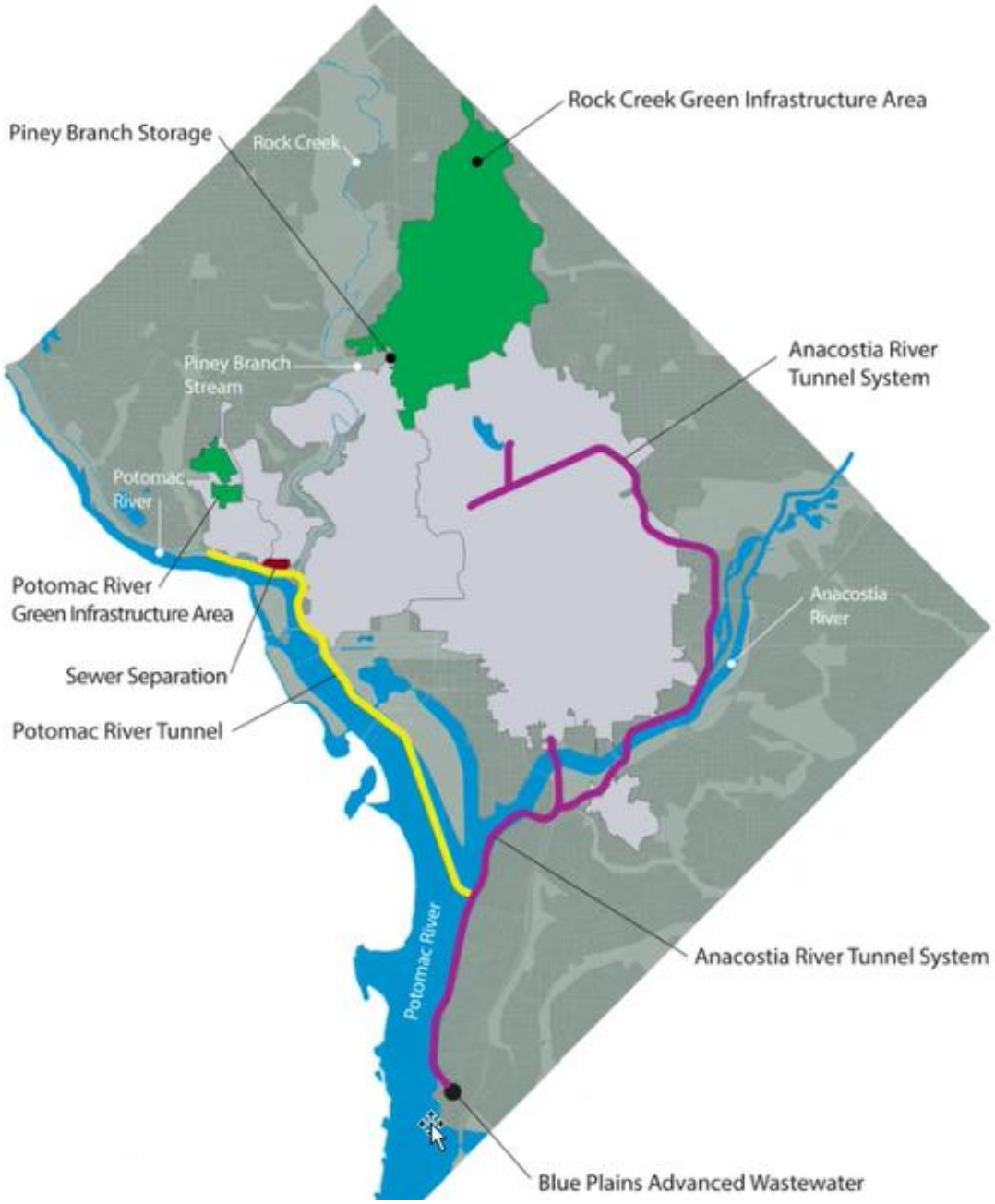


Figure 2 DC Water Clean Rivers Project location map.

The Problem Statement

Based on their dynamic nature and large scale, water treatment facilities generate an immense flow of real-time data. Monitoring such data and interpreting underlying hidden information in a timely manner is crucial for decision-making and ensuring **safe, secure, and efficient** operation at a facility (Tuptuk, 2021). Our team has developed a comprehensive three-way solution to the critical problems of **protecting (against cybersecurity threats), predicting (the system state), and optimizing** processes in the facility, using cutting-edge AI and data science technologies.

Why is AI-based security critical?

One of the problems to be addressed in intelligent water systems is how engineers can ensure the health of the data used to make decisions. In that process, it is crucial to assess whether anomalies (i.e. cyber-attacks) exist, including internal ones, from competitors, form other nations, or others non-state adversaries.

Cyber-physical attacks have increasingly targeted water treatment systems in recent years (Adepu, 2016). This is partially due to the expansion of the Internet of Things (IoT) and the proliferation of AI increasing the digitization of the decision-making processes and creating an adversarial attack opportunity following recent development in the machine learning field, which led to black-box adversarial methods that work well even with limited information of the system (Black-box Adversarial Attacks with Limited Queries and Information, 2018). Kaspersky ICS-CERT vulnerabilities report also showed that the water systems are the 3rd out of all other cyber-physical systems (CERT, 2019).

The Kemuri Water Company (KWC) (Hassanzadeh, 2020) incident in 2016 is a significant example of the risk that operators at national water infrastructures should consider. The mentioned attack has resulted in more than 2.5 million records being stolen. More importantly, the attackers were able to change control data and manipulate the water quality and flow supplied to the area. The attacks were halted before any public health damage occurred. Nonetheless, this shows how vulnerable these infrastructures are and how important it is to ensure their safety.

Another recent and critical incident was the Florida Water Supply hack in 2021 (Bergal, 2021). In this malicious attack, the hacker was able to gain remote access to the Programmable Logic Controller (PLC) unit that controls sodium hydroxide levels (also known as lye) of the water supplied to more than 15,000 people in Tampa, Florida. The hacker increased the amount of sodium hydroxide content in the water by 110-fold. Fortunately, the attack was mitigated before the poisonous levels of chemicals were diffused into the distribution network.

Both incidents demonstrate the importance of cybersecurity in water security threats with many manifestations that can't be detected unless intelligent algorithms such as Artificial Intelligence (AI) models are leveraged. The real-time management framework developed addresses these security and optimization issues. As an example, Figure 3 shows an attack scenario from a famous attack detection Secure Water Treatment (SWaT) Dataset.

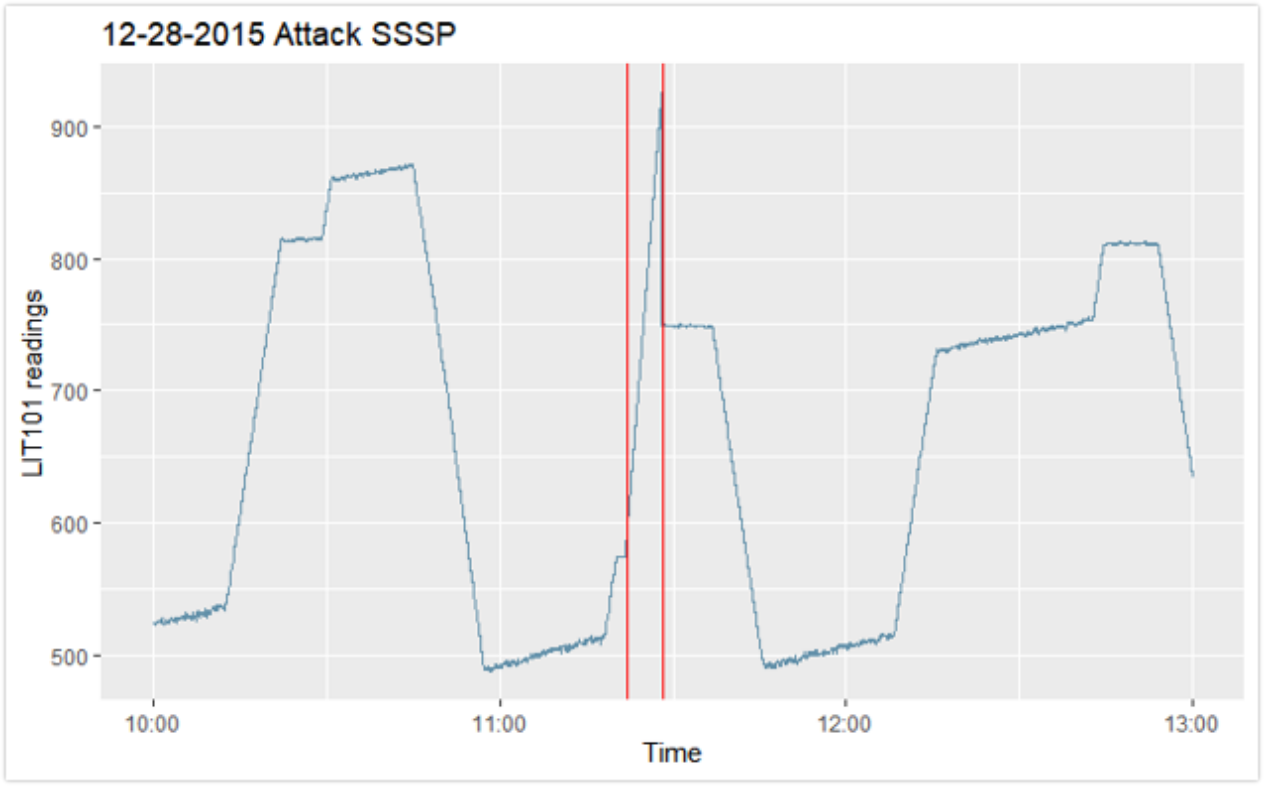


Figure 3 Attack Scenario 1: Attack on a sensor.

SWaT is a scaled-down water treatment testbed with real cyber and physical equipment to experiment with cybersecurity incidents, which was started in 2015 by the Singapore University of Technology and Design (Goh, 2016). The testbed consists of a six-stage water treatment process with modern-day components. Our initial experiments were performed using SWaT. The data collected from the testbed consists of eleven days of continuous operations, including seven days' worth of data under normal operation and four days' worth of data under attack. All network traffic, sensor, and actuator data were stored in a database supporting the models.

In the attack scenario shown in Figure 3, attackers were able to manipulate our readings from the system sensors in a marked timeframe. An experienced operator might detect this anomaly only if they monitor every single value of the system at every second. However, there are cases even under the supervision of an expert that cannot be detected. Figure 4 shows us such a case. In this attack scenario, attackers mimic the patterns of the system while manipulating the data. This is a complex attack instance for an expert to detect. Yet, AI systems are proven capable of recognizing such subtle changes in patterns. A Single Stage Single Point attack (SSSP). Most of the time, however, attackers choose to attack multiple stages of the process from multiple attack points (MSMP – Multi-Stage Multi-Point), making detecting the attacks even harder. On average, most WWTP attacks have been noticed within one week to one month after the attack happened.

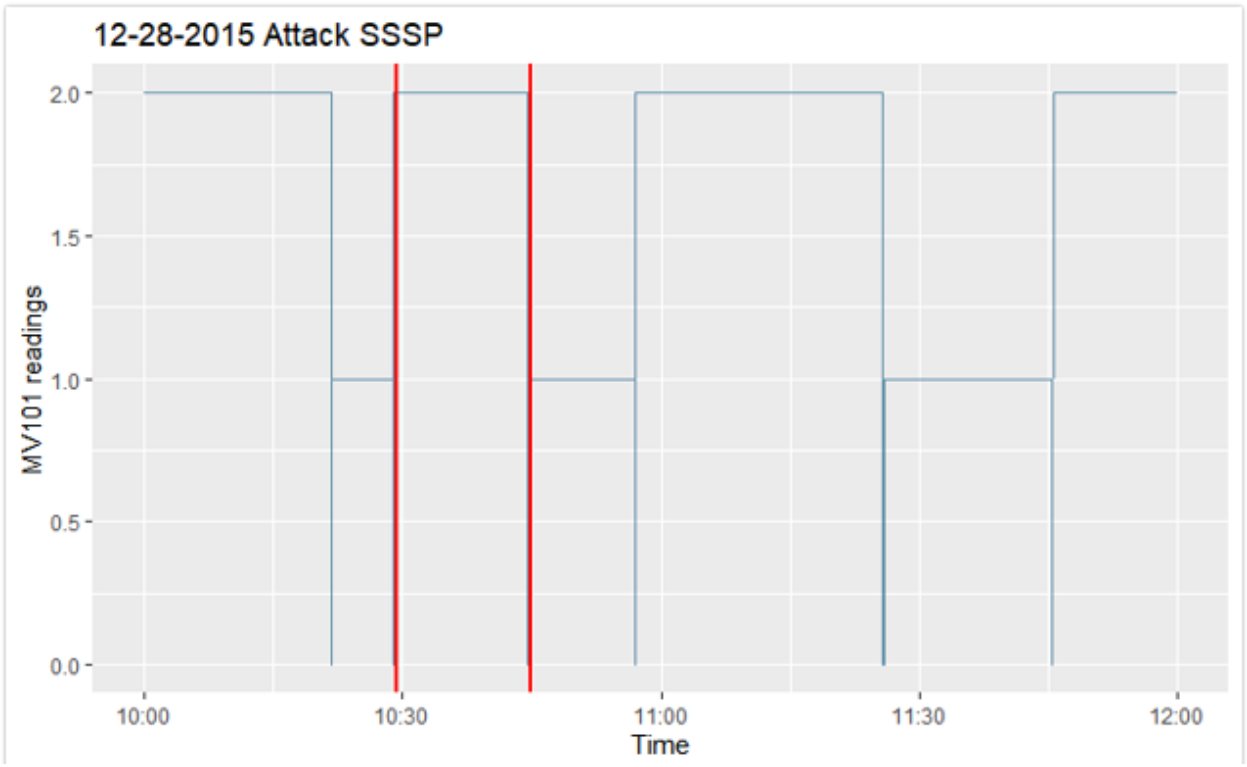


Figure 4 Attack Scenario 2: Attack on actuator data.

Why is AI-based prediction necessary?

The second problem we have identified at water facilities is the prediction of water flow and the amount of water in reservoirs. As all the operations rely on the amount of water in the system, having early information about it is crucial for taking *cost-efficient* actions while minimizing the potential overflow risks and greatly helping critical decision-making processes (such as pumping and adding chemicals). AI can help forecast (Gurrapu, 2021) and understand non-linear relationships among different system parts, including sensor values, weather, and water levels in the reservoir.

We look at other domains, including Energy and Health, and find out that those domains exploit the power of AI and save operational expenditure (OPEX) by forecasting energy demand, medical resources prediction, and allocation. We can adopt the power of AI for different downstream tasks to provide sophisticated decision intelligence support to the operators. For instance, identifying intense rainy days in a wastewater treatment plan is important to forecast water levels. Since overflow of water in the tunnel leads to higher operational costs, the overflowing water can pollute the river. By properly forecasting the water level, the operator can make decisions ahead of time for resource allocation and minimize chemical consumption. Similarly, the forecasted data can have the same energy consumption as well since energy consumption makes up 25-30% of total operation and maintenance (O&M) costs (According to the USEPA). If the water level can be forecasted, the operator can operate the most cost-efficient investment for energy savings.

Deep Learning based prediction models (including LSTM and GRU) can represent systems using historical data from a group of nodes or a single node in a Water Distribution System. Using LSTM, forecasting can be produced in real-time for the next number of hours (next 2, 4, 6, or 8 hours) to assist with capacity and

pumping plans. DC Water required a 4-hour multistep forecast to predict tunnel water level using LSTM architecture so that action can be taken in a timely manner; operators require around 3 to 5 hours to operate such decisions efficiently.

Why is AI-based optimization required?

Lastly, after ensuring the safety of the data and enriching raw data with forecasts and insights, we need to use all the information we have to make decisions regarding the facility's energy efficiency. However, it is practically impossible to completely, consistently, and accurately process all this information (i.e., Big Data) by a human operator. Therefore, it is essential to use AI for optimizing aspects such as energy utilization and operational cost.

Combined sewer overflows (CSOs) represent major water-quality threats to hundreds of cities and communities in the US that are served by combined sewer systems. CSO events cause the release of untreated stormwater and wastewater into receiving rivers, lakes, and estuaries, causing a variety of environmental and economic problems. Costs associated with CSO management are expensive. The EPA estimates the costs of controlling CSOs throughout the country are approximately $56Billion (Wise, 2010).

The optimization of the massive energy consumption of the treatment process facilitates mitigating greenhouse gas emissions, which has been considered one of the biggest global challenges in the 21st century. It is because water treatment requires intensive energy consumption. For example, wastewater treatment consumes up to 20% of the total energy by public utilities (Means, 2004) and 2-3% of the world's electricity consumption (Olsson, 2012). The accelerated population growth and urbanization further require more energy input to satisfy the higher water treatment standards. Therefore, the optimization of energy is highly desired. Since biological treatment is still the most common wastewater treatment strategy for pollutant removal, the associated pump operation, chemical addition, and aeration largely contribute to the total energy cost in the wastewater treatment plant. The real-world data collected from different WWTPs are used to develop the model in this proposed plan for future application in full-scale facilities. Understanding the factors that will most affect the energy cost allows for creating a higher energy- and cost-efficient wastewater treatment strategy.

System Goals

Our goal is to design and create a software suite that can process a high stream of real-time data and support cost-efficient, safe, and secure decision-making processes at water treatment facilities.

As shown in Figure 5, the main skeleton of RADS consists of three major AI-driven modules. The protection part assures the security of data used throughout the facility and prediction and optimization models of RADS. The prediction module forecasts the system state variables including but not limited to water levels (amount of water in different parts of the WWTP). This forecasted information is used for monitoring, better management of the WWTP resources, and being fed to the optimization module for increasing the model’s efficiency. The optimization module continuously collects sensors, actuators, pumps, PLC (Programmable Logical Unit) statuses, and forecasted data from the prediction model. It provides actionable recommendations such as when to start pumping water to the reservoir, how many pumps to operate, and with what capacity these pumps should be run for the operators.

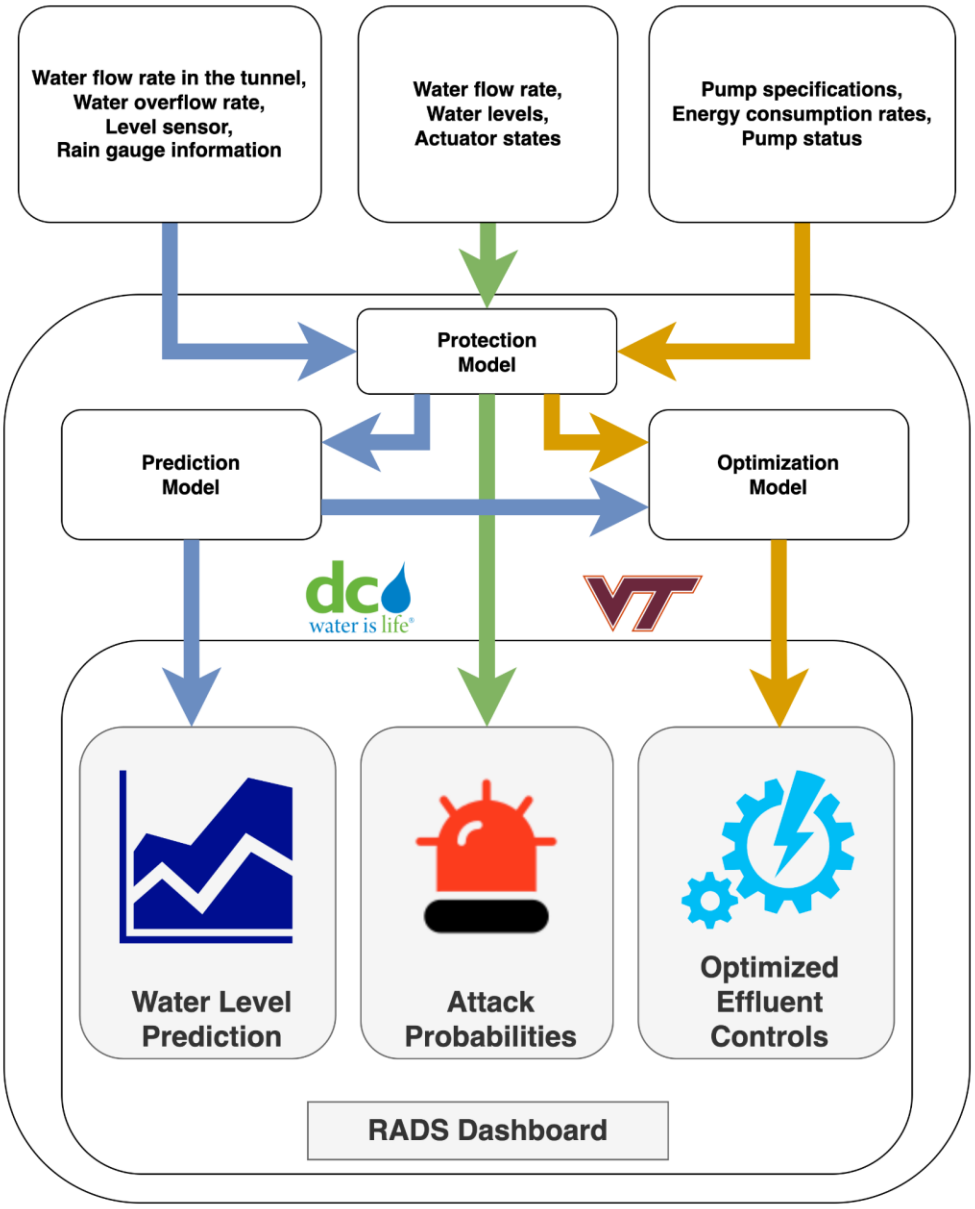


Figure 5 RADS Framework.

Protection

As mentioned earlier, the Florida Water Supply hack in 2021 showed how vulnerable water systems are (Bergal, 2021). Another (recent) intrusion happened on the water treatment plant that served parts of the San Francisco Bay area on January 15, 2021 (Collier, 2021). The hacker had the username and password of an employee's TeamViewer account. In many cases, reasons that cause a system to be unsafe include human mistakes, sensor breakdowns, and security code leakages. All these malicious incidents lead to the need to build an intelligent water protection system. In both examples, the systems were breached, yet after investigating the data flow, the authorities were able to notice the intrusions. Generally, we can see that data level protection is the last line of defense in system security. Therefore, RAD's primary goal is to

protect the system and data. For instance, a reservoir's water level would continuously grow or decline, but it would not suddenly drop. If that happens, there is usually something wrong with the data. Another point is if we see the data input just from a one-time unit when the water level suddenly drops. Therefore, we need to evaluate (using AI) the value sequence to know if there is a cyber incident.

Besides the real-life dataset provided by DC Water, we also utilize benchmark research datasets in this project. This is due to a lack of historical records of cyber-attacks at the facility. Modbus Penetration Testing Framework (Smod) (Laso, 2017) dataset is one of the most important cyber-physical attack datasets. The value of this dataset is that it records the attack situation (Table 1), which is helpful for the protection module to train a model that can detect and classify anomaly types or identify components that are under attack.

For example, a blocking object covers the sensor and creates noise for the ultrasound sensor or floating objects that cause sensor value perturbations. The physical system used to generate this dataset contains two tanks: The main tank storage and one ultrasound depth sensor to record the liquid level. The second tank has four discrete sensors to observe the liquid level and two pump state records. This collection includes 15 temporal series that record 15 different anomaly/attack situations separately (Table 1). Each situation is classified into five operational scenarios – normal, accident, breakdown, sabotage, and cyber-attacks. The duration of each scenario varies, and the frequency of the data collected is every 0.1s.

The protection model's inputs include the physical system's sensory reading (Ultrasonic level readings, humidity, pressure readings, etc.) and actuator state (pump running capacity, valve states, etc.). We use the Recurrent Neural Network Models (RNN) based Long-Short Term Memory (LSTM) architecture to build the protection model. The characteristic of LSTM is that it learns the information *on a timely basis*, unlike other AI methods. Instead of looking at one input at a time, it looks at inputs from a period to learn trends (Figure 6).

We labeled different attack scenarios into four labels: normal, intentional attack, unintentional breakage, and unknown (in cases where the algorithm is not able to classify). It is important to note that AI models are data hungry. As the number of examples from the same anomaly increases, the model's accuracy increases further as well. Suppose there is a water distribution system that serves a city that can predict the water supply and demand; if there is a failure in data collection, then the predictions can no longer be correct, and the whole city might not get enough drinking water because the system forecasts are off (Sobien et al., 2022). Also, Tuptuk et al. (Tuptuk 2021) reviewed how frequently attacks occur on water systems, showing the importance of the security of a water system. Mao et al. (Mao et al., 2022) also found the variabilities of water systems and proposed a framework to build an early warning system. Our goal aligns with their conclusion that the security of the water systems needs more attention and empirical manners to evaluate it.

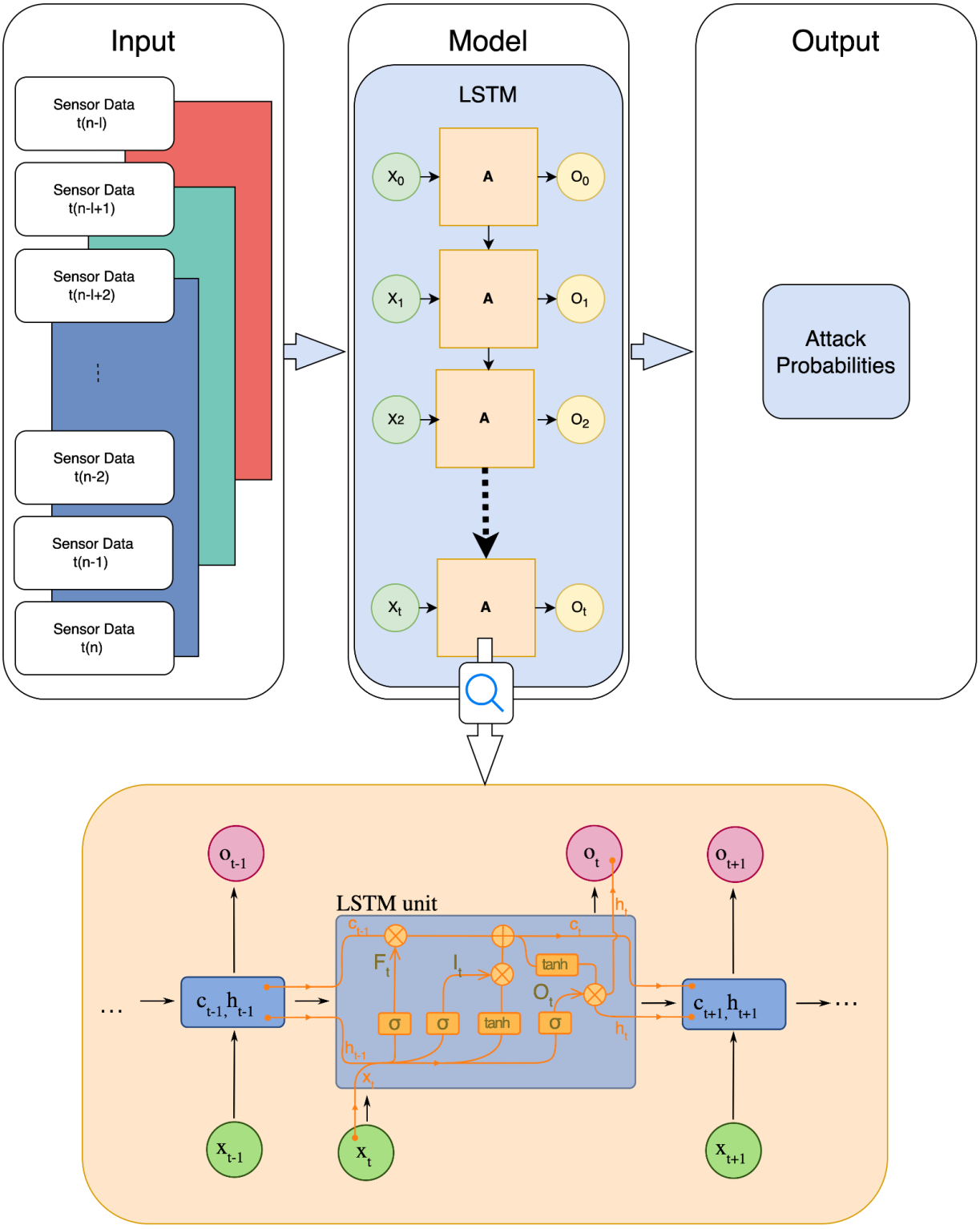


Figure 6 Protection model data flow - n: current timestamp; l: lookback window width.

Table 1 Fifteen attack situations of Smod dataset

Situation	Affected component	Operational scenario
Normal	None	Normal
Unknown sensor noise	Ultrasound sensor	Accident/ Sabotage
Blocked measure 1	Ultrasound sensor	Breakdown/ Sabotage
Blocked measure 2	Ultrasound sensor	Breakdown/ Sabotage
Floating objects in main tank (2 objects)	Ultrasound sensor	Accident/ Sabotage
Floating objects in main tank (7 objects)	Ultrasound sensor	Accident/ Sabotage
Humidity	Ultrasound sensor	Breakdown
Discrete sensor 1 failure	Discrete sensor 1	Breakdown
Discrete sensor 2 failure	Discrete sensor 2	Breakdown
Denial of service attack	Network	Cyber-attack
Spoofing	Network	Cyber-attack
Wrong connection	Network	Breakdown/ Sabotage
Leaking pipe (low intensity)	Whole subsystem	Sabotage
Leaking pipe (medium intensity)	Whole subsystem	Sabotage
Leaking pipe (high intensity)	Whole subsystem	Sabotage

Protection results

One limitation of this dataset is that the physical components that collected the data are small, and each attack situation data was collected separately, not continuously, and the time duration for each attack situation varies. Therefore, to prevent the model from learning with bias, we oversample each attack situation on a custom width lookback window to make the dataset more balanced while preprocessing the data. We tried many different approaches and architectures for classification and detection tasks. A deep-learning RNN model, LSTM (Long-Short Term Memory) has been the most successful (Figure 7). Table 2 shows that the model performs with an average of 90.2% accuracy in attack/anomaly detection tasks.

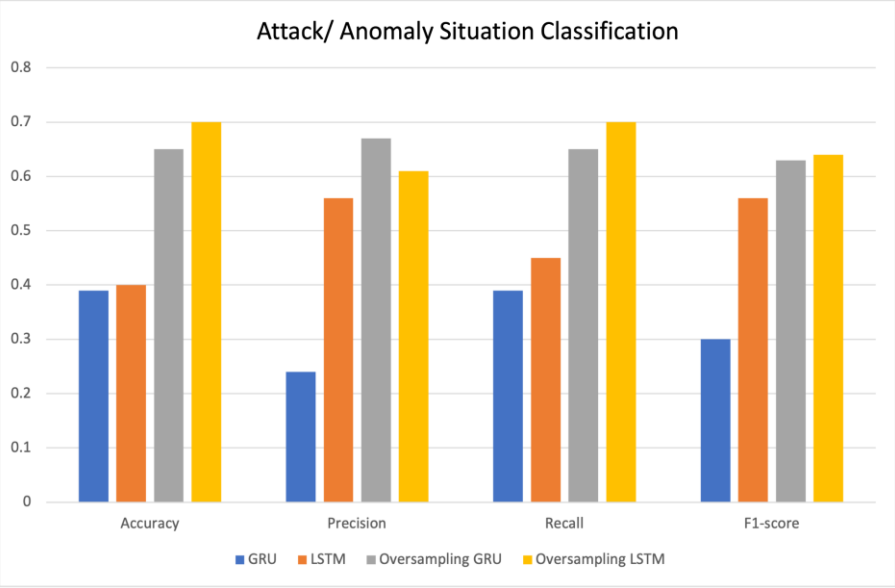


Figure 7 Attack/Anomaly Situation Classification result compared with different methods.

Table 2 Protection model metrics

Task	Accuracy
Attack/Anomaly Detection	90.2%
Intentionality Classification	74%
Attack/Anomaly Situation Classification	70%

Prediction

Accurate forecasting of WWTP's water level can improve the WWTP's reliability, reduce operational costs, and endorse overall optimization. AI models are proven helpful for data-driven applications (Sikder, 2021). In RADS, AI is used for WWTP applications to tackle the process of non-linearity and the dynamic nature of sequential system data. The proposed framework includes a regression model, a deep learning-based method, to forecast WWTP water levels. Long Short-Term Memory models are good at handling sequential time series data. Application of Recurrent Neural Networks (RNN) for learning sequential datasets, specifically LSTM networks (Zhao, 2017), to make predictions.

Additionally, we present variable importance that helps identify essential information about the system and take intelligent operational steps. Model explainability lets the operator know why and when to take an operational decision. We explain feature importance for the forecasted output using the Shapley Additive exPlanations (SHAP) tool (Lundberg, 2017). LSTM, with SHAP explanation, delivers state-of-the-art processes for an explainable AI model. Again, an LSTM network is a better architecture for learning complex trends and seasonal patterns than traditional methods, including Auto-Regressive Moving Average (ARIMA) (Box, 1994). Parametric methods, including conventional statistical learning models, require static data. The following section provides a detailed discussion of the prediction model and its application to DC Water.

BPAWWTP collects water from natural rainwater and city wastewater from the District of Columbia (DC) area and cleans it before using it for multiple purposes such as vertical farming, community gardening, and urban restoration. The diagram of the whole treatment plant is shown in Figure 9. The figure divides the treatment plan into three zones. The first zone is colored green which represents wastewater collection topology. The light orange and purple colored zone represent various operational activities for cleaning the wastewater by applying bacteria and chemical mixture. Usually, the plant saves much money by applying a natural cleaning process and exploiting natural bacteria. However, the game changes during the heavy rainy days because the tunnel quickly fills up with wastewater and overflows. In this context, the plant operators cannot rely on the natural cleaning process as this may cause an overflow of the wastewater and pollute the river. So, the facility adds additional chemicals and electrical energy to expedite the cleaning process.

To solve this overflow problem, we focus on predicting the water level in the tunnel, especially during intense rainy days. Additionally, the plant must consume more chemicals to clean the overflowed water, increasing operational expenditure during heavy rainy days. The prediction model tries to solve both problems together. For instance, the model can provide the operator with the tunnel's next 2 or 4 hours of forecasted water level data. With that information, they can make intelligent operational steps to

minimize this additional expenditure (for example, chemicals). We have access to DC Water's historical tunnel water levels data from 2018 to 2022 (Figure 10). Furthermore, we have access to all the tunnel sensor data that provide necessary information for forecasting water levels. We also incorporate weather data, including rain gauge information for predicting/forecasting water levels.

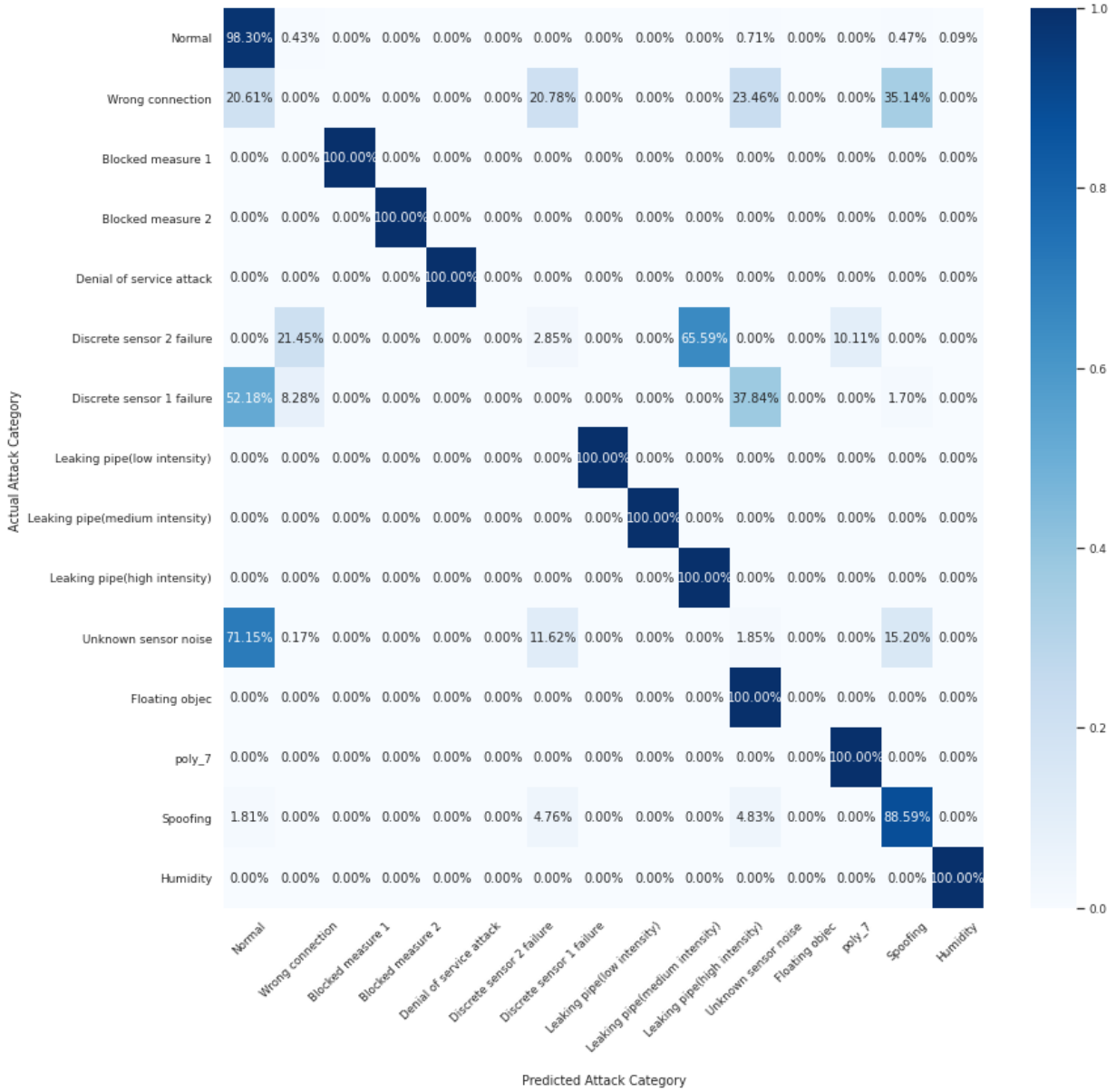


Figure 8 Attack/ Anomaly Situation Classification results.

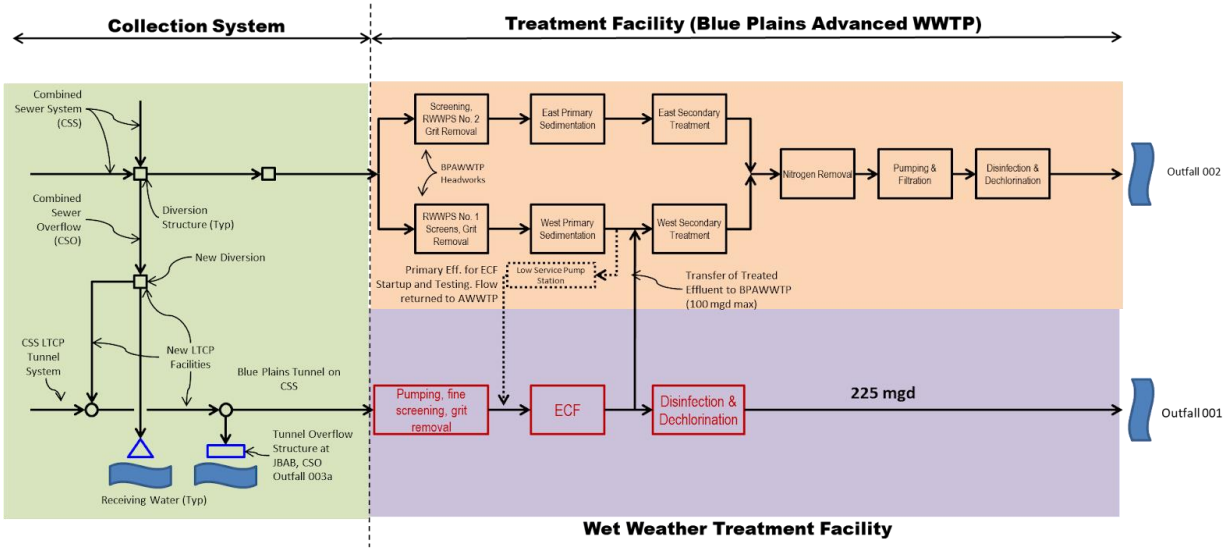


Figure 9 Blue Plains WWTP

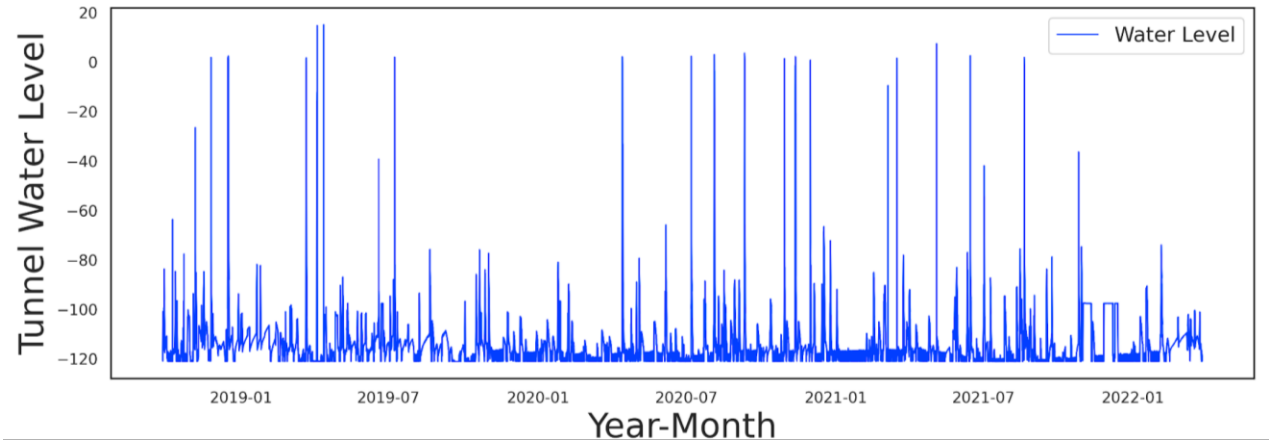


Figure 10 Blue Plains Advanced WWTP tunnel water levels over time

Figure 11 shows our initial approach to developing and testing models for the forecasting task and how we evaluated the effectiveness of the models in a pipeline. However, our proposed RADS includes LSTMs only that take sequentially processed data input and forecast multistep (2 hours) data. One of the three essential modules of RADS is to forecast or predict short-term DC tunnel water levels. The main goal of prediction is to correctly identify the potential chance of Combined Sewer Overflows (CSOs) during intense wet days.

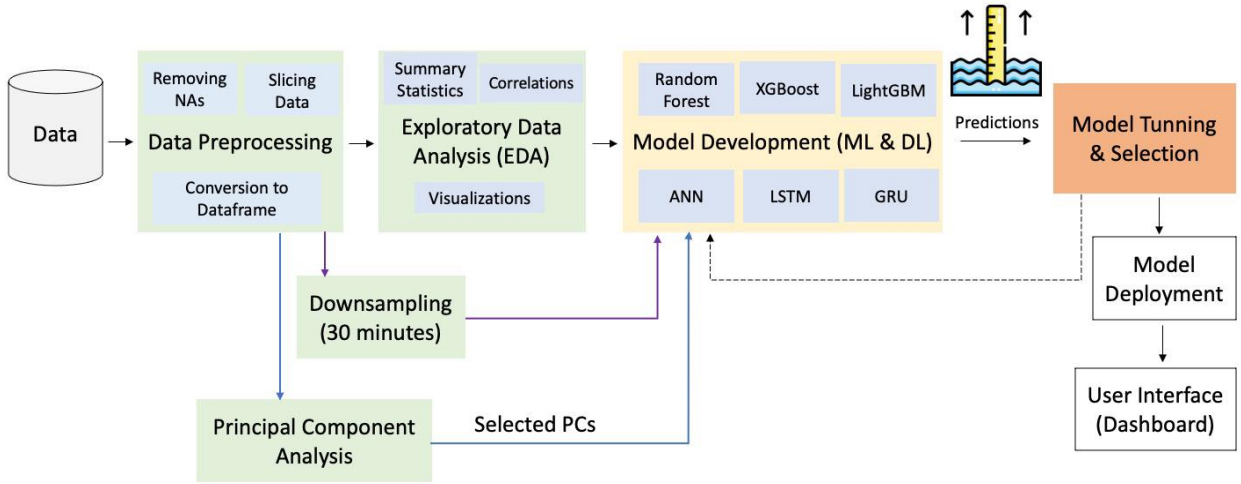


Figure 11 Water level prediction pipeline

DC Water data

The DC Water dataset consists of 30 min intervals of the plant's water level information from August 2018 to March 2022. From figure 10, it is evident that the water level data are not stationary. Usually, static data have constant standard deviation, mean, and autocorrelation. Nevertheless, Figure 10 shows non-stationary and dynamic water level behavior because of stochastic weather and precipitation; therefore, traditional parametric methods do not perform well. Thus, conventional prediction techniques, including ARMA, cannot learn the data's dependency structure over time; hence LSTMs are good alternatives. Although LSTMs are popular with natural language modeling, they can learn temporal relationships among different objects while predicting single or multiple outcomes. Additionally, LSTM models work well with complex time series (Box, 1994).

Data Processing

Adding more compelling features to the training dataset may bring good model prediction. Therefore, we feature engineer and derive three attributes from the date column, including the day of the month, the month of the year, and the hour of the day, to help predict the target feature, the water level. Since we collected raw time series data from the plant, we paid attention while wrangling the data. The trial steps incorporate data cleaning, imputation, missing value check, correlation check, PCA analysis, Pearson correlation coefficient test, and variable correlation check. Then, we split the dataset into training and test sets and rescaled all features to a standard scale between 0 and 1. When the features are on different scales, this process helps to prevent model overfitting. We observe the model's prediction performance on the test dataset only.

Next, we derive a multivariate time series model using LSTMs that can take multiple inputs for predicting multistep outputs. Before we discuss the modeling steps, we need an effective objective function (i.e., loss function) that helps to achieve the desired outcome (i.e., correctly predicting water level peaks or effective pump action generation). We aim to predict water levels as closely as possible while prioritizing water overflow incidents. We make the following considerations: choose cubic difference as a cost function (Equation 1), given that our data have outliers because of intense rainy days and water overflow situations.

$$cost/loss\ function = \frac{1}{n}\sum_{i=1}^{n} \left| y^3 - \hat{y}^3 \right| \quad (1)$$

Given some significant but rare overflow incidents in our data, cubic difference as loss is a good pick. Furthermore, we are more interested in predicting the peak of water level (i.e., overflow situations) rather than understanding the general trend of the time series.

Training and Evaluating Model

Our multivariate forecasting model (Figure 12) includes one LSTM layer and one fully connected layer for predicting the next four hours of tunnel water level. We create sequences of 24 hours in length and train the model with them, resulting in 4 hours multistep forecast. We apply the Adam optimizer to minimize our objective function (Equation 1). After the model is trained, we evaluate the outcome using the test dataset (30% of the original dataset). Next, we have to perform an inverse scale of the model's outputs to the actual scale to assess the forecast water level against the original tunnel water level.

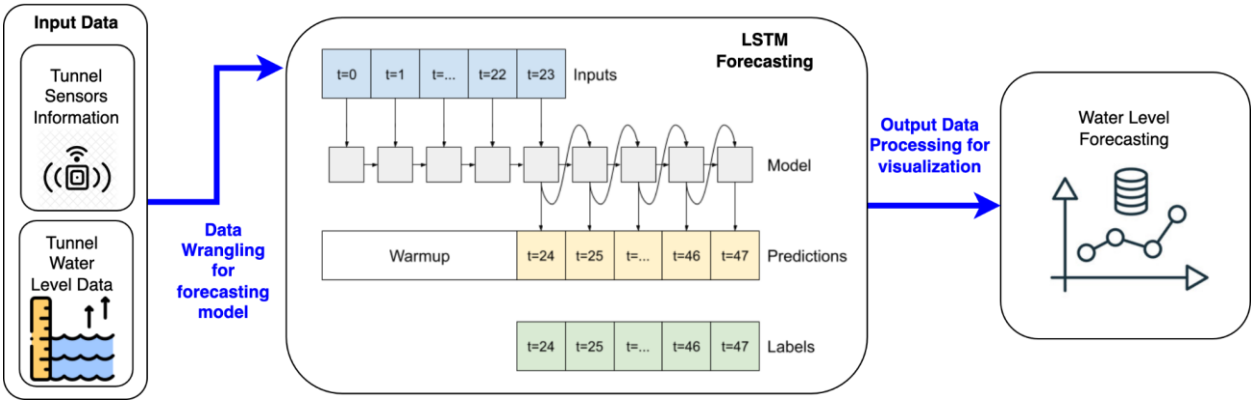


Figure 12 Water Level Forecasting LSTM Architecture

Prediction model results

Let us look at the test data's predictions (Figure 13). The model performs well in water level prediction, reflecting our primary objective of multivariate modeling. Around the middle of the test data, the model constantly underestimates the water level and fails to predict extreme scenarios. Table 3 has a few other values on other important metrics, including MAE, RMSE, and R square. Apart from these metrics, we also estimate peak detection accuracy (Equation 2).

$$Peak\ detection\ accuracy = \frac{No\ of\ predicted\ peaks}{No\ of\ actual\ peaks} \quad (2)$$

Table 3: Multivariate model performance metrics on test dataset

Performance Metric	Score
Mean Absolute Error	0.0161
Root Mean Squared Error	0.0408
R-square	0.665

Peak prediction is essential as the optimization model relies on this extreme water level information to generate optimal releasing actions. For detecting peaks, we choose threshold -47ft; therefore, we consider tunnel overflow beyond the -47ft value for model development.

Furthermore, Table 4 provides peak detection accuracy from both training and test dataset. We observe a peak detection accuracy of 87.2% on the test set and 86.5% on the training set.

Table 4: Peak detection performance

Dataset	Actual peak observed	Peak Detected	Peak Detection Accuracy
Train Dataset	223	193	86.5%
Test Dataset	39	34	87.2%
Total	262	227	86.6%

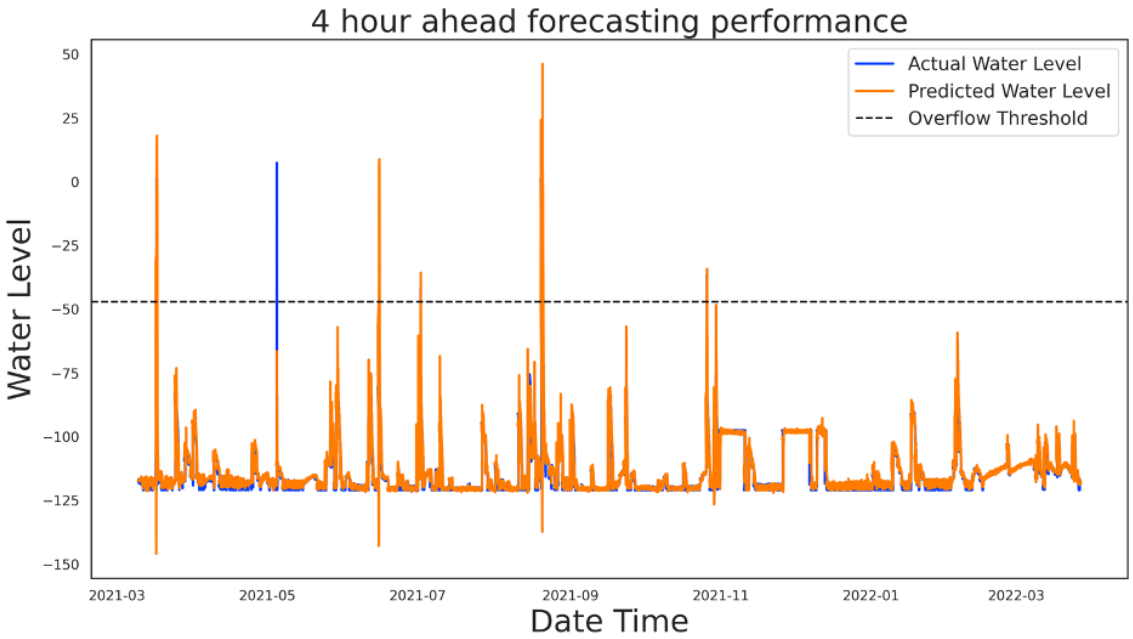


Figure 13: Predicted vs. actual tunnel water level

Sequential data, in time series forecasting, have a seasonal drift; therefore, a deployed forecasting model needs to retrain with the latest samples after a particular time step. Retraining the model with new samples helps the model to learn seasonal drift and temporal information.

Important Features

Neural networks are black boxes in nature because interpreting them is problematic. We use SHAP API to estimate Shapley values that apply a game theory. This technique assesses the model's prediction and answers the contributions of each feature for each prediction. More specifically, the process, also known as a deep explainer, decomposes each outcome of the model and backpropagates the contribution of all neurons to every feature. It then compares each neuron's activation to its reference activation. According

to the difference, it assigns contribution scores. Based on a representative dataset, once the multipliers have been computed, feature importance can be calculated using some sample inputs (900 training data samples) and ranked (100 test data samples) based on their most prominent contributions to the model's outcome. We average the weights of all 100 samples and plot them in Figure 14. The top four input features are Water Level, sensor overflow indicator, sensor total water flow, and sensor pump 5.

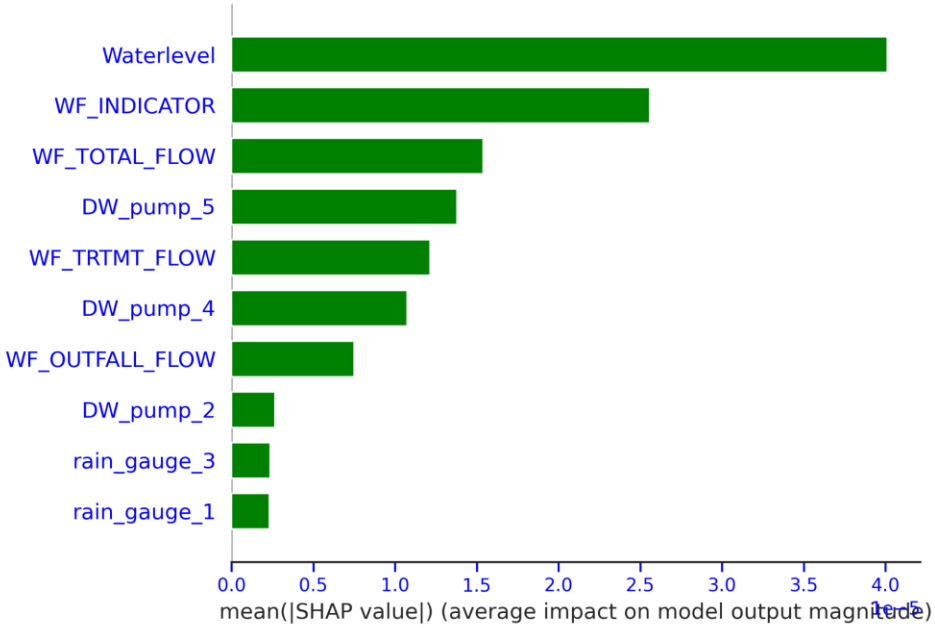


Figure 14: Feature importance summary plot

Comprehending a deep learning model's prediction is essential to interpreting the results at the DC Water plant. Using this feature importance approach, we can get insight into the DC water plant and devise action plans to promote the desired operational outcome. Furthermore, additional insights into feature importance help to optimize the operational process. For example, essential features dictate how quickly, and which pump must start early to avoid tunnel water overflow.

How the prediction module can help the facility

In full-scale WWTPs, there is always a latency phase between inflow parameters and operation adjustment, leading to less treatment efficiency with higher cost. The daily and seasonal fluctuation of influent parameters also harm the robustness of the operation. The proposed model for water level prediction and cost optimization can precisely inform the operators to control the treatment system. For example, the model can forecast the start time of pollutant concentration change caused by the increasing toilet utilization or storm. Then, the software can monitor the change and send optimization suggestions, including changing pump speed, adding more/less chemical, or increasing/decreasing inner circulation, to the operators so that the operation strategy of WWTPs can be adjusted accordingly.

Optimization

The final part of RADS is an optimization module that aims to reduce the risk of untreated wastewater overflowing the reservoir under extreme wet-weather conditions. The main goal of the Optimization

module is to prevent Combined Sewer Overflows (CSOs) in the most cost-efficient manner following the US *Environmental Protection Agency's (EPA) CSO control policy*.

RADS utilizes system state predictions to support operators and management of the WWTPs by optimizing actions using a metaheuristic algorithm inspired by natural selection called Genetic Algorithm (Hingston et al., 2008). The module calculates the optimal times and capacity for operators to start pumping untreated wastewater. This process requires wastewater to be treated by chemical means, thus introducing the energy cost associated with running the massive industrial pumps and the cost of the chemicals used in the fast treatment process instead of the most common and cheaper biological treatment, i.e., activated sludge technology. Our optimization model minimizes the usage of chemical treatment while providing secure operation perimeters for wastewater treatment processes. Hence, the developed software has the promising potential for application in other WWTPs worldwide based on activated sludge.

DC Water's Blue Plains Advanced Wastewater Treatment Plant (BPAWWTP) utilizes 24,300 linear feet of reinforced concrete tunnel as its wastewater reservoir. The tunnel can hold up to a total of 137 ft deep wastewater (17 ft above and 120 ft belove sea level). The facility has 5 large capacity industrial pumps that can move between 50 to 83.3 Mg/d (million gallons per day) and a lesser capacity pump that can move between 3 to 10 Mg/d. These pumps directly control the amount of wastewater treated by chemical means.

The model is given pump specifications and energy consumption data samples of pumps under wet-weather conditions. Inputs of the optimization model are current pump states and water level estimations from the prediction model. The model outputs optimal times to operate 6 release pumps of the facility. The optimization model also calculates the simulation results of operating under optimal actions and plots the expected level of wastewater in the tunnel.

To summarize how this module operates; RADS generates an initial population of random solutions to the problem. Then the actions of each solution are simulated, and the system state is calculated. Using a fitness function feasibility and optimality of each solution are scored. This step ensures only the non-overflowing solutions are passed on to the next generation. Pairs with the highest scores are selected for crossover to generate new solutions using a selection method. This step improves the overall fitness of the solution throughout the generations. Then each solution with a certain chance is selected for mutation. Mutation allows the algorithm to search the solution space for the optimal solution and generate variability. And then, the process is repeated from the simulation step with the new solutions until the cost of operation is improved no further.
The optimization module has been tested with wet weather records of DC Water's BPAWWTP from 2018 to 2022. These records included a total of 191 wet weather events. RADS was able to predict 95% of these extreme weather conditions. The tests show that using the optimization model prevented all the overflow incidents with very high efficiency.

WWTP records show the facility has pumped around 100 billion gallons of chemically treated wastewater to the Potomac River from 2008 to 2022. Our simulations show that the optimization model was not only able to *lower the effluent release and chemical treatment cost by 23%***,** but it also *prevented every single overflow incident* (Table 5) under extreme wet weather conditions, given our prediction model's 4 hours of water level forecasts.

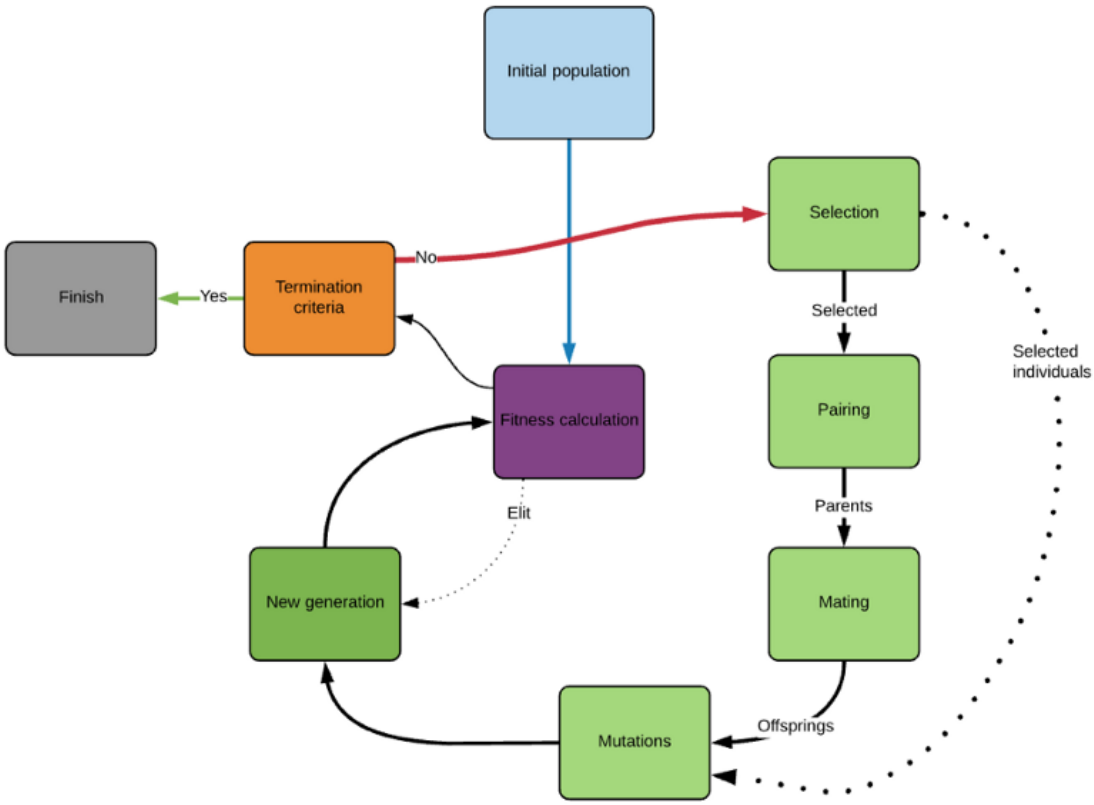


Figure 15 Genetic Algorithm flowchart

As the model used in this stage is a heuristic algorithm, the tests are evaluated 100 times. The model provides less than 3% deviation from the optimum after 50 generations with an initial solution population of 50. After 100 generations, the deviation from the optimal value becomes less than 1%.

Table 5: Optimization efficiency evaluation

	Chemical Treatment (Mg Wastewater)	Overflow (Mg Wastewater)
BPAWWTP Records	99,400	3,200
Optimization Model	76,531*	0*
RADS Efficiency	23%*	100%*
*Average over 100 repetitions		

The RADS Interface

RADS brings the presented 3 modules together into a web-based real-time monitoring interface. The application can run locally or remotely; modules and the application have been programmed on Python-3.9. The protection module uses the TensorFlow and Keras libraries to train and evaluate the LSTM model, whereas the prediction model runs on the PyTorch library. Both modules have a *millisecond* range response time. The optimization model, however, uses the PyGAD library, and evaluating optimum actions takes a few seconds (less than 10). In best-case scenarios, the average evaluation takes 1.8 seconds.
Figure 16 shows RADS graphical user interface (GUI). RADS can collect data from the WWTP and process it by the 3 modules in real-time.

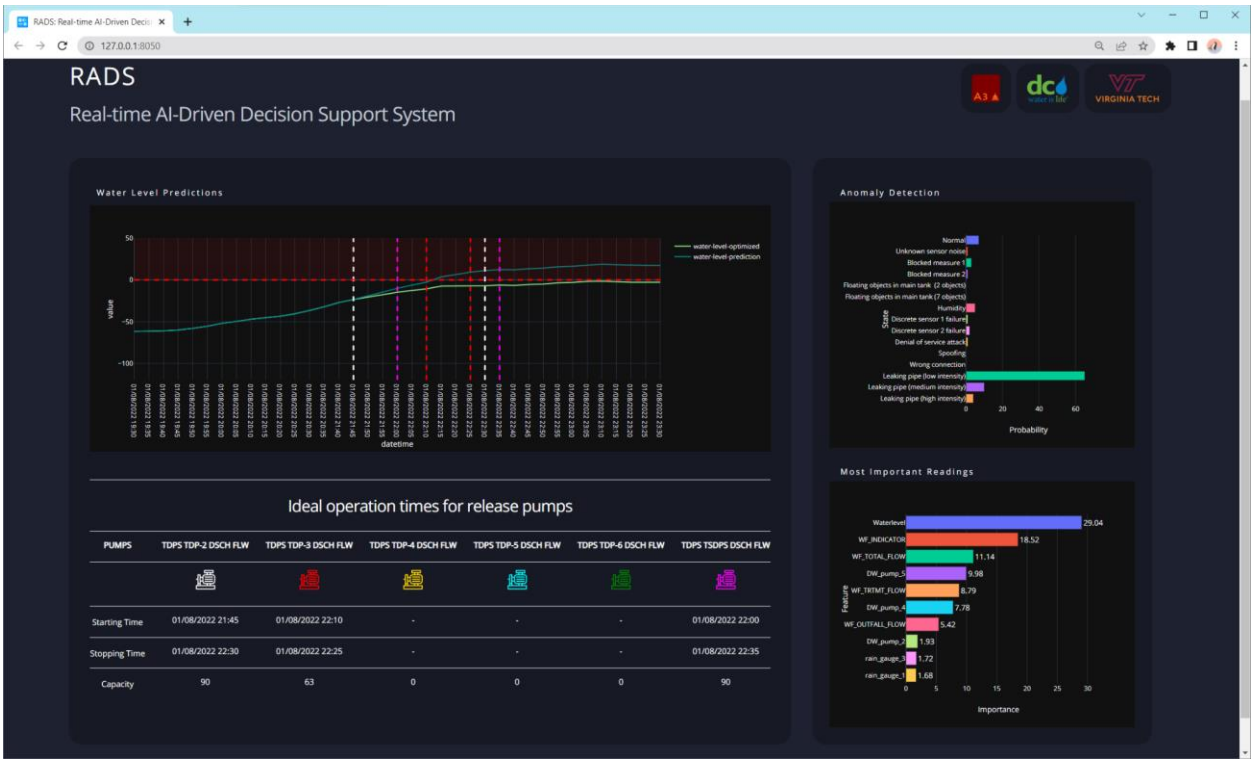


Figure 16: RADS Graphical User Interface.

The first plot on the top right indicates the wastewater level in the reservoir (Figure 17). The blue line represents the water level forecast from the prediction module. The green line represents the water level simulated following the optimal actions recommended by the optimization module. The horizontal dashed red line is sea level. RADS aims to prevent water levels from rising above sea level since the sea level is selected as the target safety level (by DC Water) even though the reservoir can hold up to 17 ft above that. The red zone above the dotted red line represents this risk associated with operating above the level. Another indicator on this graph is the optimal pump operation markers. The dashed vertical lines represent the start and stop times for release pumps. Each color represents one of the six pumps at the DC Water WWTP.

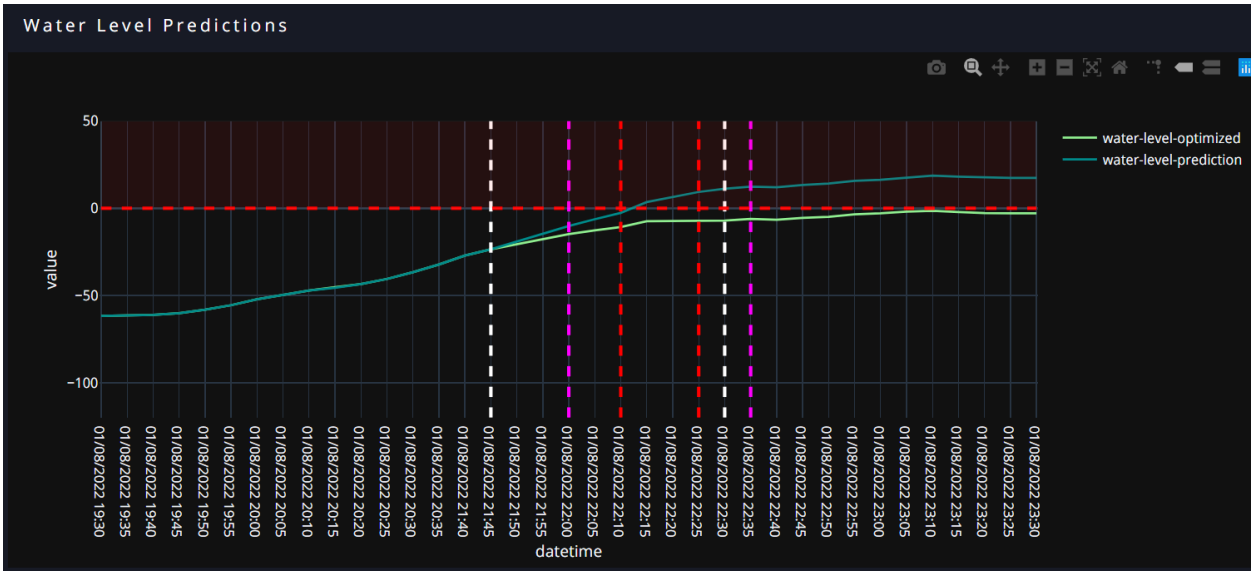


Figure 17: Water level prediction and optimum actions simulation.

The effluent control parameters are also listed on a table with further details, including start/stop times of releasing pumps and their run speed (capacities) (Figure 18); attack detection is shown in Figure 19.

Ideal operation times for release pumps

PUMPS	TDPS TDP-2 DSCH FLW	TDPS TDP-3 DSCH FLW	TDPS TDP-4 DSCH FLW	TDPS TDP-5 DSCH FLW	TDPS TDP-6 DSCH FLW	TDPS TSDPS DSCH FLW
Starting Time	01/08/2022 21:45	01/08/2022 22:10	-	-	-	01/08/2022 22:00
Stopping Time	01/08/2022 22:30	01/08/2022 22:25	-	-	-	01/08/2022 22:35
Capacity	90	63	0	0	0	90

Figure 18: Effluent release control interface.

The other plot on the top right shows the probabilities of any anomalies in the data flow of the facility. It informs the operators about the system state in real-time. It includes the 15 situations from the SMoD dataset, but the number of these can be increased by collecting more data samples from the facility.

The last part at the bottom right of the page has further details and the underlying information behind the RADS's decisions. This plot shows the operator the most critical readings in the data flow used in the decision-making process of RADS (Figure 20).

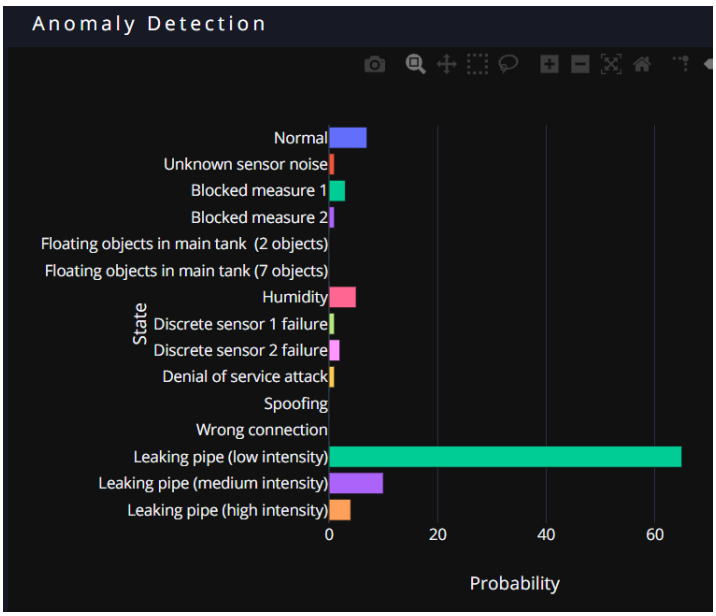


Figure 19: Attack percentage probabilities.

Figure 20: Top ten most impactful data features for decision making.

The RADS user interface is fully customizable. Any changes to the tasks we performed can be easily reflected in the GUI. New modules, plots, and indicators can be added, removed, or modified to fit any facility's needs. AI modules and the user interface code base of RADS can be found on A3's official GitHub (https://github.com/AI-VTRC/IWS-Challenge).

Adoption of RADS

RADS is a very flexible framework. Besides DC Water, RADS can be implemented at other wastewater treatment facilities (process shown in Figure 21). Our Virginia Tech team collaborates with Alexandria Renew Enterprises, a.k.a. AlexRenew (formerly Alexandria Sanitation Authority). AlexRenew operates and

maintains a new sewage treatment system to serve the city and parts of Fairfax County. Our current goal is to customize and apply RADS to provide a solution to their needs.

The timeline we envision to deploy RADS at AlexRenew can be generalized to other WWTPs. Depending on many factors such as the WWTP's size, technological infrastructure, management strategy, etc., this process can take 9 months to 2 years. Most of the time is anticipated to be spent in the deployment phase as the deployment steps may involve synchronization with the facility's workflow to minimize any disturbance to the plant's operations.

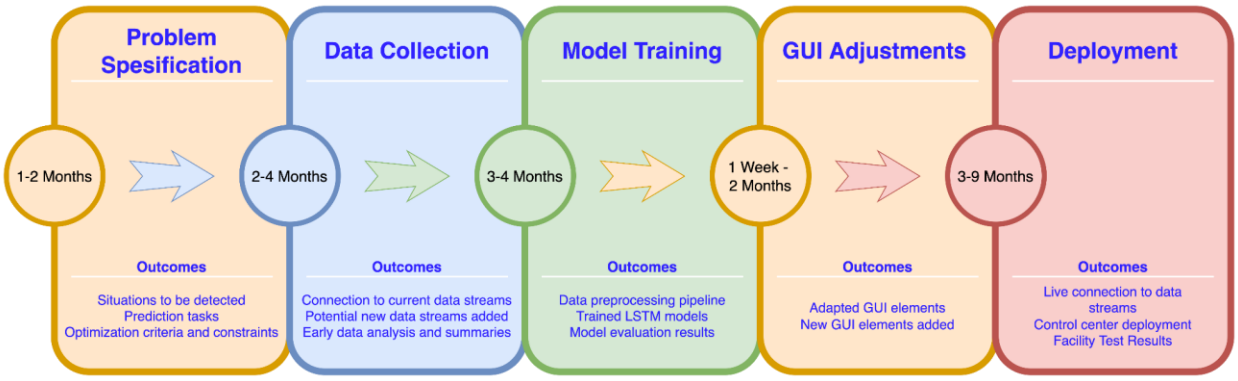


Figure 21: RADS adoption timeline.

Challenge Plan

This work was a collaboration between the Artificial Intelligence Assurance and Applications Lab (A3) at Virginia Tech and the District of Columbia Water and Sewer Authority (DC Water). The responsibilities of each researcher are listed below. Also, the project timeline is attached at the end of this document.

Responsibilities and works distribution

Batarseh, Feras A.: Responsible for communication and management of the team, quality assurance of the project, project plan, and models, evaluating the preparation of data models' tests, verification, and validation of the models, editorial of the documents, proofreading and submission.

Yardimci, Mehmet O.: Responsible for the project planning, writing, and documentation of all the materials. He was also responsible for developing and testing protection and optimization models and implementing AI Assurance into the models, visualization the results, implementing the models to the facility, and bringing submission together.

Suzuki, Ryu: Responsible for providing datasets, preparing data, and managing data access for the team. He supervised facility visits and educated the team on the operations of the WWTP. He conducted on-site facility tests for models and evaluated the test results to be reported. He was also responsible for ensuring the safety of the data and the results that have been shared.

Sikder, Md Nazmul K.: Responsible for the development, testing, and validation of the prediction model, visualization, reporting, and compiling of the prediction model results, experiment design for the prediction model, and implementation of the prediction model for the facility.

Wang, Zhiwu: Dr. Wang has joined our team after the first proposal of this project. He replaced Zhaohui An to fulfill the further need for a WWTP expert. Responsible for verifying and validating the models from a WWTP expert perspective. He also ensured data quality, analyzed results, documented findings, and wrote reports.

Mao, Wan Yi: Responsible for the development of the protection model, data collection, data preprocessing, testing models, verification and validation of the results, and documenting them as well as visualization of these results and creating related graphs.

Works Cited

1. Adepu, S. S. (2016). Introducing Cyber Security at the Design Stage of Public Infrastructures: A Procedure and Case Study. *Complex Systems Design&Management Asia, 426*, 75.
2. Andrew Ilyas and Logan Engstrom and Anish Athalye and Jessy Lin. (2018). Black-box Adversarial Attacks with Limited Queries and Information. In *Proceedings of the 35th International Conference on Machine Learning, {ICML} 2018.*
3. Bergal, J. (2021, March 10). *Florida Hack Exposes Danger to Water Systems*. Retrieved from
4. PEW: https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2021/03/10/florida-hack-exposes-danger-to-water-systems
5. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
6. CERT, K. I. (2019). *Threat landscape for industrial automation systems. Vulnerabilities identified in 2019.* Kaspersky ICS CERT.
7. Collier, K. (2021, June 17). *50,000 security disasters waiting to happen: The problem of America's water supplies*. Retrieved from NBC News Digital: https://www.nbcnews.com/tech/security/hacker-tried-poison-calif-water-supply-was-easy-entering-password-rcna1206
8. Da Costa, J. F. P., Alonso, H., & Roque, L. (2009). A weighted principal component analysis and its application to gene expression data. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 8(1), 246-252.
9. Fu, R., Zhang, Z., & Li, L. (2016, November). Using LSTM and GRU neural network methods for traffic flow prediction. In 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC) (pp. 324-328). IEEE
10. Goh, Jonathan, and Adepu, Sridhar and Junejo, Khurum Nazir and Mathur, Aditya. (2016). A dataset to support research in the design of secure water treatment systems. In International conference on critical information infrastructures security. Springer.
11. Hassanzadeh, Amin and Rasekh, Amin and Galelli, Stefano and Aghashahi, Mohsen and Taormina, Riccardo and Ostfeld, Avi and Banks, M Katherine. (2020). A review of cybersecurity incidents in the water sector. Journal of Environmental Engineering.
12. Hingston, P. F., Barone, L. C., & Michalewicz, Z. (Eds.). (2008). Design by evolution: advances in evolutionary design. Springer Science & Business Media.
13. Kalehbasti, P. R., Lepech, M. D., & Criddle, C. S. (2022). Integrated Design and Optimization of Water-Energy Nexus: Combining Wastewater Treatment and Energy System. Frontiers in Sustainable Cities, 40.
14. Laso, Pedro Merino and Brosset, David and Puentes, John. (2017). Dataset of anomalies and malicious acts in a cyber-physical subsystem. Elsevier.
15. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

16. Mao, W.-Y., Yardimci, M., Nguyen, M., Sobien, D., Freeman, L., Batarseh, F. A., Rahman, A., & Fordham, V. (2022). Trustworthy AI Solutions for Cyberbiosecurity Challenges in Water Supply Systems. FLAIRS.
17. Peña, D., & Sánchez, I. (2007). Measuring the advantages of multivariate vs. univariate forecasts. Journal of Time Series Analysis, 28(6), 886-909
18. Sikder, Md. N. K., Nguyen, M. B. T. , Chandrasekaran , J., Anuga, A., & Batarseh, F. A., (2022). DeepWater: Attack Detection in Water Distribution Systems Using Deep Learning. Submitted to International Journal of Information Security
19. Sikder, M. N. K., & Batarseh, F. A. (2021). Outlier Detection using AI: A Survey. arXiv preprint arXiv:2112.00588.
20. Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. Handbook of statistics, 24, 303-329
21. S. Gurrapu, F. A. Batarseh, P. Wang, M. N. Kabir Sikder, N. Gorentala and M. Gopinath, "DeepAg: Deep Learning Approach for Measuring the Effects of Outlier Events on Agricultural Production and Policy," 2021 IEEE Symposium Series on Computational Intelligence (SSCI), 2021, pp. 1-8, doi: 10.1109/SSCI50451.2021.9659921.
22. Tuptuk, Nilufer and Hazell, Peter and Watson, Jeremy and Hailes, Stephen. (2021). A systematic review of the state of cyber-security in water systems. Multidisciplinary Digital Publishing Institute.
23. Wise, S., Braden, J., Ghalayini, D., Grant, J., Kloss, C., MacMullan, E., ... & Yu, C. (2010). Integrating valuation methods to recognize green infrastructure's multiple benefits. In Low impact development 2010: Redefining water in the city (pp. 1123-1143).
24. Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. IET Intelligent Transport Systems, 11(2), 68-75.

IWS Challenge Project Timeline

Outline	Task Start	Task End	Batarseh	Yardimci (Protection/Optimization)	Suzuki	Sikder (Prediction)	Wang	Mao (Protection)
Challenge Planning	4/27/2022	5/3/2022	Project planning Plan documentation	Project planning, and writing Plan documentation	Data management	Prediction model diagram Hypothesis behind the model	Project plan writing Water plant details Plan terminology editorial	Protection model diagram Hypothesis behind the model
	5/4/2022	5/10/2022	Editorial check Team management Proof reading	Adding more materials to plan Improvements on the document Writing for optimization model	Editorial check	Writing for prediction model.	Editorial on other sections Checking terminology consistency	Writing for protection model.
May 16, 2022 Challenge Plan	5/11/2022	5/16/2022	Submission	Formatting the document design	Security check	Formatting the document design	Formatting the document design	Formatting the document design
Initial preperation for datasets First models development	5/17/2022	5/24/2022	Data evaluation	Data pre-procesing	Models evaluation	Data Visualization and pre-processing	Data quality evaluation	Data Pre-Processing
	5/25/2022	5/31/2022	Models evaluation	Development and test environment preparation				Development Models
	6/1/2022	6/7/2022		Reinforcement modeling		Model design and development	Data selection for optimization modeling	Compare the models result and justify the result
	6/8/2022	6/14/2022		Reinforcement Learning model training				
First tests Model improvements	6/15/2022	6/21/2022	Verification and validation	Testing and evaluating of the first models		Model testing and evaluation	Provide environmental support for modeling parts	Assurance metrics to evaluate the model
	6/22/2022	6/28/2022		Wring of the model details and first results		Adding assurance methods		
Model finalized Verification/Validation Results reported/documented	6/29/2022	7/5/2022	Document structure and editorial	Finalizing model	On-site system testing and deployment	Finalizing model	Analyzing the model outputs of optimization	Finalizing model
	7/6/2022	7/12/2022		Assurance metrics to be reported		Experiment and result Share the outcome for chemical optimization	Writing the draft of results and discussion section	Model visualization
	7/13/2022	7/19/2022	Results evaluation	Writing and documentation Model visuals to be created		Visualization of the outcome	Further analysis of results	Result visualization
Visuals Graphics Documentation of the results	7/20/2022	7/26/2022		Visuals will be refined Visuals will be added and writing		Further improvement by optimization Start writing the results Share the outcome for chemical optimization	Finalizing the results parts	Writing document of model and outcome
	7/27/2022	8/2/2022	Evaluating test results Editing and documentation	Document improvements Reading and iteration of the document	Results evaluation	Visualization of the final outcome Start writing the results Share the outcome for chemical optimization		Future discussion
Documention finalized	8/3/2022	8/9/2022		Submition refinement, reformatting, proof reading	Cross evaluation	Finaling the report	Finaling the report	Finaling the report
August 15, 2022 Challenge Solution	8/10/2022	8/15/2022	Proof reading and submission.	Formatting the document design.	Security check	Formatting the document design	Formatting the document design	Formatting the document design