Software Requirement Specification Document for Aqua Shield

Mohamed Hossam, Hanya Yasser, Youssef Mahmoud, Mostafa Soliman
Supervised by: Dr. Fatma Helmy, Eng. Yasmin Kandil

May 1, 2024

Table 1: Document version history

Version	Date	Reason for Change
1.0	14-Jan-2024	SRS First version's specifications are defined.

Github:https://github.com/Soli07/Cyber-Attack-Detection-on-Water-Treatment-System

Figure 1: Github

Contents

1 Introduction **3**
- 1.1 Purpose of this document . . . 3
- 1.2 Scope of this document . . . 3
- 1.3 System Overview . . . 4
- 1.4 System Scope . . . 5
- 1.5 Business Context . . . 6

2 Similar Systems **6**
- 2.1 Academic . . . 6
- 2.2 Business Applications . . . 11

3 System Description **12**
- 3.1 User Problem Statement . . . 12
- 3.2 Objectives . . . 12
- 3.3 User Characteristics . . . 13
- 3.4 System Context . . . 13

4 Functional Requirements **14**
- 4.1 System Functions . . . 14
- 4.2 Detailed Functional Specifications . . . 16

5 Design Constraints **19**
- 5.1 Standards Compliance . . . 19
- 5.2 Hardware Limitations . . . 19
- 5.3 Other Constraints as appropriate . . . 19

6 Non-functional Requirements **19**
- 6.1 Security . . . 19
- 6.2 Reliability . . . 20
- 6.3 Maintainability . . . 20
- 6.4 Availability . . . 20
- 6.5 Usability . . . 20

7 Data Design **20**

8 Preliminary Object-Oriented Domain Analysis **22**

9 Operational Scenarios **23**

10 Project Plan **23**

11 Appendices **24**
- 11.1 Abbreviations . . . 24
- 11.2 Supportive Documents . . . 24

Abstract

Drinking water quality has long been a source of worry. In order to ensure that the water delivered by utilities is safe for human use, independent laboratories have traditionally examined the water. Water quality has been given consideration from a security perspective since it is a component of vital infrastructure. Central to the monitoring and control of water treatment facilities are SCADA (Supervisory Control and Data Acquisition) systems, which manage and regulate various processes, and PLCs (Programmable Logic Controllers), responsible for executing specific tasks based on the SCADA system's commands. Analysis of sensor data collected at various sites and the setting of alarms when variations in quality indicators point to abnormalities that may arise from spoofing on the communication between SCADA servers and PLCS through network traffic leading to alterations in sensors and actuators readings.To further enhance cybersecurity measures against such threats, this study presents a strong cybersecurity solution by utilizing Long Short-Term Memory (LSTM) networks, a specific deep learning technology for time series data processing for detecting attacks, and performing clustering to classify the unlabeled attacks based on their characteristics. This approach aims to group the attacks, enabling detection of each specific type of threat within water treatment systems. By developing a simulation model of the six stages of the water treatment process, to achieve a comprehensive understanding and assess the overall functionality and performance of the system under various operational scenarios. The system will integrate data analysis tools for visualizing sensor readings. Through the generation of reports and graphs, the system will facilitate a clearer understanding of the data, to provide the user with the recommended action.

1 Introduction

1.1 Purpose of this document

The purpose of this document is to analyze and break down the software requirements for developing a web application for detecting attacks, providing reports, simulation of the system, and providing recommended action. It will state and describe the functional requirements alongside the expected behavior that the application should take. The document also outlines the needs and constraints of the web application. The target audience for this document are the stakeholders for the project and any developer wanting to continue to work on it.

1.2 Scope of this document

The document's scope is to tackle similar systems to AquaShield academically and business-wise, moreover, it illustrates the system overview, context, and scope, also the objectives of the AquaShield Web application, and the user characteristics are discussed too. Furthermore, this document goes through the functional and non-functional requirements of the AquaShield web application and system, the software and hardware limitations, the data design, and the object-oriented class diagram. Finally, this document also covers the operational scenarios of the system and the exact timeline of how this web application will be developed.

1.3 System Overview

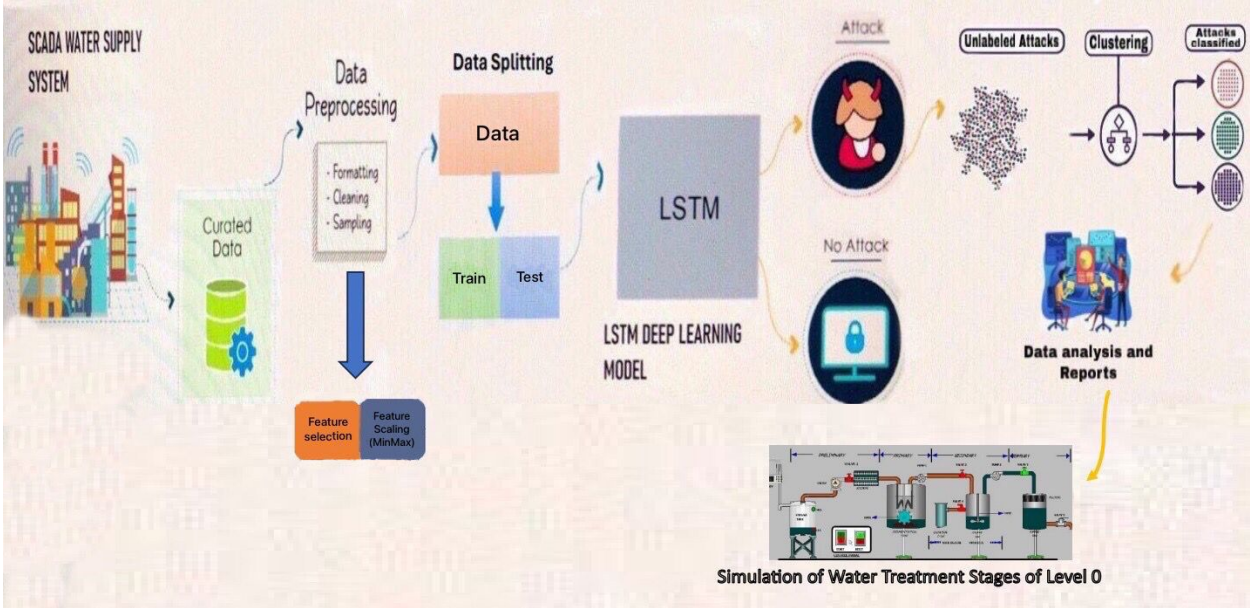

Figure 2: System Overview Diagram

Figure 2 shows the System overview of the proposed project. It contains the following steps:

- Preparing the dataset: The SWAT dataset collected from the Supervisory Control and Data Acquisition (SCADA) system is a comprehensive, curated dataset designed to research and evaluate cybersecurity in the context of facilities Water treatment. This dataset is specifically designed to simulate and analyze cyber threats, anomalies, and security measures in SCADA environments of water treatment plants.
- Data preprocessing is essential to clean, transform, and organize raw data into a format suitable for analysis. It plays an important role in ensuring the accuracy, reliability, and efficiency of subsequent analysis and modeling tasks.
- Feature selection techniques are crucial to identify a targeted subset of sensors specifically designed to enhance the accuracy of an LSTM(Long Short-Term Memory) model. This approach aims to refine the analysis process by focusing on the most relevant and informative data points, leading to improved model performance.
- Data Splitting must be performed on the dataset to ensure that the model learns from a substantial portion of the data (Training of normal data, ex: 80%) and ensures that the model generalizes well to new and unseen data (Test of normal data with some abnormal data, ex:20%).
- Long Short Term Memory deep learning model is used for cyber attack detection in water treatment by analyzing time-series data from sensors in the water treatment system, it learns the normal patterns and temporal dependency of the system's behavior then the model assesses new data and classifies it either "Safe" or "Attack".

- To enhance threat detection within the water treatment system, an integral component involves performing clustering techniques to classify unlabeled attacks based on their distinct characteristics. By grouping these attacks, this approach aims to facilitate the identification of specific threat types, contributing to a comprehensive security framework for the system.

- To enhance data interpretation and emphasize identified threats, the system will integrate data analysis tools for visualizing sensor readings. Through the generation of reports and graphs by these tools, the system will facilitate a clearer understanding of the data, aiding in the detection and highlighting of potential attacks within the water treatment system.

- Developing a simulation model encompassing the six stages of the water treatment process, the project utilizes simulation techniques to achieve a comprehensive understanding and assess the overall functionality and performance of the system under various operational scenarios.

1.4 System Scope

- Using a Real-life dataset (SWaT: Secure Water Treatment) created for research and testing purposes in the field of cybersecurity for water treatment systems.

- Ensure public health safety by detecting anomalies in the water treatment process.

- Applying preprocessing on operational data by cleaning, sampling, and scaling the data for better accuracy, and employing feature selection to curate a subset of sensors that are precisely targeted for analysis.

- The project involves utilizing a deep learning model based on LSTM (Long Short-Term Memory) to detect attacks within the water treatment system, utilizing a labeled dataset that categorizes instances as either 'attack' or 'normal' based on sensor readings.

- To optimize the approach, the project will utilize clustering techniques to classify unlabeled attacks based on distinct characteristics, enabling customized actions for each identified attack within the water treatment system.

- To facilitate comprehension and emphasize the detected attacks, the project will use data analysis tools to visualize sensor readings. This visualization will include reports and graphs generated by the data analysis tools, aiding in the easy interpretation of the data and highlighting identified attacks within the water treatment system.

- To simulate the six-stage process, the project will utilize simulation techniques, allowing for a comprehensive understanding and assessment of the entire water treatment system's functionalities and performance within diverse operational scenarios.

1.5 Business Context

Water treatment facilities are critical infrastructures that ensure that the water delivered to the public is safe and uninfected or contaminated in any way. Cyber attacks on these facilities can lead to catastrophic consequences like mass poisoning or pumps overflowing. By developing detection methods, these types of risks can be prevented and ensure that public health is safe as well as the continuity of operations of businesses that rely on secure water supply. Additionally, detecting these attacks can prevent huge economic loss due to system failure, maintenance, and potential legal consequences, and ensure that treatment facilities operate efficiently.

2 Similar Systems

2.1 Academic

1. Y. Zhang, B. Li and X. Zhang.[1]:

   Industrial control systems, or ICS, are widely utilized to regulate the functioning of production equipment in important sectors including industry, energy, transportation, and water conservation. These systems produce a lot of time series data since actuators and sensors are widely used in Industrial control systems. Currently, the lack of anomaly labels, high data volatility, and ultra-low inference time requirements pose serious hurdles to anomaly identification in ICS. In order to identify unusual behaviors in ICS, the research suggests a deep learning-based multivariate anomaly detection technique that combines Transform and GAN(Generative Adversarial Networks). The model presented in this research can outperform state-of-the-art baseline approaches in the SWaT(Secure Water Treatment) and WADI(Water Distribution) datasets gathered from industrial control systems, according to experimental results.

2. S. Ayas, A. K. Şahin, B. Özgenç, B. Çavdar, R. Ö. Doğan and M. Ş. Ayas [2]:

   Because of the possible risk to public health, water treatment facilities are among the (ICS) (Industrial Control System) where it is crucial to identify abnormalities quickly and precisely. Recent years have seen the successful application of machine learning models in the anomaly detection process thanks to advancements in computer science. This research proposes a hybrid CNN-LSTM network model for anomaly detection on the water system. The efficacy of the suggested model in identifying various forms of attacks is examined on the publicly available SWaT (Secure Water Treatment) dataset through the utilization of a statistical window-based anomaly detection methodology. The suggested model has been found to have a recall, precision, and Fl-score of 0.994, 0.973, and 0.983. These results demonstrate the model's potential for effective anomaly detection in water treatment systems.

3. Muneza, Assoumer Redempta Manzi. [3]:

   This paper talks about the usage of unsupervised machine learning techniques for detecting any anomaly in Cyber Physical Systems (CPSs) caused by cyber attacks. It specifically focuses on the use of Isolation Forest and One-Class SVM for detecting anomalies in CPSs, using data collected from the SWAT testbed. The paper covers various topics such as data

collection, data pre-processing, research methodologies, and interpretation of results. The aim of the paper is to provide insights into the potential benefits and limitations of using unsupervised machine learning for detecting anomalies in CPSs, as well as to compare the performance of the two machine learning methods used in the study. The paper concludes with recommendations for future research in this area.

Figure 3: The Secure Water Treatment (SWaT) testbed.

Figure 3 shows the physical system of the SWaT testbed.

4. Edwin Franco Myloth Josephla, Toe, Teoh Teik,and Lim Han Yi.[4]:

   This document focuses on cyber attacks detection in the processes of water treatment, specifically using real data from the SWaT(Secure Water Treatment) testbed. The SWaT testbed was developed for education and research on cybersecurity and simulated real water treatment plants behavior. The attacks in the process of SWaT involve manipulating sensor data and actuator signals at the communication level. The paper discusses cyber attacks challenges on critical infrastructure like water systems and highlights the importance of developing reliable detection systems. Most of the available attack detection systems for SWaT process utilize machine learning techniques.

5. Alabugin, Sergei K., and Alexander N. Sokolov. [5]:

   This document discusses the application of Generative Adversarial Networks(GANs) for detecting anomalies in Industrial Control System (ICS). It highlights the challenges of detecting cyber-attacks on ICS and the potential catastrophic consequences. According to the article, the task of industrial process anomaly detection may be used to reframe the intrusion detection in ICS problem. It also mentions the generation of attack examples using GANs to create a more stable dataset for testing intrusion detection systems. The paper concludes that while the results obtained with GANs may be inferior to existing methods, their advantage lies in not requiring examples of anomalous system behavior during the training stage, making them more efficient in practical use.

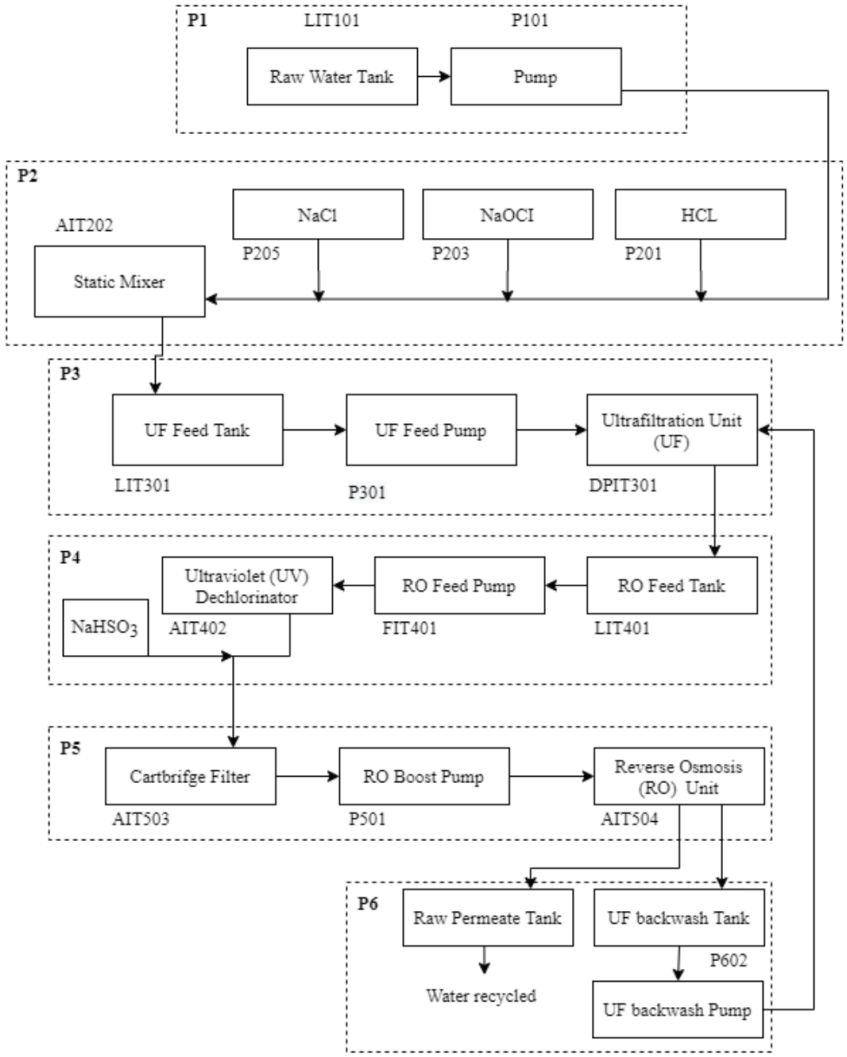

Figure 4: SWaT testbed structure

Figure 4 shows the six main processes corresponding to the physical and control components of the water treatment facility.

6. M. R.,Gauthama Raman, Nivethitha Somu, and Aditya P. Mathur. [6]:

   An essential component of any critical infrastructure is Supervisory Control and Data Acquisition (SCADA) systems. Since these systems are network linked for remote controlling and monitoring, malicious actors may be able to compromise them. Such incursions may cause the underlying physical process to behave abnormally. In order to identify abnormalities resulting from a cyberattack, this paper provides an anomaly detector based on a Probabilistic Neural Network (PNN). The dataset from SWaT (Secure Water Treatment), an operating water treatment testbed, was used for the experimental validation. An analysis was conducted to determine how the smoothening parameter affected the Probabilistic Neural Network (PNN) anomaly detector's performance. In comparison to numerous rival detectors, experimental evaluations show the relevance of the Probabilistic Neural Network (PNN) anomaly detector in terms of detection, F-score, precision, and false alarm rate.

Attack ID	Type	Target	Duration (Secs.)	Expected impact	Unexpected impact
1	SSSP	MV-101	539	Tank overflow	
2	SSSP	LIT-101	300	Tank Underflow; damage P-101	
3	SSSP	MV-504	300	Halt RO shut down sequence; reduce life of RO	
4	SSSP	DPIT-301	500	Backwash process is started again and again; normal operation stops; Decrease in water level of tank 401. Increase in water level of tank 301	
5	SSSP	AIT-504	200	RO shut down sequence starts after 30 min. Water should go to drain	RO did not shut down; water does not drain
6	SSMP	MV-101, LIT-101	501	Tank overflow	
7	MSMP	P-602, DIT-301, MV-302	251	System freeze	
8	MSSP	P-101, LIT-301	251	Stop inflow of tank T-401	
9	MSSP	AIT-402, AIT-502	251	Water enters the drain due to overdosing	Water does not drain
10	SSSP	LIT-302	501	Tank overflow	Rate of decrease of water level reduced after 1:33:25 PM

Figure 5: Attacks considered in experiments.

Figure 5 shows the type of attacks, targeted sensors, duration of attack, and expected/unexpected impact upon attacks.

7. Al-Dhaheri, Mohammed, , Dina Mikhaylenko and Ping Zhang . [7] :

   This study delves into identifying cyber threats within critical infrastructure—specifically, the water treatment procedures. It utilizes authentic data from SWaT (Secure Water Treatment) housed at the iTrust Centre. Over 30 distinct attacks occurred within the communication network connecting the sensors/actuators and programmable logic controller (PLC). Upon grasping the fundamental operational mechanisms of the SWaT system, the researchers methodically crafted an attack detection system. This system comprises three core components: a model-based and a data-driven monitoring unit and a value limit safety and check rule unit. To handle fluctuations in certain water chemical properties,(RPCA) in the data-driven segment. The study illustrates how effectively segmenting the SWaT process enables the integration of data-driven monitoring and model-based. A comparative study shows that the suggested attack detection system is superior to other currently in use systems created for the secure water treatment process.

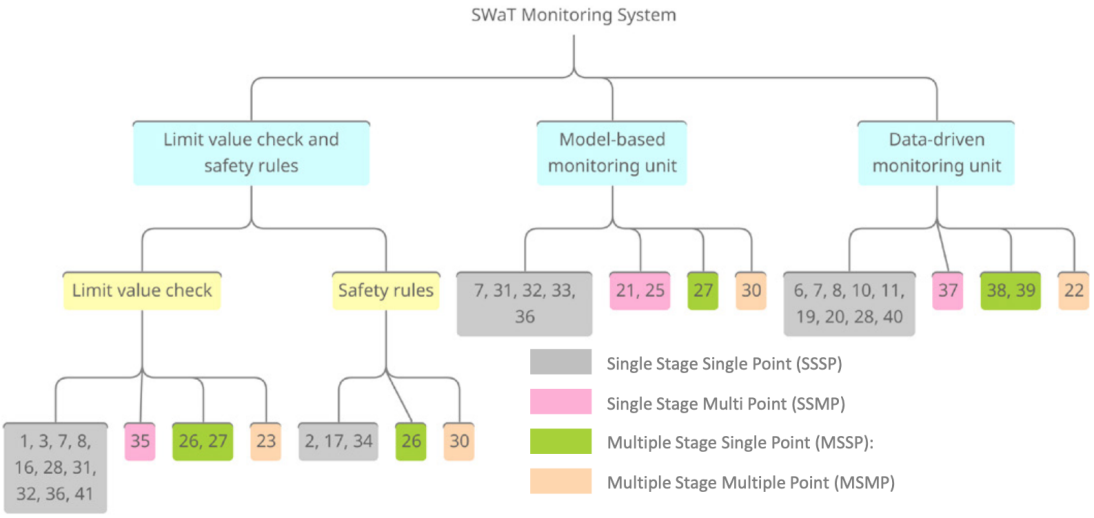

Figure 6: Attack monitoring system Overview

8. Semwal, Prabhat, and Akansha Handa. [8]:

   Cyber-Physical Systems are those in which technology components and computation govern the physical processes. These CPS data have been accurately analyzed for attack detection using machine learning techniques. This paper used four different supervised machine learning algorithms (KNN), (SVM), (DT), and (RF) to build models for cyber attack detection on a CPS water treatment facility and the results are compared. The results show that the Decision Tree model performs a higher accuracy of 99.9%.

Model	Accuracy	TPR	FPR
KNN	99.65	99.9	0.008
SVM	98.70	99.03	0.018
DT	99.9	99.0	0.006
RF	96.3	98.6	0.013

Figure 7: Obtained accuracy, TPR and FPR values

   Figure 7 shows all the models trained on the processing of the dataset with their results.

9. Perales Gómez, Ángel Luis, et al. [9]:

   In order to address the issue of attack detection in Industrial Control Systems (ICS), this study presents the MADICS (technique for Anomaly Detection in Industrial Control Systems) technique, which is intended to identify cyberattacks in ICS settings. The approach models the behavior of ICS using deep learning techniques and is based on a semi-supervised anomaly detection approach. The five primary processes of the MADICS methodology—preprocessing the dataset, feature extraction, feature filtering, choosing and training the best model, and testing model performance—are described in the article. Additionally, the authors offer a thorough justification of how MADICS was implemented using the SWaT

dataset. The tests show that MADICS obtained a recall of 0.750, an F1-score of 0.851, and an accuracy of 0.984, which is higher than usual, proving that the proposed methodology is applicable for use in real scenarios of ICS.

10. Xue, Feng, et al. [10]:

    The authors discuss the use of deep learning for anomaly detection in industrial systems, where data used for detecting anomalies are primarily time-series sensor measurements. The paper presents a case study on the Secure Water Treatment (SWaT) testbed dataset, which uses a self-supervised learning approach to train a deep neural network autoregressive model. and discusses the challenges of employing deep learning for industrial system detecting anomalies, including proper problem formulation for neural network training, data preprocessing, and choosing a preferred backbone neural network architecture. The paper also explains that deep anomaly detection methods have two broad settings: indirect (2-step) and direct (1-step), and discusses the challenges of employing deep learning for industrial system detecting anomalies.

Attack #	Start Time	End Time	Attack Point	Start State	Attack	Actual Change
1	28/12/2015 10:29:14	10:44:53	MV-101	MV-101 is closed	Open MV-101	Yes
2	28/12/2015 10:51:08	10:58:30	P-102	P-101 is on where as P-102 is off	Turn on P-102	Yes
3	28/12/2015 11:22:00	11:28:22	LIT-101	Water level between L and H	Increase by 1 mm every second	No
4	28/12/2015 11:47:39	11:54:08	MV-504	MV-504 is closed	Open MV-504	Yes
5	28/12/2015 11:58:20		No Physical Impact Attack			

Figure 8: Sample attacks from SWaT dataset

Figure 8 shows the start time, end time, targeted sensors, start state of the sensors, and what the attack is.

2.2 Business Applications

A cutting-edge feature called Bitdefender App Anomaly Detection has been added to the Bitdefender Malware Scanner to give an extra degree of security by constantly monitoring, identifying, and alerting the user to any unusual activity.

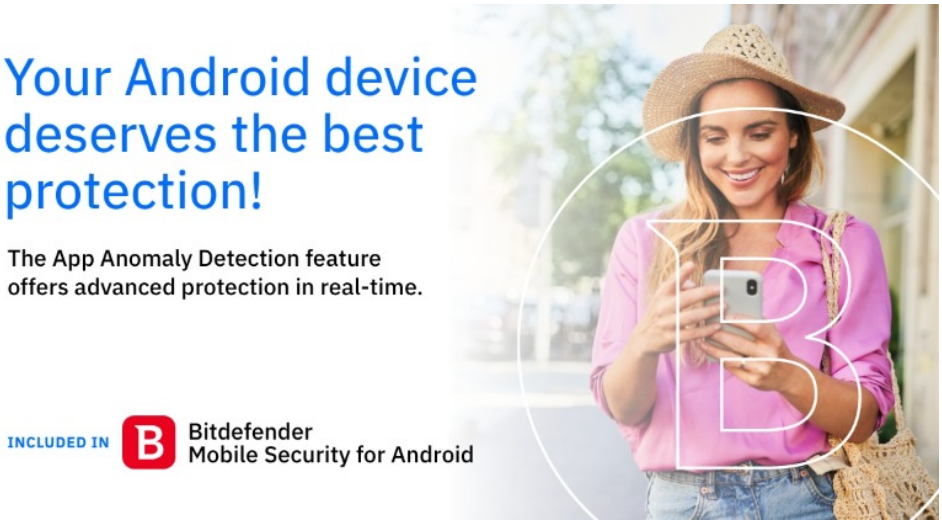

Figure 9: Bitdefender App

3 System Description

3.1 User Problem Statement

The project aims to address the specific challenge of minimizing cyber-physical attacks on water treatment processes, where attackers manipulate sensor readings and compromise Programmable Logic Controllers (PLCs) responsible for monitoring chemical levels or pH levels in the water and also Supervisory Control and Data Acquisition (SCADA) servers that potentially lead to water contamination or other adverse effects, introducing harmful chemicals into the water supply that pose a threat to public health.

3.2 Objectives

The objectives of this cybersecurity project for water treatment processes are:

- Enable detection of cyber-physical attacks aiming to manipulate sensor readings, introduce harmful chemicals, or alter pH levels in the water supply.
- Enhance the security infrastructure to safeguard public health by promptly identifying and minimizing potential threats arising from cyber-physical attacks on water treatment systems.
- Develop and implement an unsupervised clustering framework specifically designed to classify unlabeled attacks based on their characteristics, enabling more informed and targeted responses for each identified cluster
- To simulate the six-stage process, Enabling a thorough evaluation of the functionality and performance of the water treatment system across various operational situations.
- To visually highlight detected attacks, we'll use data analysis tools to create easy-to-understand graphs and reports of sensor readings in the water treatment system.

3.3 User Characteristics

- Since English will be the user interface's default language, the user must have a basic understanding of the language.
- The user must have some basic knowledge of how to use a computer.
- A fundamental comprehension of SCADA Server readings is required of the user.
- The admin should have some basic knowledge of Administrator roles (CRUD operations)

3.4 System Context

As shown in figure 10, The User uploads time series data to the system. The system takes unlabeled data and utilizes the LSTM model to detect anomalies then uses clustering to classify attacks and group them based on their characteristics. The system then provides the user with a report and simulation of the data.

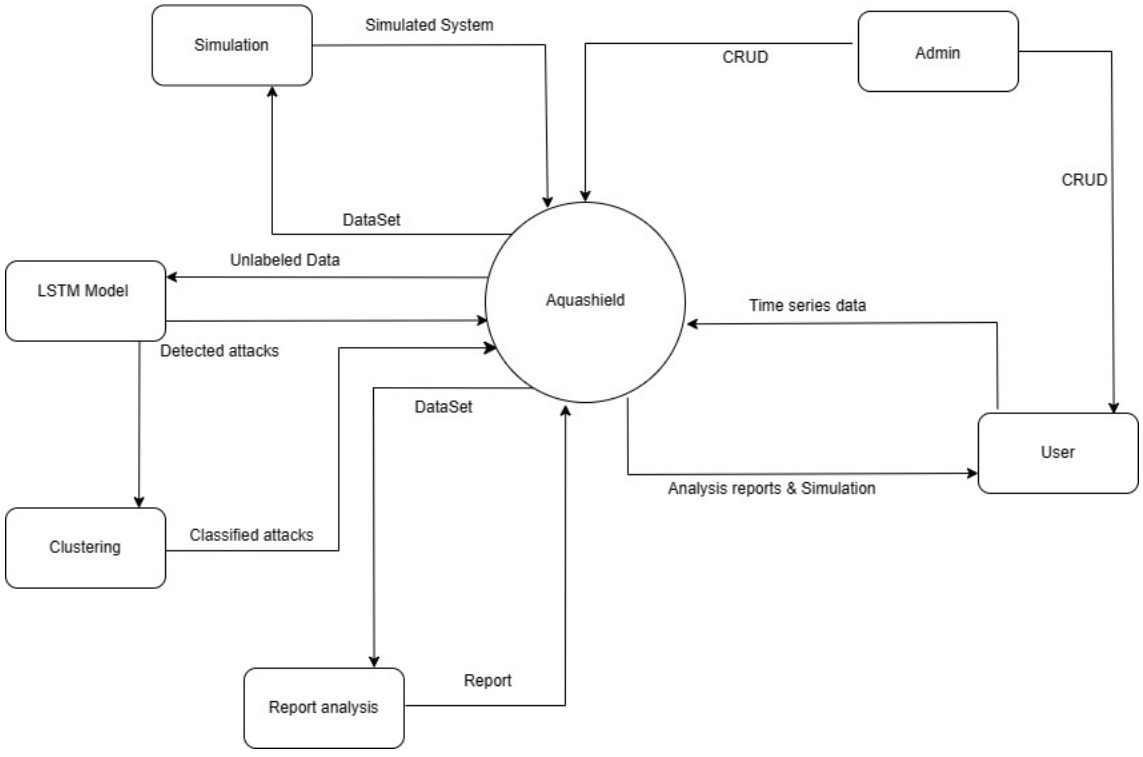

Figure 10: Context Diagram

4 Functional Requirements

4.1 System Functions

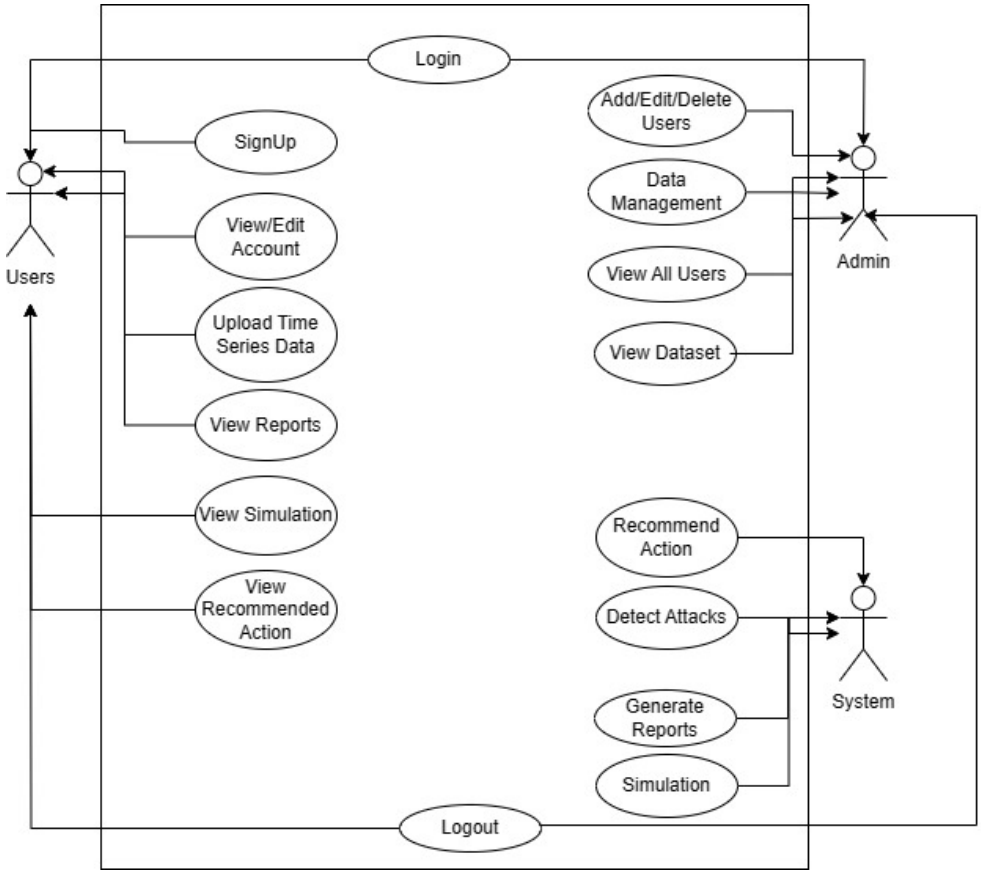

Figure 11: User Use Case

1. Admin shall login to the system.
2. Admin shall add users.
3. Admin shall edit users.
4. Admin shall delete users.
5. Admin shall manage data.
6. Admin shall search users.
7. Admin shall view all users.
8. Admin shall accept payments.
9. Admin shall decline payments.

10. Admin shall view all feedbacks.
11. Admin shall send a reply to the user's requests.
12. User shall sign up for a new account.
13. User shall sign in to the system.
14. User shall view account.
15. User shall send contact us form.
16. User shall provide payment details.
17. User shall edit account details.
18. User shall upload their time series data.
19. User shall view the generated reports.
20. User shall test their data using simulation.
21. User shall view the whole simulation.
22. User shall view the recommended actions.
23. User shall log out from the system.
24. User shall provide a feedback.
25. System acknowledges that data is uploaded.
26. System shall ask users to re-upload the data if an error happens.
27. System shall allow users to edit their data.
28. System shall preprocess the data.
29. System shall detect attacks.
30. System shall classify attacks.
31. System shall recommend actions.
32. System shall analyze data.
33. System shall generate reports.
34. System shall simulate process.

4.2 Detailed Functional Specifications

Table 2: Login Function Description

Name	Login
Code	Fn1
Priority	Extreme
Critical	essential for the user to use all services on the website.
Description	It searches in the database for the entered username and password to whether it's valid or not.
Input	Email and Password.
Output	Boolean (found or not)
Pre-condition	User must already have an account.
Post-condition	If valid, User is redirected to the home page. If not, notify the user that the entered email or password are incorrect.
Dependency	Fn2
Risk	Interrupted connection, lost data, and the requirement for the user to fill out the form again

Table 3: SignUp Function Description

Name	SignUp
Code	Fn2
Priority	Extreme
Critical	Data entered must be checked if it has any error.
Description	Function shall allow users to create new accounts.
Input	Full Name, Email, Password, Mobile Number.
Output	Boolean: True if account created otherwise False
Pre-condition	None
Post-condition	A successful account created notification is sent to the user and redirected to the home page
Dependency	None
Risk	Interrupted connection, lost data, and the requirement for the user to fill out the form again

Table 4: Edit Account Function Description

Name	Edit Account
Code	Fn3
Priority	Medium
Critical	None
Description	Allows user to edit account
Input	Email, Full Name, Password, Mobile Number.
Output	Boolean (data updated or not)
Pre-condition	User must be logged in.
Post-condition	If the account was updated successfully, redirect the user back to the home page. If not, the user will remain on the same page of editing the account.
Dependency	Fn1
Risk	Connection interruption, the data is lost and the user needs to re-fill the field

Table 5: Upload Time series data Function Description

Name	Upload Time Series Data
Code	Fn4
Priority	High
Critical	None
Description	User will upload time series data in order to be able to use the services provided.
Input	Dataset
Output	Sensors and Actuators readings
Pre-condition	User asked to upload the dataset
Post-condition	if Dataset uploaded in an excel form, readings will be viewed.
Dependency	Fn1
Risk	Connection interruption, the data is lost and the user needs to re-fill the field

Table 6: Detect Attacks Function Description

Name	Detect Attacks
Code	Fn5
Priority	Extreme
Critical	None
Description	Function will detect any attack that occurred using the LSTM model.
Input	The Dataset uploaded by the user.
Output	Attacks detected by the model.
Pre-condition	User must have uploaded a dataset
Post-condition	The attack detected
Dependency	Fn4
Risk	None

Table 7: View Reports Function Description

Name	View Reports
Code	Fn6
Priority	Medium
Critical	None
Description	Allows user to view reports about detected attacks
Input	The Uploaded Dataset
Output	Analysis Reports viewed
Pre-condition	User must have uploaded a dataset
Post-condition	Clear findings and actionable recommendations
Dependency	Fn4
Risk	None

Table 8: View Simulation Function Description

Name	View Simulation
Code	Fn7
Priority	High
Critical	To be able to see a simulation of the water treatment process.
Description	Function creates a simulation of the six-phase water treatment process.
Input	The Uploaded Dataset
Output	Simulation of the whole process.
Pre-condition	None
Post-condition	A Simulation is created.
Dependency	Fn4
Risk	None

Table 9: Recommend action Function Description

Name	Recommend Action
Code	Fn8
Priority	High
Critical	Essential for recommending actions to the detected attacks.
Description	Recommended actions by the system towards each and every attack.
Input	Detected Attacks
Output	Recommended action for each attack.
Pre-condition	A list of attacks must be provided.
Post-condition	A list Of all possible actions for any possible attack that might occur.
Dependency	Fn4
Risk	None.

Table 10: View Recommended Action Function Description

Name	View Recommended Action
Code	Fn9
Priority	Medium
Critical	User must view all recommendations in order to choose the relevant/suitable one.
Description	Function allows the user to view the list of recommendations.
Input	None
Output	A list of all attacks correspondingly to their recommended action.
Pre-condition	None
Post-condition	Actions are viewed.
Dependency	Fn1
Risk	None.

5 Design Constraints

5.1 Standards Compliance

- The web application must comply with industry-standard security protocols (e.g., SSL/TLS) for secure data transmission.
- Incorporation of security measures ensuring the secure handling and processing of sensitive information within the water treatment system.

5.2 Hardware Limitations

- Users accessing the web application require a device (e.g., laptop, smartphone, tablet) with a compatible web browser for accessing the application interface.
- Compatibility with standard hardware configurations for running the simulation tool and data analysis software (e.g., CPU, RAM requirements).

5.3 Other Constraints as appropriate

- The web application must be platform-independent and accessible across multiple web browsers (e.g., Chrome, Firefox, Safari) for user convenience.
- Compatibility with varying internet connection speeds to ensure usability in different network environments (e.g., 3G, 4G, Wi-Fi).

6 Non-functional Requirements

6.1 Security

Ensure data encryption, access control, and other security measures safeguard sensitive information and prevent unauthorized access or attacks.

6.2 Reliability

Design the system to be fault-tolerant, implement redundancy measures, and establish continuous monitoring to ensure consistent and dependable system performance, even during failures or attacks.

6.3 Maintainability

Build the system such that upgrades, modifications, and maintenance are simple to complete. ensuring that changes or improvements can be implemented without disrupting system functionalities.

6.4 Availability

Implement measures to ensure that the system remains available and operational, minimizing downtime or disruptions, and ensuring consistent access to critical functions and data.

6.5 Usability

Design the system with a user-friendly interface and intuitive functionalities, ensuring ease of use, learnability, and efficiency for system users and administrators.

7 Data Design

The dataset has the following characteristics listed: 11 days of nonstop operation, consisting of 4 days of attack scenarios and 7 days of routine operation; network traffic was gathered; all 51 sensors' and actuators' values were acquired; and data were labeled based on normal and abnormal behaviors.

Timestamp	FIT101	LIT101	MV101	P101	P102	AIT201	AIT202	AIT203	FIT201	MV201	P201	P202	P203
22/12/2015	2.470294	261.5804	2	2	1	244.3284	8.19008	306.101	2.471278	2	1	1	2
22/12/2015	2.457163	261.1879	2	2	1	244.3284	8.19008	306.101	2.468587	2	1	1	2
22/12/2015	2.439548	260.9131	2	2	1	244.3284	8.19008	306.101	2.467305	2	1	1	2
22/12/2015	2.428338	260.285	2	2	1	244.3284	8.19008	306.101	2.466536	2	1	1	2
22/12/2015	2.424815	259.8925	2	2	1	244.4245	8.19008	306.101	2.466536	2	1	1	2
22/12/2015	2.425456	260.0495	2	2	1	244.5847	8.19008	306.101	2.465127	2	1	1	2
22/12/2015	2.472857	260.2065	2	2	1	244.5847	8.19008	306.101	2.464742	2	1	1	2
22/12/2015	2.513532	260.5991	2	2	1	244.5847	8.19008	306.101	2.468331	2	1	1	2
22/12/2015	2.559972	261.0309	2	2	1	244.5847	8.19008	306.101	2.469612	2	1	1	2
22/12/2015	2.598085	261.1093	2	2	1	244.809	8.19008	306.101	2.470894	2	1	1	2
22/12/2015	2.630753	261.7766	2	2	1	244.809	8.19008	306.101	2.470894	2	1	1	2
22/12/2015	2.649329	261.7766	2	2	1	244.809	8.19008	306.101	2.472175	2	1	1	2
22/12/2015	2.654133	261.8944	2	2	1	244.8731	8.19008	306.101	2.474097	2	1	1	2
22/12/2015	2.646446	261.6589	2	2	1	244.8731	8.19008	306.101	2.474097	2	1	1	2
22/12/2015	2.625949	261.2664	2	2	1	245.0333	8.19008	305.8703	2.474097	2	1	1	2
22/12/2015	2.61602	260.8346	2	2	1	245.0333	8.19008	305.8703	2.474097	2	1	1	2
22/12/2015	2.609935	261.0309	2	2	1	245.0333	8.19008	305.8703	2.474097	2	1	1	2
22/12/2015	2.602889	261.1093	2	2	1	245.0333	8.19008	305.8703	2.474097	2	1	1	2
22/12/2015	2.587516	260.9916	2	2	1	245.0333	8.19008	305.8703	2.474097	2	1	1	2
22/12/2015	2.573103	261.3056	2	2	1	245.0333	8.19008	305.8703	2.473457	2	1	1	2
22/12/2015	2.556769	261.6589	2	2	1	245.4499	8.19008	305.8703	2.471663	2	1	1	2
22/12/2015	2.543958	261.9729	2	2	1	245.4499	8.19008	305.8703	2.471663	2	1	1	2
22/12/2015	2.519617	262.0514	2	2	1	245.4499	8.19008	305.8703	2.472559	2	1	1	2
22/12/2015	2.502002	262.3654	2	2	1	245.4499	8.1904	305.8703	2.474482	2	1	1	2
22/12/2015	2.486308	262.4832	2	2	1	245.4499	8.193604	305.8703	2.474482	2	1	1	2
22/12/2015	2.470294	262.2084	2	2	1	245.4499	8.193604	305.8703	2.474482	2	1	1	2
22/12/2015	2.452359	262.0121	2	2	1	245.4499	8.193604	305.6396	2.473457	2	1	1	2

Figure 12: DATASET EXAMPLE 1

?04	P205	P206	DPIT301	FIT301	LIT301	MV301	MV302	MV303	MV304	P301	P302	AIT401
1	2	1	20.79839	2.235275	327.4401	1	2	1	1	2	1	0
1	2	1	20.79839	2.234507	327.4401	1	2	1	1	2	1	0
1	2	1	20.8432	2.233354	327.4401	1	2	1	1	2	1	0
1	2	1	20.8432	2.233354	327.2799	1	2	1	1	2	1	0
1	2	1	20.8432	2.233354	327.1597	1	2	1	1	2	1	0
1	2	1	20.8432	2.235147	326.9194	1	2	1	1	2	1	0
1	2	1	20.8432	2.235147	326.7592	1	2	1	1	2	1	0
1	2	1	20.8368	2.235147	327.0396	1	2	1	1	2	1	0
1	2	1	20.824	2.233994	327.0796	1	2	1	1	2	1	0
1	2	1	20.80799	2.233994	326.9594	1	2	1	1	2	1	0
1	2	1	20.80799	2.233738	326.3586	1	2	1	1	2	1	0
1	2	1	20.80799	2.233738	326.1584	1	2	1	1	2	1	0
1	2	1	20.80799	2.233738	326.3586	1	2	1	1	2	1	0
1	2	1	20.80799	2.235147	326.4388	1	2	1	1	2	1	0
1	2	1	20.90403	2.235147	327.0396	1	2	1	1	2	1	0
1	2	1	20.92003	2.235147	327.1197	1	2	1	1	2	1	0
1	2	1	20.92003	2.235147	327.3199	1	2	1	1	2	1	0
1	2	1	20.86241	2.235147	327.8807	1	2	1	1	2	1	0
1	2	1	20.8272	2.235147	328.121	1	2	1	1	2	1	0
1	2	1	20.8208	2.235147	328.4815	1	2	1	1	2	1	0
1	2	1	20.8208	2.235147	328.4815	1	2	1	1	2	1	0
1	2	1	20.85281	2.235147	328.2411	1	2	1	1	2	1	0
1	2	1	20.85601	2.235147	328.4414	1	2	1	1	2	1	0
1	2	1	20.85601	2.235147	328.9221	1	2	1	1	2	1	0
1	2	1	20.79839	2.235147	328.842	1	2	1	1	2	1	0
1	2	1	20.77278	2.235147	328.9621	1	2	1	1	2	1	0

Figure 13: DATASET EXAMPLE 2

P401	P402	P403	P404	UV401	AIT501	AIT502	AIT503	AIT504	FIT501	FIT502	FIT503
1	1	1	1	1	7.308575	176.8265	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437
1	1	1	1	1	7.308575	176.8008	263.6504	12.68905	0.00166624	0.00140899	0.00166437

Figure 14: DATASET EXAMPLE 3

8 Preliminary Object-Oriented Domain Analysis

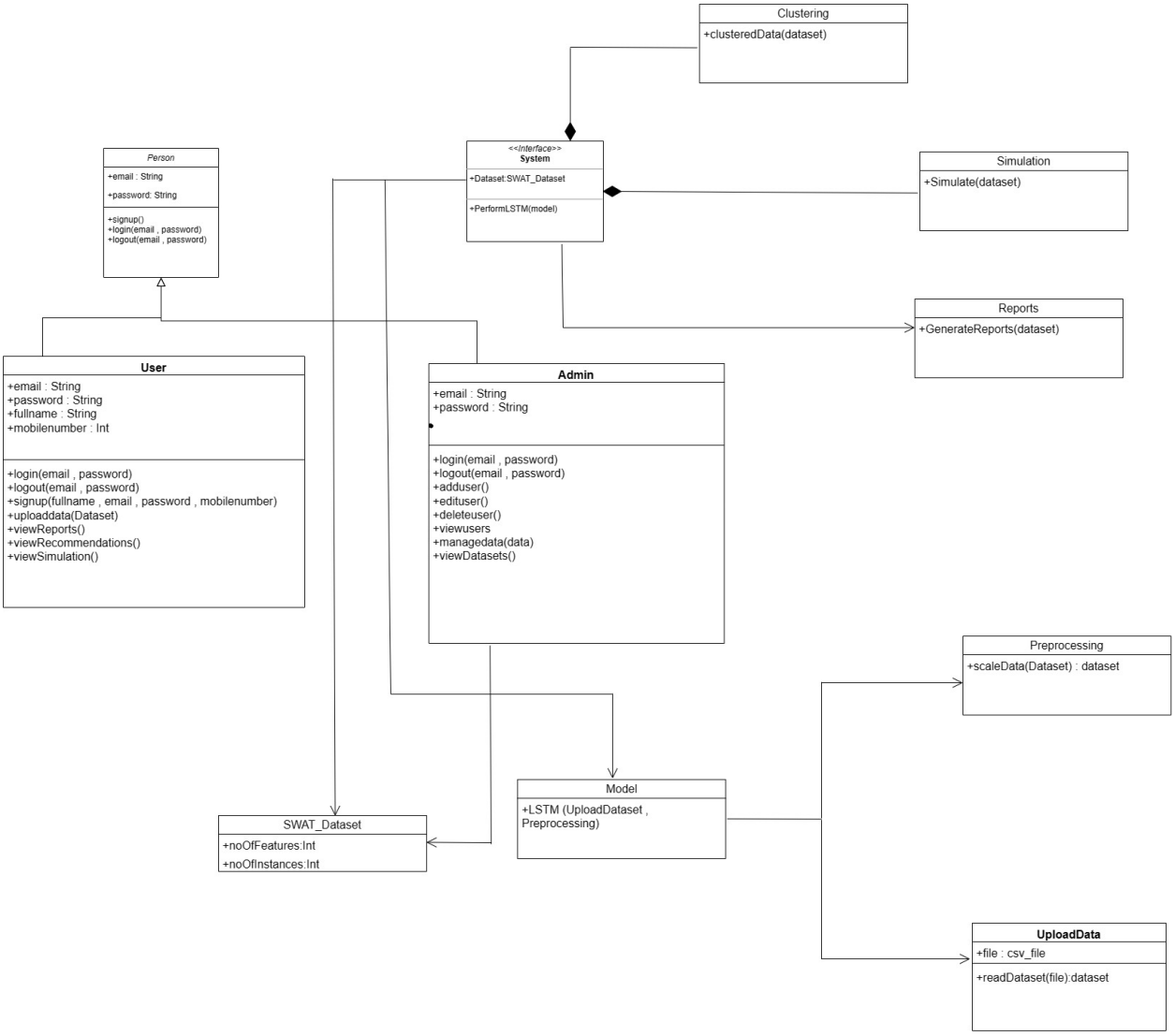

Figure 15: Class Diagram

9 Operational Scenarios

1. **Scenario 1:** The user logs into the web application and uploads data of sensor readings for attack detection. The user reviews the preprocessed and clustered data. They also examine generated reports, recommended actions, and simulations of the system's behavior.

2. **Scenario 2:** After uploading sensor data and receiving predictions, the user can modify measurements if needed. They explore the recommendation system for suggested actions based on the analysis results. The user interacts with the system to adjust input data and explore suggested actions accordingly.

3. **Scenario 3:** The admin logs into the system with administrative privileges. Manages the database by overseeing client lists and their respective data analysis. The admin ensures the functionality of the system and may validate or add new features as appropriate.

4. **Scenario 4:** The database stores client data, system information, and attributes used for model training and classifier creation. It ensures the availability of historical data for potential retraining and comparison with new input data. The stored information allows the system to classify and contrast input data for analysis and detection purposes.

10 Project Plan

Task	**Start Date**	**End Date**	**Duration**	**Members**
Idea and Supervisor	10/10/2023	12/11/2023	32 days	All Team Members
Information Collection and researches	12/11/2023	16/11/2023	4 days	All Team Members
Survey and Proposal preparation	12/11/2023	15/11/2023	3 days	All Team Members
Acquiring dataset	10/11/2023	15/11/2023	4 days	All Team Members
Preprocessing stage	13/11/2023	5/12/2023	22 days	All Team Members
Proposal presentation	15/11/2023	15/11/2023	1 day	All Team Members
SRS Documentation	20/12/2023	10/1/2024	21 days	All Team Members
SRS Diagrams	21/12/2023	27/12/2023	6 days	All Team Members
Working on the model	25/12/2023	10/1/2024	16 days	Mohamed Hossam, Mostafa Mohamed
User Interface	25/12/2023	10/1/2024	16 days	Hanya Yasser, Youssef Mahmoud
SRS Powerpoint presentation	10/1/2024	13/1/2024	3 days	All Team Members
Submission of SRS	13/1/2024	-	-	-
SDD Documentation	1/2/2024	-	-	-

Table 11: Time plan table

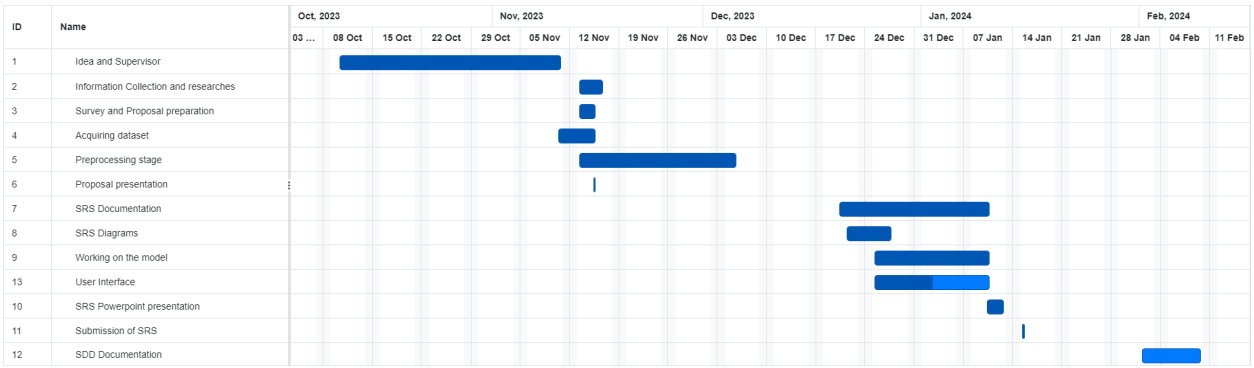

Figure 16: Gantt Chart

11 Appendices

11.1 Abbreviations

Term	Stands For
SCADA	Supervisory Control and Data Acquisition
PLC	Programmable Logic Controllers
CNN	Convolution neural network
SWaT	Secure Water Treatment
LSTM	Long Short-Term Memory
SVM	Support Vector Machine
GAN	Generative Adversarial Networks
ICS	Industrial control system
CPS	Cyber-Physical System

11.2 Supportive Documents

1. Dataset: For this section, we accessed the iTrust website to obtain the SWaT dataset. The following features of the dataset are listed: (11 days of nonstop operation, including 7 days of regular operation and 4 days of attack scenarios; network traffic was collected; all values from all 51 sensors and actuators were obtained; and data were labeled based on normal and abnormal behaviors).

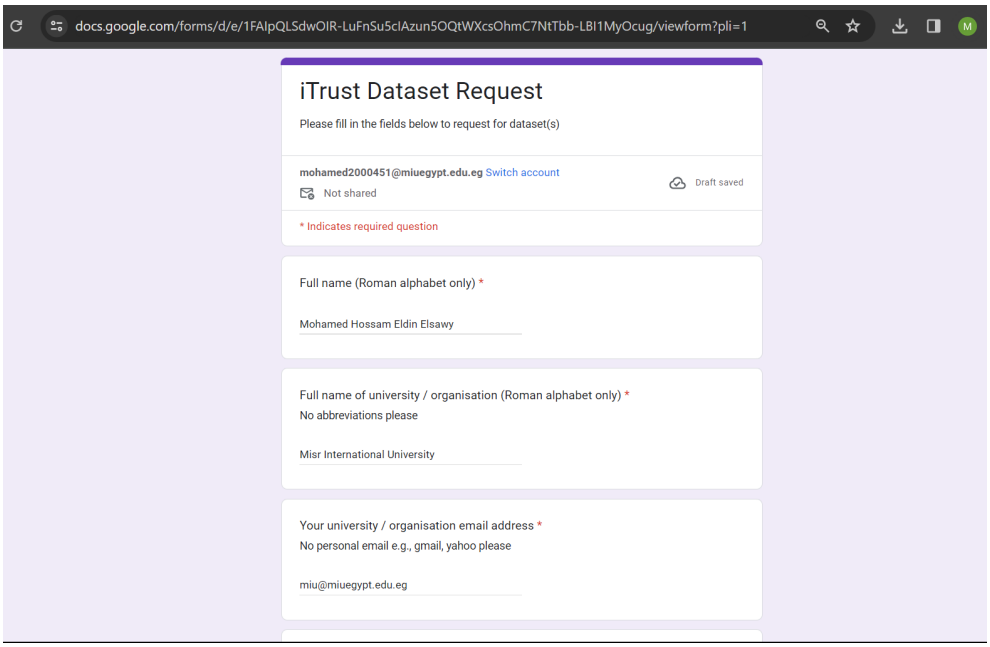

Figure 17: Requesting the dataset

Figure 17 shows a request for iTrust for the SWaT database.

The dataset link is: Dataset Request

2. Survey

   According to the survey we made "Cyber Attacks in Water Treatment Systems", We have concluded few points.

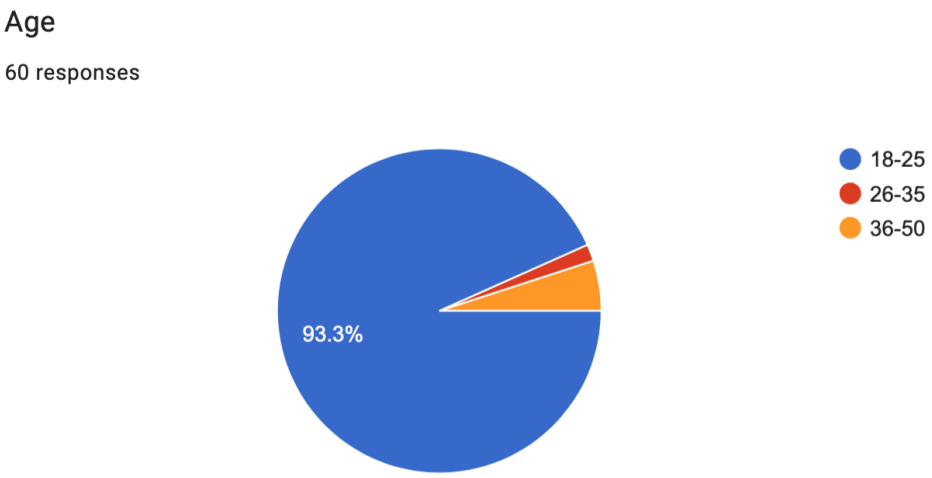

Figure 18: Aquashield Statistics

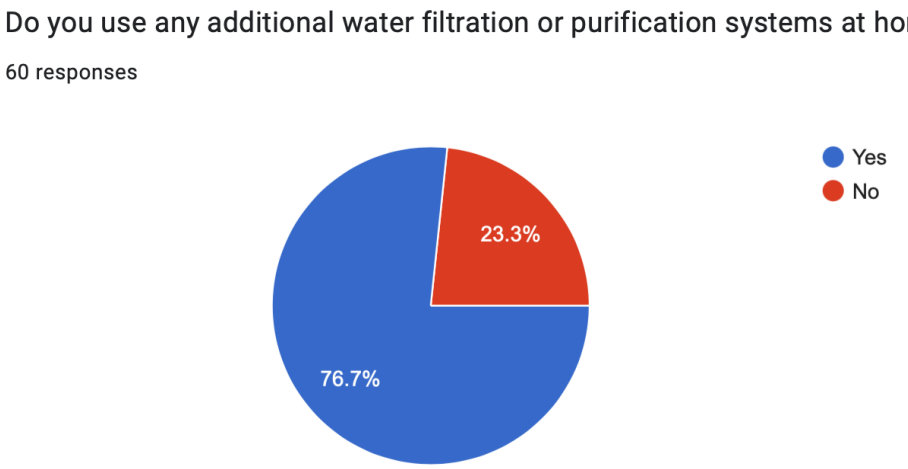

Figure 19: Aquashield Statistics

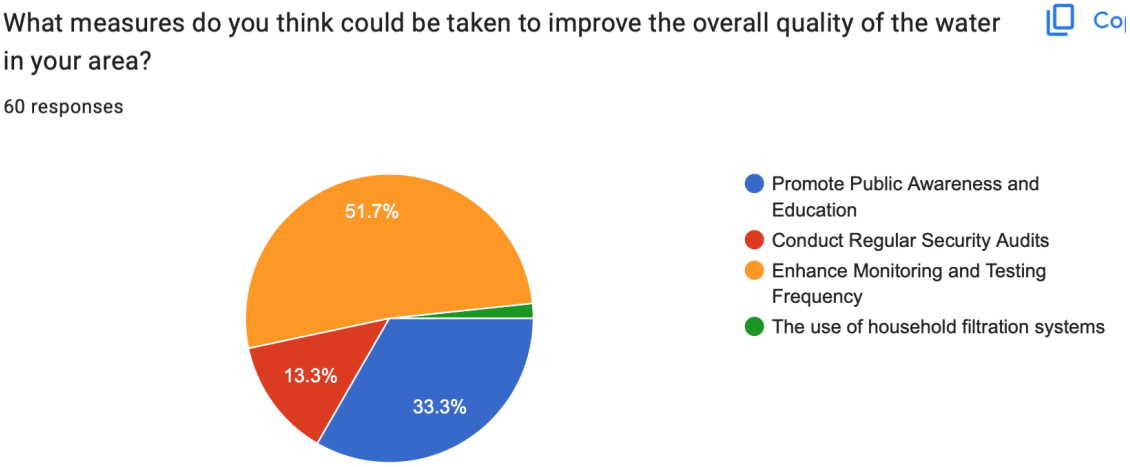

Figure 20: Aquashield Statistics

How concerned are you about the potential health impacts of water quality in your community?

60 responses

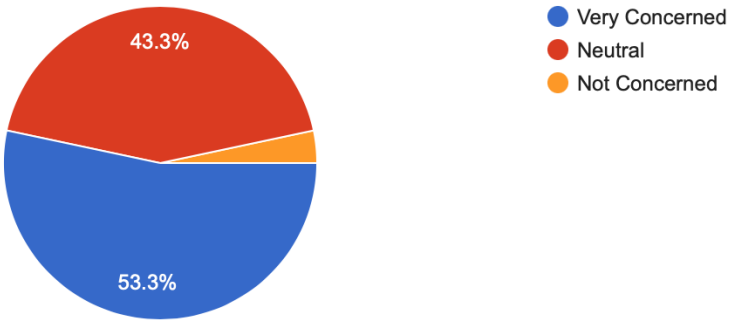

Figure 21: Aquashield Statistics

Have you or someone in your household experienced any health issues that you believe could be related to water quality?

60 responses

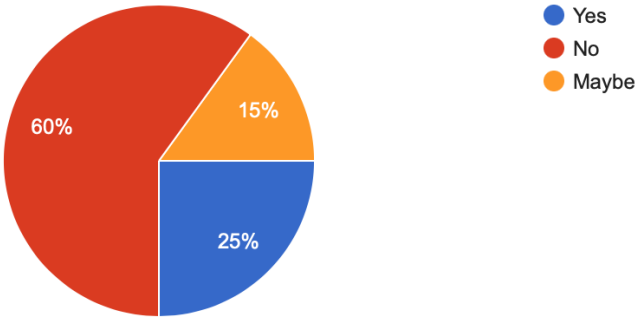

Figure 22: Aquashield Statistics

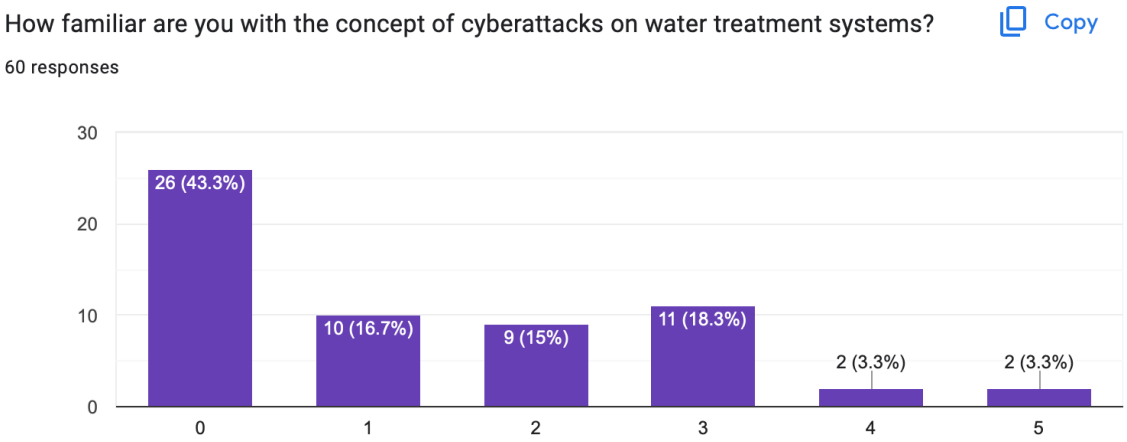

Figure 23: Aquashield Statistics

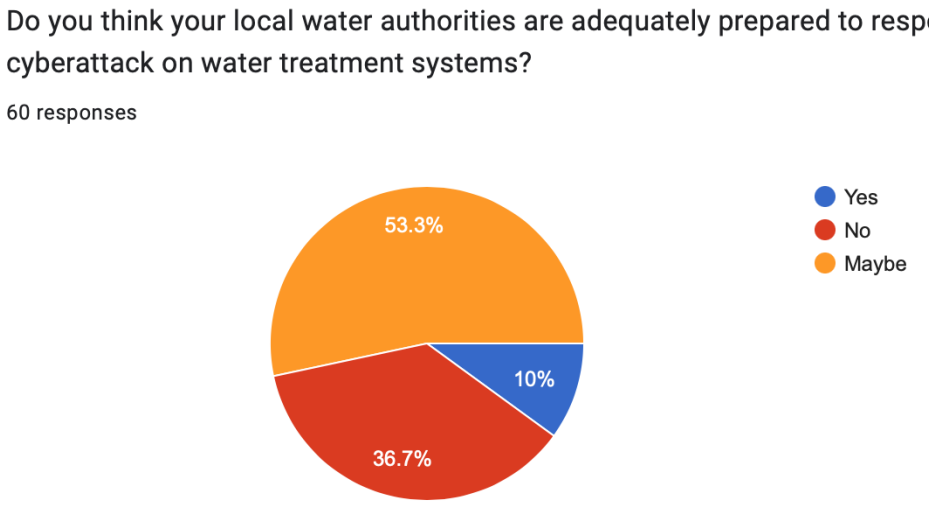

Figure 24: Aquashield Statistics

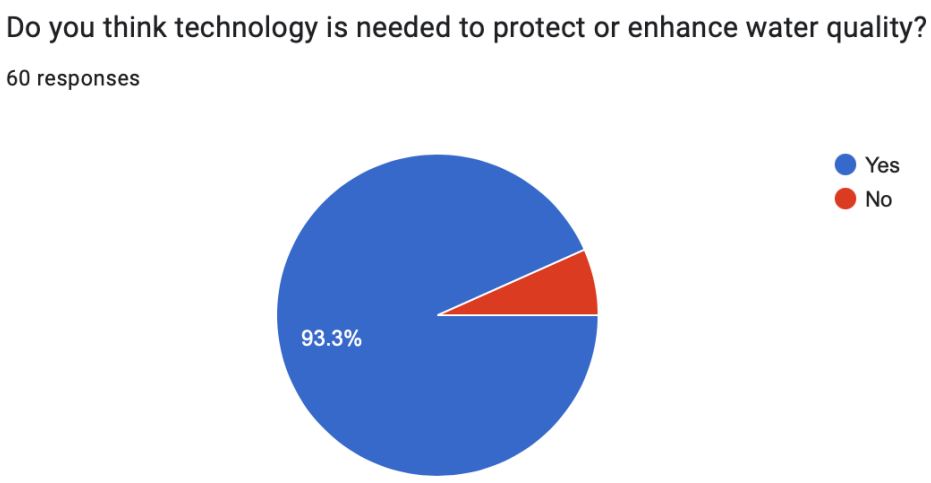

Figure 25: Aquashield Statistics

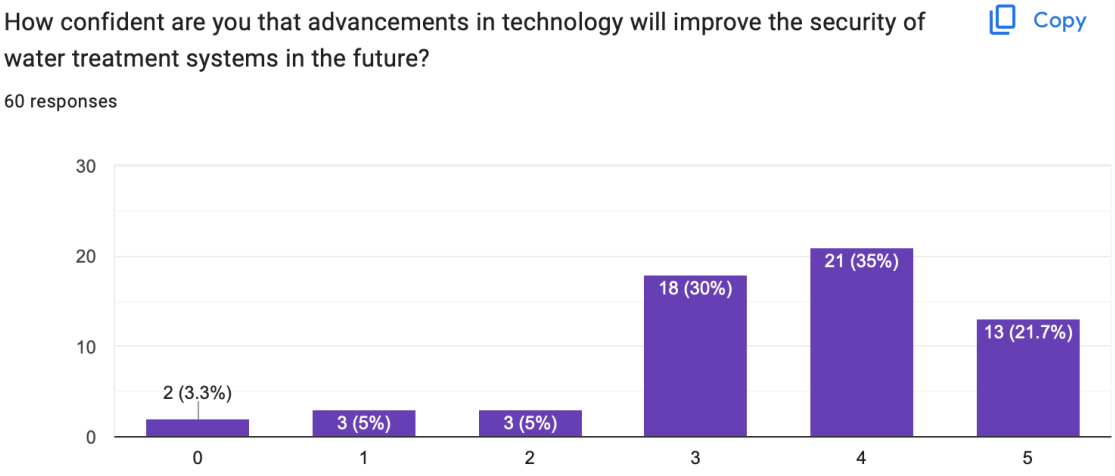

Figure 26: Aquashield Statistics

References

[1] Yong Zhang, Bingjie Li, and Xinqi Zhang. "Deep Learning-Based Anomaly Detection for Time-Series Data in Industrial Control Systems". In: *2023 42nd Chinese Control Conference (CCC)*. IEEE. 2023, pp. 8825–8829.

[2] Büşra Özgenç et al. "Anomaly Detection in Predicted Water Treatment Data Using Hybrid CNN-LSTM Network Model". In: *2023 31st Signal Processing and Communications Applications Conference (SIU)*. IEEE. 2023, pp. 1–4.

[3] Assoumer Redempta Manzi Muneza. "The Application of Unsupervised Machine Learning Techniques to Anomaly Detection, to Identify Cyber Attacks on Cyber-Physical Systems". In: (2021).

[4] Teoh Teik Toe, Lim Han Yi, and Edwin Franco Myloth Josephlal. "Advanced predictive techniques for detection of cyber-attacks in water infrastructures". In: *2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. IEEE. 2020, pp. 1–6.

[5] Sergei K Alabugin and Alexander N Sokolov. "Applying of generative adversarial networks for anomaly detection in industrial control systems". In: *2020 Global Smart Industry Conference (GloSIC)*. IEEE. 2020, pp. 199–203.

[6] MR Gauthama Raman, Nivethitha Somu, and Aditya P Mathur. "Anomaly detection in critical infrastructure using probabilistic neural network". In: *Applications and Techniques in Information Security: 10th International Conference, ATIS 2019, Thanjavur, India, November 22–24, 2019, Proceedings 10*. Springer. 2019, pp. 129–141.

[7] Mohammed Al-Dhaheri, Ping Zhang, and Dina Mikhaylenko. "Detection of cyber attacks on a water treatment process". In: *IFAC-PapersOnLine* 55.6 (2022), pp. 667–672.

[8] Prabhat Semwal and Akansha Handa. "Cyber-attack detection in cyber-physical systems using supervised machine learning". In: *Handbook of Big Data Analytics and Forensics* (2022), pp. 131–140.

[9] Ángel Luis Perales Gómez et al. "Madics: A methodology for anomaly detection in industrial control systems". In: *Symmetry* 12.10 (2020), p. 1583.

[10] Feng Xue et al. "Deep anomaly detection for industrial systems: a case study". In: *Annual Conference of the PHM Society*. Vol. 12. 1. 2020, pp. 8–8.