IoT-Enabled Water Quality Monitoring and Alum Dose Determining System

HHC Gunathilaka[*], HT Kodikara[†], GRMNND Ranaweera[‡], DWNT Wijethunga[§],
Charithri Yapa[¶], Bhathiya Pilanawithana[‖], Nishan Dharmaweera[**]

[*][†][‡][¶][‖][**]Electrical and Electronic Engineering Department, University of Sri Jayewardenepura, Sri Lanka

[§]National Water Supply and Drainage Board

Email: [*]chamathharitha@gmail.com, [†]hirushkodikara@gmail.com, [‡]nethmi.ranaweera12@gmail.com,
[§]nuwaneew@gmail.com, [¶]charithriyapa@sjp.ac.lk, [‖]bpilanawithana@sjp.ac.lk,
[**]nishanmd@sjp.ac.lk

Abstract—The water treatment process involves a series of physical, chemical, and biological techniques and methods designed to remove contaminants, impurities, and undesirable substances from raw or untreated water to make it safe, clean, and usable. Adverse weather conditions, such as rainy seasons and floods, present significant challenges, and difficulties to the water treatment process. Alum is mainly used for reducing the turbidity of the water in the treatment process. In Sri Lanka, all water treatment plants use the Jar Test to determine the Alum dosage manually. They conduct Jar Tests to measure raw water quality in their laboratories. The Alum addition valve is set to a pre-determined quantity and the value is decided based on the experience of the technicians. Since there is no specific formula, the presence of an experienced technician is required to determine alum dosage using the Jar test. Thus, this study intends to propose an automated system for determining the optimum Alum quantity to be added to a given raw water sample. This is achieved by designing an Internet of Things (IoT)-enabled sensing device for measuring the water quality, which feeds data to a Machine Learning Model trained to predict the Alum quantity to be added. The task of the Machine learning model is to identify the pattern of raw water parameters and Alum dosage. By doing this, a could be provided a guideline for determining alum dosage instead of a manual Jar test and removed the necessity of experienced technician support always. A user-friendly web interface will allow the operators to customize the parameters according to the requirement and display a detailed analysis report. The predetermined quantity of Alum is transmitted to the output Alum adding value, to enhance its performance as an automated control rather than doing the addition process manually. This is a conceptual study on how to determine the Alum dosage using Raw water parameters and past experiment results.

Index Terms—Alum dosage, Automated, IoT-enabled, Machine Learning Model, Water quality, Water treatment

I. INTRODUCTION

The water treatment process is essential for preserving the environment, promoting various industries, protecting public health, and protecting the healthiness and sustainability of communities and economies. There are several water treatment plants in Sri Lanka such as Ambalthale, Biyagama, Kandana, Kalatuwawa, Labugama etc. Ambathale water treatment plant is the largest water treatment plant among them [1]. Coagulation, flocculation, sedimentation, filtration, and disinfection are the steps in the water treatment process that are used in public water systems in Sri Lanka as shown in Fig.1. After taking raw water from the distribution chamber, it goes through a coagulation process. Coagulation is the chemical water treatment process used to remove solids from water, by manipulating electrostatic charges of particles suspended in water [2]. This process introduces small, highly charged molecules into the water to destabilize the charges on particles, colloids, or oily materials in suspension [2]. Mainly, water treatment plants use Alum as a coagulant for this process. The flocculation process is used to remove large clusters (which means after adding irons of the impurities with Alum) from the water. The sedimentation process is used to remove solids that settle in the water under the effect of gravity [3]. The filtration process uses various materials (like sand, gravel, and charcoal) to filter the water after the sedimentation process [4]. The disinfection process uses various chemicals like Chlorine and Lime to purify the water before the distribution process.
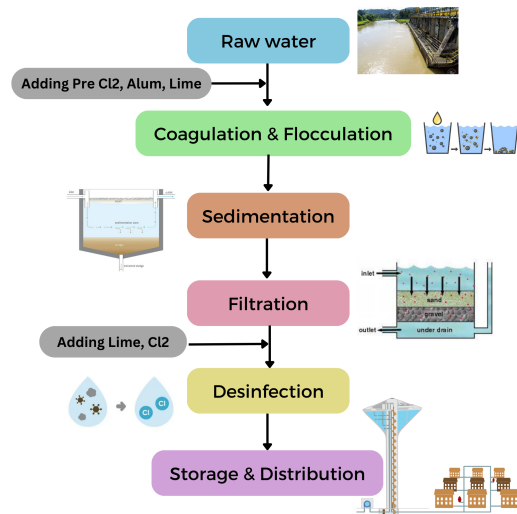

Fig. 1: Overall process in a water treatment plant

Especially in rainy seasons and when there are floods, water has a high amount of suspended solids. It

means the turbidity of the water increases. To reduce the turbidity of the water, Alum is used. Adding Alum is done in the process of coagulation and all the water treatment plants in Sri Lanka use the Jar test to determine the optimum Alum dosage. Two pumps are dedicated to the laboratory to get raw water samples in the Ambathale water treatment plant. In the plant, they conduct three manual tests for measuring raw water quality in terms of turbidity, conductivity, and pH.

This research paper mainly focuses on proposing an automated system for determining and releasing alum from the Ambathale water treatment plant. This solution is described under four sections; developing a real-time water quality measuring device under Section V-A, developing a Machine Learning model for determining optimum Alum dosage under Section V-B, designing a user-friendly web page under Section V-C and automating Alum adding valve under Section V-D in this research paper.

II. State-of-the-Art

A visit to the Ambathale Water Treatment Plant provided the opportunity to witness the plant's current water treatment process. The block diagram in Fig. 2 shows the process followed by the existing system in the Ambathale water treatment plant until the state of determining the Alum dosage.

Currently, the Ambathale Water Treatment Plant uses the Jar Test for deciding the alum dosage. At the beginning of the water treatment process, raw water is taken into the plant from Kaleniya River using an intake well. From the intake well, raw water is pumped into the distribution chamber. In the distribution chamber, the first post-treatment process begins. Alum and lime will be introduced to raw water. Lime is responsible for controlling the pH value of the water. Currently, lime is not used in the Ambathale Water Treatment Plant since raw water has an adequate pH value in the water. However, alum is introduced into the raw water, to control the turbidity of the water [5]. In this phase the Jar Test is conducted in the Ambathale Water Treatment Plant, to find the optimal alum dosage. Every hour they need to perform the Jar Test manually.

In the Jar Test, the procedure uses Alum, a chemical employed for coagulation/ flocculation within the context of water treatment. Six standard beakers, each having a 1-liter capacity, are employed for this process. Initially, 1 liter of untreated raw water is introduced into each beaker. The laboratory is equipped with two dedicated pumps responsible for sourcing samples of this raw water. In the treatment plant, a series of three manual tests are conducted to assess the initial quality of the raw water. These tests focus on parameters such as turbidity, conductivity, and pH, as documented in reference [6].

Following the assessment of the raw water, predetermined quantities of alum are introduced into each beaker, and the solution's concentration is progressively increased. Subsequently, the stirrers are activated to initiate the coagulation/ flocculation process. This stage of the procedure mimics the conditions experienced within the treatment plant. Initially, the stirrers operate at a higher revolutions per minute (RPM) rate, gradually reducing their speed over 22 minutes.

Upon completion of this stirring phase, a visual inspection of the beakers is conducted to determine the treatment effectiveness. If satisfactory results are not achieved, the alum dosage is increased, and the aforementioned steps are repeated, as per the guidance provided in reference [6].

It is important to note that throughout the Jar Test, various parameters, including turbidity, conductivity, and pH, are thoroughly monitored for both the untreated raw water and the water treated with alum. These recorded parameters serve as crucial data points in the assessment of the treatment process.
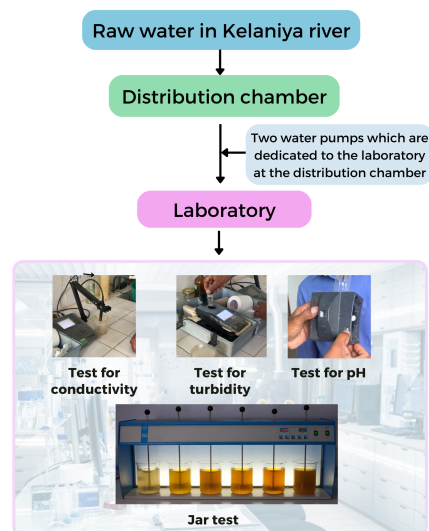

Fig. 2: Existing system in the Ambathale water treatment plant until determining the Alum dosage

III. Related work

Several related works have suggested a reliable real-time monitoring system for raw water involving multi-parameter measurement instruments with embedded devices, a wireless communication system, and an interface to present the acquired data to the operators in an organized manner. For parameter measurements remote sensing devices are often used [7]. Multi-parameter sensors can be more costly than the single-parameter sensor, but the multi-parameter sensor has less broadcasting information cost [8]. Basically, temperature, turbidity, conductivity, and pH parameters have been mostly focused on the real-time monitoring system. In our work, parameters will be input to the Machine Learning Model to predict alum amount. The alum dosage shows a strong correlation with turbidity

and moderate correlations with temperature and pH [9].

In the related works, several types of models have been utilized. The Artificial Neural Network (ANN) is known as an excellent estimator of nonlinear relationships between accumulated input and output numerical data. Using this nature of the ANN, the optimal coagulant dosing rate can be predicted from the operating data with accuracy and in time [10]. The results show that these two types of support vector machine regression techniques have good predictive capabilities for large and medium Water Treatment Plants (WTPs) as compared to small water systems [9]. The comparison shows that the K-Nearest Neighbour (KNN) has similar performances as the Support Vector Machine (SVR) for the large and medium-sized Water Treatment Plants (WTPs) and performs better for two small-sized WTPs [9].

The acquired data has to be transferred via IoT-based communication to the server for processing and to the devices to represent the process data [7]. The results show that these two types of support vector machine regression techniques have good predictive capabilities for large and medium Water Treatment Plants as compared to small water systems.

In their related work energy consumption, IoT technologies used in the system, memory restriction, data accuracy, and sensor localization are some of the key factors that were considered [8] [7].

A. Motivation

As mentioned above in the introduction, two pumps at the distribution chamber supply raw water to the laboratory, and their motors are working 24 hours. These motors are drawing large amounts of power. They separately conduct three tests, manually, to measure the pH, turbidity, and conductivity of the raw water. In there, sometimes human errors can occur, and those three tests take much time to complete. Jar test has the drawbacks of requiring manual intervention and no adaptation to sudden changes in water characteristics. Hence, operators typically adjust the coagulant dosage based on their past experiences and knowledge. Since the entire process is manual, it wastes expensive chemicals, fails water quality goals, and reduces the effectiveness of the sedimentation and filtration processes. They entered their Jar test results and raw water quality parameters manually. Therefore, many papers are involved, and sometimes data analysis part is not easy. Normally, the Alum adding valve is controlled manually with the determined Alum dosage. Hence, it decreases the efficiency of the whole procedure as well and causes time consumption. Sometimes human errors can occur in this scenario. Hence, this research paper proposes a method to automate the Alum dose-determining and releasing system.

IV. PROPOSED SOLUTION

To develop an automated solution for the Alum adding part, the solution can be solved under four main tasks. They are developing a real-time water quality measuring device, developing a Machine Learning model for determining optimum Alum dosage, designing a user-friendly web page, and automating Alum adding valve as shown in Fig.3. A real-time water quality machine measures the raw water quality in terms of turbidity, conductivity, pH, and temperature as a mounted device in the distribution chamber. This sensing device sends the sensed data to the Machine Learning model to determine the optimum Alum dosage. The Machine Learning model uses dosage predictor to determine optimum Alum dosage. Machine Learning models that is stored in the cloud-based platform will be able to take those raw water parameter values and predict the optimum Alum dosage. That determined Alum dosage and sensed data will be displayed in a user-friendly web page. By representing those data on an interface that could be easily read and understood. That determined Alum dosage value is sent to the automated valve to control it.
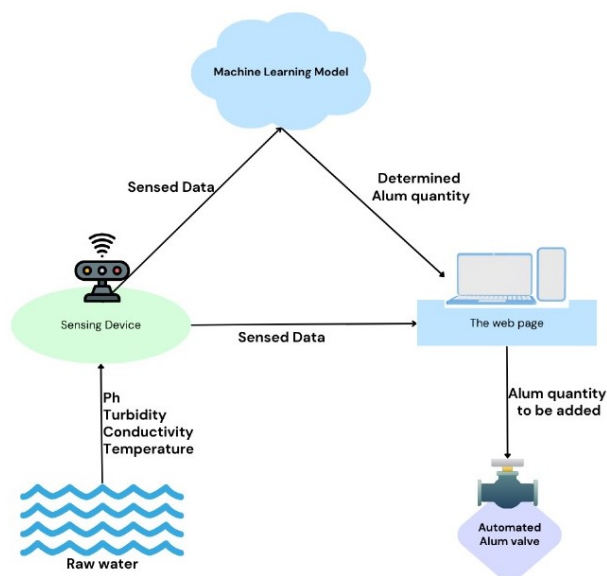

Fig. 3: Proposed solution

V. METHODOLOGY

Building an automated Alum dosage-determining system can be discussed under four subsections as follows.

A. Developing a real-time water quality measuring device

IoT sensing device is required to be designed in order to get real-time data of raw water parameters. Here, pH value, turbidity, conductivity, and temperature values of raw water are taken as the sensing parameters. This device is mounted in the distribution chamber.

The four sensors for pH, turbidity, conductivity, and temperature are used for the device. Since this device is an IoT-based device, data can be sent to the Machine Learning module and to the user-friendly web page through the cloud.
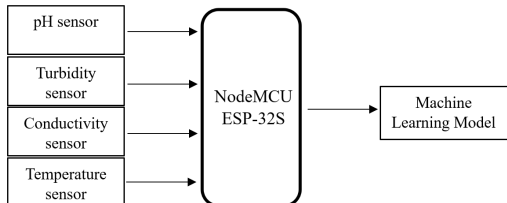

Fig. 4: Connecting sensors to Machine Learning Module

The following **table** shows the chosen sensor details for the device by considering the data set after conducting a market survey.

TABLE I: The sensor details for the device by considering the dataset after conducting a market survey

Sensor	Range	Accuracy	Working Temperature
PH Sensor (PH-4502C)	0-14 pH	± 0.2	0-80 °C
Turbidity Sensor (MD0591)	0-3000 NTU	± 10	5-90 °C
Conductivity Sensor (EC/TDS transmitter)	0-4400 μs/cm	± 0.01	0-100 °C
Temperature Sensor (DS18B20)	-55–125 °C	± 0.5	-55–125 °C

B. Developing a Machine Learning model for determining optimum Alum dosage

In order to prepare a Machine Learning Model historical data for pH, turbidity, conductivity, and the optimal alum dosage that was used for the water treatment plant have to be collected. Afterward, the data needs to be organized in a structured format, such as a Comma-Separated Values (CSV) file. Afterwards, the data set needs to be processed. In this step, null values need to be removed and data should be normalized if it's necessary. The data set will split into two subsets, training and testing set. A suitable machine learning model must be selected for prediction. The model should be able to be trained with an available data set and an adequate accuracy level should expected for the model. A trained model will be deployed for real-time predictions. The model can be stored in a cloud platform and it will ease the maintenance and monitoring of the model performance. Utilization of a cloud platform will simplify the machine learning model integration into a user-friendly web page for interfacing.

C. Designing a user-friendly web page

This web page acts as an interface for the user to show the data of parameters of raw water which can be measured by the IoT sensing device and to show determined Alum dosage by Machine Learning model. This web page can used to gain optimum Alum dosage by using data from the sensing device and by adding water water quality parameters manually.

D. Automating Alum adding valve

We can use the Machine Learning model's output to operate the alum addition valve, which will automatically change its state and add the right quantity of Alum to the raw water in the distribution chamber.

VI. MACHINE LEARNING MODELS

Since our goal is to predict the alum dosage needed for water treatment, it's a regression problem, and we want to predict a continuous numeric value. No Free Lunch Theorem essentially tells us that there is no universal algorithm that can outperform all others across every possible problem and dataset [11]. Rather than presuming that a particular algorithm will work definitively, we should try every possible algorithm and method to see what is best. This means systematically testing a range of algorithms and methods that are suitable for the problem's structure, dataset, and goals. The following algorithms have been suggested by looking at the nature of the algorithms and the history of utilization by observing related works [9] [10].

A. Linear Regression

Start with simple linear regression to see if there is a linear relationship between the input features (pH, turbidity, conductivity of raw water) and the alum dosage. It's a good baseline model. Multiple Linear Regression can be used to predict the continuous target variable based on two or more predictor variables. The result from this algorithm mostly will depend on the linear relationship between pH, turbidity, and conductivity parameter values with the optimal alum dosage that is required for water treatment.

B. Decision Trees for Regression

Decision trees can capture non-linear relationships and don't require the features to be normally distributed. Decision Tree is one of the most used, practical approaches for supervised learning. It can be used to solve both Regression and classification problems. In our case, we required for regression task. In the decision tree model, we can visualize the tree structure and measure its depth and complexity. Pruning can help control overfitting by limiting the depth of the tree or setting minimum sample leaf sizes. This also will reduce the complexity of the model.

C. Support Vector Regression (SVR)

SVR is an adaptation of support vector machines for regression tasks. It can work well when you have a small to medium-sized dataset. Instead of finding a hyperplane that separates classes, SVR aims to find a hyperplane (or function) that best fits the data points in a way that minimizes the error between predicted and actual values. The results show that these two types of support vector machine regression techniques have good predictive capabilities for the large and medium WTPs as compared to small water systems [9].

D. Artificial Neural Network (ANN)

Depending on the size of the dataset and the need to capture very complex relationships, the ANN can consider regression. The ANNs can model highly non-linear relationships between input features and the target variable, allowing for more accurate predictions when relationships are complex. Using this nature of the ANN, the optimal Alum dosing rate can be predicted from the operating data with accuracy and in time [12]. The ANN has a learning process like that of the biological neural network. Also, ANNs can automatically extract relevant features from raw data, reducing the need for manual feature engineering. However, the ANN model required a large amount of data set for training and this could be one drawback.

E. Comparison of ML models

As shown in Table II all four machine-learning algorithms have pros and cons when predicting alum dosage in water treatment with pH, turbidity, and conductivity as input parameters. K-Nearest Neighbours (KNN) can adapt to non-linearity by considering the nearest neighbours. However, it's computationally intensive with large datasets and sensitive to outliers. Support Vector Machine (SVM) Regression, employing kernel functions, can handle non-linear relationships effectively. It is robust to outliers but can be computationally expensive. Artificial Neural Networks (ANNs), while highly flexible and capable of capturing intricate relationships, lack interpretability and may require substantial data preprocessing. They perform well with large datasets but come with higher computational demands.

However, the final decision should align with our specific dataset size and computational constraints. Besides these few base estimators, we would need to try every possible algorithm and method and find the best one.

However, we understand that these few base estimators are capable and versatile, but they may not always deliver the desired level of performance, and it will depend on the datasets. In such cases, relying solely on a single base estimator might lead to suboptimal results, including reduced accuracy, overfitting, or underfitting. This is where ensemble methods come into play. Ensembles offer a strategic approach to enhance model performance by leveraging the strength of diversity. Instead of putting all effort into a single algorithm, ensembles combine predictions from multiple base estimators. By doing so, we could capture complex relationships within the parameters and achieve higher predictive accuracy.

VII. Conclusion

The main objective of this research paper is to design a fully automated system for adding Alum to raw water in the distribution chamber of the plant before adding raw water to the coagulation process. It will reduce many drawbacks in the overall manual process in the plant for the Alum-adding system to raw water. Hence, this research paper proposes to design an IoT-based sensing device to measure raw water quality parameters like turbidity, pH, conductivity, and temperature of raw water in the distribution chamber and to determine the optimum Alum dosage from the Machine Learning model from those parameters. These raw water quality parameters and Alum dosage will show on a user-friendly web page and by providing optimum Alum dosage for the automated valve, it controls the Alum quantity that is wanted to add to the raw water.

VIII. Future work

This research paper proposes a solution and gives an overview of the concept of the project. By doing a wide comparison of the technologies and components that can be used for the project, with optimum solutions, the project will be implemented within 8 months as a final year project. After completing the project, it will be tested at the Ambathale water treatment plant. With the evaluation and acceptance of the relevant parties, the project will be executed at the Ambathale water treatment plant.

References

[1] "National water supply and drainage board." Accessed: October 12, 2023. (), [Online]. Available: https://nsdi.gov.lk/national-water-supply-and-drainage-board.

[2] "Wastewater coagulation." accessed Oct. 9, 2023, Dober. (), [Online]. Available: https://www.dober.com/water-treatment/resources/wastewater%20coagulation.

[3] "Sedimentation processes." accessed Oct. 9, 2023, IWA Publications. (), [Online]. Available: https://www.iwapublishing.com/news/sedimentation%20processes.

TABLE II: Feature Comparison of Machine Learning Algorithms for Alum Dosage Prediction

Feature	Linear Regression	KNN	SVM Regression	ANN
Model Complexity	Low	Low to High	Low to High	High
Interpretability	High	Low	Low	Low
Data Preprocessing	Sensitive to Scaling	Sensitive to Scaling	Sensitive to Scaling	Sensitive to Scaling
Feature Engineering	Extensive	Limited	Limited	Extensive
Handling Outliers	Sensitive	Sensitive	Sensitive	Sensitive
Performance with Large Data	Efficient	Inefficient	Inefficient	Efficient
Non-linearity Handling	Linear	Can handle non-linearity to some extent	Kernel trick can be used	High capacity for non-linearity
Ease of Model Update	Easy	Requires retraining	Requires retraining	Easy (Online Learning possible)
Robustness to Noisy Data	Low to Moderate	Sensitive	Moderate to High	Moderate
Scalability	Very High	Moderate	Moderate	Moderate to High

[4] "Water treatment." accessed Oct. 10, 2023, Centers for Disease Control and Prevention. (), [Online]. Available: https://www.cdc.gov/healthywater/drinking/public/water-treatment.html.

[5] Q. H. Malik, "Performance of alum and assorted coagulants in turbidity removal of muddy water," *Applied Water Science*, vol. 8, no. 1, p. 40, 2018. [Online]. Available: https://www.dober.com/water-treatment/resources/wastewater-coagulation.

[6] P. E. Z. Satterfield, *Jar testing*, The National Environmental Services Center, 2005.

[7] F. Ungureanu, R. Lupu, A. Stan, I. Craciun, and C. Teodosiu, "Towards real-time monitoring of water quality in river basins," *International Journal On Recent Innovative Trend In Technology*, 2010.

[8] S. Sooryansh, K. Avinash, and B. A. P. Nitish, "Iot-based water quality monitoring system," *International Journal On Recent Innovative Trend In Technology*, p. 6, 2016.

[9] K. Zhang, G. Achari, H. Li, A. Zargar, and R. Sadiq, "Machine learning approaches to predict coagulant dosage in water treatment plants," *International Journal of System Assurance Engineering and Management*, vol. 4, no. 2, pp. 205–214, May 2013. DOI: 10.1007/s13198-013-0166-5.

[10] A. Aiswarya, N. Kurup, and N. Johnson, "Ann-based modeling for coagulant dosage in drinking water treatment plant," *Int Res J Eng Technol (IRJET)*, vol. 6, no. 5, pp. 6154–6158, 2019.

[11] "No free lunch theorem for machine learning." Accessed: Oct. 10, 2023, Machine Learning Mastery. (), [Online]. Available: https://machinelearningmastery.com/no-free-lunch-theorem-for-machine-learning/.

[12] G.-D. Wu and S.-L. Lo, "Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 8, pp. 1189–1195, Dec. 2008. DOI: 10.1016/j.engappai.2008.03.015.