Explanation and optimization of ML for coagulant dosing in Water treatment plant using XAI(eXplainable AI)

Hongeun Park[1*], Yunhwan Nam[2†], Hohyun Lee[3†], Sungyun Kim[4†]

[1*]Hangang River Basin Head Office, Korea Water Resources Corporation, Gyoyugwon-ro, Gwacheon-si, 13841, Gyeonggi-do, Republic of Korea.
[2]Southeast Gyeonggi Office, Korea Water Resources Corporation, Manhyeon-ro, Yongin-si, 16935, Gyeonggi-do, Republic of Korea.
[3]K-water Research Institute, Korea Water Resources Corporation, Yuseong-daero, Yuseong-gu, 34045, Daejeon, Republic of Korea.
[4]Geum River Weir Office, Korea Water Resources Corporation, Bukpo-ro, Buyeo-gun, 33127, Chungcheongnam-do, Republic of Korea.

*Corresponding author(s). E-mail(s): hong89@kwater.or.kr;
Contributing authors: namyh@kwater.or.kr; lhh@kwater.or.kr; moldau@kwater.or.kr;
†These authors contributed equally to this work.

Abstract

In the water treatment process, determining the appropriate dosing of coagulants is crucial for both ensuring water quality and optimizing operational costs. Traditional methods predominantly rely on heuristic approaches or manual calibrations. However, the actual reactions between various water qualities and chemicals are intricate and nonlinear, often making it challenging to stably secure the targeted water quality. In particular, due to industrialization-induced water pollution and rapid climate change, conventional automated water treatment systems alone are insufficient to adequately respond. This research introduces an innovative approach for clarifying and optimizing the dosing of coagulants in water treatment plants using Explainable Artificial Intelligence (XAI), specifically the SHAP and ICE techniques. These methods allow for a nuanced understanding of feature contributions and individual predictions. Our findings suggest that the XAI-driven approach not only improves the accuracy of coagulant dosing but also offers valuable insights to operators, promoting more informed decision-making. The integration of machine learning and XAI in the water treatment field promises a future of more efficient, transparent, and responsible operations.

Keywords: Water treatment process, Plant automation, Machine learning, XAI, Explainable artificial intelligence, SHAP, ICE, Trustworthy-AI

1 Introduction

Water treatment processes are essential for promoting public health and improving the quality of life by producing clean drinking water, and their stability must always be ensured. However, due to industrialization and climate change, the quality of water sources is increasingly fluctuating, intensifying the challenges in water treatment operations.

K-water, along with other water utilities, is pursuing the construction of smart water treatment plants based on artificial intelligence, starting with the incorporation of ICT technologies for process automation. As a result, they have introduced an autonomous operating system for determining the coagulant dosing rate, which is a core component of the AI algorithm.

However, the high volatility, uncertainty, complexity, and ambiguity of the operating environment undermine the reliability of autonomous operations in processes based on artificial intelligence. Moreover, the black box[1] issue of artificial intelligence is a major factor that hampers the sustainability of AI in water treatment plants. Consequently, these problems pose a risk of reducing the trust of operators in the new digital-based operating system of water treatment plants, and incorrect decision-making by AI can lead to severe consequences.

This study aims to address these challenges and to overcome the black box problem of the existing autonomous operation algorithm, ensuring user-centered trustworthiness. Additionally, we aim to mitigate uncertainties and optimize decision-making based on data. To achieve this, we have utilized Explainable Artificial Intelligence (XAI)[2]. Specifically, to provide explanations for the predictions of the coagulant dosing rate influenced by water quality factors, we have implemented the SHAP (Shapley Additive exPlanations) algorithm. Furthermore, we employed the ICE (Individual Conditional Expectation) technique for process optimization.

2 Related Works

2.1 Coagulant Dosing Process

In water treatment plants, coagulation, sedimentation, and filtration processes are used to remove turbidity from raw water. In the coagulation-sedimentation process, which removes most of the turbidity present in the raw water, coagulants are added to create conditions that allow impurities to coagulate with each other. The turbidity is then removed by the sedimentation of the enlarged flocs due to their own weight. he traditional methods for determining coagulant dosing rates can be summarized as follows[3]:

1. Reference Table Operation based on Raw Water Turbidity: Using a reference table based on the turbidity of the raw water has the advantage of allowing operators to respond quickly to rapidly changing water quality during the summer. However, since it relies on fixed values based on experience, continuous calibration is required.
2. Jar-Test: Typically, a jar-test is conducted on raw water 1-2 times a day to determine the coagulant dosing rate. However, it has the disadvantage of being time-consuming, making it impossible to respond quickly to rapidly changing water conditions, and being discontinuous.
3. Regression Polynomials: This method involves regression analysis based on statistical techniques. It requires continuous calibration due to the nonlinear changes in water quality and the associated reactions.

To overcome the limitations of the aforementioned methods and to reflect the nonlinearity of physicochemical reactions and the interrelationships between various water qualities, AI-based methods have been devised that can determine the optimal coagulant dosing rate based on changes in water quality and operational conditions.

The Ministry of Environment of South Korea has formalized a project to construct an AI-Based smart water treatment plant as part of the Green New Deal initiative in the water management sector[4].

Following this, K-water established an artificial intelligence autonomous operation strategy for the water treatment plant[5], and developed a system to predict the coagulant dosing rate by training water quality data using the K-means and GBR algorithm[6]. As a result, K-water's Hwaseong Water Purification Plant was selected as a part of the 2023 Global Lighthouse Network at the World Economic Forum, local time on December 14. [7]

2.2 Explainable Artificial Intelligence

XAI (Explainable Artificial Intelligence) is a technique that, unlike traditional machine learning with its black-box characteristics, provides the rationale and validity for machine learning decisions, thereby increasing trustworthiness. By utilizing this, we aim to provide the basis for ensuring the reliability of autonomous operation algorithms. In this paper, we seek to visualize the relationship between input data and AI decision-making by applying SHAP and ICE, which are among the most widely used model-agnostic methods for model interpretability analysis.

2.2.1 SHAP(Shapley Additive exPlanations)

SHAP (Shapley Additive exPlanations) is a methodology developed to enhance the explainability of predictive models, based on the Shapley Value[8] from cooperative game theory developed by Lloyd Shapley, who won the Nobel Prize in Economics in 2012. The Shapley value for a feature is calculated as follows:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (v(S \cup \{i\}) - v(S))$$

- ϕ_i is the Shapley value for feature i.
- N is the set of all features.
- S is a subset of N not containing i.
- $v(S)$ is the prediction model's output when considering only the features in subset S.
- The sum is taken over all subsets S of N not containing i, and it calculates the weighted average of the marginal contribution of feature \$i\$ across all possible subsets.

This equation assesses the fair contribution of each player (or each feature in a machine learning model) by evaluating how much value that player adds to the game (or to the model's prediction).

Lundberg and Lee (2017)[9] introduce the SHAP (Shapley Additive exPlanations) technique, asserting that this methodology offers an integrated approach to interpreting predictions from various models. They developed a method that utilizes SHAP values to quantify the contribution of each input variable to the output of machine learning models, based on the concept of Shapley Value developed by Lloyd S. Shapley.

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!} [f_x(z') - f_x(z' \setminus \{i\})] \tag{1}$$

- $\phi_i(f, x)$: This is the Shapley value for feature i in the context of a prediction model f and a particular input x.
- $\sum$: The sum over all subsets z' of the feature set x'.
- $z' \subseteq x'$: Each subset z' of the feature set x', where x' is the set of all features.
- $|z'|$: The cardinality of the subset z', which is the number of features in the subset.
- M : The total number of features in the set x'.
- $|z'|!$: The factorial of the number of features in z', representing the number of permutations of $z'$with other features.
- $(M - |z'| - 1)!$: The factorial of the number of features not in the subset z', excluding feature i.
- $M!$: The factorial of the total number of features, representing the total number of permutations of the feature set.
- $f_x(z')$: The model prediction when the feature set z' is present.
- $f_x(z' \setminus \{i\})$: The model prediction when the feature set z' is present minus the feature i.

The SHAP technique is model-agnostic and can be applied to a wide variety of machine learning models without being limited to specific model types. In particular, the application of various machine learning techniques to the prediction model targeted in this study showed that tree-based ensemble models performed the best. To explain and interpret the tree-based ensemble models, Tree SHAP[10] was used to optimize the calculation of SHAP values. Additionally, methods that not only provide local explanations of the prediction model but also allow for an understanding of the model's global operating principles were referenced[11] and applied.

2.2.2 ICE(Individual Conditional Expectation)

As mentioned earlier, the black box issue in the field of machine learning and artificial intelligence is a major challenge to overcome in the autonomous operation of water treatment plants based on artificial intelligence. Goldstein et al. (2015)[12] proposes a solution to this problem by devising the ICE technique, which makes the internal operations of the model more transparent.

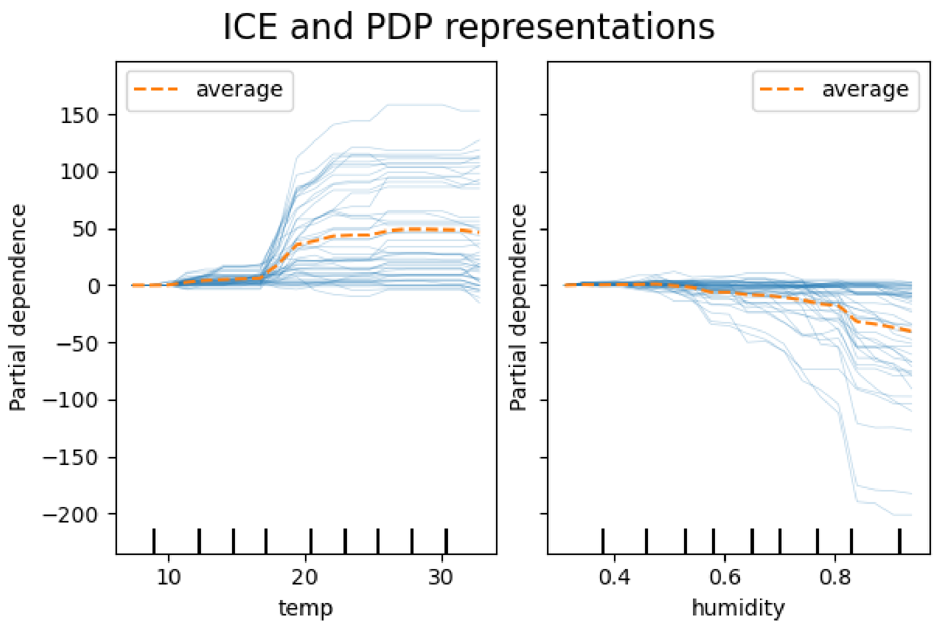

Fig. 1 ICE plot from the Scikit-learn documentation on partial dependence plots (Scikit-learn, n.d.).[13]

As shown in the ICE plot (Scikit-learn, n.d., Fig.1), the ICE plot visualizes how the model's prediction changes with the variation of values for each specific feature for each data point. In this study, the ICE technique was applied to the prediction model to analyze and visualize the impact of feature changes on model predictions, and to suggest directions for process optimization.

3 Method

The collected data are preprocessed, and machine learning is utilized to predict the coagulant dosing rate and the turbidity of settled water. The dosage adjustment is optimized through ICE analysis, and explanations for the model's predictions are provided using the SHAP algorithm.

- **Research Purpose and Overview :** This study aims to enhance the efficiency of the water treatment process by predicting the coagulant dosing rate and turbidity of sediment water using collected water quality data. To achieve this, various types of machine learning techniques are applied, and the explainability and interpretability of predictive models are improved through ICE analysis and the SHAP algorithm, promoting the optimization of the process.

- **Research Design :** The research is based on water quality data collected from actual water treatment processes. The study focuses on data preprocessing, the development and evaluation of machine learning models, and the interpretation of predictive models. Exploratory data analysis based on model interpretation and explanation techniques is also conducted to visualize the complex correlations between water quality and treatment processes.

- **Data Collection Method :** The data used in the research include water quality data collected from water treatment plants, encompassing variables related to coagulant dosing rates and turbidity of settled water. The data were measured at various stages of the water treatment process.

- **Basic Data Preprocessing and Analysis Strategy :** The collected data are initially processed through missing value handling, outlier removal, and normalization. Subsequently, appropriate feature selection and data splitting are conducted to train the machine learning models. Each model applies optimal hyperparameters found through grid-search techniques, and their performance is evaluated by comparing r-squared and RMSE values.

- **Analysis Tools and Software :** Python programming language and major libraries such as scikit-learn and TensorFlow are used for the development and evaluation of machine learning models. For the implementation of model explanation techniques, the PDP library is utilized for ICE analysis, and the SHAP library is employed for calculating SHAP values.

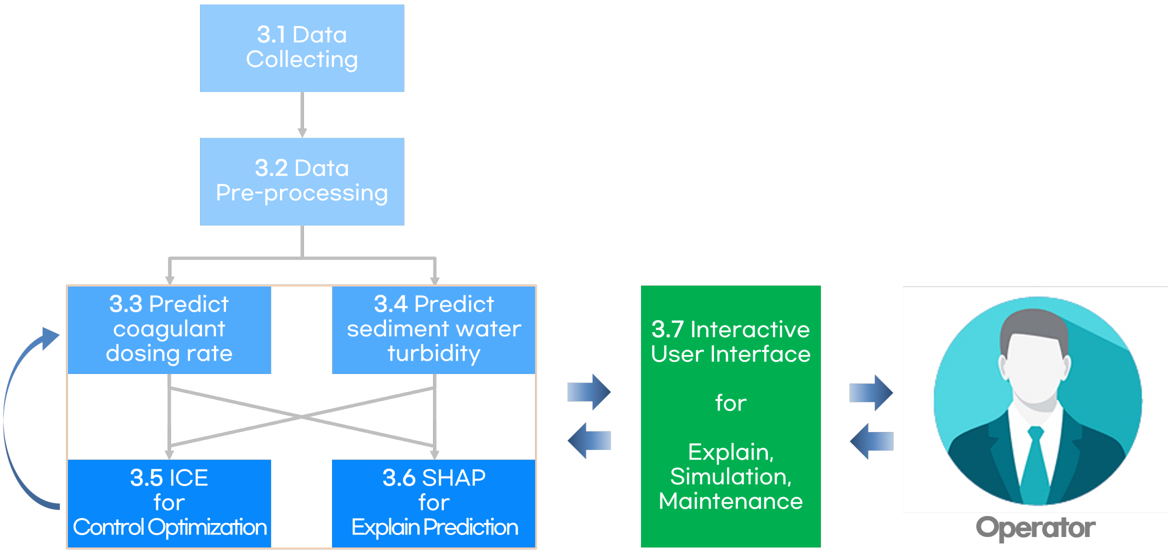

Fig. 2 Overall Research and Analysis Methodology Diagram

4 Experiments

4.1 Data Collecting

We collected data for analysis from the water treatment plant's SCADA server's Historian DB using queries, and the data used are as shown in Tables 1 .

Table 1 Status of training data collection

Data	Type	Time-Resolution	Source of data	Collection period
Raw water turbidity	Input variable			
Raw water conductivity	Input variable			
Raw water alkalinity	Input variable			
Raw water temperature	Input variable			
Raw water pH	Input variable	Minute	iWater Historian	2020.01 - 2022.12
Raw water flow-meter	Input variable			
Coagulant type	Input variable			
Raw water quality Cluster	Input(Derived) variable			
Coagulant dosing rate	Output variable[1]			
Sediment water turbidity	Output variable[2]			
Sediment water pH	-			

[1]Target data of Machine learning for prediction model-3.3(Predict coagulant dosing rate)

[2]Target data of Machine learning for prediction model-3.4(Predict sediment turbidity)

4.2 Pre-processing

Due to the presence of a large amount of missing data and abnormal data caused by facility operation issues in the raw data generated by the measurement sensors at the water treatment plant, preprocessing was performed as follows to improve the quality of the data. Improving data quality through such preprocessing can enhance the appropriateness of data analysis results and the performance of learning models.

Data preprocessing was conducted on six items:

1. Noise removal : Remove of hunting or random noise from datasets
2. Anomaly handling : Interpolation of bad quality data in datasets
3. Log scale transformation : Improving non-linearity of correlation between variables
4. Normalization : Min-Max scaling
5. Correlation analysis(Fig.3) : Pearson correlation analysis for variable selection
6. Clustering(Fig.4) : Labeling of status assessment through raw water quality by K-means

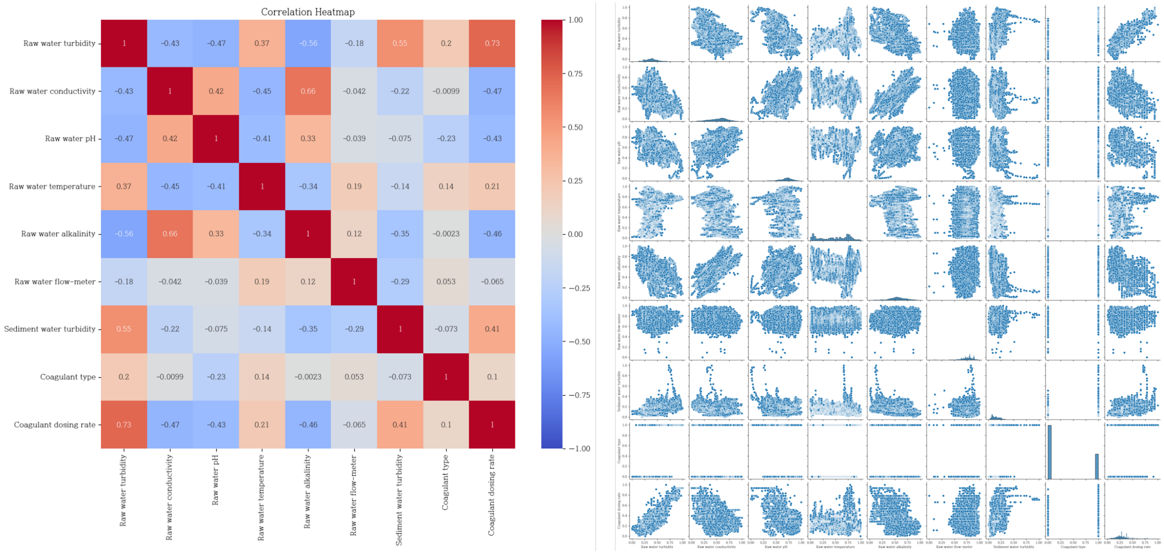

Fig. 3 Pearson correlation analysis for preprocessed input-output variables

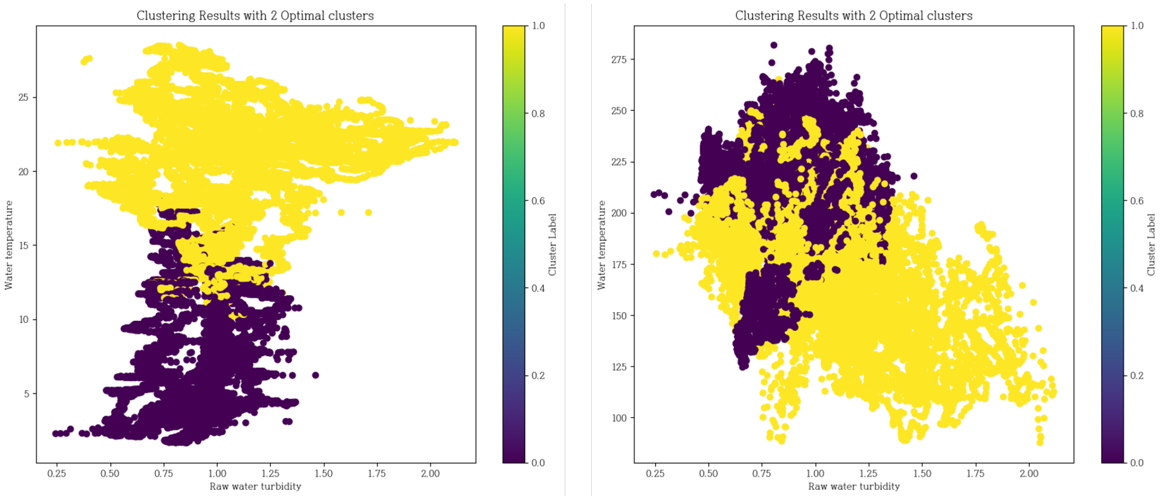

Fig. 4 K-means clustering for raw water quality 5 parameters (Optimal clusters = 2)

4.3 Machine Learning for Prediction

4.3.1 Coagulant dosing rate prediction model

A model was implemented using supervised learning based on the past three years of data from the operation results by operators at a water treatment plant. Using raw water quality and the type of coagulant as inputs and the coagulant dosing rate as the output, the data set was divided into 70The prediction training and analysis results showed that the tree-based ensemble model performed excellently. Based on the performance evaluation results from Tables 2 , the coagulant dosing rate prediction model was selected as Random Forest(R-sq = 0.97, RMSE = 1.08). Each models in Table 2 were tuned by searching for the optimal hyper-parameters using the Grid-Search technique.

Table 2 Summary of the performance of the coagulant dosing rate prediction model

Model	Ridge	Lasso	XGboost	**Random-forrest**	Decision tree	Keras DL	GRU
R squared	0.63	0.63	0.94	**0.97**	0.94	0.85	0.88
RMSE	3.59	3.60	1.40	**1.08**	1.44	2.30	2.00
Selection				**O**			

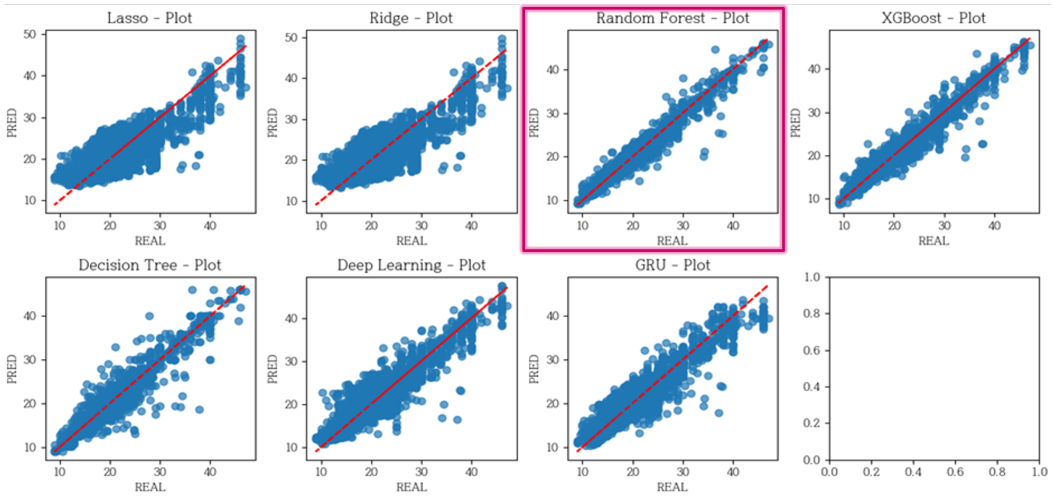

Fig. 5 Coagulant dosing rate prediction model result comparison plot

4.3.2 Sediment water turbidity prediction model

After the injection of the coagulant, there is a time difference between the quality of the raw water and the quality of the sediment water due to a certain reaction time and the residence time caused by the incoming flow rate. This time difference was calculated through WTLCC(Windowed Time Lagged Cross Correlation) analysis between the raw water quality (pH) and the sediment water quality (pH). Subsequently, a linear regression analysis model was designed to predict the time difference based on the flow rate, which was then incorporated to reflect the time difference between data during the learning phase of the sediment turbidity prediction model based on raw water quality.

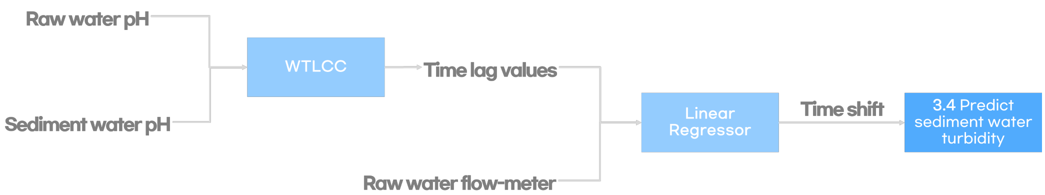

Fig. 6 Process for calculating the time lag between raw water and sediment water for the training dataset of the sediment water turbidity prediction model

The linear regression analysis results for time lag due to flow rate (R-sq: 0.937, RMSE: 5.98 min, (Fig.7)) were used to resolve the time lag between raw water and sediment water data, and to construct the training dataset for the sediment water turbidity prediction model.

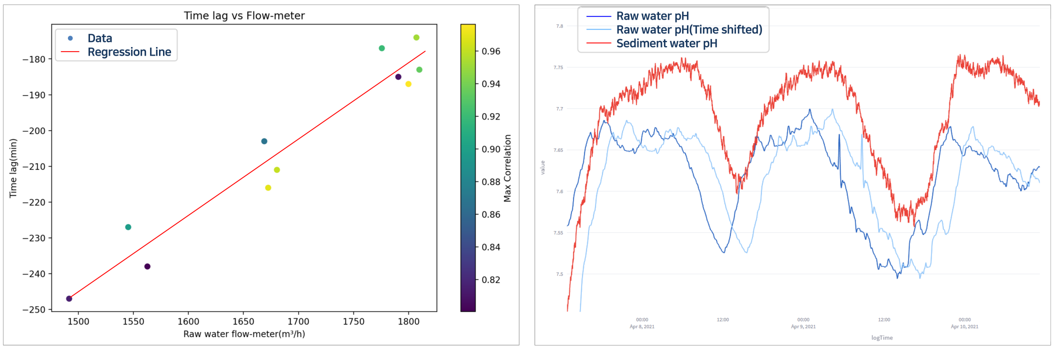

Fig. 7 Linear regression model plot for time delay (left) and comparative plot of time series data by time shift (right)

The subsequent procedure, similar to the coagulant dosing rate prediction model, involved using raw water quality as the input and sediment water turbidity as the output. The dataset was divided into 70% for training and 30% for testing to implement and validate the model. Based on the performance evaluation results from Tables 3 , the sediment water turbidity prediction model was selected as XGboost(R-sq = 0.96, RMSE = 0.04). Each models in Table 3 were tuned by searching for the optimal hyper-parameters using the Grid-Search technique.

Table 3 Summary of the performance of the sediment water turbidity prediction model

Model	Ridge	Lasso	**XGboost**	Random-forrest	Decision tree	Keras DL	GRU
R squared	0.45	0.15	**0.96**	0.94	0.91	0.75	0.83
RMSE	0.16	0.19	**0.04**	0.06	0.06	0.11	0.09
Selection			**O**				

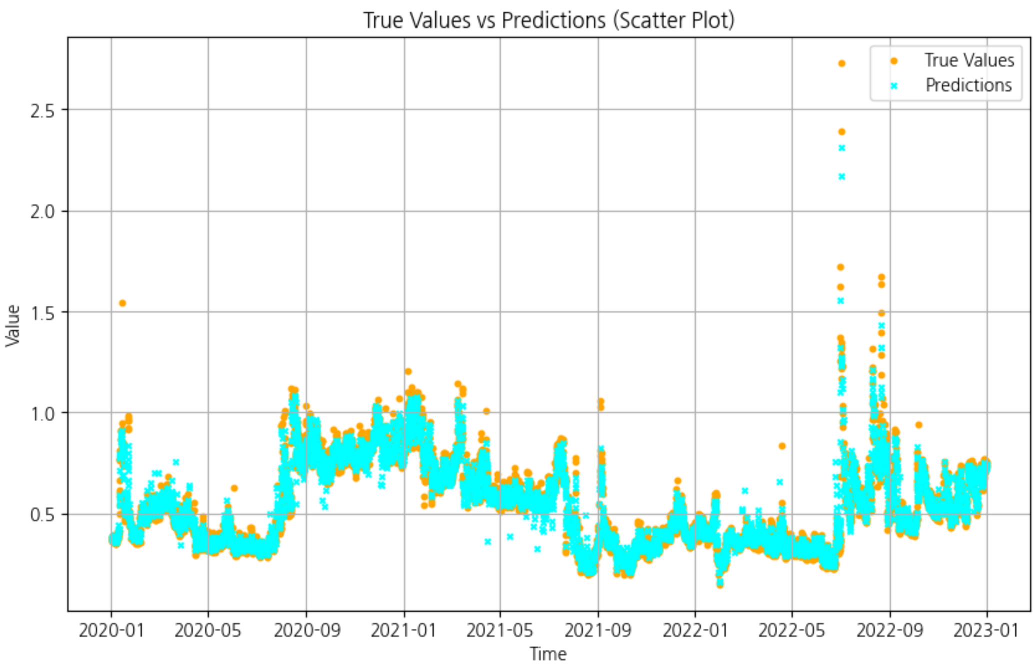

Fig. 8 Comparison plot of actual vs. predicted time series data for sediment water turbidity

Through the two prediction models, it is possible to autonomously control the dosing of coagulants in response to real-time raw water quality. Additionally, 3-4 hours after coagulant injection, the future quality of sediment water can be predicted, serving as a decision-making indicator for the operation of subsequent filtration processes.

4.4 SHAP Algorithm

Figure.9 depicts the Shapley Value of each variable for the coagulant dosing rate prediction model. As can be seen in the graph, the variable with the highest contribution to the coagulant dosing rate prediction is raw water turbidity. The contributions of the other variables are relatively even, and for water quality factors other than turbidity, an increase in their values contributed to a decrease in the predicted coagulant dosing rate.

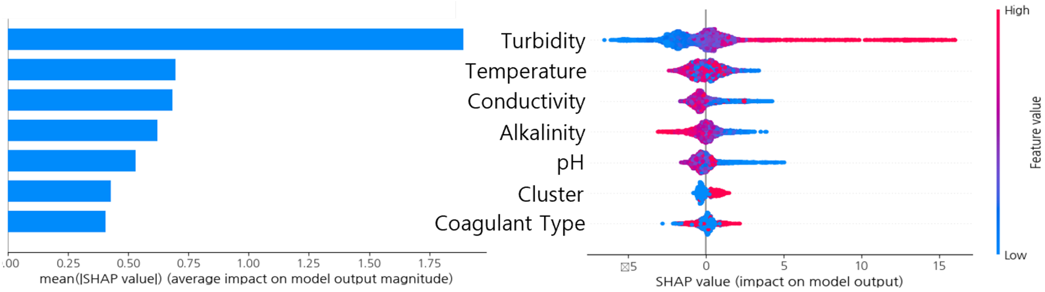

Fig. 9 Contribution analysis of variables based on SHAP for coagulant dosing rate prediction

To clarify the correlation between the data, we conducted exploratory data analysis based on the SHAP algorithm. The left side of Figure.10 shows that as raw water turbidity increases, the contribution to the coagulant dosing rate prediction linearly increases, while the electrical conductivity shows a decreasing pattern. The contribution slope of the raw water turbidity increases the most in the 40-60The central chart of Figure.10 indicates that as the raw water alkalinity decreases, the prediction of the coagulant dosing rate linearly increases. Additionally, the slope varies depending on the type of coagulant (PAC10, PAC12), which allows us to observe the differences in alkalinity consumption characteristics of the coagulants. The chart on the right side of Figure.10 shows that selecting PAC10 in areas with high conductivity and PAC12 in areas with low conductivity results in a decrease in the predicted dosing rate due to differences in coagulation effectiveness.

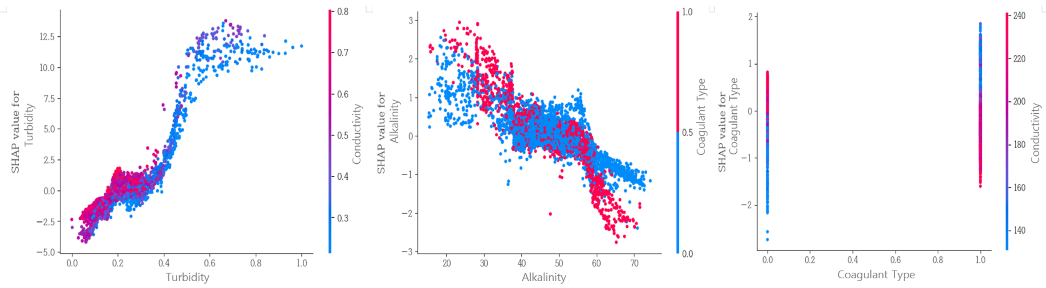

Fig. 10 Visualization of detailed interrelationships among variables based on SHAP

The chart in Figure.11 displays the spectrum of changes in the contribution to the predicted coagulant dosing rate based on the trends of each variable. As electrical conductivity decreases, the predicted coagulant dosing rate increases. At the same time, there is an upward trend in turbidity and downward trends in alkalinity and pH.

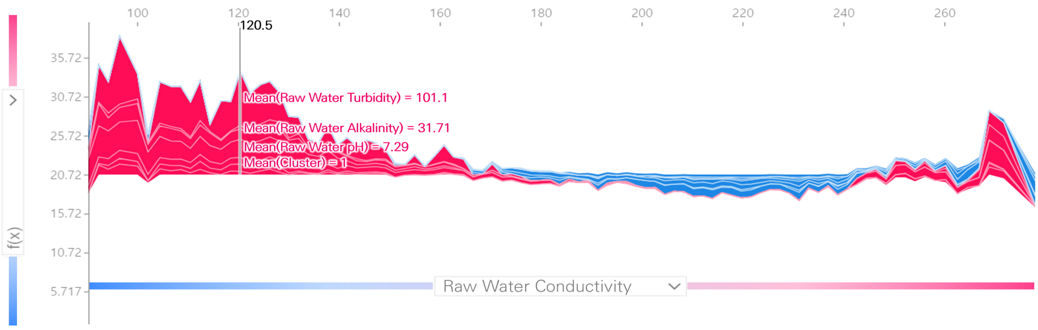

Fig. 11 Spectrum analysis of water quality and other variable contributions based on SHAP

Figure.12 is a contribution chart of prediction results for each input variable for the sediment water turbidity prediction model. The variable with the highest contribution to the prediction of sediment water turbidity was the turbidity of the raw water. Indeed, since the coagulant dosing process involves coagulating the turbidity of the raw water to form large flocs and then settling them, operators consider the turbidity of the raw water as the most important reference for determining the coagulant dosing rate. After raw water turbidity, the variable that has the most significant impact is the raw water temperature. As the water temperature decreases, it contributes to increasing the sediment water turbidity prediction value. Chemically, an increase in water temperature indeed accelerates reaction rates and particle mobility, enhancing coagulation efficiency.

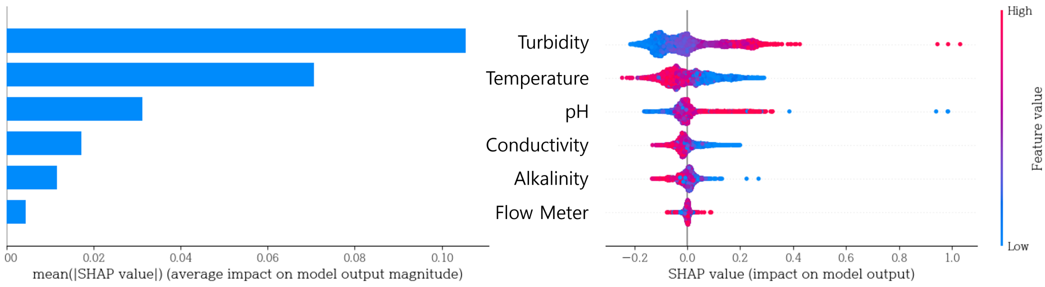

Fig. 12 Simulator explaining the prediction results of the sediment water turbidity

An explainable interface for operators was implemented with Streamlit, a Python web visualization application (Figure.13). By entering arbitrary data from the distribution of past data using a slider, it returns the coagulant dosing rate and the predicted quality of settled water a few hours later, providing justification for the prediction results based on the Shapley Value. This allows for local interpretation or understanding of the model's performance over a broad range of inputs. Additionally, the analysis results of the model on past data were visualized as a heatmap based on the Shapley Value. Through interpretation and history management of the algorithm's analysis results, it is possible to ensure the model's transparency and secure the utility of maintenance.

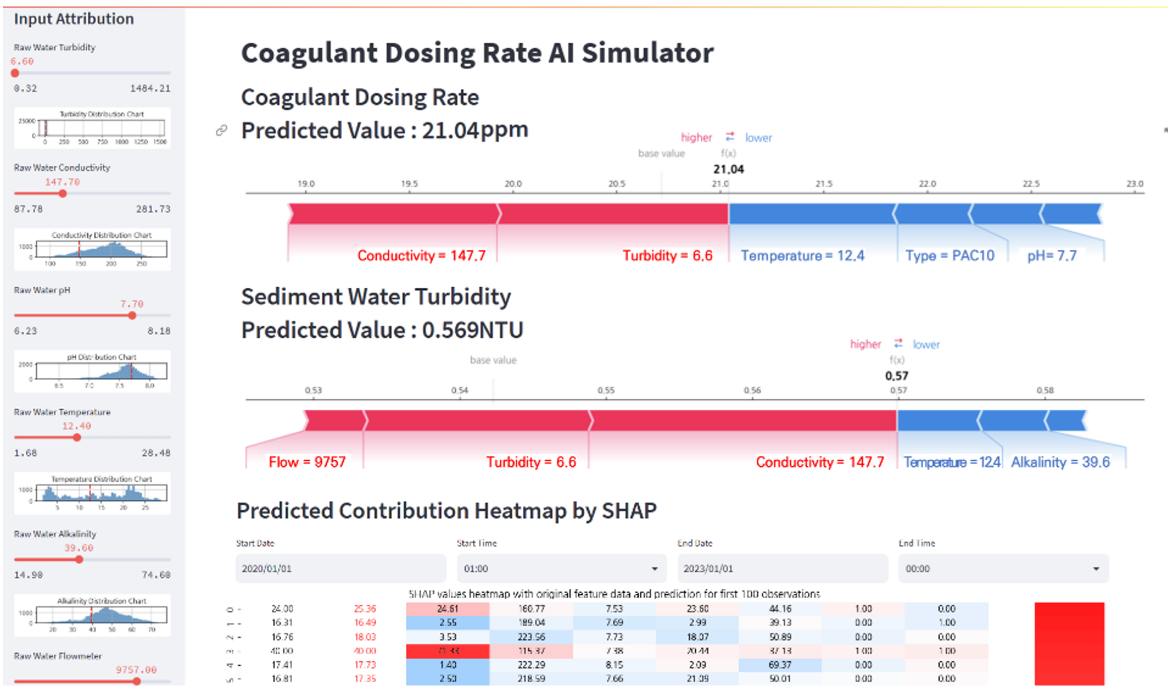

Fig. 13 Simulator explaining the prediction results of the coagulant dosing rate and the sediment water turbidity

4.5 ICE analysis for Control Optimization

To derive the operational band and optimize the control value, the ICE (Individual Conditional Expectation) was applied. Among tens of thousands of prediction cases, this methodology focuses on the raw water turbidity with the highest correlation to estimate the optimal coagulant dosing rate for the current water quality. For the predicted raw water turbidity at a given point in time, the possible operational band and the sediment water turbidity prediction value are automatically derived as ICE values. The current sediment water turbidity is then converted into a percentile within the operational band. This is subsequently translated into a percentile within the coagulant dosing rate prediction operational band to infer the optimal control value.

Figure.14 presents analysis results for data at 1pm on July 12, 2023. The predicted sediment water turbidity based on the raw water quality (turbidity 11.2NTU) was 0.71NTU. The actual sediment water turbidity was 0.83NTU, suggesting a decline in coagulation efficiency compared to average water quality conditions. Therefore, based on the basic predicted coagulant dosing rate of 22.1ppm, the optimal dosing rate is determined to be 25.79ppm, which is derived from all possible prediction values at a raw water turbidity of 11.2NTU, considering the current water quality and coagulation efficiency.

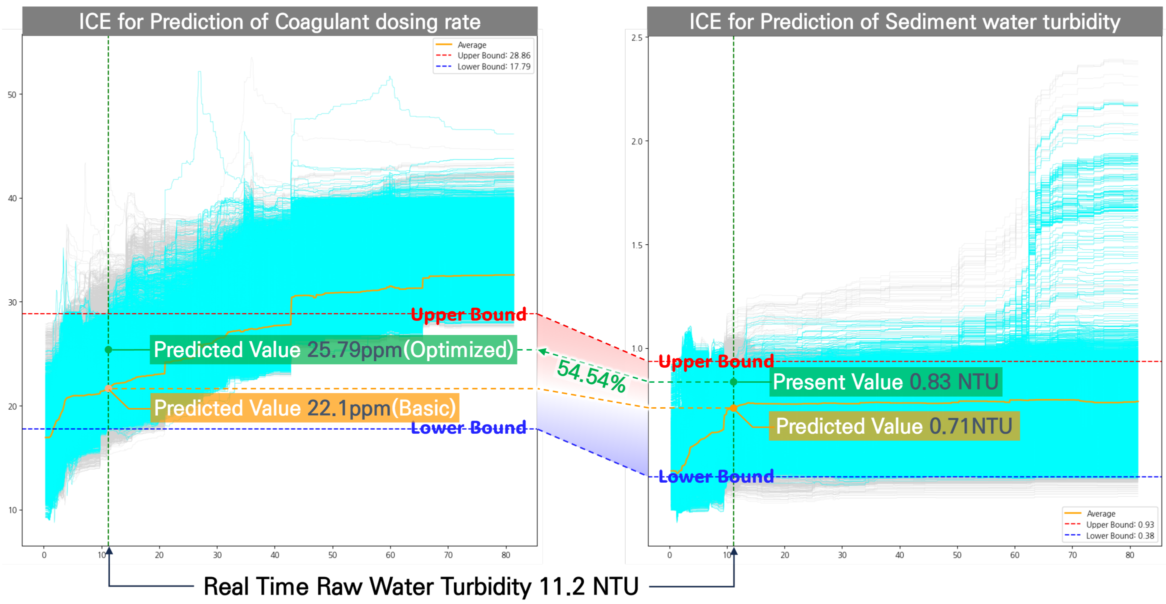

Fig. 14 Optimal coagulant dosing rate inference method based on ICE plot

Based on the 2022 test data, when comparing the performance of the coagulant dosing rate prediction model before and after optimization using ICE, the R-squared value of the basic injection rate prediction model was 0.92, and the R-squared value of the optimized model was 0.95. As a result, the reliability of the response to high turbidity during the summer months of July to September has increased.

5 Conclusion

This study aims to address complex water treatment processes by vividly revealing the correlations embedded in the data and providing them to users, ultimately optimizing decision-making in the water treatment process. Additionally, by explaining the model's prediction results, we sought to build a trust relationship between humans and AI. As a methodology to achieve this, we utilized SHAP and ICE algorithms, prominent techniques in the field of eXplainable Artificial Intelligence (XAI). Based on the data interpretation grounded in these algorithms, numerous elements aligned with physicochemical truths were identified. However, it should be emphasized that XAI primarily functions to explain and interpret model predictions and does not prove scientific facts. Misunderstandings about this distinction must be avoided. Some explanations and data interpretations are still not user-friendly. Explanations should not be substituted for genuine understanding of phenomena; hence, the interface for AI-human interaction requires further research.

References

[1] Zednik, C.: Solving the black box problem: A normative framework for explainable artificial intelligence. Philosophy & technology **34**(2), 265–288 (2021) https://doi.org/10.1007/s13347-019-00382-7

[2] Gunning, D.: Explainable artificial intelligence (xai). Defense advanced research projects agency (DARPA), nd Web **2**(2), 1 (2017)

[3] Jun, U.-P., Oh, S.-Y.: The method of optimum operation of coagulant dosage facility. 유체기계공업학회: 학술대회논문집, 275–281 (2004)

[4] Environment, M.: 그린뉴딜 밑그림 그린다. . . 광역상수도 스마트관리 본격 구축[Laying the groundwork for the Green New Deal... Full-scale construction of smart management for metropolitan water supply begins.]. https://www.me.go.kr/home/web/board/read.do?menuId=10525&boardMasterId=1&boardCategoryId=39&boardId=1377620 (2020)

[5] 신동기[Shin, D.: 4 차 산업 기술 기반의 미래형 스마트 정수장 구축전략[future smart water purification plant construction strategy based on industry4.0 technology]. 저널 물 정책· 경제[Journal of Water Policy and Economy] **33**, 75–85 (2020)

[6] Kim, J., Kang, B., Jung, H.: Determination of coagulant input rate in water purification plant using k-means algorithm and gbr algorithm. Journal of the Korea Institute of Information and Communication Engineering **25**(6), 792–798 (2021) https://doi.org/10.6109/jkiice.2021.25.6.792

[7] K-water News: Water, Nature and Humankind. http://k-waterwebzine.com/data/202401_en/sub-03-19.html (2024)

[8] Shapley, L.S., et al.: A value for n-person games, 307–317 (1953)

[9] Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. Advances in neural information processing systems **30**, 4768–4777 (2017)

[10] Lundberg, S.M., Erion, G.G., Lee, S.-I.: Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888 (2018) https://doi.org/10.48550/arXiv.1802.03888

[11] Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I.: From local explanations to global understanding with explainable ai for trees. Nature Machine Intelligence **2**(1), 2522–5839 (2020) https://doi.org/10.1038/s42256-019-0138-9

[12] Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E.: Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. journal of Computational and Graphical Statistics **24**(1), 44–65 (2015) https://doi.org/10.1080/10618600.2014.907095

[13] Scikit-learn contributors: Partial Dependence Plots. https://scikit-learn.org/stable/modules/partial_dependence.html. Accessed: 2023-02-20 (n.d.)

Declarations

- Funding : The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.
- Competing Interests : The authors have no relevant financial or non-financial interests to disclose.
- Author Contributions : All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Hongeun Park], [Yunhwan Nam], [Hohyun Lee] and [Sungyun Kim]. The first draft of the manuscript was written by [Hongeun Park] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.
- Data Availability: The datasets generated during and/or analysed during the current study are not publicly available due to security concerns associated with critical national information and communication infrastructure, but are available from the corresponding author on reasonable request.