Clean Technol., Vol. 29, No. 2, June 2023, pp. 145-150

청정환경기술

Stochastics and Artificial Intelligence-based Analytics of Wastewater Plant Operation

Sung-Hyun Kwon[1] and Daechul Cho[2]*

[1]Department of Marine Environmental Engineering, Engineering Research Institute
Gyeongsang National University Tongyong, 53064, Korea
[2]Department of Energy&Environmental Engineering, Soonchunhyang University Asan, 31538, Korea

(Received for review May 15, 2023; Revision received June 2, 2023; Accepted June 6, 2023)

Abstract

Tele-metering systems have been useful tools for managing domestic wastewater treatment plants (WWTP) over the last decade. They mostly generate water quality data for discharged water to ensure that it complies with mandatory regulations and they may be able to produce every operation parameter and additional measurements in the near future. A sub-big data group, comprised of about 150,000 data points from four domestic WWTPs, was ready to be classified and also analyzed to optimize the WWTP process. We used the Statistical Product and Service Solutions (SPSS) 25 package in order to statistically treat the data with linear regression and correlation analysis. The major independent variables for analysis were water temperature, sludge recycle rate, electricity used, and water quality of the influent while the dependent variables representing the water quality of the effluent included the total nitrogen, which is the most emphasized index for discharged flow in plants. The water temperature and consumed electricity showed a strong correlation with the total nitrogen but the other indices' mutual correlations with other variables were found to be fuzzy due to the large errors involved. In addition, a multilayer perceptron analysis method was applied to TMS data along with root mean square error (RMSE) analysis. This study showed that the RMSE in the SS, T-N, and TOC predictions were in the range of 10% to 20%.

Keywords : Artificial intelligence, Environmental big data, Wastewater plants, Stochastic analysis, Manpower saving

1. Introduction

Every wastewater treatment plant (WWTP) in South Korea needs well trained manpower and sophisticated skills for a regulation-meeting operation. Sorts of big data from wastewater treatment plants have been generated for past decade and kept in some central data control facilities through official environmental data sites. Tele-metering system along with quality sensors, being operated in most domestic areas collects core data on water quality in effluent streams including suspendid solids (SS), total nitrogen (T-N), total phosphorus (T-P), total organic carbon (TOC) and pH. These TMS data are constantly accumulated to reach 120 per day or 44,000 every year at least [1,2].

The ministry of Envoronment in South Korea has developed so called 'Smart Water Management Initiative' based on ICT (information communication technology) since 2021. The objective to this development is to allow a few selected WWTPs to optimize their processes simultaneously meeting the regitimate limits of effluent waters using a lot of information and best adapted communication technology *in situ*. Besides the five core data, WWTPs use pollutants concentrations in influent, water temperature, DO in unit tanks, and sludge production not only to maintain their operation, but also to analyze, control and predict the best performance as indicated in early works [3,4]. They claimed this tool in aspects of pros and cons for big data analytics, and applied to a typical activated sludge process and a membrane assisted biological process in the name of AI (artificial intelligence) modeling. Water monitoring sensors played a critical role in collecting big data, but the intrinsic instability in sensors like durability and error-conceiving problems had us to focus on modeling and control of a specific device such as a blower and pumps. Some worked on anaysis of time varying concentrations or loads of a specific pollutant in influents so that a distinct distribution of the pollutant over the

* To whom correspondence should be addressed.
E-mail:daechul@sch.ac.kr; Tel: +82-41-530-1341; Fax: +82-41-530-4728

doi: 10.7464/ksct.2023.29.2.145 pISSN 1598-9712 eISSN 2288-0690

entire process [5]. That result led to successful prediction of the pollutant behavior in the effluent streams. Machine learning was adopted in a research [6] to find a way of prophecy, how the influent status could change over time.

This work presents four domestic WWTPs with time-line data for water quality. Then, the work continues to analyze the whole data using multilayer perceptron technique, an AI method incorporated in the SPSS package. The analytes were SS, TOC, T-N, T-P, and electric power usage.

2. Materials and Method

2.1 Characteristics and Data Produced at Sampled WWTPs

Total four plants were selected for study (see Table 1). All plants were fed with extra nitrogen source (mostly livestock wastes and wastewater). Oxidation ditch and membrane separation were extraordinarily equipped for performance improvement at least for 10 years.

2.2 Data Preparation

Data being handled in this work were classfied into two and 130 thousand TMSS data were ordered in time line. After this each daily data were averaged so that over 1000 standardized data were prepared. Also, plant provided water data and temperature, and recycled rate at the beginning were matched to the TMSS data. Some abnormal measurements were corrected before analysis.

2.3 Stochastic Analysis with the SPSS tool

Influent waster quality data and process variables like DO (dissolved oxygen), temperature, recycled rate/ratio were set as independents while TMSS data – SS, T-N, T-P, TOC) were set as dependents. Two sets of variables were analyzed through the Pearson correlation method. Time-line analysis revealed daily water temperatures which might be changeable and a certain trend among the independents and dependents including total power consumption during the operation. In addition to that regression analysis between dependents and independents were conducted successfully. The newest version of SPSS 25 licensed through our institition was used in the entire work.

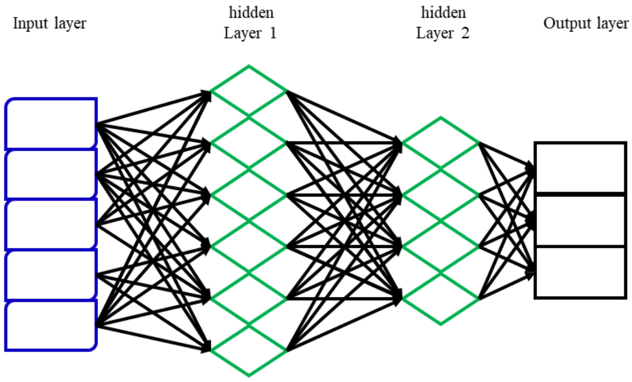

Figure 1. A typical layout explaining MLP (multilayer perceptron).

2.4 Multi Layer Perceptron Algorithm (MLP)

As a way of artificial neural network, multi-layer perceptron method was drawn in Figure 1. This analytic is working much like human neurons in its brain. The independent variables are displayed in the input layer while the expected (final) variables are shown in the output layer. In the middle, there are one or more hidden layers. First, calculation is going on between input and hidden layers and the calculated results are seen in the output layer. On each node a temporary weighted value is given, and an activated function is kept being calculated until no significant gaps occur with the weighted values being changed at every iteration [7].

MLP method provides with two roles: learning process and validation process. If you use m% of the data in the learning process, the rest – (100-m)% - would be used in the validation stage. For example, among 100 measurements 70 of data could be used in the learning process meanwhile 30 of data are assigned in calculations for validation, which means you may change the ratios among learning and validation processes. To keep them consistent, we tried these processes for each independent variable 5 through 10 times.

2.5 Regression Analysis and RMSE Evaluation for Prediction

All analytical results were shown as x-y graphs: x axis-real value (measurement) versus y axis-predicted values. You can refer to the diagonal line for comparision between reality and ideality (prediction). Each calculation is displayed as distributed dots, thus giving a distribution of calculations and a probable

Table 1. Characteristics of 4 selected wastewater treatment plants

Plant	Applied Process	Primary influent(m3/d)	Auxiliary influent(m3/d)	Max. influent (m3/d)
A	DNR/DeNiPho	23000	6000	30000
B	Oxidation Ditch	17000	-	17000
C	DeNiPho	5000	-	5000
D	Oxic/MBR	8000	-	8000

trend. To extract a precise trend out of the scattered data, regression and error analysis were followed. For four pollutants above analyses were carried out. All necessary formula and equations can be found in the previous work[8].

3. Result and Discussion

3.1 Stochastic Analysis via SPSS

Pearson correlation was conducted via pollutant concentrations in inlet and outlet. Two groups of variables were mostly correlated at super standard level. If you have "-1", correlation coefficient the plot shows a diagonally negative correlation between the two. In contrast, "1" coefficient means they are postively correlated (strong). Figure 2 shows a strong postive correlation between inlet flow rate and inlet T-N. This indicates that the linear increase manifests one property in input is correlated other ones in the same stream. On the other hand, inlet temperature and effluent T-N show a considerable, negative correlation (-0.55) which meant the warmer environment make nitfication and/or denitrification activate more to the lower T-N both in the process and in effluent (Figure 3; time-line data and correlation chart).

3.2 AI-like analysis

MLP function in SPSS 25 was used to interpret TMSS supported data (Figure 4) which are scattered and widely distributed, resulting in no tendency. These data are set as independent. Analytic effect was investigated on classified groups of dependents and independents as seen in Figure 5. The first inset (A) shows an inconsistent prediction when the inlet flow rate and temperature were set as independents and SS in the effluent was set as dependent. This preliminary test wa not successful. Next trial was done with more number of independents such as 5 to 9. The insets (B) and (C) show the measurement and prediction for SS with a higher precision. Obviously, other parameters such as recycling rate, aeration rate, flocculent and so on might have provided with more reliable prediction.

All the layers for neural network calculations are shown in Figure 6. Nine repetitive computations were conducted for each item. The daily data on inlet flow rate, water temperature and both concentrations of SS, T-N, T-P and TOC were assigned for learning process. The eighty percent of data were used for learning and the rest was allotted for data validation.

Four main parameters – SS, T-N, T-P, TOC – were investigated. Some plants show linear relationships, which means predictions and measurements are mutually correlated

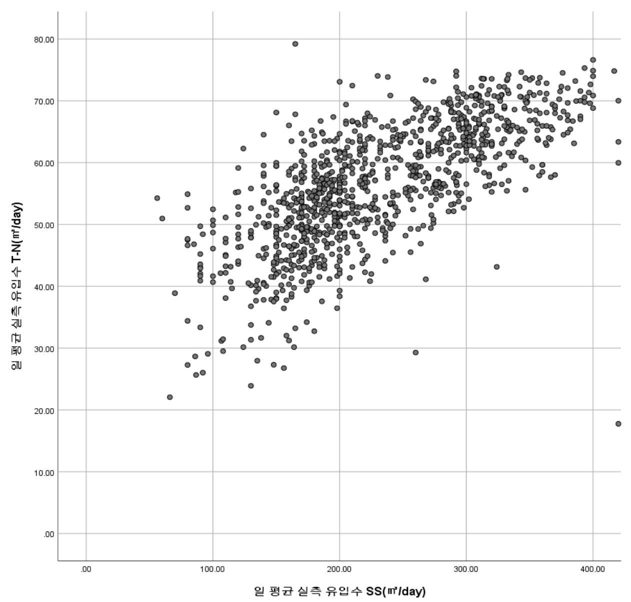

Figure 2. Scattered data distribution between SS and T-N in an inlet stream.

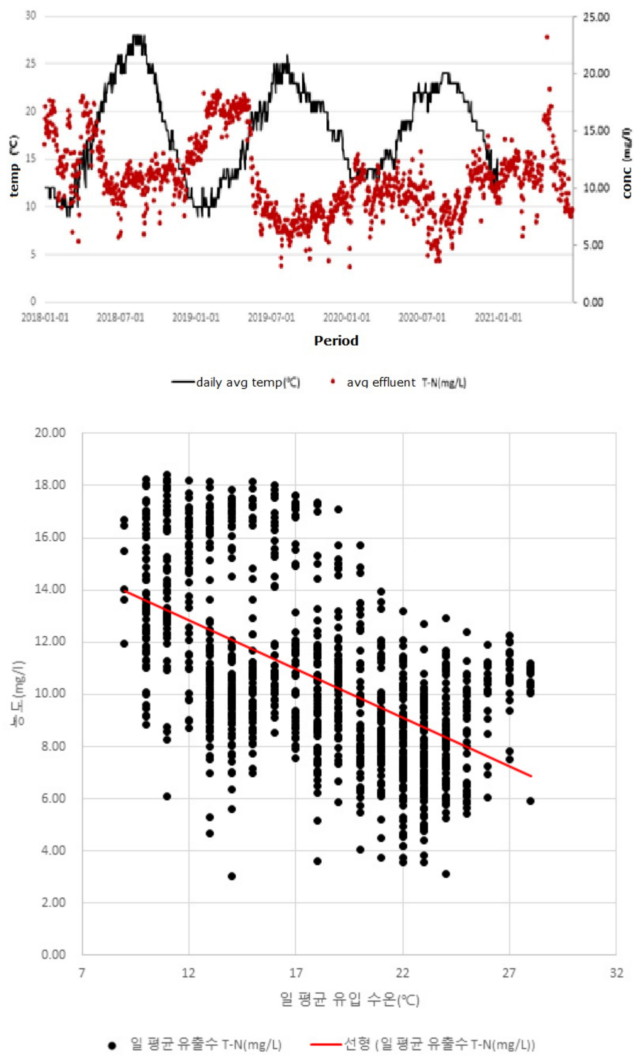

Figure 3. Temperature versus T-N: top, time-line T-N and temperature changes; bottom, Pearson correlation between temperature and effluent T-N.

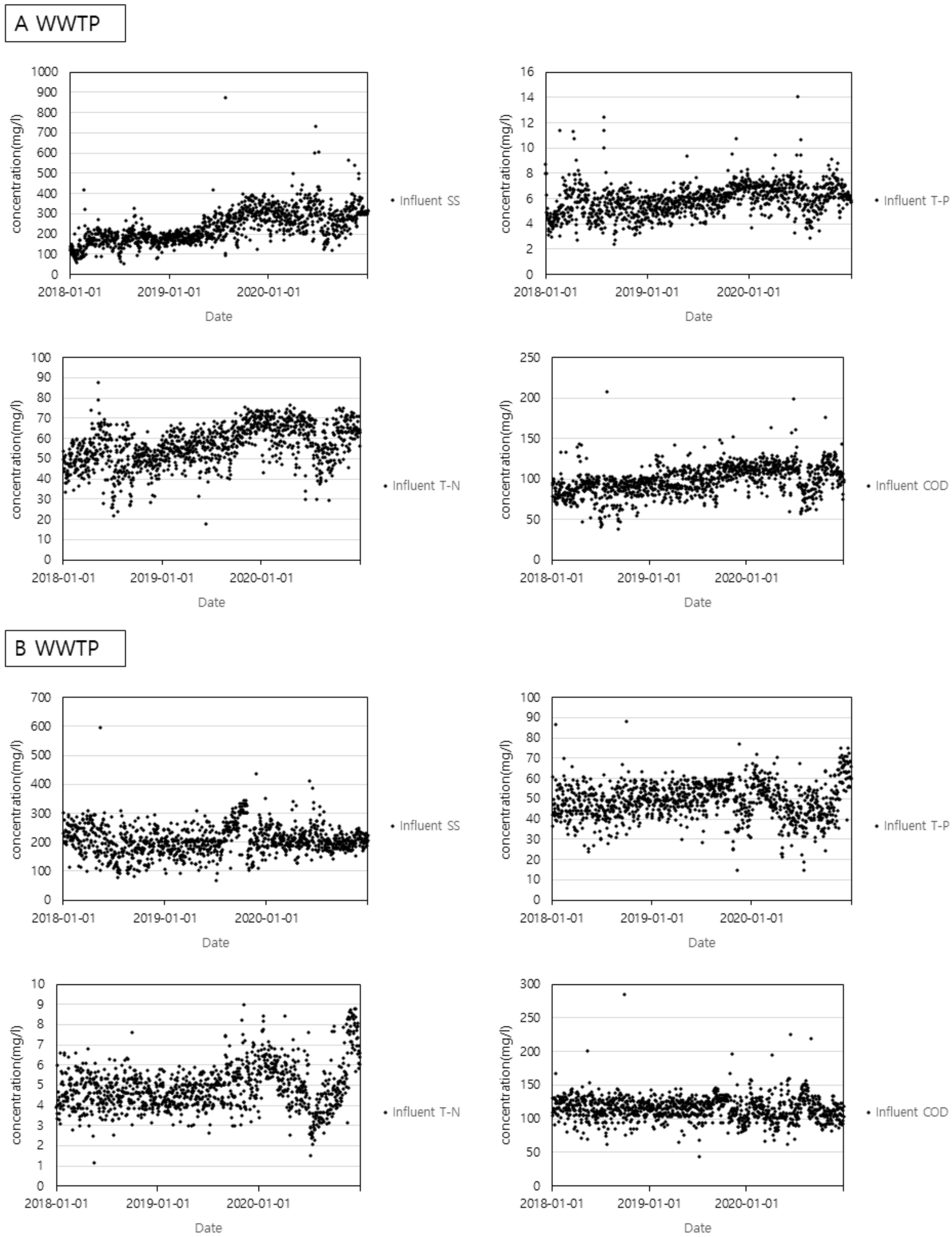

Figure 4. Time-line data plots for two typical plants.

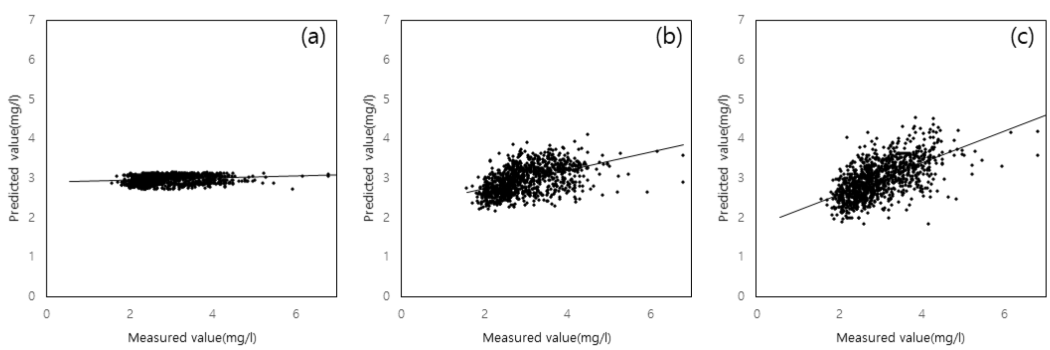

Figure 5. A preliminary test for multi-layer perceptron analysis: (a) independent variable (IV)=1, (b) IV=5, (c) IV=9.

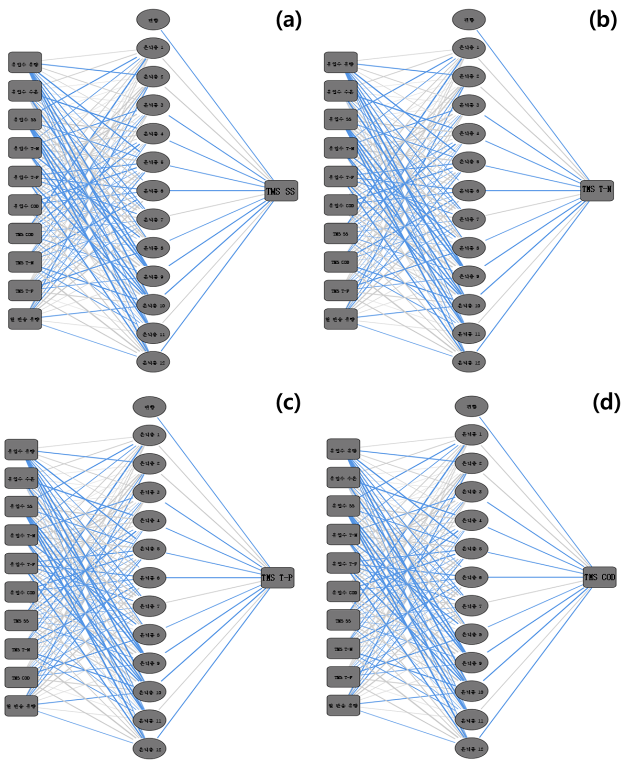

Figure 6. Schematic diagram depicting couple of trials for MLP analysis: (a) SS, (b) T-N, (c) T-P, (d) TOC.

with fine p-values (less than 0.05). The results tell that for every parameter, measurements were higher than predictions. That implies it may probable to take a significant risk of unsafe operation in the conventional biological process employed by A plant. Even though plant B shows more variant data distribution than plant A. AI predicts that more or less oxidation ditch process (B) could be more stable than the traditional one (A) despite its widened data scattering.

One way to deal with relatively large RMSEs in the AI analysis is to increase data number. In this case, if we used real hourly data instead of ones derived from a daily data the bigger data set and larger chances of learning may lead to higher prediction.

4. Conclusion

Domestic WWTPs were analyzed through two tools: stochastic analysis and artificial intelligence based approach. 1) This group of data as a sub-big data, comprising of about 150,000 per plant, was ready to be classified and also analyzed for optimizing a WWTP process. We used the SPSS 25 (statistical product and service solutions) package in order to statistically treat those data with linear regression and correlation analysis. Four domestic WWTPs were selected for this work. Major independent variables for analysis were water

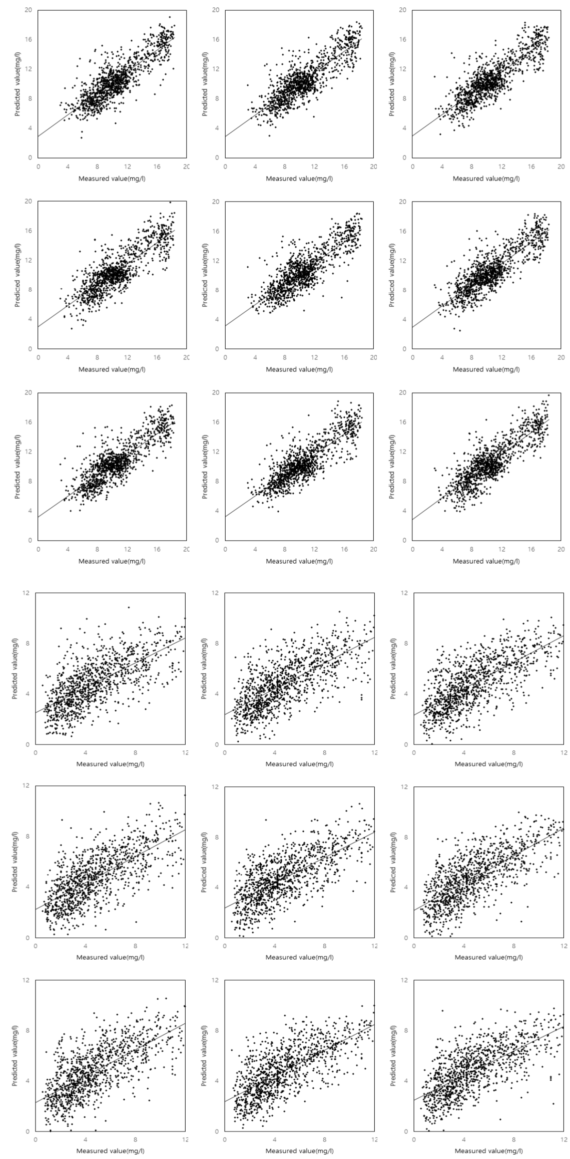

Figure 7. MLP learning results after 9th repetitive calculations for T-N in WWTP A (top) and WWTP B (bottom).

temperature, sludge recycle rate, electricity used, and water quality of the influent while dependent variables represented water quality of the effluent including total nitrogen, the most emphasized index of the discharged flow in all plants. Water temperature and consumed electricity showed a strong correlation with total nitrogen but the other indices meanwhile mutual correlations among other variables were found to be fuzzy due to the large errors involved. 2) An AI method like a multilayer perceptron analysis method was applied to TMS data along with a tool, RMSE (root mean square error) analysis. Overall, influent water data may predict effluent water quality in a specific plant. Even power consumption rate could be improved through MLP analysis. This study showed that RMSE in SS, T-N, and TOC predictions were in the range of 10% to 20%.

Acknowledgment

This paper was supported by the Soonchunhyang University. Authors thank for it.

References

1. SOOSIRO, 2022, http://www.soosiro.or.kr
2. Koprea Environment Corporation, 2022, http://www.keco.or.kr
3. Woongburi, P. and Park, J. K., "Big Data Analytics from a Wastewater Treatment Plant," *Sustainability*, **13**, 12383 (2021).
4. Jawad, J., Hawari, A. H., and Zaidi, S. J., "Artificial Neural Network Modeling of Wastewater Treatment and Desalination using Membrane Process: A Review," *Chem. Eng. J.*, **419**, 129-540 (2021).
5. Im, J. M. and Joo, H. K., "Big Data Analysis and Forecasting of Sewage Treatment Facilities using Machine Learning," *Kor. Inst. Ind. Eng.*, **22**, 1316-1322 (2021).
6. Jeong, H. S., Lee, S. H., Shin, H. S., and Song, E. Y., "Analysis and Prediction of Sewage Components of Urban Wastewater Treatment Plant using Neural Network," *J. Kor. Soc. Environ. Eng.*, **28**, 3038-315 (2006).
7. Seiffert, U., "Multiple Layer Perceptron Training using Genetic Algorithms," *Eur. Symp. Art. Neur. Network*, 159-164 (2021).
8. Lee, S. M., Park, K. D., and Kim, I. K., "Comparison of Machine Learning Algorithms for chi-a prediction in the middle of Nakdong River," *J. Kor. Soc. Water Wastewater*, **34**, 277-288 (2022).