



기계학습 기반 응집제 주입농도 예측모델의 파일럿 규모 혼화·응집·침전 공정 적용성 평가

Field application of a machine learning-based coagulant dosage prediction model in a pilot-scale coagulation · flocculation · sedimentation process

오창석¹ · 안주석^{1,2} · 연승재¹ · 이복진¹ · 오현제^{1,3,*}

Changseog Oh¹ · Jusuk An^{1,2} · Seungjae Yeon¹ · Bokjin Lee¹ · Hyun Je Oh^{1,3,*}

¹한국건설기술연구원 환경연구본부

²연세대학교 건설환경공학과

³과학기술연합대학원대학교

¹Korea Institute of Civil Engineering and Building Technology

²Yonsei University, Department of Civil Environmental Engineering

³University of Science & Technology

Received 16 March 2026, revised 23 April 2026, accepted 30 April 2026.

*Corresponding author: Hyun Je Oh (E-mail: hjoh@kict.re.kr; Fax: 031-910-0291, Tel. 031-910-0295 (010-3340-0295))

1 오창석 (박사후연구원) / Changseog Oh (Postdoctoral Researcher)

경기도 고양시 일산서구 고양대로 283, 10223

283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do 283 10223, Republic of Korea

1,2 안주석 (전임연구원) / Jusuk An (Researcher)

경기도 고양시 일산서구 고양대로 283, 10223

283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do 283 10223, Republic of Korea

서울특별시 서대문구 연세로 50, 03722

50, Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

1 연승재 (박사후연구원) / Seungjae Yeon (Postdoctoral Researcher)

경기도 고양시 일산서구 고양대로 283, 10223

283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do 283 10223, Republic of Korea

1 이복진 (박사후연구원) / Bokjin Lee (Postdoctoral Researcher)

경기도 고양시 일산서구 고양대로 283, 10223

283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do 283 10223, Republic of Korea

1,3 오현제 (선임연구위원/교수) / Hyun Je Oh (Senior Researcher/Professor)

경기도 고양시 일산서구 고양대로 283, 10223

283, Goyang-daero, Ilsanseo-gu, Goyang-si, Gyeonggi-do 283 10223, Republic of Korea

대전광역시 유성구 가경로 217, 34113

217, Gajeong-ro, Yuseong-gu, Daejeon 34113, Republic of Korea

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

pp. 347-365

pp. 367-378

pp. 379-390

pp. 391-402

pp. 403-425

pp. 427-433

pp. 435-447

pp. 449-464

요약문

정수처리 공정에서 혼화·응집·침전 공정은 원수 탁도 제거와 처리수 수질 안정성을 결정하는 핵심 단위 공정이다. 그러나 최근 기후와 유역 환경 변화로 인해 원수 수질 변동성이 증가함에 따라 기존 운영자 경험에 의존한 응집제 주입농도 결정 방식에는 한계가 있다. 본 연구에서는 파일럿 규모의 혼화·응집·침전 공정을 대상으로 인공지능 기계학습 기반 응집제 주입농도 예측 모델의 적용 가능성을 평가하고자 하였다. 이를 위해 유입수의 수온, 탁도, pH, 전기전도도, 알칼리도, 유량 및 24시간 전 응집제 주입농도를 입력 변수로 활용하여 Ridge Regression, Lasso Regression, Decision Tree, Random Forest, ANN(Artificial Neural Network), DNN(Deep Neural Network), 그리고 RNN(Recurrent Neural Network) 등 총 7개의 기계학습 모델을 포함하는 Model Bank 형태의 예측 방안을 적용하였다. Model Bank에서의 모델 성능 평가는 RMSE(Root Mean Square Error)를 기준으로 수행하였으며, 각 시점에서 RMSE가 가장 낮은 모델을 선택하여 응집제 주입농도를 산정하였다. Model Bank 기반 분석 결과, Random Forest 모델이 평균 RMSE 0.866로 가장 우수한 예측 성능을 나타냈으며, 모델 선택 빈도에서도 가장 높은 비율을 보였다. 예측모델 기반 산정된 응집제 주입농도를 적용한 결과, 평균 응집제 주입농도는 기존 정수장 운영 조건의 10.67 mg/L에서 7.03 mg/L로 감소하여 약 34.1%의 응집제 사용량 절감 효과가 나타났다. 파일럿 공정에서 측정된 침전 처리수 탁도는 평균 0.33 NTU로 실제 정수장의 평균 0.26 NTU와 유사한 수준을 유지하였으며, 대부분 0.5 NTU 이하로 안정적으로 유지되는 것으로 확인되었다. 이러한 결과는 인공지능 기계학습 기반 예측 모델이 파일럿 규모 응집·침전 공정에서 응집제 주입농도 제어와 침전 공정의 처리수 탁도 관리에 효과적으로 적용될 수 있음을 보여준다. 또한 Model Bank 기반 접근은 공정 조건 변화에 따라 최적 모델을 선택함으로써 단일 모델 기반 예측 값에 비해 보다 안정적인 공정 운영을 가능하게 할 수 있음을 확인하였다.

주제어: 응집제 주입, 주입농도 예측, 정수처리공정, 기계학습, 모델뱅크

ABSTRACT

In water treatment processes, the coagulation-flocculation-sedimentation process is a critical unit operation that determines raw water turbidity removal and the stability of treated water quality. However, increasing variability in raw water quality due to climate change and watershed environmental changes has reduced the effectiveness of conventional coagulant dosage determination methods that rely on operator experience. In this study, a machine learning-based prediction model for coagulant dosage was developed, and its applicability was evaluated using a pilot-scale coagulation-flocculation-sedimentation process. Water temperature, turbidity, pH, electrical conductivity, alkalinity, flow rate, and the coagulant dosage applied 24 hours prior were used as input variables. A Model Bank-based approach incorporating seven machine learning models: Ridge Regression, Lasso Regression, Decision Tree, Random Forest, ANN(Artificial Neural Network), DNN(Deep Neural Network), and RNN(Recurrent Neural Network) was applied. Model performance was evaluated using the RMSE(Root Mean Square Error), and comparative analysis was conducted to identify the most suitable model. The results showed that the Random Forest model exhibited the best overall prediction performance, with an average RMSE of 0.866 and the highest selection frequency. When applied to the pilot process, the predicted coagulant dosage reduced the average dosage from 10.67 mg/L under conventional operation to 7.03 mg/L, corresponding to a reduction of approximately 34.1%. In addition, the turbidity of the sedimentation effluent in the pilot system was maintained at an average of 0.33 NTU, comparable to 0.26 NTU observed in the full-scale plant, and remained below 0.5 NTU for most of the operational period. These results demonstrate that machine learning-based prediction models can be effectively applied to optimize coagulant dosing and maintain stable effluent turbidity in pilot-scale water treatment processes.

Key words: Coagulant dosing, Dosage prediction, Drinking water treatment process, Machine learning, Model bank

1. 서론

정수처리 공정에서 혼화·응집·침전 단계는 원수 내 부유물질 및 미세입자를 제거하여 탁도를 저감하는 핵심 단위 공정으로, 전체 정수처리 공정의 처리 효율

과 최종 처리수 수질을 결정하는 중요한 역할을 한다. 특히 응집 공정에서의 응집제 주입농도는 원수 수질 특성에 따라 처리 성능에 직접적인 영향을 미치며, 주입농도가 적절히 제어되지 않을 경우 침전 효율 저하, 처리수 탁도 증가, 슬러지 발생량 증가 등 다양한 운



영상의 문제가 발생할 수 있다 (Kim and Parnichkun, 2017; Lin et al., 2023a).

최근 기후 변화에 따른 극한 강우의 증가와 유역 환경 변화로 인해 정수장 유입 원수의 수질 변동성이 증가하고 있다. 특히 원수 탁도는 단기간 내 급격한 변동을 보이는 대표적인 수질 인자로, 이러한 변동성 증가는 기존의 경험 기반 운전 방식만으로 안정적인 공정 제어를 수행하는 데 어려움을 야기하고 있다. 원수 수질 변화에 적절히 대응하지 못할 경우 응집제의 과주입 또는 과소주입이 발생할 수 있으며, 이는 처리수 수질 저하뿐만 아니라 약품 사용량 증가 및 운영비 상승으로 이어질 수 있다 (Hyun and Kim, 2024; McKelvey et al., 2024).

이러한 문제를 해결하기 위해 최근 정수처리 공정에서 데이터 기반 분석 및 기계학습(ML, Machine Learning) 기법을 적용하려는 연구가 활발히 수행되고 있다. 다양한 연구에서는 인공신경망(ANN, Artificial Neural Network), 랜덤 포레스트(Random Forest), 딥러닝 모델 등을 활용하여 응집제 주입농도 또는 처리수 탁도를 예측하는 연구 결과를 보고하였다 (Gonzales Medina et al., 2023; Jayaweera et al., 2019; Kim et al., 2024; Lin et al., 2023b; Lee et al., 2024). 이러한 접근은 공정 변수 간의 비선형 관계를 효과적으로 반영할 수 있다는 장점을 가지며, 정수처리 공정 운영의 데이터 기반 의사결정 지원 도구로 활용될 가능성을 보여주고 있다 (Bachri et al., 2025; Feng et al., 2025).

그러나 기존 연구의 상당수는 특정 기계학습 모델을 단일하게 적용하거나 실험실 규모의 제한된 수질 조건에서 수행된 경우가 많아, 실제 정수장에서 관측되는 복잡한 수질 변동 특성과 운전 환경을 충분히 반영하지 못한 한계를 가진다 (Li et al., 2021; Wang et al., 2025). 또한 예측모델의 정확도 평가에 초점을 둔 연구가 많아, 다양한 기계학습 모델 간의 성능 비교를 통해 공정 조건에 적합한 모델을 선택하는 접근은 상대적으로 제한적으로 이루어졌다 (Mahesh et al., 2024). 더 나아가 파일럿 규모 공정을 활용하여 예측모델을 실제 운영 조건에서 검증한 연구 역시 매우 제한적인 실정이다 (Namane et al., 2025; Park et al., 2024).

이와 같은 한계를 고려하여 본 연구에서는 다양한 기계학습 알고리즘을 기반으로 하는 모델뱅크(Model Bank) 접근을 적용하였다. 모델뱅크는 Random Forest, Decision Tree, 회귀모델, 인공신경망 등 여러 기계

학습 모델을 동시에 구축한 후, 성능 평가지표를 기반으로 최적의 예측모델을 선택하는 방식이다. 본 연구에서는 RMSE를 평가지표로 활용하여 공정 조건에 적합한 모델을 선정하고 이를 통해 침전 공정에서의 탁도와 응집제 주입농도를 예측하였다 (An et al., 2025).

또한 실제 정수장 원수를 이용하여 혼화·응집·침전 공정을 모사한 파일럿 규모 공정을 구축하고, 정수장 운영 데이터와 파일럿 계측 데이터를 기반으로 예측모델의 성능과 적용 가능성을 평가하였다.

본 연구는 다양한 기계학습 알고리즘을 활용한 모델뱅크 기반 예측모델을 구축하고 이를 실제 정수장 유입 원수를 대상으로, 파일럿 규모 응집·침전 공정에 적용하여 성능을 검증하였다는 점에서 기존 연구와 차별성을 갖는다. 이러한 연구 결과는 향후 정수처리 공정에서 학습 데이터 기반 공정 운영 및 응집제 주입 최적화를 위한 기초 자료로 활용될 수 있을 것으로 기대된다.

2. 연구방법

2.1 데이터 수집

본 연구에 사용된 데이터는 G 정수장의 실제 유입 원수를 이용하여 구축된 파일럿 규모 혼화·응집·침전 공정과 정수장 SCADA(Supervisory Control and Data Acquisition) 시스템으로부터 수집하였다. 파일럿 공정은 정수장의 원수를 직접 유입 받아 혼화·응집·침전 공정을 모사하도록 구성되었으며, 공정 운영 과정에서 주요 수질 및 운전 변수를 계측하였다.

연구에 활용된 주요 변수는 유입수의 수온, 탁도, pH, 전기전도도, 알칼리도, 유량 및 24시간 전 응집제 주입농도로 구성하였다. 이러한 변수들은 정수처리 공정에서 응집 반응 및 침전 효율에 영향을 미치는 주요 운전 인자로 알려져 있으며, 이를 기계학습 기반 예측모델의 입력 변수로 활용하였다. 사용된 주요 수질 및 운전 변수의 구성과 단위는 Table 1에 정리하였다.

2024년 11월부터 2025년 11월 까지 1년 간의 정수장 SCADA 시스템으로부터 확보한 실제 운영데이터를 바탕으로 초기모델 학습을 수행하였고, 2025년 11월부터 약 3개월 동안 파일럿 공정 운영을 통해 계측된 데이터와 비교하여 모델의 성능을 검증하였다. 각

pp. 347-365

pp. 367-378

pp. 379-390

pp. 391-402

pp. 403-425

pp. 427-433

pp. 435-447

pp. 449-464

Table 1. Input variables used in the machine learning model

Variable	Description	Unit	Data Range
Temperature	Temperature of influent raw water	°C	3.55 – 14.8
Turbidity	Influent turbidity of raw water	NTU	0.45 – 10.2
pH	Raw water pH	-	8.32 – 8.6
Electrical conductivity	Raw water electrical conductivity	μS/cm	191 – 279
Alkalinity	Raw water alkalinity	mg/L as CaCO ₃	23.28 – 51.27
Flow rate	Flow rate of influent raw water	mL/min	2.47 – 4.34
Coagulant dosage (t-1day)	Coagulant dosage applied on the previous day	mg/L	2.16 – 9.54

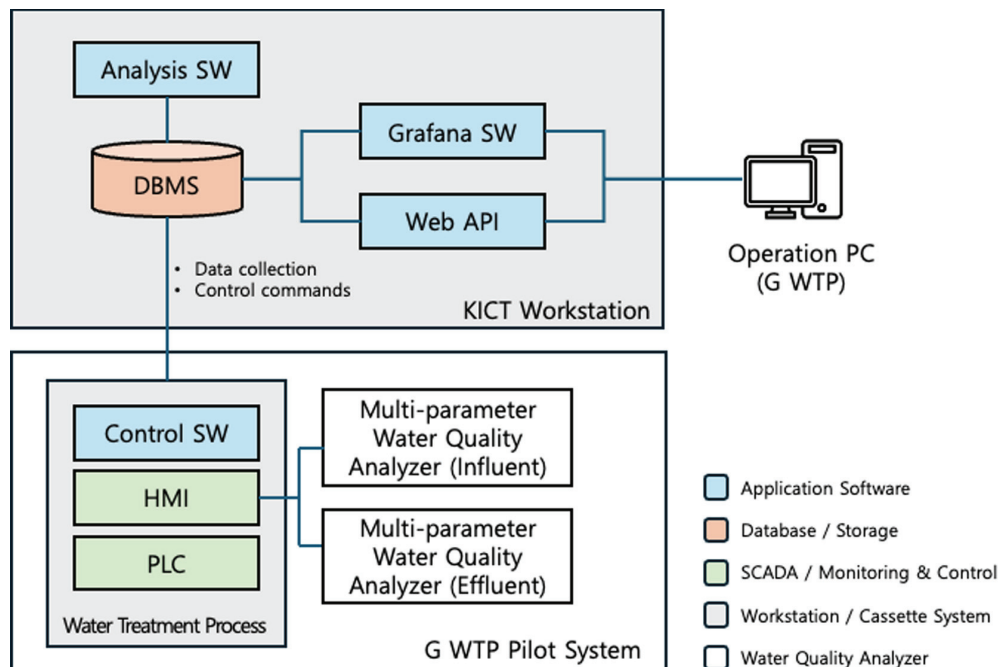
변수는 1시간 간격으로 수집된 시계열 데이터로 구성하였다.

구축된 데이터셋은 기계학습 기반 예측모델의 학습 및 검증에 활용되었으며, 모델 예측 결과는 실제 정수장 운영 데이터와 파일럿 공정 계측 데이터를 비교·분석하여 모델 성능 및 공정 적용 가능성을 평가하는데 사용하였다. Fig. 1은 본 연구에서 구축한 데이터 수집 및 공정 제어 시스템의 구성도를 나타낸 것이다. Fig. 1. 파일럿 장치에서는 PLC와 CIMON 기반 HMI (Human-Machine Interface)를 이용하여 공정 운전 및 계측 데이터 수집이 이루어지며, 제어 소프트웨어를

통해 공정 운전 상태를 관리한다. 수집된 공정 데이터는 KICT Workstation으로 전송되어 DBMS(Data Base Management System)에 저장되며, 이후 분석 소프트웨어를 통해 데이터 처리 및 예측모델 학습에 활용된다. 또한 Grafana 기반 모니터링 시스템과 Web API를 통해 공정 데이터의 실시간 조회 및 제어 명령 전달이 가능하도록 시스템을 구성하였다.

2.2 예측모델 구성

정수처리 공정에서의 응집제 주입농도를 예측하기 위해 다양한 기계학습 알고리즘을 기반으로 한 Model

**Fig. 1.** Architecture of the data acquisition and process control system.



Bank를 적용하였다. Model Bank는 서로 다른 구조와 특성을 갖는 여러 기계학습 모델을 동시에 구축하고, 각 모델의 예측 성능을 비교하여 최적의 모델을 선택하는 방식으로 구성된다. 이러한 접근 방식은 단일 모델에 의존하는 방법보다 데이터의 복잡한 패턴과 비선형 관계를 보다 효과적으로 반영할 수 있다는 장점을 가진다.

본 연구에서는 총 7개의 기계학습 모델을 Model Bank에 포함하였으며, Fig. 2와 같이 나타내었다. Fig. 2. 선형 회귀 기반 모델로는 Ridge Regression과 Lasso Regression을 적용하였으며, 트리 기반 모델로는 Decision Tree와 Random Forest를 활용하였다. 또한 비선형 관계를 효과적으로 학습하기 위해 ANN(Artificial Neural Network), DNN(Deep Neural Network), RNN(Recurrent Neural Network) 모델을 추가적으로 적용하였다. 각 모델은 동일한 입력 변수 집합을 기반으로 학습하도록 구성하였으며, 모델의 출력 변수는 응집제 주입농도로 설정하였다. 또한 예측된 응집제 주입농도는 실제 공정 운영에 적용할 수 있도록 응집제 주입 펌프의 회전 속도로 변환하여 공정 운전에서 적용할 수 있도록 하였다. 이를 통해 모델 예측 결과가 단순한 농도 예측에 그치지 않고, 실제 혼화·응집 공정에서 응집제 주입 펌프의 운전 제어에 직접 활용될 수 있도록 하였다.

또한 본 연구에서는 정수처리 공정의 실제 운전 특성을 반영하기 위해 시계열 데이터 기반의 예측 구조를 적용하였다. 이를 위해 현재 시점의 수질 및 운전

변수를 입력 데이터로 활용하고, 일정 시간 이후의 응집제 주입농도를 예측 대상으로 설정하였다. 특히 공정 운영의 일 단위 의사결정 구조를 반영하기 위해 전일 24시간 동안의 수질 및 운전 데이터를 기반으로 다음날 24시간 동안의 응집제 주입농도를 예측하도록 모델을 구성하였다.

각 모델은 동일한 입력 데이터를 기반으로 예측을 수행하며, 이후 성능 평가 지표를 통해 모델별 예측 성능을 비교하여 최적 모델을 선정한다. 선정된 모델의 예측 결과는 실제 공정 운전에서 적용 가능한 응집제 주입농도 및 응집제 주입 펌프 회전 속도 산정에 활용되도록 구성하였다. 이러한 Model Bank 기반 예측 구조는 정수처리 공정의 수질 변동성에 대응하여 보다 안정적인 응집제 주입 전략을 도출하는 데 활용될 수 있다.

2.3 데이터 전처리 및 모델 학습

수집된 데이터는 모델 학습에 활용하기 전에 전처리 과정을 수행하였다. 먼저 센서 측정 과정에서 발생할 수 있는 결측값을 보완하기 위해 공정 운전 범위를 벗어나는 비정상적인 데이터는 이상치로 판단하여 제거하였으며, 이후 선형 보간(Interpolation)을 적용하여 전처리된 데이터셋을 기반으로 모델 학습에 활용할 입력 변수와 예측 대상 변수를 구성하였다.

모델 학습 과정에서는 입력 변수 간의 비선형 관계를 반영하기 위해 회귀 기반 모델에 대해 다항 특성

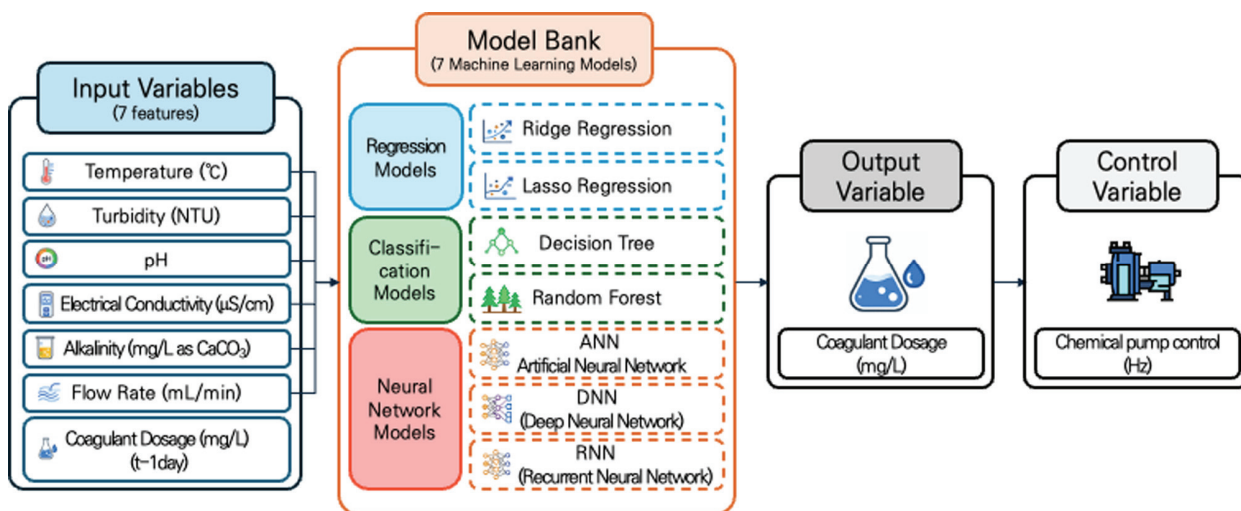


Fig. 2. Structure of the model bank framework used for predicting coagulant dosage and chemical pump control based on multiple machine learning models.

변환(Polynomial Feature Transformation)을 적용하였다. 이후 변수 간 스케일 차이에 따른 영향을 최소화하기 위해 표준화(Standardization)를 수행하였다. 신경망 기반 모델의 경우 입력 변수에 대해 동일한 표준화 과정을 적용한 후 학습에 활용하였다.

구성된 데이터셋은 모델 학습과 성능 평가를 위해 학습(Train), 시험(Test) 및 검증(Validation) 데이터로 6:2:2 분할하였으며, 시계열 순서를 유지하여 분할하였다. 학습 데이터는 모델 파라미터 학습에 사용되었으며, 시험 데이터는 최종 모델의 예측 성능을 평가하기 위한 독립적인 데이터셋으로 활용하였다. 검증 데이터는 모델 학습 과정에서 과적합(Overfitting)을 방지하고 모델의 일반화 성능을 확인하기 위한 용도로 활용하였다.

각 기계학습 모델은 동일한 데이터셋과 동일한 입력 변수 구조를 기반으로 학습되었으며, 모델 학습 이후 동일한 입력 데이터에 대해 응집제 주입농도를 예측하였다. 이후 모델뱅크에 포함된 각 모델의 예측 결과를 비교하여 최적 모델을 선정하고 이를 공정 운전에 적용하였다.

2.4 모델 구현

제안한 예측모델은 Python 3.8 기반 환경에서 구현하였다. 데이터 처리 및 전처리에는 Pandas 2.2.3과 Numpy 2.1.3 라이브러리를 활용하였으며, 기계학습 모델 구현에는 Scikit-learn 1.6.1 라이브러리를 사용하였다. 또한 신경망 기반 모델(ANN, DNN, RNN)의 구현 및 학습에는 TensorFlow 2.19.0, Keras 3.10.0 프레임워크를 활용하였다.

모델 실행은 하루 1회 수행되며, 전일 24시간 동안의 수질 및 운전 데이터를 입력 변수로 사용하여 다음날 24시간 동안의 응집제 주입농도를 예측하도록 구성하였다. 이러한 방식은 정수처리 공정의 실제 운전 특성을 반영하여 일 단위 운전 의사결정에 활용할 수 있도록 하였다.

각 모델은 동일한 입력 데이터를 기반으로 예측을 수행하며, 이후 성능 평가 지표를 이용하여 모델별 예측 성능을 비교하였다. Model Bank에 포함된 모델 중 가장 우수한 성능을 보이는 모델을 최적 모델로 선정하고 해당 모델의 예측 결과를 공정 운전에 적용하도록 구성하였다.

예측된 응집제 주입농도는 실제 약품 주입 펌프의 운전 조건에 적용하기 위해 선형 변환식을 통해 제어 신호로 변환하였다. 이를 통해 예측된 응집제 요구량을 펌프 운전 주파수 등 공정 제어 변수로 변환하여 약품 주입 펌프의 운전을 자동으로 조정할 수 있도록 하였다.

2.5 성능 평가 지표

Model Bank에 포함된 각 기계학습 모델의 예측 성능을 정량적으로 평가하기 위해 평균제곱근오차(RMSE)를 활용하였다. RMSE(Root Mean Square Error)는 예측값과 실제값 간의 오차의 제곱 평균의 제곱근으로 정의되며, 예측 오차의 절대적인 크기를 평가하는 지표이다. RMSE 값이 낮을수록 모델의 정확도가 높은 것을 의미한다. RMSE는 다음과 같이 계산된다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

여기서 y_i 는 실제 응집제 주입농도, $\hat{y}_i$ 는 모델 예측값, 그리고 n 은 데이터의 개수를 의미한다.

본 연구에서는 Model Bank에 포함된 각 기계학습 모델에 대해 RMSE 값을 산출하여 예측 성능을 비교·분석하였다. 각 시점에서 계산된 RMSE 값을 기반으로 모델 성능을 평가하였으며, 가장 낮은 RMSE 값을 나타내는 모델을 해당 시점의 최적 모델로 선정하였다.

선정된 최적 모델의 예측 결과는 응집제 주입농도 산정에 활용되었으며, 이를 기반으로 파일럿 공정에서의 응집제 주입 제어에 적용하였다. 이러한 성능 평가 및 모델 선택 과정은 다양한 기계학습 모델 중 공정 조건에 가장 적합한 모델을 선택하기 위한 Model Bank 기반 접근의 핵심 절차이다.

2.6 파일럿 플랜트 공정 구성

실제 정수처리 공정의 혼화·응집·침전 단계를 현장 조건에 맞게 재현하고 다양한 운전 조건에서의 공정 거동을 실증적으로 평가하기 위해 파일럿 규모 혼화·응집·침전 공정을 구축하였다. 구축된 파일럿 시스템은 혼화 및 응집 공정을 담당하는 모듈과 침전 공정을 담당하는 모듈로 구성된 2단 구조로 설계되었



으며, 시스템의 구성 및 외관은 Fig. 3에 나타내었다. (Fig. 3).

혼화·응집 모듈은 원수와 응집제가 혼합되어 플록이 형성되는 과정을 모사하도록 설계되었다. 해당 모듈은 혼화조 1기와 3단의 응집조가 직렬로 배치된 구조로 구성되었다. 응집제는 G정수장에서 사용하고 있는 17% PAC 용액을 파일럿 설비 규모에 맞도록 희석해서 사용하였다. 혼화조 유입부에 설치된 주입 라인을 통해 주입되며, 주입농도는 PLC 기반 제어 시스템을 통해 조절되도록 구성하였다. 각 반응조의 교반 조건은 실제 정수장 운전 조건을 고려하여 혼화조 50 rpm, 응집조 단계별로 35 rpm 및 20 rpm으로 설정하였다. 또한 반응조 내부에서의 플록 형성 과정을 확인할 수 있도록 각 반응조는 투명 아크릴 재질로 제작하였다.

침전 모듈은 응집 공정에서 형성된 플록의 침전 거동을 모사하도록 설계된 침전조로 구성하였다. 침전조 유입부에는 유입수의 유속을 균일하게 분배하기 위한 분배 장치를 설치하였으며, 침전조 내부 체류시간은 약 103.68 min이 유지되도록 설계하였다. 침전조 출구부에는 다항목 수질 측정 장치를 설치하여 처리수의 탁도 및 주요 수질 항목을 연속적으로 계측할 수 있도록 하였다.

파일럿 공정은 현장 설치 및 운전의 유연성을 확보하기 위해 이동식 구조로 제작되었으며, 각 모듈은 일체형 프레임 구조로 구성하여 실제 정수장 현장 조건에 따라 시스템 배치 및 운전 조건을 유연하게 조정할 수 있도록 하였다. 또한 본 파일럿 공정은 환경부 정수시설 설계 기준을 준수하여 설계·제작되었으며, 주요 설계 및 운전 조건은 실제 정수장 설계와 운전 조건을 반영하여 설정하였다.

공정 운전 및 데이터 수집은 PLC 및 HMI 기반 제어·모니터링 시스템을 통해 수행되었다. 공정 각 지점에 설치된 센서로부터 취득된 수질 및 운전 데이터는 로컬 데이터베이스에 실시간으로 저장되었으며, 이후 예측모델 구축 및 성능 검증을 위한 데이터로 활용되었다.

3. 결 과

3.1 기계학습 모델 성능 비교

정수처리 공정에서의 응집제 주입농도 예측 성능을 비교하기 위해 Model Bank에 포함된 여러 기계학습 모델의 예측 성능을 평가하였다. Fig. 4는 Model Bank에 포함된 각 기계학습 모델의 예측 오차(RMSE)를 시간 순서에 따라 비교한 결과와 각 시점에서 선택된

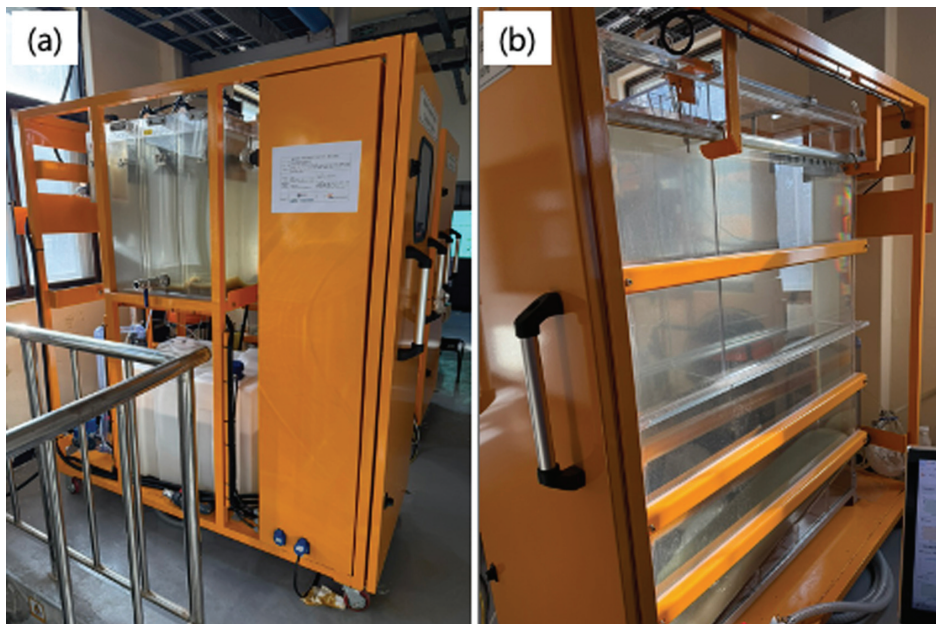


Fig. 3. (a) Coagulation module, (b) Sedimentation module.

최적 모델을 함께 나타낸 것이다. Fig. 4(a)는 모델별 RMSE 값을 색상 분포 형태로 표현한 것으로, 적색에 가까울수록 높은 예측 오차를, 녹색에 가까울수록 낮은 예측 오차를 의미한다. 그림을 통해 동일한 공정 조건에서도 모델별 예측 성능이 시간에 따라 서로 다르게 나타나는 경향을 확인할 수 있었다. 특히 일부 기간에서는 Random Forest와 신경망 기반 모델(ANN, DNN, RNN)이 상대적으로 낮은 RMSE 값을 나타내는 구간이 관찰되었으며, 반면 Decision Tree 모델은 전반적으로 높은 RMSE 값을 나타내는 경향을 보였다. 이러한 결과는 동일한 입력 조건에서도 모델 구조에 따라 예측 성능이 시간적으로 변동할 수 있음을 보여준다.

Fig. 4(b)는 각 시점에서 RMSE가 가장 낮은 모델을 선택한 결과를 나타낸 것이다. 분석 결과 Random Forest 모델이 총 1,514회로 가장 높은 빈도로 최적 모델로 선택된 것으로 나타났다. 또한 RNN 모델은 626

회, Ridge Regression 모델은 560회 선택되었으며, ANN과 DNN 모델은 각각 214회와 186회 선택된 것으로 확인되었다. 반면 Lasso Regression 모델은 39회로 상대적으로 낮은 선택 빈도를 보였으며, Decision Tree 모델은 전체 기간 동안 최적 모델로 선택되지 않는 것으로 나타났다. 이러한 결과는 공정 조건 변화에 따라 모델별 예측 성능이 시간적으로 변동할 수 있음을 보여준다.

각 모델의 전체 예측 성능을 정량적으로 비교하기 위해 RMSE를 주요 성능 평가 지표로 사용하였으며, 전체 기간에 대한 평균 RMSE와 표준편차를 산출하였다. 그 결과는 Table 2에 정리하였다. Table 2. 분석 결과 Random Forest 모델이 평균 RMSE 0.866 (± 0.249)로 가장 낮은 값을 나타내어 전체적으로 가장 우수한 예측 성능을 보이는 것으로 확인되었다. DNN과 RNN 모델 또한 각각 0.867 (± 0.260)과 0.868 (± 0.259)로 유사한 수준의 RMSE 값을 나타내며 비교적 안정적인

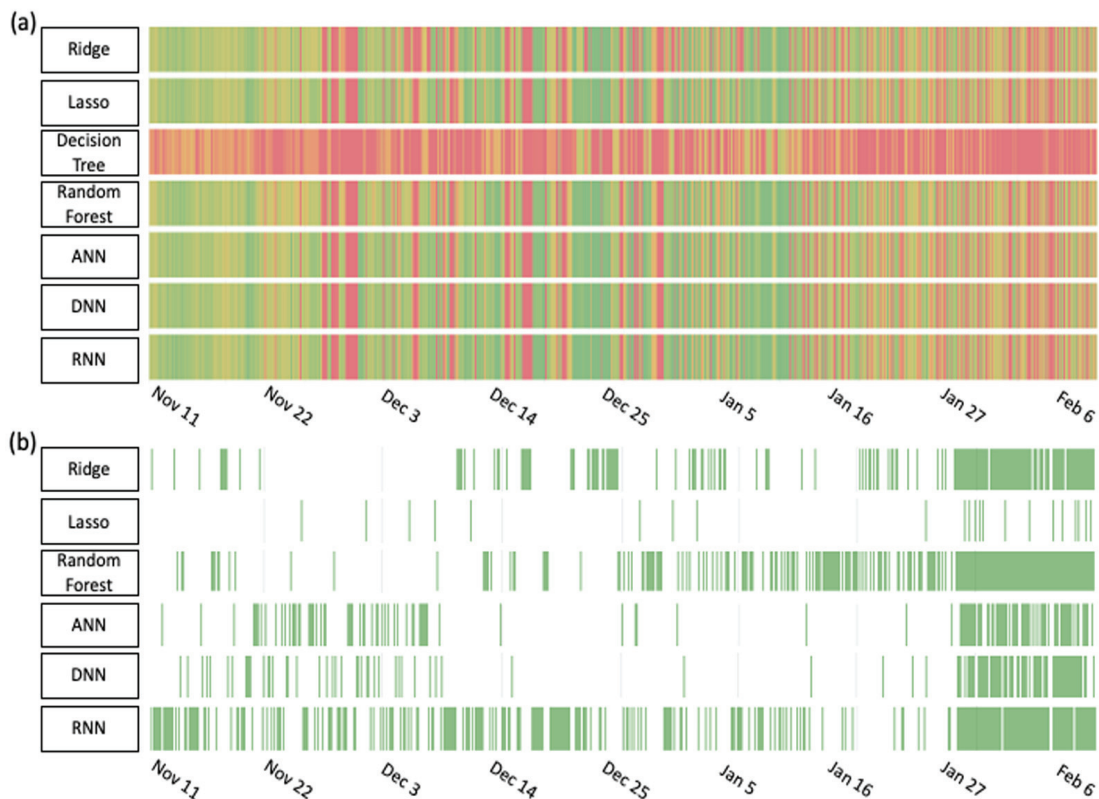


Fig. 4. Performance comparison of machine learning models in the Model Bank: (a) Temporal distribution of RMSE values (red indicates higher prediction error, while green indicates lower prediction error), (b) Optimal model selection based on the lowest RMSE.



Table 2. Performance comparison of machine learning models in the model bank

Model	Mean RMSE ($\pm$ SD)	Selection Frequency
Ridge Regression	0.875 ($\pm$ 0.272)	560
Lasso Regression	0.872 ($\pm$ 0.260)	39
Decision Tree	1.196 ($\pm$ 0.293)	-
Random Forest	0.866 ($\pm$ 0.249)	1,514
ANN	0.870 ($\pm$ 0.260)	214
DNN	0.867 ($\pm$ 0.260)	186
RNN	0.868 ($\pm$ 0.259)	626

예측 성능을 보였다. 반면 Decision Tree 모델은 평균 RMSE 1.196 ($\pm$ 0.293)으로 가장 높은 예측 오차를 나타냈으며, 이는 단일 트리 구조가 복잡한 수질 데이터의 비선형 관계를 충분히 반영하지 못했기 때문으로 판단된다.

이와 함께 평균 RMSE 기준으로 가장 우수한 성능을 보인 모델과 실제 모델뱅크에서 가장 많이 선택된 모델이 동일하게 나타난 점 또한 확인되었다. Random Forest 모델은 평균 RMSE 기준에서도 가장 낮은 값을 나타냈으며, 동시에 모델 선택 빈도에서도 가장 높은 비율을 차지하였다. 이는 Random Forest 모델이 다양한 공정 조건에서 상대적으로 안정적인 예측 성능을 유지했음을 의미한다. 한편 Ridge Regression과 RNN 모델 역시 일부 구간에서 최적 모델로 선택되며 모델뱅크 내에서 보조적인 역할을 수행하는 것으로 나타났다.

이러한 결과는 정수처리 공정과 같이 원수 수질 및 운전 조건이 시간에 따라 변화하는 환경에서는 단일 모델이 모든 조건에서 항상 최적의 성능을 보장하기 어렵다는 점을 보여준다. 반면 모델뱅크 접근 방식은 다양한 기계학습 모델을 동시에 활용하고 각 시점에

서 가장 우수한 성능을 나타내는 모델을 선택함으로써 공정 조건 변화에 보다 유연하게 대응할 수 있다. 따라서 모델뱅크 기반 예측 구조는 단일 모델 기반 접근 방식보다 응집제 주입농도 예측 정확도를 향상시키는 데 효과적인 방법으로 판단된다.

3.2 최적 모델 기반 응집제 주입농도 제어 적용 결과

기계학습 기반 예측모델을 활용하여 산정된 응집제 주입농도의 적용 가능성을 평가하였다. 평가는 정수장에서 실제로 운전된 응집제 주입농도(실제값)와 예측모델을 통해 산정된 응집제 주입농도(검증값)을 비교하는 방식으로 수행하였다. 분석에는 파일럿 공정 운영 기간 동안 수집된 데이터를 일평균 값으로 정리하여 활용하였다.

Fig. 5는 파일럿 공정 운영 기간 동안의 실제 응집제 주입농도와 예측모델 기반 제어를 통해 산정된 응집제 주입농도의 변화를 비교한 결과를 나타낸 것이다. Fig. 5. 실제 정수장 운영에서는 응집제 주입농도가 대부분 약 10~11 mg/L 수준에서 비교적 일정하게 유지되는 경향을 보였다. 이는 원수 수질 변동에 대한 불확실성을 고려하여 일정 수준 이상의 응집제를 지속적으로 주입하는 기존 정수장 운영 방식의 특성을 반영한 결과로 해석된다.

반면, 예측모델 기반 응집제 주입농도는 전반적으로 약 7 mg/L 수준에서 안정적으로 유지되는 경향을 보였다. 실제 운영 데이터와 예측모델 결과를 평균값 기준으로 비교한 결과, 실제 정수장 운영 시 평균 응집제 주입농도는 약 10.67 mg/L로 나타났으며, 예측모델 기반 제어 결과는 평균 약 7.03 mg/L 수준으로 나타났다. 이는 기존 정수장 운영 대비 약 34.1%의 응집제 사용량 절감 효과에 해당한다.

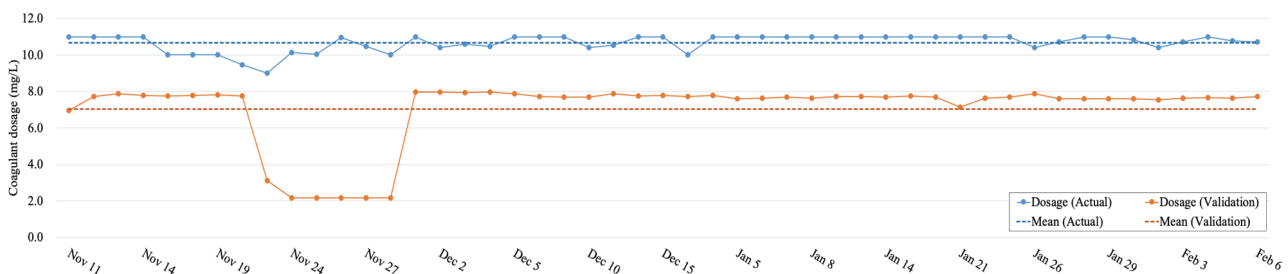


Fig. 5. Comparison of coagulant dosage between the G plant operation and the model-based control.

또한 일부 기간에서는 예측모델 기반 응집제 주입 농도가 기존 운영 대비 크게 감소하는 구간이 확인되었다. 특히 11월 하순의 특정 구간에서는 검증 주입농도가 일정 수준 이하로 낮아지면서 동일한 값이 반복적으로 나타나는 경향이 관찰되었다. 이러한 패턴은 단순한 수질 변화에 따른 연속적 감소라기보다, 예측값을 실제 주입 제어값으로 변환하는 과정에서 적용된 최소 주입 한계, 펌프 운전 하한값, 또는 제어 로직상의 제한 조건이 반영된 결과일 가능성을 시사한다. 그럼에도 불구하고 해당 기간 동안 침전조 처리수 탁도는 안정적인 수준을 유지하는 것으로 나타나, 저주입 조건에서도 침전 공정의 처리 성능이 크게 저하되지 않았음을 확인할 수 있었다.

이와 같은 결과는 기계학습 기반 예측모델을 활용한 응집제 주입농도 제어 전략이 공정 안정성을 유지하면서도 약품 사용량을 효과적으로 저감할 수 있음을 보여준다. 기존 정수처리 공정에서는 원수 수질 변동에 대한 불확실성을 고려하여 비교적 보수적인 약품 주입 전략이 적용되는 경우가 많다. 반면 본 연구에서 제안한 기계학습 기반 예측 및 제어 접근은 원수 수질 조건과 공정 상태를 반영하여 보다 합리적인 응집제 주입농도를 제시할 수 있으며, 이를 통해 정수처리 공정의 약품 사용 효율을 향상시킬 수 있는 가능성을 확인하였다.

3.3 모델 적용 전후 처리수 탁도 비교 평가

기계학습 기반 응집제 주입 전략이 적용된 파일럿 시스템의 침전 공정 성능을 평가하기 위해 G 정수장에서 측정된 실제 침전조 처리수 탁도와 파일럿 시스템에서 측정된 처리수 탁도를 비교·분석하였다. 실제값은 G 정수장에서 측정된 침전조 처리수 탁도를 의미하

며, 검증값은 동일한 원수를 이용하여 운전된 파일럿 시스템에서 측정된 침전조 처리수 탁도를 의미한다.

Fig. 6는 파일럿 운영 기간 동안 G 정수장에서 측정된 실제 침전조 처리수 탁도와 파일럿시스템에서 측정된 처리수 탁도의 변화를 시간 순서에 따라 나타낸 것이다. Fig. 6. 분석 결과 G 정수장의 실제 침전조 처리수 탁도는 평균 0.26 NTU (± 0.10) 수준으로 나타났으며, 파일럿시스템에서 측정된 처리수 탁도는 평균 0.33 NTU (± 0.14) 수준으로 나타났다. 이는 파일럿 공정과 실제 정수장 간 규모 및 유동 조건의 차이에 기인한 것으로, 실제 정수장 대비 약 0.07 NTU 높은 값을 나타내는 경향이 확인되었다.

실제 정수장의 침전조 처리수 탁도는 대부분 0.2~0.3 NTU 범위에서 안정적으로 유지되는 경향을 보였으며, 파일럿시스템에서도 전반적으로 유사한 수준의 탁도 범위를 나타냈다. 다만 일부 구간에서는 파일럿 시스템에서 측정된 처리수 탁도가 실제 정수장보다 다소 높은 값을 나타내는 경향이 관찰되었다. 이러한 차이는 파일럿시스템의 공정 규모 및 운전 조건 차이에 따른 것으로 판단되며, 특히 응집제 주입농도 감소에 따른 플록 형성 조건 변화가 일부 영향을 미친 것으로 해석된다.

그럼에도 불구하고 파일럿시스템의 침전조 처리수 탁도는 대부분 0.5 NTU 이하의 안정적인 수준을 유지하는 것으로 나타났다. 이는 국내 먹는물 수질기준에서 제시하는 0.5 NTU의 탁도 기준을 만족하는 수준이다. 따라서 파일럿시스템에서 적용된 기계학습 기반 응집제 주입 전략은 약품 사용량 감소에도 불구하고 침전 공정의 처리 성능을 크게 저하시키지 않으면서 안정적인 수질을 유지할 수 있음을 확인하였다.

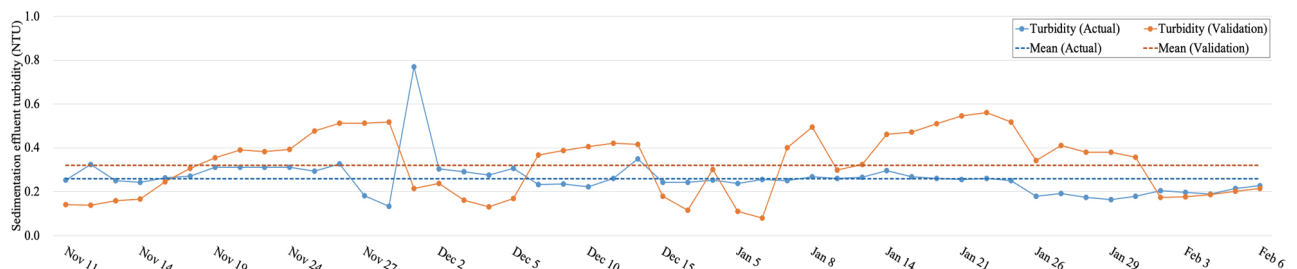


Fig. 6. Comparison of sedimentation effluent turbidity between the G plant and the pilot system.



4. 결 론

본 연구에서는 파일럿 규모의 혼화·응집·침전 공정을 대상으로 기계학습 기반 응집제 주입농도 예측 모델의 적용 가능성을 평가하였다. 이를 위해 다양한 기계학습 모델을 포함하는 Model Bank 기반 예측 구조를 구축하고, 실제 정수장 원수를 이용한 파일럿 공정 실증을 통해 공정 성능과 운영 효과를 분석하였다. 주요 연구 결과는 다음과 같다.

첫째, Model Bank에 포함된 7개의 기계학습 모델(Ridge Regression, Lasso Regression, Decision Tree, Random Forest, ANN, DNN, RNN)의 예측 성능을 비교한 결과, Random Forest 모델이 평균 RMSE 0.866, DNN 모델이 평균 RMSE 0.867, RNN 모델이 평균 RMSE 0.868로 나타났다. 그러나, 선택 빈도를 보면 Random Forest 모델이 1,514회로 약 48.2%, RNN 모델이 626회로 약 19.9%, Ridge Regression 모델이 560회로 약 17.8% 선택되어 특정 단일 모델이 모든 시점에서 항상 최적의 예측 성능을 나타내는 것은 아님을 확인할 수 있었다. 이는 데이터 특성 및 공정 조건에 따라 모델 성능이 달라질 수 있음을 의미하며, 단일 모델에 의존하기보다는 Model Bank 기반 접근을 통해 상황별로 최적 모델을 선택하는 방식이 보다 안정적인 예측 결과를 제공할 수 있음을 보여준다.

둘째, 최적모델 기반 응집제 주입농도 제어를 적용한 결과, 평균 응집제 주입농도는 기존 정수장 운영 조건의 10.67 mg/L에서 7.03 mg/L로 감소하여 약 34.1%의 응집제 사용량 절감 효과가 나타났다. 특히 일부 구간에서는 예측모델 기반 제어에 따라 주입농도가 크게 감소하는 현상이 관찰되었으며, 이러한 저주입 조건에서도 침전조 처리수 탁도는 안정적인 수준을 유지하는 것으로 확인되었다.

셋째, 파일럿 공정에서의 침전조 처리수 탁도를 분석한 결과, 평균 0.33 NTU 수준으로 나타났으며, 이는 실제 정수장의 평균 탁도 0.26 NTU와 유사한 수준을 보였다. 또한 대부분의 처리수 탁도가 0.5 NTU 이하로 유지되어 국내 먹는물 수질기준을 만족하는 것으로 확인되었다. 이는 기계학습 기반 응집제 주입 전략이 적용된 조건에서도 침전 공정의 처리 성능이 안정적으로 유지될 수 있음을 보여준다.

종합적으로 본 연구 결과는 기계학습 기반 예측 모

델이 파일럿 규모의 정수처리 공정에서 응집제 주입농도 예측과 침전 공정의 처리수 탁도 관리에 효과적으로 적용될 수 있음을 보여준다. 또한 Model Bank 기반 접근 방식은 공정 조건 변화에 따라 최적 모델을 선택함으로써 단일 모델 기반 예측에 비해 보다 안정적인 예측 성능을 확보할 수 있는 장점을 가진다. 향후 이러한 데이터 기반 예측 및 제어 기술은 정수처리 공정의 자동화 및 지능형 운영 기술 개발에 활용될 수 있으며, 장기적으로는 정수장 운영 효율 향상과 약품 사용량 절감을 동시에 달성할 수 있는 유효한 방법이 될 것으로 기대된다.

사 사

본 결과물은 환경부의 재원으로 한국환경산업기술원의 상하수도 혁신 기술개발사업의 지원을 받아 연구되었습니다(2020002700003).

References

- An, J., Park, J., Yeon, S., Oh, C., Lee, B., Jung, W., Yun, J. and Oh, H.J. (2025). Machine learning-based prediction of coagulant dosing in drinking water treatment plants using polynomial regression with lasso regularization, *Processes*, 13(2), 3829.
- Bachri, A., Mahamad, A.K., Utomo, W.M., Saon, S., and Laksono, A.B. (2025). Optimization of coagulation process in water treatment plant using artificial intelligence: A systematic literature review and framework, *J. Robot. Control*, 6(5), 2471-2483.
- Feng, L., Zhang, Y., Wei, X., Wang, M., Niu, Z. and Wang, C. (2025). Interpretable prediction of coagulant dosage in drinking water treatment plant based on automated machine learning and SHAP method, *J. Water Process Eng.*, 75, 107925.
- Gonzales Medina, R.I., Mendoza, J.A.C., Zuñiga Goyzueta, E.J., Morán-Silva, R.M., and López-Gonzales, J.L. (2023). Random forest model for optimizing coagulant doses in drinking water treatment: Application at the Miguel de la Cuba Ibarra plant, *Environments*, 10, 69.
- Hyun, J.W. and Kim, C.G. (2024). Selection of the best fitted machine learning algorithms for turbidity prediction using water quality parameters in non-point source pollution

pp. 347-365

pp. 367-378

pp. 379-390

pp. 391-402

pp. 403-425

pp. 427-433

pp. 435-447

pp. 449-464

- monitoring network, *J. Environ. Anal. Health Toxicol.*, 27(4), 249-256.
- Jayaweera, C.D., Othman, M.R., and Aziz, N. (2019). Improved predictive capability of coagulation process by extreme learning machine with radial basis function, *J. Water Process Eng.*, 32.
- Kim, C.M. and Parnichkun, M. (2017). Prediction of settled water turbidity and optimal coagulant dosage in drinking water treatment plant using a hybrid model of k-means clustering and adaptive neuro-fuzzy inference system, *Appl. Water Sci.*, 7, 3885 - 3902.
- Kim, J., Hua, C., Kim, K., Lin, S., Oh, G., Park, M.-H. and Kang, S. (2024). Optimizing coagulant dosage using deep learning models with large-scale data, *Chemosphere*, 350, 140989.
- Lee, S.M., Kang, Y.J., Song, J.W., Kim, J.H., Kim, H.S., and Kim, S.J. (2024). Study on water quality prediction in water treatment plants using AI techniques, *J. Korea Water Resour. Assoc.*, 57(3), 151-164.
- Li, L., Rong, S., Wang, R., and Yu, S. (2021). Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review, *Chem. Eng. J.*, 405, 126673.
- Lin, S., Kim, J., Hua, C., Kang, S., and Park, M.-H. (2023a). Comparing artificial and deep neural network models for prediction of coagulant amount and settled water turbidity: Lessons learned from big data in water treatment operations, *J. Water Process Eng.*, 54, 103949.
- Lin, S., Kim, J., Hua, C., Park, M.-H., and Kang, S. (2023b). Coagulant dosage determination using deep learning-based graph attention multivariate time series forecasting model, *Water Res.*, 232, 119665.
- Mahesh, N., Babu, J.J., Nithya, K., and Arunmozhi, S.A. (2024). Water quality prediction using LSTM with combined normalizer for efficient water management, *Desalin. Water Treat.*, 317, 100183.
- McKelvey, S., Abassi, A., Nataraj, C. and Duran, M. (2024). Data-driven modeling techniques for prediction of settled water turbidity in drinking water treatment, *Front. Environ. Eng.*, 12, 1401180.
- Namane, P.I., Letshwenyo, M.W. and Yahya, A. (2025). Evaluation of plant-based coagulants for turbidity removal and coagulant dosage prediction using machine learning, *Environ. Technol.*, 46(14), 2570-2585.
- Park, K., Lee, Y.J., Noh, H., Heo, J. and Jung, S.H. (2024). Machine learning based coagulant rate decision model for industrial water treatment plant, *J. Korean Soc. Ind. Syst. Eng.*, 47(3), 68-74.
- Wang, Y.Q., Wang, H.C., Xiao, Z.J., Bu, L.J., Li, J., Feng, X.C., Liang, B., Liu, W.Z., Sun, F.Y., Zhou, S.Q. and Wang, A.J. (2025). Machine learning strategy secures urban smart drinking water treatment plant through incremental advances, *Water Res.*, 280, 123541.