

Vol. 26, No. 4, pp. 1031-1040, Apr. 2025



# 수처리공정에서 AI 기반 자율 운영 모델 연구

김 학 성<sup>1</sup>·김 창 배<sup>2</sup>·장 혁<sup>2</sup>·정 회 경<sup>3\*</sup> <sup>1</sup>배재대학교 컴퓨터공학과 박사과정 <sup>2</sup>배재대학교 스마트ICT융합학과 박사과정 3배재대학교 컴퓨터공학과 교수

# Research on AI-Based Autonomous Operation Model in Water Treatment Process

Hak-Seong Kim<sup>1</sup> · Chang-Bae Kim<sup>2</sup> · Hyeok Chang<sup>2</sup> · Hoe-Kyung Jung<sup>3\*</sup>

<sup>1</sup>Doctor's Course, Department of Computer Engineering, Paichai University, Daejeon 35345, Korea <sup>2</sup>Doctor's Course, Department of Smart ICT Convergence, Paichai University, Daejeon 35345, Korea <sup>3</sup>Professor, Department of Computer Engineering, Paichai University, Daejeon 35345, Korea

#### [요 약]

본 논문에서는 정수장 수처리공정 운영 시 감시제어시스템에서 취득한 데이터에 대해 빅데이터 분석과 인공지능 알고리즘을 융합하여 자율운영 모델을 개발하고 시뮬레이션하여 검증하였다. 약품공정은 N-Beats 모델을 선택하였으며, 약품 사용량을 절감 (-1.91 ℓ/h)하여 응집제 주입량의 자율운영 가능성을 확인하였다. 침전공정은 LightGBM 모델을 선택하였고, 슬러지 수집기 운행 스케줄에 대해 슬러지 발생량 예측 결과에 따라 연동되도록 제어 로직을 개발하였다. 해당 모델의 현장 시뮬레이션 결과에서 슬 러지 발생량이 예측값과 실제값이 거의 유사하였다. 또한, 기존 운영 스케줄보다 슬러지 수집기 운행 시간 단축(43시간)에 따른 전 기에너지 소비량 감소 효과와 자율운영을 통한 운영근무자의 의존도 감소와 근무 부담 경감 효과를 확인하였다.

# [Abstract]

In this paper, we developed and validated an autonomous operation model by integrating big data analysis and artificial intelligence algorithms. Simulations were conducted using data obtained from the monitoring and control system during the operation of a water treatment process at a water purification plant. The N-BEATS model was selected for the chemical process, and confirmed its potential for autonomous operation by optimizing the coagulant injection, resulting in a reduction in chemical usage (-1.91 l/h). The LightGBM model was used for the sedimentation process. Control logic was developed to link the sludge collector operation schedule with the predicted sludge generation. In the model simulation results based on field data, the predicted and actual sludge generation amounts were closely aligned. Additionally, reducing the sludge collector's operating time by 43 hours led to lower electrical energy consumption compared to that of the existing operation schedule. The autonomous operation model also demonstrated a reduction in worker dependence and workload.

**색인어**: 감시제어시스템, 딥러닝, 빅데이터, 수처리공정, 자율운영

Keyword : Supervisory Control System, Deep Learning, Big Data, Water Treatment Process, Autonomous Operation

#### http://dx.doi.org/10.9728/dcs.2025.26.4.1031



use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 15 January 2025; Revised 14 February 2025 Accepted 26 February 2025

\*Corresponding Author; Hoe-Kyung Jung

Tel: +82-42-520-5640 E-mail: hkjung@pcu.ac.kr

# I.서 론

경제협력개발기구(Organization for Economic Cooperation and Development)에 따르면 기후변화와 수자원관리 문제로 인한 물 부족이 심화하여 물 부족 지역의 경우 2030년 무렵 에는 39억여 명이 물 부족에 시달릴 것으로 예상하고 있다. 이러한 물 부족 문제에 대처하기 위해 한정적인 수자원에 대 한 체계적인 운영관리와 효율적인 이용에 대한 중요성과 필 요성은 더욱 중요시되고 있다. 2022년 상수도통계에 따르면 국내에서 상수도를 공급하는 수도사업자는 161개 지방상수 도 사업자와 1개 광역상수도 사업자로 전국 기준 정수장 478 개소를 운영하여 수돗물을 생산하여 공급하고 있다. 정수장에 서는 여러 단계의 공정을 거쳐 수돗물이 생산·공급되고 있고, 수처리공정은 약품공정, 착수공정, 침전공정, 혼화응집공정, 소독공정, 여과공정 등의 단위공정으로 구성되어 있다[1]. 정 수장을 중단없이 실시간으로 운영하기 위해 설치된 계측제어 설비는 네트워크, 감시제어설비, 다양한 계측 센서, HMI (Human Machine Interface), 통신설비 등으로 구성되어 수 처리공정으로부터 수량, 설비 상태, 수질 등을 얻고, HMI를 이용하여 시각화하고 근무자가 정수장 운영이 가능하도록 정 보를 제공한다. 이와 같은 기존 시스템은 일반적으로 실시간 모니터링 기반으로 수집한 데이터를 분석하여 이상 상태를 예측하거나 운영의 최적화 기능이 부족하여 수량 부족, 수질 이상 등의 문제를 미리 감지하지 못하고. 대응이 늦을 수가 있었다. 또한, 수동적 감시제어 방식으로 사고 상황이나 급격 한 운영 변화에서 빠른 대응이 어려운 시스템이다. 이와 같은 문제점을 해결하기 위하여 인공지능 및 빅데이터 등의 도입 이 필요하다. 이를 통해, 문제 발생 시 빠른 대응을 가능하게 하고 수처리공정의 효율성을 높일 수 있는 전반적인 시스템 에 대해 지능화가 가능하다.

이에 본 논문에서는 정수장 수처리공정을 운영할 때 근무 자의 노하우와 경험 중심 감시제어로 발생되는 오류 및 운영 편차 방지하고, 안정성 향상 및 정수장 운영 효율과 수돗물 생산원가를 절감할 수 있는 방안을 제시한다. 이를 위해 기존 감시제어시스템의 데이터에 대해 빅데이터 분석과 인공지능 기술을 융합하고 정수장 수처리공정 중 침전공정 및 약품공 정 자율적 운영 모델을 개발 및 검증하였다. 이를 통해 수처 리공정 중 약품공정에서는 검증 기간동안 약품 사용량을 절 감(-1.91  $\ell/h$ )하고, 원수 수질 변화에 따른 응집제 주입량의 자율운영 가능성을 확인하였다. 또한, 침전공정에서는 슬러지 수집기 운행 시간을 단축(43시간)하고, 슬러지 발생량 예측을 통한 슬러지 수집기 운행 스케줄 연동으로 자율운영 가능성 을 확인하였다.

본 논문은 제Ⅱ장에서는 정수장 수처리시스템 및 정수장 수처리공정 인공지능 적용 인공지능 기술을 분석하였고, 제Ⅲ 장에서는 자율운영 모델을 개발하고 운영하기 위해 제안한 인공지능 기반 정수장 자율운영 시스템을 설명하였고, 제Ⅳ장 에서는 정수장 수처리공정 자율운영 모델 개발 및 검증 결과 를 제시하였고, 제V장에서는 연구 결과를 정리하였다.

# Ⅱ. 정수장 수처리공정 인공지능 적용 관련 연구

본 장에서는 물 관리 분야에서 사용되고 있는 인공지능 알 고리즘 및 관련 기술과 정수장 수처리시스템에 대해 기술한다.

#### 2-1 정수장 수처리시스템

정수장 수처리공정의 착수정에서 혼화 및 응집, 약품투입, 침전, 응집, 소독, 여과 등 공정에서 각 설비를 감시하고 자동 적으로 공정을 제어하기 위하여 그림 1과 같은 시스템을 운영 하고 있고, 이중화 시스템으로 장애 시 정상적으로 가동되도 록 구성하고 있다[2]. 정수장 운영을 위해 수처리공정별로 설 치된 계측기기와 설비로부터 감시 데이터를 현장감시제어설 비에서 취득하여 중앙조정실 운영근무자가 HMI를 통해 모니 터링하고, 설비를 제어하여 각 수처리공정을 최적으로 운영한 다. 그리고, 현장감시제어설비인 RCS(Remote Cotnrol Station)나 PLC(Programmable Logic Controller)에는 부 분적으로 수처리공정 단위 프로세스를 자동적으로 운영하도 록 제어로직이나 시퀀스가 탑재되어 운영된다. 광역상수도에 서는 HMI인 표준수운영시스템에 표준 제어로직을 탑재하여 더욱 향상된 수준의 자동화를 운영하고 있다. 하지만 표준제 어로직도 급격한 수질 변화나 운영 여건의 변화에 빨리 대응 이 가능하지 못해서, 이를 반영하기 위해 추가적인 프로그램 작업을 수행해야 하는 문제점이 발생한다.



그림 1. 광역상수도 정수장 감시제어시스템 구성

Fig. 1. Configuration of a regional water purification plant monitoring and control system

### 2-2 정수장 수처리공정 적용 인공지능 기술

Heddam 등은 일단위 자료 725개 세트를 활용하여 응집제 인 Alum의 주입률 결정 모델을 개발하였다[3]. 별도의 입력 변수 시나리오 없이 원수의 탁도, 전기전도도, pH, 수온, 용존 산소농도, UV(Ultraviolet) 흡광도를 입력변수로 활용하였다. RBF(Radial Basis Function)와 GRNN(Generalized Regression Neural Network) 모델을 개발하여 비교한 결과 GRNN 모델 이 더욱 우수한 성능을 나타내었다. 동일한 데이터 세트를 활 용하여 Alum 주입률에 대한 ANFIS (Adaptive Neuro Fuzzy Inference System) 모델도 개발하였다[4].

Kim은 2017년 1월부터 2018년 11월까지 시간단위 데이 터 세트를 활용하여, 평시 PAC(Poly Aluminum Chloride) 주입률 결정모델과 고탁도시 PAHCS(Poly Aluminum Hydroxy Chloro Sulfate) 주입률 결정모델을 개발하였다. 개발시 원 수 탁도, 수온, pH, 전기전도도는 필수 입력변수로 고정하고, PAC는 유입유량, iPDA(i Photometric Dispersion Analyser) 값, 침전지 유출수 탁도, PAHCS는 유입유량, iPDA 값, TOC(Organic Carbon), 암모니아성 질소, 침전지 유출수 탁 도에 대한 입력변수 시나리오를 설정하였다. 중회귀모델, SVM(Support Vector Machine), MLP(Multi-Layer Perceptron), LSTM(Long-Short Term Memory) 모델을 적용한 결과 시 계열 모델인 LSTM이 가장 우수한 결과를 나타내었다[5].

이에 관련 기존 연구에서는 공통적인 모델 개발에 이전에 기본적으로 과거 응집제 주입에서 운영 결과가 최적 상태인 것으로 가정하여 모델을 학습시켰고, 모델 개발 과정에서는 하이퍼파라미터를 결정하는데 연구자의 판단에 따라 고정 입 력변수를 이용하였다. 그리고, 하이퍼파라미터 시나리오를 설 정한 경우에 일부 하이퍼파라미터만 설정하여 수행하는 등의 한계점이 있어 추가적인 연구가 필요하다[5],[6].

이에 본 논문에서는 시계열 데이터의 특성을 잘 고려한 모 델을 개발하기 위해 기계학습 모델 중에서도 LSTM(Long Short Time Memory), GRU(Gated Recurrent Unit), N-BEATS(Neural Basis Expansion Analysis for Interpretable Time Series) 등과 같은 알고리즘을 정수장 수처리 공정별 운영 요소 결정 및 목표값 예측에 활용하였다.

# Ⅲ. 인공지능에 기반한 정수장 자율운영 시스템

본 장에서는 제안하는 정수장 자율운영시스템의 구성을 다 룬다.

### 3-1 빅 데이터 플랫폼

인공지능에 기반한 정수장 수처리공정 자율운영을 위한 모 델 개발을 위하여 빅데이터 플랫폼은 데이터를 수집하고, 저 장, 처리, 분석, 모델 학습 및 배포를 종합적으로 지원하여 인 공지능 모델 개발 및 운영하는 데 필요한 여러 기능을 제공한 다. 이 플랫폼은 머신러닝 워크플로우과 데이터 파이프라인을 효율적으로 관리가 가능하도록 그림 2와 같이 구성하였다.

정수장 수처리공정에서 운영 중인 설비 및 센서로부터 취 득한 원시 데이터를 SCADA(Supervisory Control And Data Acquisition) 서버에서 CSV(Comma-Separated Values) 형태의 태그데이터 원본 파일을 수집·전송하기 위해 Fluentd를 활용하였다. Fluentd는 SCADA 서버의 iRDC (iWater Realtime Data Collector)에서 생성한 CSV 파일을 읽어서 데이터 스트림을 JSON(JavaScript Object Notation) 형식으로 변환한 후 빅데이터 플랫폼 서버의 Apache Kafka 에 데이터를 전송하도록 구성하였다. 자료 수신 시스템은 Apache Kafka로 Fluentd에서 Kafka의 특정 토픽으로 데이 터를 전송하고, Kafka 토픽은 데이터의 카테고리나 스트림을 정의하여 Kafka 컨슈머를 사용하여 토픽의 메시지를 읽도록 구성하였다.





데이터 전처리 시스템인 Nifi는 Kafka에 저장된 데이터를 입력 프로세서를 사용하여 실시간으로 적재하고, 변환 프로세 서를 이용하여 데이터를 변환하여, 아피치(Apache) Hive에 데이터를 전송한다. 또한, RDBMS(Relational DataBase Management System)인 MariaDB에 원시 데이터를 전송하 여 임시로 저장하고, MariaDB에 저장된 익일에 하둡 (Hadoop)의 HDFS(Hadoop File System)에 저장한다. 아 파치 Hive는 Kafka에서 수신한 데이터에 대해 분석용 파일 형식인 parquet로 변환하여 하둡의 저장소인 HDFS에 저장 하는 기능을 수행한다. Hive는 큰 양의 데이터 처리 및 분석 을 위한 하둡의 서브 프로젝트로 개발되어 활용된다[7],[8].

데이터 웨어하우스에는 각 정수장에서 수집한 대량의 데이 터가 저장되므로 데이터의 실시간 처리 및 분산 저장을 위해 고가용성 시스템이 필요하여, 하둡의 분산 파일 시스템을 구 축하였다[9].

Hive는 하둡의 HDFS에 저장된 데이터에 대해 쿼리하고, 분석하기 위해 데이터 웨어하우스 시스템으로 HiveQL을 사 용하며, HDFS의 데이터에 대해 전처리하며, Hive 테이블을 정의하고, 데이터를 적재한 후 HiveQL 쿼리를 이용하여 데 이터를 변환 및 분석한다. Hue(Hadoop User Experience) 는 하둡 시스템에서 여러 가지의 데이터 분석 및 처리 작업을 효과적으로 수행가능 하도록 도와주는 웹 기반의 인터페이스 로 사용자가 하둡 클러스터를 손쉽게 관리하고 데이터 작업 수행이 가능하다. Spark, 하둡, Hue, Hive를 통합하여 큰 규 모의 데이터를 분석하고 처리하는 환경을 구축해 데이터 저 장, 수집, 분석, 시각화, 처리 시각화까지 전 과정을 효율적으 로 수행이 가능하도록 하였다.

# 3-2 인공지능 플랫폼

제안하는 인공지능 플랫폼은 인공지능 기술을 개발하고, 훈련 및 배포하며, 관리하기 위한 통합된 환경을 제공한다. 인 공지능 플랫폼은 정수장에 인공지능 플랫폼 서버와 운영 서 버를 구축하고, 유역본부에도 인공지능 운영 서버를 그림 3과 같이 구성하였다. 인공지능 플랫폼 서버는 무중단으로 운영하 기 위하여 Active-Active 방식으로 이중화로 구성하여 안정 성을 확인하였다.



그림 3. 인공지능 플랫폼 구성도

빅데이터 플랫폼과 동일하게 정수장에서 원시 데이터는 Fluentd가 SCADA 서버의 iRDC에서 생성된 CSV 파일을 읽어 데이터 스트림을 JSON 형식으로 변환하여 인공지능 플 랫폼 서버인 아파치 Kafka에 데이터를 전송할 수 있도록 하 였다. Fluentd에서 Kafka의 특정 주제로 데이터를 전송하고, Kafka 주제는 데이터의 스트림이나 카테고리를 정의하여 Kafka Consumer를 사용하여 주제의 메시지를 읽도록 구성 하였다.

데이터 저장 시스템으로 MairaDB를 활용하였다. 실시간 데이터 처리 플랫폼인 Kafka는 MariaDB라는 데이터 마트에 데이터를 송수신하는 중간 역할을 하도록 구성하였다. Kafka Producer는 실시간으로 데이터를 수집하여 Kafka에 전송하 고, JSON 등과 같은 형식으로 Kafka에 저장하며, 이 데이터 는 MariaDB에 삽입되기 이전에 변환된다. Kafka Consumer 는 MariaDB에 데이터를 저장하도록 Kafka에서 데이터를 읽 어와 MariaDB에 삽입하는 SQL 쿼리를 사용해 데이터를 저 장한다.

훈련 데이터는 모델을 학습하기 위해 원시 데이터가 MariaDB에 저장된다. 인공지능 모델의 버전, 성능지표, 학습 상태와 같은 메타데이터를 저장하며, 인공지능 모델이 생성한 예측 결과와 이 결과를 바탕으로 분석 데이터를 저장한다. 자 율운영 모델을 분석한 결과인 제어 데이터가 MariaDB에서 Kafka로 전송되며, 다음으로 SCADA 서버의 PDB(Process Database)에 저장한 후 표준준수 운영시스템 제어 태그를 통 해 명령이 실행된다. 인공지능 플랫폼의 데이터 처리에 대한 흐름도는 그림 4와 같다.



그림 4. 인공지능 플랫폼 데이터 처리 플로우 Fig. 4. Al platform data processing flow

# Ⅳ. 정수장 수처리공정 자율운영 모델 개발·검증

# 4-1 대상지역 선정

본 연구에서 연구 대상은 댐에서 취수하여 정수를 생산 및 공급하는 정수장을 선정하였다. 연구 대상 정수장은 하 루 용량 700,000m<sup>3</sup>, 2022년 기준 하루평균 486,456m<sup>3</sup>의 수돗물을 생산해 2개 군과 4개 시에 상수도를 공급하고 있 다[4],[10]. 정수장은 착수, 응집, 혼화(약품 투입), 여과, 침 전 등의 표준처리 공정 및 최종적으로 염소로 소독하여 정수 를 생산하는 공정이다.

# 4-2 약품공정 자율운영 모델

#### 1) 공정 개요 및 목표 설정

수처리공정 중 약품공정에서 투입된 응집제는 혼화지에서 혼화기의 동력에 의해 원수와 섞는 과정 이후 응집기에 의해 플록(Floc) 형태로 응집된다. 수처리공정에서 약품공정 다음 공정의 효율성을 높이고, 플록 형성을 위한 응집 효율 증대를 위해 원수 수질 변화에 따라서 적절한 응집제 주입률이 결정 될 필요가 있다. 응집제 주입률은 대부분 운전 근무자의 경험 에 따라 쟈 테스트(jar test) 또는 판단과 같은 수질 시험을

Fig. 3. Artificial intelligence platform configuration diagram

통해서 작성된 조견표를 참조하여 수동으로 결정된다.

약품공정의 자율운영 모델 목표는 원수 수절 등 응집제 주 입률에 관련된 주요 인자의 상관관계 및 특성 분석을 통하여 주입률을 예측하고, 응집제를 자동으로 주입하는 인공지능 기 반 자율운영을 수행한다.

# 2) 데이터 전처리 및 구성

약품공정에서 응집제 주입률에 대한 예측을 위한 주요 인 자로는 원수 pH, 원수 전기전도도, 원수 탁도, 원수 온도, 원 수 알카리도가 있다. 각 주요 인자에 따른 데이터는 시스템에 서 수집하여 빅데이터 플랫폼에 저장한 데이터를 이용하였고, 데이터 기간은 2018. 1. 1 ~ 2022. 12. 31까지로 주요 인자 별 태그 중 총 건수가 2,624,573건이다.

이상치 탐지 및 처리 기법으로 Z-Score, IQR(Interquartile Range), 시계열 분해를 비교 사용하였다. 이와 같이 전처리 기법을 이용해 약품 투입 및 원수 수질과 관련되는 데이터의 이상치를 제거하고 그래프로 시각화하여 비교하고, 히스토그 램으로 시각화하여 원수 수질 관련 데이터인 전기전도도, 탁 도, pH, 알카리도, 수온과 응집제 주입률 모두 이상치 제거 기법으로 시계열 분해 기법을 선정하였다.

#### 3) 데이터 분석

상관분석은 응집제 주입률 예측에 이용되는 관련 변수 및 입력 후보와 응집제 주입률과의 상관분석을 수행하였다. 상관 분석을 이용한 예측 변수는 응집제 주입률이며, 입력 변수는 원수 전기전도도, 원수 탁도, 원수 pH, 원수 온도, 원수 알카 리도이며, 관련 변수는 활성탄 투입량, 원수 망간, 친전수 탁 도이다. 피어슨 상관계수에 대해 히트맵을 이용하여 수치를 분석하고 시각화한 결과, 응집제 주입률에 대해 각 데이터는 다음과 같이 상관관계를 가진다. 입력 후보 변수 중 알카리도 및 원수 탁도는 양의 상관관계, 원수 전기전도도, pH 및 온도 는 음의 상관관계를 보였다. 상관 계수와 응집제 주입률이 높 은 입력 후보 변수는 원수 전기전도도(-0.53), 원수 알칼리도 (0.59), 원수 탁도(0.78) 순으로 높은 선형적 상관관계를 가 진다. 특히 원수의 탁도가 응집제 주입률과의 상관계수가 0.78로 강하게 양의 상관관계를 보이며, 원수 탁도값이 상승 할 수록 응집제 주입률이 증가함을 알 수 있다.

상관분석을 이용하여 데이터를 분석한 결과 결정된 입력변 수는 원수 전기전도도, 원수 탁도, 원수 알카리도, 원수 pH, 원수 온도, 출력변수는 응집제 주입률이다.

응접제 주입률에 영향을 끼치는 원수 수질에서의 분포 특 성에 의해 분류하며, 군집별로 응집제 주입률의 분포 차이를 확인하여 원수 수질 분류에 의한 응집제 주입률 예측 성능의 통계적 유효성을 검증하기 위하여 Elbow 곡선 방법을 이용 하여 군집분석을 수행하였다. 정수장의 경우에는 원수 수질 특성에 의해 2개의 군집으로 분류하고, 1군집과 2군집은 탁 도 9 NTU(Nephelo-metric Turbidity Unit)를 기준으로 나 누어지므로, 9 NTU 이상이면 고탁도, 미만이면 저탁도로 판 단하였다.

# 4) 자율운영 모델 개발

약품공정 자율운영 모델 개발을 위해 훈련 데이터 셋 60%, 학습 모델의 성능을 측정하여 최종 모델 선정을 위해 검증 데 이터 셋 20%, 선택된 최종 모델 성능을 측정하기 위해 시험 데이터 셋 20%로 구성하였다.

분석 데이터 셋의 각 변수들에 대한 척도의 차이가 존재하 므로 모든 데이터에 대해 Minmax Scaler를 적용하여 정규 화 하였다. 약품공정 자율운영 모델에서의 학습 과정에서 과 적합(Overfitting)을 방지위해 미리 학습을 중단하는 방법인 학습 조기 종료를 수행하였다.

모델 검증에는 K-fold Cross Validation 기법을 사용하였 다. 4개 모델에 대한 성능을 검증한 결과, 성능 평가 지표인 RMSE(Root Mean Squared Error), R2 Score(Coefficient of Determination), MSE(Mean Squared Error), MAE (Mean Absolute Error) 값은 표 1과 같다. N-BEATS 모델 의 평가측도(R2)가 0.87로 가장 높아서 최종 모델로 선택하 였다. N-BEATS 모델은 시계열 예측에 적합하고[11], 응집 제 주입률 실제값과 예측값 그래프는 그림 5와 같고, 모델 평 가 측도(R<sup>2</sup>)값(0.87)에 의해 실제값과 예측값의 차이가 발생 한다. 그림 5에서 y축이 응집제 주입률(단위 ppm)이며, 실제 값과 예측값을 구분하기 위하여 예측 구간 값을 일정 주기마 다 점으로 표시하였다. x축은 입력 변수인 원수 수질의 데이 터 셋 번호로 수질의 변화에 따라 출력 변수인 응집제 주입률 예측값이 실제값과 비슷한 추세로 변경되는 것을 확인할 수 있었다.

표 1. 약품공정 자율운영 모델 성능 평가지표 Table 1. Pharmaceutical process autonomous operation model performance evaluation indicators

	1				
Model	Data Type	MAE	MSE	RMSE	R <sup>2</sup>
N-BEATS	Test	0.68	2.37	1.54	0.87
GRU	Test	0.82	2.22	1.49	0.80
LSTM	Test	0.78	2.13	1.42	0.83
GBR	Test	0.86	2.24	1.50	0.81



-ig. 5. Comparison graph of flocculant injection rate predicted values and test data actual values

약품공정 자율운영을 위하여 선정한 N-Beats 모델에 대한 네트워크 레이어는 15개이고, 총 파라미터 수는 1,328,724개 이다. 약품공정 자율운영 모델의 흐름도는 그림 6과 같고, 응 집제 주입률의 예측 부분과 저탁도와 고탁도를 구분하여 예 측한 주입률을 보정하고, 최대 주입률과 최소 주입률을 벗어 나지 않게 조정하는 부분으로 나누어진다. 모델을 통하여 최 종 결정된 응집제 주입률에 의해 약품공정 자율운영이 이루 어진다.



그림 6. 약품공정 자율운영 모델 플로우차트



# 5) 자율운영 모델 시뮬레이션

시뮬레이션은 A계열을 2024.01.22~01.31일까지 10일 간, B계열을 2024.01.15~01.28일까지 14일간 진행하였다. 약품공정 자율운영 모델에 따른 응집제 주입률 운영 실제값 과 예측값을 비교한 그래프는 A계열은 그림 7, B계열은 그림 9와 같다. 응집제 주입률 운영 실제값은 쟈테스트 또는 조견 표에 의하여 운영근무자가 수동으로 입력한 값을 사용하여 변경 전까지 동일한 값으로 고정하여 운영하기 때문에 시뮬 레이션 기간동안 그림 7, 그림 9에서와 같이 일정한 값으로 고정되어 있다. 하지만 예측값은 응집제 주입률 영향 인자인 원수 수질의 그래프인 A계열 그림 8, B계열 그림 10에서와 같이 수질 변화 추세에 따라 변동되는 추세를 보여주었다. 단, 시뮬레이션 기간이 짧아 운영 실제값의 변화에 따른 예측값 변동을 비교할 수 없어 향후 장기간 운영을 통한 검증이 필요 하다. A계열과 B계열은 시설만 분리되어 있고 원수는 동일하 므로 시뮬레이션 기간 차이만 있을 뿐 그림 8과 10에서와 같 이 동일 기간에는 변화 추세가 동일함을 확인할 수 있다.



\*screen capture 그림 7. A계열 응집제 주입률 비교(AI 예측값 VS 운영 실제값) Fig. 7. Comparison of A-series coagulant injection rates (AI predicted value VS operational actual value)













또한 응집제 주입량도 운영 실제값과 예측값을 그림 11과 같 이 비교하였는데, 모델 성능은 R<sup>2</sup>값이 A계열의 경우 0.9349, B 계열의 경우 0.9312로 향상되었다. 시뮬레이션 동안 A계열 에서 실제 응집제 주입률 변화가 없이 14ppm으로 고정되었 고, 응집제 주입률의 예측하는 값은 영향 인자인 원수 수질 변화에 의해 13~15ppm 범위 내에서 변화하였으며, 응집제 주입량이 평균 1.59 l/h 만큼 차이가 발생하였다. B계열은 실 제 응집제 주입률의 변화가 없이 14ppm으로 고정되었고, 응 집제 주입률이 예측되는 값은 영향 인자인 원수 수질 변화에 의해 12~15ppm 범위내에서 변화하였으며, 응집제 주입량이 평균 -3.50 l/h 차이가 발생하였다.

<Comparison of coagulant injection amount for A-series>



<Comparison of coagulant injection amount for B-series>



#### \*screen capture

Fig. 11. Comparison of A-series and B-series coagulant operation actual values and injection amount Al predicted values

#### 4-3 침전공정 자율운영 모델

#### 1) 공정 개요 및 목표 설정

침전은 현탁액을 청정액과 농축된 현탁액으로 분리시키는 공정으로, 침전지는 응집지에 형성된 플록을 중력에 따라 제 거 및 침강분리하고 상등수는 여과지로 유입시킨다. 침전된 슬러지를 제거하기 위하여 인발밸프와 슬러지 수집기가 보통 연동되어 운전된다.

침전공정의 자율운영 모델 목표는 슬러지의 발생량을 예측 하며, 이를 바탕으로 최적의 슬러지 수집 운영 스케줄을 결정 하는 것이다. 운영 스케줄에 따라 침전지별 슬러지 수집기가 운영되고, 슬러지 수위 및 슬러지 인발밸브를 고려하여 제어 하는 로직을 구현한다. 침전지별 인공지능 기반 자율운영이 선택적으로 가능하도록 개발한다.

### 2) 데이터 전처리 및 구성

침전공정에서 슬러지 발생량에 대한 예측을 위한 주요 인 자로는 원수 유입유량, 원수 탁도, 응집제 및 응집보조제 주입 률, 회수유량이 있다. 슬러지 수집기 운영 스케줄 결정을 위하 여 주요 인자로 침전지별 인발밸브 닫힘 및 열림 여부, 슬러 지 수집기 운전 시간, 슬러지 수집기 운전 여부, 응집지 유입 밸브 열림 및 닫힘 여부가 있다. 각 주요 인자별 데이터는 표 준 수운영 시스템에 의해 수집되어 빅데이터 플랫폼에 저장 된 데이터를 이용하였다. 이상치 처리 기법은 약품공정과 동 일하게 사용하였다.

#### 3) 데이터 분석

상관분석은 정수장 슬러지 발생량 계산식을 기반으로 각 단계별에 따라 슬러지 발생량을 계산해 슬러지 발생량 예측 에 활용된 입력 변수와 상관분석을 수행하였다.

피어슨 상관계수를 시각화를 위해 히트맵을 통해 보이고, 수치를 분석한 결과와 이론적 슬러지 발생량(ws)에 대해 각 데이터는 다음의 상관관계를 가진다. 분말활성탄은 0.58(+), 유입유량은 0.11(+), 가성소다는 0.59(+), 응집제 주입률은 0.7(+), 원수 탁도는 0.72(+)이다. 응집제 주입률, 유입유량, 원수 탁도, 분말활성탄 투입량, 가성소다 주입률 모두 슬러지 발생량과는 양의 상관관계를 보인다. 특히, 응집제 주입률과 원수 탁도는 슬러지 발생량과 높은 양의 상관관계를 가지는 데, 이는 응집제의 주입력이 증가하거나 원수의 탁도가 높아 짐에 따라 슬러지 발생량이 증가하거나 원수의 탁도가 높아 짐에 따라 슬러지 발생량이 증가하거나 원수의 탁도가 높아 집에 따라 슬러지 발생량이 증가하거나 원수의 탁도가 높아 집에 따라 슬러지 발생량이 증가하면 할수록 슬러지 발생량이 상승하는 것을 나타낸다. 그래서, 원수 탁도, 유입유 량, 응집제 주입률, 원수 탁도, 분말활성탄, 가성소다 등의 변 수가 높아질수록 슬러지 발생량이 상승하는 것으로 관단된다.

상관분석을 통하여 데이터 분석한 결과 결정된 입력변수는 응집보조제 주입률, 원수 탁도, 원수 유량, 응집제 주입률이 며, 출력변수는 슬러지 발생량이다. 자율 운영 모델링에는 원 수유량, 원수 탁도, 주입률을 독립변수로 하며, 종속변수로는 슬러지 발생량을 사용하고 슬러지 발생량에 대해서는 계산식 으로 산출한다.

#### 4) 자율운영 모델 개발

침전공정에 대한 자율운영 모델 개발은 슬러지 발생량이 시계열적인 연속형 값이므로 XGboost, GRU, Catboost, LightGBM과 같은 회귀 모델의 성능을 비교해 최적의 모델을 선택하여 구축한다.

자율운영 모델을 정의하는 1단계에서는 이론적 슬러지 발 생량 계산식을 이용하여 발생량을 산출하고, 이를 기반으로 실제 응집보조제 주입률 및 응집제와 원수 유입유량으로 슬

그림 11. A계열 및 B계열 응집제 운영 실제값과 주입량 AI 예측값 비교

러지 발생량을 예측한다. 2단계에서 응집기 운전 상태를 이용 하여 침전지 사용 여부를 판단하며, 침전지별 슬러지 수집기 운전 상태로 출발 시각 및 전·후진을 판단한다. 3단계에서는 예측한 슬러지 발생량 및 슬러지, 인발밸브, 수집기 운전 상태 등을 고려해 스케줄을 작성한다. 4단계에서는 슬러지 수집기 운전을 제어함으로서 슬러지 배출 주기를 결정해 침전공정 자율운영이 가능하도록 하였다.

침전공정 자율운영 모델 구축을 위하여 데이터 셋은 앞 절 에서 사용한 데이터를 활용하였다. 활용 변수는 원수 탁도, 원 수 유입유량, 응집제 주입률, 분말활성탄 주입량이다.

침전공정 자율운영 모델의 학습 과정에서 과적합을 방지하 기 위하여 학습을 미리 중단하는 방법인 학습 조기 종료도 동 일하게 수행하였다. 4개 알고리즘의 성능을 검증한 결과는 표 2 와 같다. 시험 결과를 보면 XGBoost 모델의 평가측도(R<sup>2</sup>)가 0.9898으로 가장 높으나, LightGBM 모델의 평가측도(R<sup>2</sup>)가 훈련, 검증, 시험 모두 균등하게 높으므로 최종 모델로 선택하 였다. LightGBM 모델은 회귀와 분류 등 다양한 분야에 사용 되며[12], 시험 데이터 슬러지 발생량 예측값과 실제값 그래 프는 그림 12와 같이 극히 일부 구간 오차가 발생하였으나 전 구간 변화 추세가 비슷함을 확인할 수 있었다.

표 2. 침전공정 자율운영 모델 예측 결과 Table 2. Sedimentation process autonomous operation model prediction results

Model	Data Type	MAE	MSE	RMSE	R <sup>2</sup>
LightGBM	Train	0.024545	0.002259	0.047529	0.9904
	Validation	0.018482	0.000660	0.025690	0.9727
	Test	0.029409	0.003878	0.062274	0.9884
XGBoost	Train	0.027625	0.002572	0.050715	0.9891
	Validation	0.021984	0.001149	0.033897	0.9525
	Test	0.032956	0.003390	0.058224	0.9898
CatBoost	Train	0.028514	0.002853	0.053413	0.9879
	Validation	0.022766	0.001298	0.036028	0.9464
	Test	0.035778	0.004727	0.068753	0.9858
GRU	Train	0.0475	0.005791	0.0761	0.994
	Validation	0.0287	0.001681	0.0410	0.953
	Test	0.0561	0.008519	0.0923	0.922

침전공정 자율운영 모델의 흐름도는 그림 13과 같고, 크게 슬러지 수집기 운행 스케줄을 결정하는 부분과 슬러지 발생 량 예측하는 부분으로 나눌 수 있다. 침전 공정의 슬러지 발 생량 예측을 통해 최종적으로 슬러지 수집기 스케줄 제어를 통하여 침전공정 자율운영을 구현한다. 슬러지 수집기 운영 스케줄 제어를 위하여 필요한 데이터는 대차 운행기준 슬러 지양, 슬러지 수집기 총 운전거리(편도거리), 집기 총 운전시 간, 대차 운행 대기 최대일 수로 정수장 시설 현황 데이터를 바탕으로 결정하였다.



그림 12. 침전공정 슬러지 발생량 AI 예측값 및 운영 실제값 비교

Fig. 12. Comparison of sedimentation process sludge generation AI predicted values and operational actual values



그림 13. 침전공정 슬러지 자율운영 모델 플로우차트 Fig. 13. Sedimentation process sludge autonomous operation model flowchart

# 5) 자율운영 모델 시뮬레이션

슬러지 발생량에 대한 시뮬레이션은 2024.01.22부터 10일 간 진행하였고, 슬러지 수집기 스케줄 시뮬레이션은 2024.01.27. 16:58분~2024.02.06. 00:48분까지 진행하였으며, 자율운 영은 01. 31. 23:59분 이후부터 실시하여 실제 운영과 침전 공정 자율운영 결과를 그림 14, 15와 같이 비교하였다. 이의 모델 성능은 R<sup>2</sup> 값이 0.9257로 다소 낮았지만 슬러지 발생량 은 거의 비슷하였고, 슬러지 수집기에 대한 스케줄이 슬러지 배출량 0.5m<sup>3</sup>/일 기준으로 3일 15:39분과 5일 10:09분으로 대략 1일 19시간 차이를 보였다. 실제로 운영할 때에는 지정 된 시간을 기준으로 운영하였고, 자율운영 모델은 슬러지 발 생량과 슬러지 수집기 최대 대기일을 기준으로 운영하므로 슬러지 수집기 스케줄의 차이가 발생하는 것으로 판단된다.



<sup>\*</sup>screen capture

그림 14. 침전공정 실제 운영 및 자율운영 결과 슬러지 발생량 비교

Fig. 14. Comparison of sludge generation results from actual and autonomous operation of the sedimentation process



\*screen capture

그림 15. 침전공정 실제 운영 및 자율운영 결과 슬러지 수집기 스케줄 비교

Fig. 15. Comparison of sludge collector schedules with actual and autonomous operation results of sedimentation process

# V.결 론

정수장 수처리공정 운영 시 기존 감시제어시스템의 한계와 운영근무자 중심의 운영 방식으로 인한 문제점 등을 해소하 기 위하여 빅데이터 및 인공지능 플랫폼 기반의 자율운영 시 스템을 구축하였다. 그리고, 기존 감시제어시스템에서 정수장 수처리 공정을 운영할 때 얻은 데이터에 대해 인공지능 알고 리즘과 빅데이터 분석을 융합하여 정수장에서의 수처리 공정 에 대해 자율운영 모델을 개발하고 검증하였다.

약품공정 응집제 주입률 예측을 위해 자율운영 모델 개발 에 대한 주요 변수인 원수의 알카리도, 탁도, 전기전도도, pH, 수온과 응집제 주입률 이전 데이터에 대하여 전처리 후 피어 슨 상관계수를 사용하여 상관분석을 수행하였고 히트맵으로 시각화하였다. 예측 모델의 최종 시험 데이터 적용 결과로 모 델에 대한 성능 평가 측도인 R<sup>2</sup>값이 0.87로 제일 높은 N-Beats 모델을 선택하였다. 모델에 대한 시뮬레이션 결과 모델 성능은 R<sup>2</sup>값이 A계열의 경우 0.9349, B계열의 경우 0.9312로 성능이 향상되었다. 예측 주입률 기반 실제값과 응 집제 주입량을 비교 결과 A계열은 평균 1.59 l/h, B계열은 평 균 -3.50 l/h 차이가 생기므로 전체적으로 -1.91 l/h 차이가 발생하였다. 짧은 기간의 시뮬레이션에서 기간이 짧았지만 약 품 사용을 절감하여 수돗물 생산비의 절감 효과를 확인하였다. 침전공정은 슬러지 수집기 자율운영을 위하여 모델 수행결 과 슬러지 발생량과 응집제 주입률의 상관계수는 0.7, 원수 탁도의 상관계수는 0.72로 양의 상관관계를 보여 응집제 주 입률과 원수 탁도값에 비례해 슬러지 발생량이 상승하는 것 을 알 수 있었다. 최종 성능 평가 결과의 모델 성능 평가 척도 인 R<sup>2</sup>값이 XGBoost 모델의 경우 시험 데이터 0.9898로 가 장 높았으나, LightGBM 모델의 경우 훈련 데이터 0.9904, 검증 데이터 0.9727, 시험 데이터 0.9884로 다 같이 균등하 게 높아서 최종 모델로 선택하였다. 슬러지 수지기 운행 스케 줄은 대차 운행기준 슬러지양, 슬러지 수집기 총 운전거리(편 도거리), 수집기 총 운전시간, 대차 운행 대기 최대일 수를 고 려해 슬러지 발생량 예측 결과와 연동하게 제어 로직을 개발 하였다. 시뮬레이션 결과 모델 성능은 R<sup>2</sup>값이 0.9257로 약간 낮았지만 슬러지 발생량은 거의 비슷하였다. 짧은기간 동안 시뮬레이션 동안 슬러지 수집기 운행 시간이 단축되어 전기 소비량을 줄일 수 있고, 수돗물 원가를 절감하며, 자율 운영을 통하여 운영 근무자의 의존도 감소와 근무 부담 경감효과를 확인하였다.

향후에 약품공정에서 자율운영 모델의 경우 장기간 운영을 통해 원수 고탁도와 조류 유입과 같은 급격한 수질 변화에 대 하여 시뮬레이션하여 성능을 검증하고 보정이 필요하다.

# 감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신 기획평가원의 지원을 받아 수행된 지역지능화혁신인재양성사 업임(IITP-2025-RS-2022-00156334).

# 참고문헌

- Ministry of Environment, 2022 Waterworks Statistics, Author, Sejong, 2023.
- [2] J. Kim, A Study on the Determination of Coagulant Input Rate in Water Purification Plants Based on Artificial Intelligence, Ph.D. Dissertation, Pai Chai University, Daejeon, February 2021.
- [3] S. Heddam, A. Bermad, and V. Dechemi, "Applications of Radial-Basis Function and Generalized Regression Neural Networks for Modeling of Coagulant Dosage in a Drinking Water-Treatment Plant: Comparative Study," Journal of Environmental Engineering, Vol. 137, No. 12, pp. 1209-1214, December 2011. https://doi.org/10.1061/(ASCE) EE.1943-7870.0000435
- [4] S. Heddam, A. Bermad, and V. Dechemi, "ANFIS-Based Modelling for Coagulant Dosage in Drinking Water Treatment Plant: A Case Study," *Environmental Monitoring*

*and Assessment*, Vol. 184, pp. 1953-1971, April 2012. https://doi.org/10.1007/s10661-011-2091-x

- [5] T. H. Kim, A Study on Development of Optimal Process Management Model for Water Treatment Plant Using Deep Learning Based on Big Data, Ph.D. Dissertation, University of Seoul, Seoul, February 2021.
- [6] J. S. Hyung, Development of Operation Diagnosis and Optimal Decision Making Model Based on Big Data for Drinking Water Treatment Process, Ph.D. Dissertation, University of Seoul, Seoul, February 2022.
- [7] H. Lee, Y.-W. Kim, and K.-Y. Kim, "Study of In-Memory based Hybrid Big Data Processing Scheme for Improve the Big Data Processing Rate," *The Journal of Korea Institute of Information, Electronics, and Communication Technology*, Vol. 12, No. 2, pp. 127-134, April 2019. http://dx.doi.org/10.17661/jkiiect.2019.12.2.127
- [8] I.-H. Joo, "Spatial Big Data Query Processing System Supporting SQL-Based Query Language in Hadoop," *The Journal of Korea Institute of Information, Electronics, and Communication Technology*, Vol. 10, No. 1, pp. 1-8, February 2017. http://dx.doi.org/10.17661/jkiiect.2017.10.1.1
- [9] M. Merrouchi, M. Skittou, and T. Gadi, "Popular Platforms for Big Data Analytics: A Survey," in *Proceedings of 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, Kenitra, Morocco, pp. 1-6, December 2018. http://doi.org/10.1109/ICECOCS.2018.8610652
- [10] H. S. Kim, A Study on Autonomous Operation Model for Water Treatment Process in Water Purification Plant Based on Artificial Intelligence, Ph.D. Dissertation, Pai Chai University, Daejeon, February 2025.
- [11] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting," arXiv:1905.10437, 2020. https://doi.org/10.48550/arXiv.1905.10437
- [12] X. Ma, J. Sha, D. Wang, Y. Yu, Q. Yang, and X. Niu, "Study on a Prediction of P2P Network Loan Default Based on the Machine Learning LightGBM and XGboost Algorithms according to Different High Dimensional Data Cleaning," *Electronic Commerce Research and Applications*, Vol. 31, pp. 24-39, September-October 2018. https://doi.org/10.1016/j.elerap.2018.08.002



# 김학성(Hak-Seong Kim)

1998년 : 전북대학교 (공학사-전자공학) 2010년 : 충남대학교 대학원 (공학석사 -전자정보통신공학) 2025년 : 배재대학교 대학원 (공학박사 -컴퓨터공학)



# 김창배(Chang-Bae Kim)

1995년 : 대전대학교 (공학사-컴퓨터공학) 2024년 : 배재대학교 대학원 (공학석사 -스마트ICT융합학과)

2025년~현 재: 배재대학교 스마트ICT융합전공 박사과정 2019년~현 재: 주식회사 인포비정보기술 근무 \*\*관심분야 : IoT센싱(Internet of Things Sensinf), 안전관리 (Safety Management), 인공지능(AI), 제어 (Control) 클라우드 시스템 등



# 장혁(Hyeok Chang)

1999년 : 한발대학교 (공학사-전자공학과) 2024년 : 배재대학교 대학원 (공학석사 -스마트ICT융합학과)

2025년~현 재: 배재대학교 스마트ICT융합전공 박사과정 2024년~현 재: ㈜큐심플러스 근무 2003년~2024년: ㈜다임즈 근무 \*\*관심분야 : 양자암호통신(Quantum Key Distribution Network), 양자얽힘통신(Quantum

Entanglement) 등



#### **정회경(Hoe-Kyung Jung)** 1985년 : 광운대학교

(공학사-컴퓨터공학) 1987년 : 광운대학교 대학원 (공학석사 -컴퓨터공학) 1993년 : 광운대학교 대학원 (공학박사 -컴퓨터공학)

1994년~현 재: 배재대학교 컴퓨터공학과 교수 ※관심분야 : 머신러닝(Machine learning), 빅데이터(Big data), 임베디드 시스템(Embedded system), IoT 등