



Research Paper

Comparing the Performance of Machine Learning and Deep Learning Algorithms in Wastewater Treatment Process

Jaeil Kim^{1,2} · Hyo Sub Lee³ · Jinuk Jang⁴ · Yongtae Ahn¹ · Seo Jin Ki⁵ · Hyun-Geoun Park^{5†}

¹Department of Energy Engineering, Gyeongsang National University, Republic of Korea

²Environmental Management Corporation, Republic of Korea

³Residual Agrochemical Assessment Division, National Institute of Agricultural Sciences, Rural Development Administration, Republic of Korea

⁴Department of Energy System Engineering, Gyeongsang National University, Republic of Korea

⁵Department of Environmental Engineering, Gyeongsang National University, Republic of Korea

(Received October 12, 2023; Revised November 27, 2023; Accepted November 28, 2023)

Abstract : This study assessed the performance of single and modified algorithms based on machine learning and deep learning for wastewater treatment process. More specifically, this study adopted support vector machine (SVM), random forest (RF), and artificial neural network (ANN) for machine learning as well as long short-term memory (LSTM) for deep learning. The performance of these (single) algorithms were compared with that of modified ones processed through hyperparameter tuning, ensemble learning (only for machine learning), and multi-layer stacking (i.e., two layers of LSTM units). The daily effluent of wastewater treatment process observed between 2017 and 2022 in the Cheong-Ju National Industrial Complex was used as input to all tested algorithms, which was evaluated with respect to mean squared error. For the model performance evaluation, discharge and biochemical oxygen demand are selected as dependent variables out of nine measured parameters. Results showed that the performance of any machine learning algorithms was superior to their competitor LSTM. This is mainly attributed to a small amount of input data provided to the LSTM algorithm and unstable effluent wastewater characteristics. Meanwhile, hyperparameter tuning improved the performance of all tested algorithms. However, ensemble learning for machine learning and two-layer stacking for LSTM generally resulted in performance degradation as compared to that of single algorithms, regardless of dependent variables. Therefore, this calls for a careful design and evaluation of modified algorithms, specifically for model architecture and performance improvement processes.

Keywords : Machine learning, Deep learning, Hyperparameter tuning, Ensemble learning, Multi-layer stacking

The Korean text of this paper can be translated into multiple languages on the website of <http://jksee.or.kr> through Google Translator.

† Corresponding author

envpark@gnu.ac.kr

Tel: 055-772-3344

© 2023, Korean Society of Environmental Engineers



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

연구논문

폐수처리 공정에서의 기계학습 및 심층학습 알고리즘의 성능 비교

김재일^{1,2} · 이효섭³ · 장진욱⁴ · 안용태¹ · 기서진⁵ · 박현건^{5*}

¹경상국립대학교 에너지공학과

²환경시설관리주식회사

³농촌진흥청 국립농업과학원 잔류화학평가과

⁴경상국립대학교 에너지시스템공학과

⁵경상국립대학교 환경공학과

요약: 본 연구는 폐수처리 공정에서의 기계학습 및 심층학습 기반의 단일 및 변형 알고리즘의 성능을 평가하였다. 보다 구체적으로, 본 연구에서는 기계학습 알고리즘으로 서포트 벡터 머신(Support vector machine, SVM), 랜덤 포레스트(Random forest, RF) 및 인공 신경망(Artificial neural network, ANN), 그리고 심층학습 알고리즘으로는 장단기 메모리(Long short-term memory, LSTM)를 채택하였다. 이러한 단일 알고리즘의 성능은 초매개변수 조정, 앙상블 학습(기계 학습만 적용), 및 다중 레이어 스택킹(2층의 LSTM 레이어 구조)을 거친 변형된 모델의 성능과 비교되었다. 평가에 사용된 모든 알고리즘은 2017년부터 2022년까지 청주 국가산업단지 폐수처리 공정의 일별 유출수를 입력 자료로 사용하였으며, 모델의 성능은 평균제곱오차(Mean squared error, MSE)를 기준으로 평가되었다. 모델 성능 평가를 위해 9개의 관측 항목 중 유량과 생물화학적 산소요구량은 독립 변수로 선택되었다. 평가 결과, 모든 기계학습 알고리즘의 성능은 경쟁 알고리즘인 LSTM에 비해 보다 우수한 것으로 조사되었다. LSTM에 제공된 작은 입력 데이터 양과 불안정한 방류수 폐수 특성에 기인한 것으로 판단된다. 한편, 초매개변수 조정은 평가에 사용된 모든 알고리즘의 성능을 향상하였다. 그러나, 기계학습에 적용된 앙상블 학습과 2층 구조의 LSTM은 독립 변수에 관계없이 단일 알고리즘들의 성능과 비교하여 성능의 저하를 야기하였다. 따라서, 이러한 결과들은 변형 알고리즘(특히 모델의 구조 및 성능 향상 과정들)의 주의 깊은 설계와 평가를 요구하는 것으로 판단된다.

주제어: 기계학습, 심층학습, 초매개변수 조정, 앙상블 학습, 다중 레이어 스택킹

1. 서론

폐수처리시설은 도시의 물 인프라 중 하나로서, 시민의 건강을 보호하고 공공수역의 수질 및 수생태계 건강성을 증진하는데 중요한 역할을 하고 있다.¹⁾ 수처리시설은 사업장별 유입 폐수의 특성 및 처리 요구사항에 따라 다양한 단위 공정들의 조합으로 구성될 수 있으며, 물리적, 화학적, 생물학적 처리시설이 고도처리시설과 함께 주요 처리 공법으로 적용될 수 있다.^{2,4)} 수치해석 모델(예시: 전산유체역학 프로그램)은 이러한 다양한 단위 공정들로부터 효율적인 폐수처리 공정을 설계하고 처리 목표에 적합한 최적의 운전 조건을 도출하기 위해서 활용될 수 있으며, 다만 유입 부하의 변동과 더불어 폐수처리 공정의 복잡성 및 비선형성의 특성으로 인하여 모든 공정 매커니즘의 모의 및 해석에는 다소 한계가 있는 것으로 평가되고 있다.⁵⁻⁸⁾

한편, 최근 인공지능 기반의 기계학습 및 심층학습 알고리

즘은 입력 및 출력 데이터의 복잡하고 비선형적인 관계의 학습을 통해 예측 성능이 우수한 모델 개발이 가능한 것으로 알려져 있다.^{9,10)} 예를 들면, Wang 등¹¹⁾은 기계학습 및 심층학습 알고리즘을 조합하여 폐수처리 공정간 시간 지연의 영향을 설명할 수 있는 방법을 개발하였으며, 이를 통해 폐수처리 공정의 제어 정확도를 높이고 운영 비용을 낮출 수 있는 전략을 개발하였다. 또한, Safder 등¹²⁾은 기계학습 모델을 적용하여 폐수처리 공정의 유출수의 총질소를 사전에 예측할 수 있는 모델을 개발하였으며, 평가 결과 1시간 및 3시간 전 예측 시 98.1% 및 96.3%의 높은 정확도를 가지는 것으로 평가되었다. Alali 등¹³⁾은 23개의 기계학습 모델을 폐수처리 공정의 에너지 소비 예측에 활용하였으며, 시차 지연 변수를 포함하는 모델이 예측 성능을 보다 향상하는 것으로 조사되었다. 이외에도 최근 선형 연구 결과에 따르면, 심층학습 모델은 폐수처리 공정에서 공정 시뮬레이션, 공정 제어 및 최적화, 직접 측정이 어려운 변수의 상태 진단, 데이터 품질 관리(고장 감지, 진단,

결측치 대체 등)에도 광범위하게 활용되고 있는 것으로 조사되었다.¹⁴⁾ 이외에도 앙상블 모델을 통해 폐수처리 공정의 예측 정확도를 향상하는 연구가 제시된 바 있으며, 초매개변수 조정을 통해 모델 예측 성능이 향상이 부분적으로 가능한 것으로 알려져 있다.^{13),15),16)}

이상의 선행연구 사례들에서 제시된 바와 같이 기계학습 및 심층학습 모델이 폐수처리 공정에 광범위하게 적용될 수 있는 반면, 현재까지 폐수처리 공정에서 관측되는 데이터의 다양한 시공간 패턴 및 해상도 등으로 인하여 두 유형의 모델간 성능 및 변형 알고리즘의 성능을 비교한 연구는 제한적인 것으로 판단된다. 따라서, 본 연구에서는 폐수처리 공정의 일별 유출 수 자료를 이용하여 기계학습 및 심층학습의 성능을 일차적으로 비교분석하고, 이후 초매개변수 조정 및 변형 알고리즘을 적용하여 성능 변화를 관찰하고자 수행되었다. 본 연구를 통해 도출된 결과는 향후 폐수처리시설의 단위 또는 전체 공정에 대한 복합(Hybrid) 예측 모델 개발 또는 평가에 유용하게 활용될 수 있을 것으로 예상된다.

2. 실험방법

2.1 폐수처리 공정

본 연구에서는 기계학습 및 심층학습 알고리즘의 성능을 비

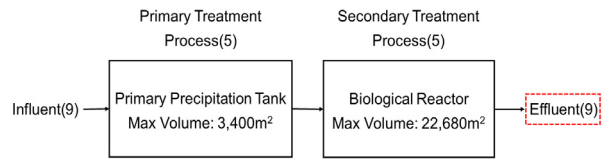


Fig. 1. Schematic diagram of wastewater treatment processes in the Cheong-Ju National Industrial Complex.

교하기 위하여 청주시의 국가산업단지에 위치한 폐수종말처리장을 선정하였다. 선정된 폐수종말처리장의 시설용량은 31,000 m³/일이며 전자부품, 인쇄회로기판, 반도체, 섬유·의복 등의 업종에서 발생하는 폐수를 처리한다. 일차 침전, ASTR Control System (ACS) 고도처리공법, 가압부상, 오존(동결기)을 통해 폐수처리를 하고 있으며, 계획 및 운영 수질은 Table 1에 제시하였다. 선정된 폐수종말처리장의 개략적인 공정 개요도는 Fig. 1에 제시되어 있으며, 주요 처리공정은 1차 침전공정과 2차 생물학적 처리공정으로 구성되어 있다. 유입 수, 1차 공정, 2차 공정 및 유출수에서 각각 측정되는 모니터링 항목의 개수는 공정별로 괄호 안에 제시되어 있다. 대상 폐수종말처리장의 유출수로부터 최근 6년간 일별로 측정된 자료는 Table 2에 제시되어 있으며, 본 평가에서는 일부 결측된 자료를 제외하고 총 1,527개의 데이터가 사용되었다.

Table 1. The effluent quality targets during planning and operating phases for wastewater treatment plant in the Cheong-Ju National Industrial Complex.

	BOD	COD	SS	T-N	T-P	TC	Ecotoxicity
in the planning stage							
Influent	288	249	285	53	5	250,000	-
Effluent	8 or less	20 or less	10 or less	20 or less	0.3 or less	3,000 or less	1 or less
in the operational stage							
Influent	187.1	107.4	64.4	31.447	1.631	241,888	3.3
Effluent	2.9	10.5	3.5	11.605	0.072	357	0

Table 2. Descriptive statistics for 9 effluent quality parameters observed on a daily basis between 2017 and 2022 in the selected wastewater treatment plant.

	Methods ^{a)} or equipment	Units	n	Mean	Coefficient of variation
Discharge	Pashall flumes	m ³ /day	1,527	21,758.12	0.17
Water temperature	Digital/mercury thermometer	°C	1,527	23.31	0.21
pH	pH meter	-	1,527	6.75	0.05
Biochemical oxygen demand (BOD)	ES 04308.2b	mg/L	1,527	2.93	0.51
Chemical oxygen demand (COD)	ES 04315.1a	mg/L	1,527	11.76	0.27
Suspended solids (SS)	ES 04303.1b	mg/L	1,527	3.36	0.30
Total nitrogen (T-N)	ES 04363.1a	mg/L	1,527	11.03	0.47
Total phosphorus (T-P)	ES 04362.1b	mg/L	1,527	0.07	0.52
Total coliforms (TC)	ES 04701.3b	cfu/100 mL	1,527	324.83	0.90

^{a)} Official test method for water quality pollution (Notice No. 2014-163, Ministry of Environment, Republic of Korea)

2.2 단일 알고리즘

본 연구에서는 폐수처리 공정의 유출수 자료를 사용하여 기계학습 및 심층학습 기반의 단일 및 변형 알고리즘들의 성능을 비교하였다(Fig. 2). 성능 평가를 위해 기계학습 기반의 단일 알고리즘으로는 서포트 벡터 머신(Support vector machine, SVM), 랜덤 포레스트(Random forest, RF) 및 인공 신경망(Artificial neural network, ANN)이 사용되었다(Fig. 2a). 또한, 심층학습 기반의 알고리즘으로는 장단기 메모리(Long short-term memory, LSTM)가 사용되었다(Fig. 2b). 참고로, 본 연구에서 사용된 3가지 단일 기계학습 모델의 경우 분류 및 회귀 분석에 연구자들에게 광범위하게 활용되고 있으며, 심층학습 모델 장단기 메모리의 경우 시계열 예측모델에 개발에 우수한 성능을 보이는 것으로 잘 알려져 있다. 단일 알고리즘의 성능을 보다 향상하기 위하여 초매개변수 조정(Hyperparameter tuning)이 추가적으로 수행되었으며, 기계학습 및 심층학습 모두 사용자가 정의한 그리드 검색(Grid search)을 통해 최적의 매개변수가 결정되었다(Fig. 2a 및 2b). 보다 구체적으로 그리드 검색에 적용된 기계학습 알고리즘의 초매개변수로는 C(1~15)가 사용되었고 심층학습 알고리즘의 초매개변수로는 unit(32~256) 및 dropout(0.2~0.3)이 사용되었다.

2.3 변형 알고리즘

본 연구에서는 단일 알고리즘 및 초매개변수 조정 이외에 추가적인 성능 향상 가능성을 확인하기 위하여 앙상블 학습과 다층 레이어 기법을 도입하였다(Fig. 2c 및 2d). 앙상블 학습은 기계학습 모델의 성능 향상을 위해 적용된 기법으로서, 본 연구에서는 보팅(Voting), 배깅(Bagging), 부스팅(Boosting), 스택킹(Stacking) 등 다양한 학습 유형 중 스택킹을 사용하였다.

또한, 스택킹의 최종 모델(메타 모델)로는 확률적 경사 부스팅(Stochastic gradient boosting, SGB) 알고리즘이 사용되었다. 참고로, 대표적인 앙상블 학습별로 다양한 세부 알고리즘들이 제공되고 있지만, 본 연구에서는 모델간 비교 평가를 단순화하기 위하여 스택킹 학습 기법만을 제한적으로 사용하였다(Fig. 2c). 한편, 다층 레이어 기법은 심층학습 모델의 성능 향상을 위해 활용되었으며, 본 연구에서는 2층으로 구성된 LSTM 알고리즘을 평가에 사용하였다(Fig. 2d). 본 연구에서 활용된 단일 및 변형 알고리즘들의 성능 평가는 R(version 4.1.1) 프로그램을 기반으로 수행되었으며, 기계학습 모델은 Caret(version 6.0.94) 패키지가, 심층학습 모델은 Keras(version 2.13.0) 패키지가 각각 활용되었다. 또한, 세부적으로 사용자 정의 초매개변수 조정을 위해 기계학습 및 심층학습 모델 모두 expand.grid() 함수가 사용되었으며, 특히 심층학습 모델에서는 다수의 매개변수를 동시에 조정할 수 있는 tuning_run() 함수가 추가적으로 활용되었다. 알고리즘의 성능 평가에 활용된 예측 변수로는 유량(Discharge)과 생물화학적 산소요구량(Biochemical oxygen demand, BOD)이 설정되었으며, 손실함수로는 평균제곱오차(Mean squared error, MSE)가 사용되었다.

3. 결과 및 고찰

3.1 단일 알고리즘 성능 비교

Fig. 3은 기계학습 및 심층학습 기반의 단일 알고리즘간의 성능을 비교한 결과를 도시하고 있다. 성능 예측은 유량과 생물화학적 산소요구량을 대상으로 수행되었다. 평가 결과, 3가지 기계학습 알고리즘 중 랜덤 포레스트(RF)의 성능이 다른 기계학습 모델(SVM 및 ANN)의 성능에 비해 예측 변수에 상관없이 가장 우수한 것으로 평가되었으며, 다만 알고리즘간 예측 성능은 다소 차이는 있지만 큰 차이는 발생하지 않는

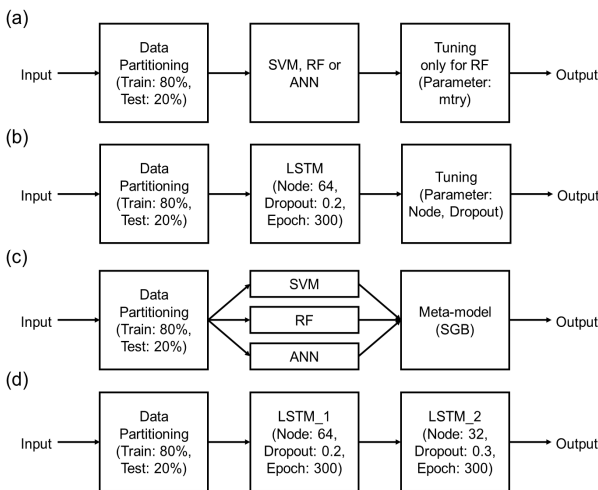


Fig. 2. Single and modified machine learning and deep learning algorithms examined in this study: single algorithms with hyperparameter tuning for (a) machine learning and (b) deep learning, and modified algorithms with (c) ensemble learning and (d) multi-layer stacking.

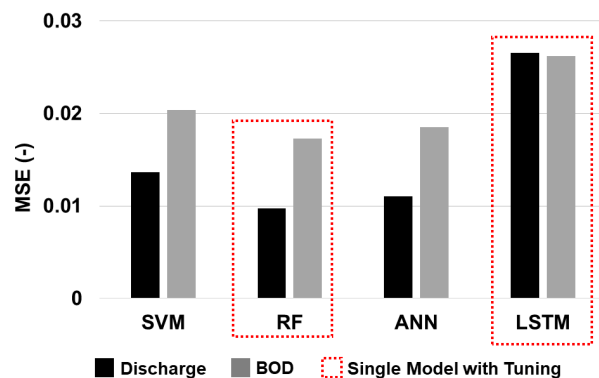


Fig. 3. The accuracy of single machine learning and deep learning algorithms assessed for two dependent variables discharge and BOD in terms of MSE.

Table 3. The accuracy of RF and LSTM algorithms with hyperparameter tuning assessed for two dependent variables discharge and BOD in terms of MSE.

Prediction algorithms	Variables	
	Discharge	BOD
RF(T)	0.01 (+1.03%)	0.02 (+1.16%)
LSTM(T)	0.02 (+40.60%)	0.03 (+3.81%)

것으로 평가되었다. 그러나, 초기 예상과 달리 심층학습 알고리즘 장단기 메모리(LSTM)의 오차는 평가에 사용된 3가지 기계학습 모델에 비해 다소 높은 것으로 조사되었으며, 이는 평가에 활용된 입력 데이터의 수가 상대적으로 적고 국가산업단지 폐수처리 공정의 유출수가 일정한 시계열 패턴을 갖지 않는 것에 기인한 것으로 판단된다. 참고로, 그림에서 점선은 추가적인 초매개변수 조정이 수행된 2개 알고리즘을 나타내고 있으며, 이는 1) 기계학습 알고리즘 중 랜덤 포레스트(RF)의 성능이 가장 우수하고 2) 추가적인 초매개변수 조정 이후 두 유형의 알고리즘간의 성능 변화를 보다 세부적으로 관찰하기 위해 수행되었다.

3.2 매개변수 조정 후 성능 비교

Table 3은 초매개변수 조정 이후 기계학습 및 심층학습 기반의 단일 알고리즘간의 성능 변화를 나타내고 있다. 표에서 알고리즘명 이후 괄호 안의 영문 대문자 T는 초매개변수가 조정된 알고리즘을 나타내며, 오차 값 이후 괄호안의 백분율은 초매개변수 조정 전후의 성능 향상 정도를 나타내고 있다. 즉, 초매개변수 조정 전에 비해 초매개변수 조정 이후 평균제곱오차(MSE)가 낮아질 경우 알고리즘의 성능은 향상된 것으로 평가될 수 있으며, 이 경우 백분율은 양의 방향을 가지게 된다. 평가 결과, 두 유형의 알고리즘 간 오차는 큰 차이는 없지만 기계학습 알고리즘 랜덤 포레스트(RF)의 성능은 심층학습 알고리즘 장단기 메모리(LSTM)에 비해 예측 변수에 관계없이 다소 높은 것으로 조사되었다. 다만, 초매개변수 조정 후 장단기 메모리(LSTM)의 오차는 랜덤 포레스트(RF)에 비해 보다 높게 향상되는 것으로 평가되었으며, 다만 초매개변수 조정 이후에도 예측 성능은 여전히 랜덤 포레스트(RF)에 비해 낮은 것으로 조사되었다.

3.3 변형 알고리즘 성능 비교

Table 4는 앙상블 모델과 심층학습 기반의 2층으로 구성된 장단기 메모리(LSTM)의 성능 변화를 비교하고 있다. 여기에서, 앙상블 모델이란 기계학습 기반의 알고리즘들의 조합을 통해 개발된 모델을 의미하며(**Fig. 2c**), 알고리즘명 이후 괄호 안의 영문 대문자 M은 다층으로 구성된 알고리즘을 나타내고 있다(**Fig. 2d**). 또한, 오차 값 이후 괄호안의 백분율은 기존

단일 알고리즘과 비교하여 성능 향상 정도를 나타내고 있다. 즉, 앙상블 모델의 경우 3가지 단일 알고리즘들의 평균 오차와 비교하여 앙상블 모델의 성능 향상 정도를 나타내고 있으며, 장단기 메모리(LSTM)의 경우 단층과 2층으로 구성된 모델들의 성능 향상 정도를 평가하고 있다. 평가 결과, 기계학습 기반의 앙상블 모델은 장단기 메모리(LSTM)의 성능에 비해 우수한 성능을 나타내는 것으로 평가되었다. 그러나, 장단기 메모리(LSTM)를 사용하여 생물화학적 산소요구량을 예측하는 경우를 제외하고, 앙상블 모델과 장단기 메모리(LSTM) 알고리즘 모두 단일 알고리즘의 성능에 비해 모델의 오차는 대체로 낮아지는 것으로 조사되었다. 이는 일반적으로 앙상블 모델과 다층으로 구성된 심층학습의 알고리즘의 성능이 단일 알고리즘에 비해 향상되는 것으로 알려져 있지만, 모델 구조에 대한 세부적인 설계와 검증이 수반되지 않는다면 예측 성능 향상이 용이하지 않는다는 사실을 반증하는 결과로 판단된다.

4. 결론

본 연구에서는 청주 국가산업단지 폐수처리 공정의 일별 유출수 자료를 이용하여 기계학습 및 심층학습 기반의 단일 및 변형 알고리즘의 성능을 평가하였다. 단일 알고리즘으로는 3가지 기계학습 모델(SVM, RF, ANN)과 1개의 심층학습 모델(LSTM)이 사용되었으며, 단일 알고리즘의 성능 변화를 파악하기 위하여 초매개변수 조정과 앙상블 학습 및 다층 구조의 알고리즘이 적용되었다. 본 연구를 통해 도출된 주요 결과는 다음과 같다.

- 3가지 기계학습 모델(SVM, RF, ANN)과 1개의 심층학습 모델(LSTM)의 성능을 비교한 결과 기계학습 모델들의 성능이 심층학습 모델에 비해 다소 우수한 것으로 평가되었다. 심층학습 알고리즘의 성능 저하의 주요 사유로는 입력 데이터의 개수 및 일정하지 않는 유출수 패턴이 주요 사유로 추정된다.
- 초매개변수 조정은 모델 성능 향상을 위해 반드시 필요한 과정으로 조사되었다. 초매개변수 조정 이후 심층학습 모델(LSTM)의 성능은 기계학습 모델(RF)에 비해 크게 향상된 것으로 평가되었지만, 다만 심층학습 모델(LSTM)의

Table 4. The accuracy of two modified algorithms, the one developed from ensemble learning and the other derived from two LSTM layers, assessed for two dependent variables discharge and BOD in terms of MSE.

Prediction algorithms	Variables	
	Discharge	BOD
Ensemble	0.01 (-28.94%)	0.02 (-13.90%)
LSTM(M)	0.03 (-2.63%)	0.03 (+4.20%)

성능은 기계학습 모델의 오차에 비해 여전히 다소 높은 것으로 조사되었다.

- 기계학습 기반의 앙상블 모델 및 다층 심층 알고리즘의 적용은 단일 알고리즘의 성능을 오히려 낮추는 역할을 하는 것으로 조사되었다. 따라서, 변형 알고리즘 적용 시 모델 구조에 대한 세심한 검토가 필요한 것으로 판단되며, 향후 초매개변수 조정 과정과 순차적으로 적용되어 성능 평가가 수행되어야 할 것으로 판단된다.

Acknowledgement

본 논문은 농촌진흥청 연구사업(세부과제번호: PJ015845)의 지원에 의해 이루어진 것임.

References

1. Ministry of Environment, Water Environment Conservation Act(2022).
2. Lee, Jong-Tak, Analysis of wastewater inflow rate in wastewater treatment plants and review of improvements in plant capacity calculation method, *J. Korean Soc. Water Wastewater*, 28(6), 747-753(2014).
3. Mi Soo Shin, Hey Suk Kim, Jin Young Joh, Jun Ho Choi, Dong Soon Jang, Numerical Fluid Dynamic Study for Improvement of Mixing Efficiency in the Contactor, *J. Korean Soc. Environ. Eng.*, 28(8), 860-865(2006).
4. Onodera T., Komatsu K., Kohzu A., Kanaya G., Mizuochi M., and Syutsubo K., Differences in the isotopic signature of activated sludge in four types of advanced treatment processes at a municipal wastewater treatment plant, *J. Environ. Manage.*, 286, 112264(2021).
5. Koyilath Nandakumar V., Palani S. G., and Raja Raja Varma M., Interactions between microplastics and unit processes of wastewater treatment plants: A critical review, *Water Sci. Technol.*, 85(1), 496-514(2021).
6. Elaissaoui Elmeliani M., El Amine, Aguedal H., Iddou A., Alaoui C., Benaissa B., Belhadj M. E. A., Nguyen T., Sun M., and Terashima M., Optimizing the disinfection inactivation efficiency in wastewater treatment: A computational fluid dynamics investigation of a full-scale ozonation contactor, *Chem. Eng. Technol.*, 46, 1-11(2023).
7. Wang H., Ohta M., Anzai H., and Ji J., Computational fluid dynamics simulation to investigate diffuser outlet factors in anaerobic membrane bioreactors treating wastewater, *Sustainability*, 15(15), 11959(2023).
8. Khatri N., Singh M., Pokhriyal S., and Rene E. R., Computational fluid dynamics modelling of primary sludge classification in an activated sludge process based wastewater treatment plant: Simulating the hydrodynamic behaviour and experimental verification of the classification efficiency, *Chem. Eng. J.*, 464, 142475(2023).
9. Oliveira P., Fernandes B., Analide C., and Novais P.,

- Forecasting energy consumption of wastewater treatment plants with a transfer learning approach for sustainable cities, *Electronics*, 10(10), 1149(2021).
10. Kang H., Yang S., Huang J., and Oh J., Time series prediction of wastewater flow rate by bidirectional LSTM deep learning, *Int. J. Control Autom. Syst.*, 18(12), 3023-3030(2020).
 11. Wang D., Thunell S., Lindberg U., Jiang L., Trygg J., Tysklind M., and Souihi N., A machine learning framework to improve effluent quality control in wastewater treatment plants, *Sci. Total Environ.*, 784, 147138(2021).
 12. Safder U., Kim J., Pak G., Rhee G., and You K., Investigating machine learning applications for effective real-time water quality parameter monitoring in full-scale wastewater treatment plants, *Water*, 14(19), 3147(2022).
 13. Alali Y., Harrou F., and Sun Y., Unlocking the potential of wastewater treatment: Machine learning based energy consumption prediction, *Water*, 15(13), 2349(2023).
 14. Alvi M., Batstone D., Mbamba C. K., Keymer P., French T., Ward A., Dwyer J., and Cardell-Oliver R., Deep learning in wastewater treatment: A critical review, *Water Res.*, 245, 120518(2023).
 15. Nourani V., Asghari P., and Sharghi E., Artificial intelligence based ensemble modeling of wastewater treatment plant using jittered data, *J. Clean. Prod.*, 291, 125772(2021).
 16. Brownlee, J., *Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models and Work Projects End-To-End*, Machine Learning Mastery(2016).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors and Contribution Statement

Jaecil Kim

Department of Energy Engineering, Gyeongsang National University, Graduate student, ORCID[®] 0009-0000-8276-1636: Writing - original draft, Writing - review and editing.

Hyo Sub Lee

Residual Agrochemical Assessment Division, National Institute of Agricultural Sciences, Rural Development Administration, Researcher, ORCID[®] 0000-0002-7021-2514: Project administration, Methodology, Writing - review and editing.

Jinuk Jang

Department of Energy System Engineering, Gyeongsang National University, Graduate student, ORCID[®] 0000-0002-6540-9080: Data analysis, Validation, Visualization.

Yongtae Ahn

Department of Energy Engineering, Gyeongsang National University, Associate Professor, [ORCID](#) 0000-0001-6294-710X: Conceptualization, Methodology, Writing - review and editing.

Seo Jin Ki

Department of Environmental Engineering, Gyeongsang National University, Associate Professor, [ORCID](#) 0000-0001-7056-9217: Funding acquisition, Project administration, Supervision, Writing - original draft, Writing - review and editing.

Hyun-Geoun Park

Department of Environmental Engineering, Gyeongsang National University, Professor, [ORCID](#) 0000-0001-8654-5017: Supervision, Writing - review and editing.