



공정 간 연계성을 반영한 Merged LSTM 기반 하수처리장 수질 예측 모델 구축

Development of a Merged LSTM Model for Wastewater Treatment Plants

Incorporating Inter-Process Connectivity

김진호, 박건하, 박범진

서울시립대학교 환경공학부

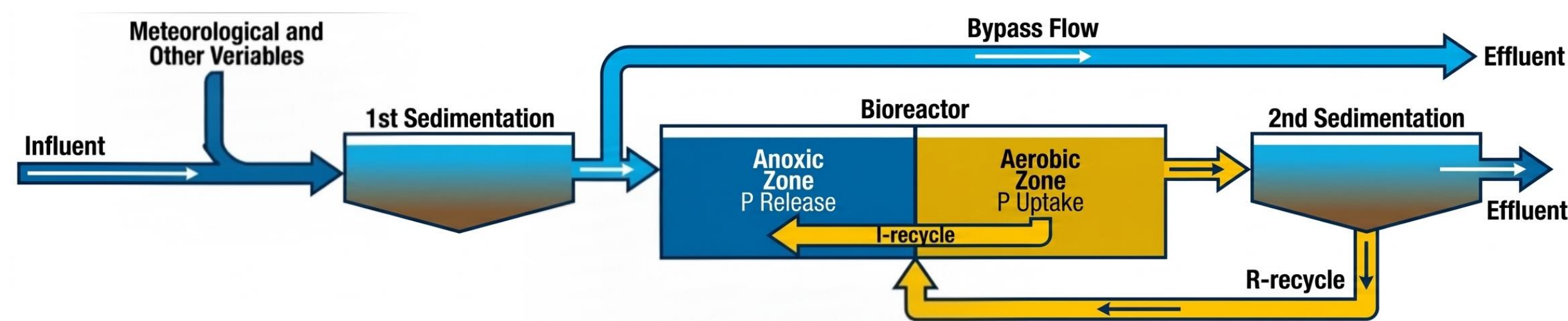
Introduction

Background

- 하수처리장에서 유입되는 과도한 질소-인 부하(TN 등)는 부영양화를 유발하며, 안정적인 제거 및 예측 제어를 어렵게 만드는 주요 요인임
- 기존 연구에서는 ASM, ARIMAX, RNN, LSTM 등의 모델로 생물학적 질소 제거 과정을 설명력 있게 모사 및 예측 시도를 하였지만, 공정 간의 연관성을 고려하지 않고 특정 공정에서만 예측하거나 외부 요인을 고려하지 않아 공정 단계 간 상호의존성-연결 구조를 충분히 반영하지 못하는 한계가 존재하며 재보정이 필요했음

Objective

- 본 연구는 MLE 공정을 유입수(IN), 1차 침전지(1st SD), 생물반응조-2차 침전지(BIO-2nd SD)의 3단계로 구분하고, 공정간 연계성을 반영한 Merged LSTM 기반 TN 예측 프레임워크를 제안함
- 하위 공정에서 상위 공정의 영향을 학습하고, 공정별 변수 중요도 분석을 통해 주요 인자를 도출함으로써 지능형 공정 제어-의사결정 지원이 가능한 자료 기반 운영 체계의 기반을 마련하고자 함



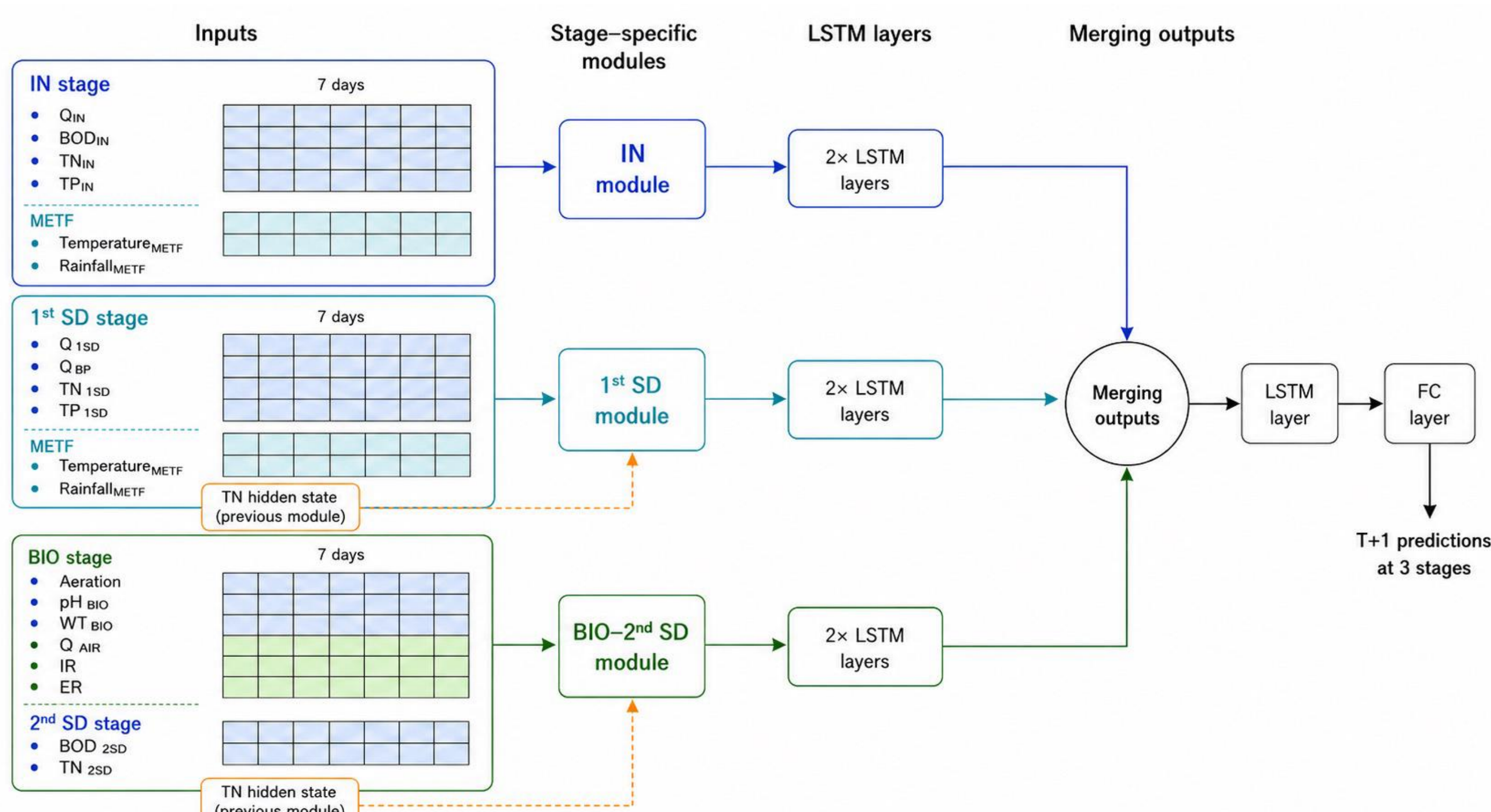
Materials and Method

Study region & Data description

- 연구지역: 서울물재생시설 서남센터
- 활용 데이터 기간: Train (2019.01 - 2021.06), Validation (2022.07 - 2023.03), Test (2023.04 - 2023.12)
- 공정별 활용 입력변수
 - IN module: 유입수(유량, BOD, TN, TP, 수온), 기상(기온, 강수량)
 - 1st SD module: 1차 침전지(유량, 바이패스 유량, TN, TP), 기상(기온, 강수량)
 - 2nd SD module: 2차 침전지(TN, BOD), 생물반응조(pH, 수온, aeration)

Model development

- Data preprocessing
 - 입력변수 선정: 각 공정별 TN과의 Pearson 상관관계 기반 입력변수 선정
 - 이상치 처리: Isolation Forest 기반 이상치 제거
 - 결측치 보간: Unscented Kalman Filter 기반 결측 데이터 보완
 - 표준화: Min-Max 정규화
- Model development
 - 제안 모델: Merged LSTM
 - 비교 모델: Local-LSTM, Upstream-LSTM
 - Local-LSTM: 각 모듈의 자체 입력변수만 사용하여 개별적으로 학습
 - Upstream-LSTM: 하위 공정 예측 시 상위 공정의 모든 입력변수까지 함께 사용하여 학습
- Model interpretation
 - 모듈 단위: Merged-LSTM에서 학습된 스칼라 가중치를 이용해 상대적 기여도 산정
 - 변수 단위: Integrated gradients 기반으로 입력 변수별 상대적 중요도 분석



Results and Discussion

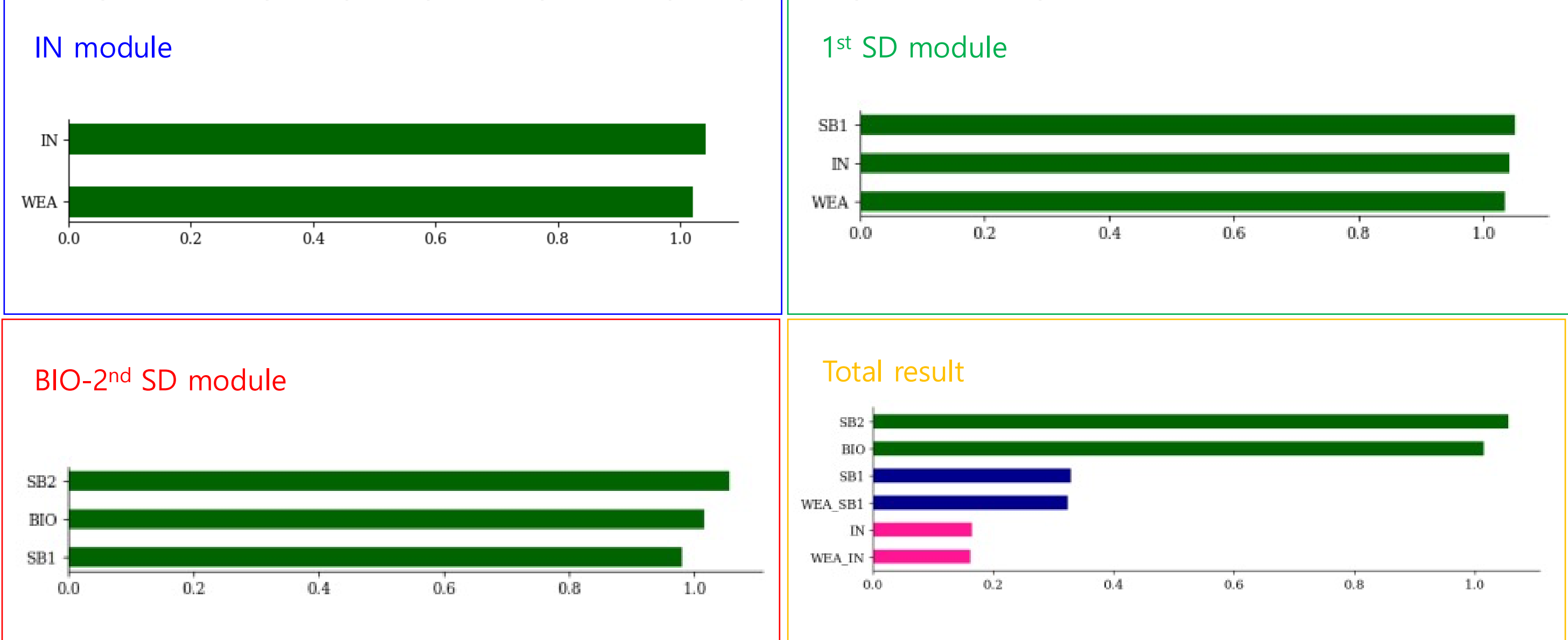
Model performance

- 선행한 EDA(Exploratory Data Analysis) 결과를 바탕으로 공정별 예측 모델의 입력 변수 구성을 최적화 하기 위해 결측치 비율과 상관계수를 기준으로 세 가지 Case의 변수 조합을 도출함
- 각 Case에서 SD2 오차제어는 Case 2의 성능이 가장 우수하지만, Case 3가 IN, SD1에서 가장 높은 예측 효율을 보이며, 종합적으로 고려하였을 때 가장 성능이 안정적으로 좋다고 판단하여 Case 3를 선정
- Merged-LSTM은 선정된 Case 3 모듈에서 테스트 기간(2023.04-2023.12)동안 안정적인 예측 성능을(NSE 0.516-0.707) 보임
- SD2 모듈에서는 Local-LSTM의 예측 성능이 우수했는데, 이는 SD2 공정이 생물학적 반응 및 내부 운전 조건에 의해 상대적으로 안정적인 상태를 유지하는 구간일 가능성으로 시사됨
- Merged-LSTM은 공정 단계별 TN을 예측하고 비선형 관계와 공정 간의 유기적 연계를 모델 내부 구조에 반영하고 있어 단위 공정만 예측하는 기존 모델보다 운영 의사결정 지원에 유리함

Model	Evaluation metric	Test			Test		
		Case 1	Case 2	Case 3	Merged - LSTM (Case 3)	Local - LSTM	Upstream - LSTM
IN	NSE	0.705	0.694	0.706	0.706	0.697	0.697
	RMSE	2.860	2.913	2.841	2.841	3.002	3.002
	MAE	2.103	2.115	2.067	2.067	2.152	2.152
SD1	NSE	0.705	0.700	0.707	0.707	0.698	0.708
	RMSE	2.860	3.024	2.988	2.988	3.047	2.996
	MAE	2.100	2.161	2.162	2.162	2.152	2.131
SD2	NSE	0.516	0.526	0.516	0.516	0.539	0.485
	RMSE	1.239	1.225	1.240	1.240	1.183	1.249
	MAE	0.978	0.972	0.980	0.980	0.869	0.992

Model Interpretation

- 각 공정별 독립적인 LSTM 브랜치 내 변수 가중치 분석 결과, 모든 공정에서 수질 데이터뿐 아니라 외부 기상 요인도 예측에 유의미한 영향을 미치는 것으로 나타남
- 전체 통합 가중치 분석 결과, BIO 및 SD2 공정 데이터의 기여도가 상대적으로 높게 나타났으며, 이는 상류 공정의 영향이 하류 공정으로 단계적으로 전달되는 구조적 특성이 예측 과정에 반영되었음을 시사함
- IN 공정에서는 유입수 자체 데이터와 기상 데이터가 유사한 수준의 기여도를 보였으며, 이는 유입 부하가 외부 기후 변화에 영향을 받는 특성이 예측 과정에 반영되었음을 시사함
- SD1 공정에서 해당 공정 데이터와 IN 공정 모두 높은 비중을 차지하여 공정 간 연속성 반영 확인
- SD2 공정에서는 SD2 자체 데이터와 BIO 공정 데이터의 기여도가 상대적으로 높게 나타나 최종 예측 단계에서 두 공정의 영향이 크게 반영됨을 확인함
- 이러한 결과는 공정 간 종속 관계가 예측 과정에 반영되었음을 시사하며, 단순한 선형 상관관계를 넘어 공정 간 구조적 연계 정보가 모델 학습에 활용되었음을 보여줌



Conclusion

- 본 연구는 MLE 공정을 IN-1st SD-BIO/2nd SD로 모듈화한 Merged-LSTM 프레임워크를 구축하여, 공정 간 연계성을 반영한 TN 농도 예측을 수행함
- 주로 단일 공정만 예측하는 기존 모델과 달리, 본 프레임워크는 공정별 입력 시계열을 독립적으로 학습한 뒤, 이를 통합하는 구조로 설계하여 공정 간 연속성과 외부 요인을 동시에 반영하는 예측 모델을 구축함
- 입력 변수의 상대적 중요도를 정량적으로 해석할 수 있고 하수처리 공정 간 구조적 영향 관계를 정량적으로 설명 가능한 시계열 기반 수질 예측 모델 적용 가능성을 확인할 수 있음
- 향후 연구에서는 여러 처리장 대상으로 적용성 검증과 장기 시계열 예측 구조 및 다양한 딥러닝 기반 예측 모델과의 비교 분석을 통해 모델 확장 가능성을 검토할 예정이며, 이를 실시간 운영 자료와 연계하여 운영 의사결정 지원 도구로 활용될 가능성이 있음