

하천 부영양화 평가를 위한 해석가능한 딥러닝 모델의 적용

Application of Explainable Deep Learning Model for River Eutrophication Assessment

양수현, 이해든, 이건형, 차윤경†

서울시립대학교 환경공학부

† Corresponding author, E-mail: ykcha@uos.ac.kr

Introduction

Background

- 전세계적으로 기후 변화와 인간의 활동으로 인해 하천에서의 부영양화 발생 빈도와 강도가 증가하고 있는 추세임
- 하천 부영양화는 수생태계에 악영향을 미치며 인간의 건강에도 위협을 주기 때문에 이를 관리하기 위한 보다 고도화된 모니터링이 요구되고 있음
- 최근 센서 네트워크를 활용해 클로로필 *a*와 같이 하천 부영양화 관리에 중요한 정보를 보다 높은 시간해상도로 확보할 수 있게 됨
- 이렇게 수집된 대규모 환경 데이터를 효과적으로 분석하여 하천 부영양화 관리에 활용할 필요가 있음

Objectives

- 고해상도 환경 데이터에 대한 해석가능한 딥러닝 모델인 RETAIN-D의 적용 가능성을 평가하고자 함
- 하천 부영양화에 영향을 미치는 다양한 환경 변수의 변수별, 시점별 기여도를 분석하고, 이를 통해 하천 부영양화 관리에 기여하고자 함

Materials and Method

Study site and data description

- 1) 연구 대상 지역
: 금강 4개 지점 (남면, 미호강, 갑천, 현도)
- 2) 데이터 수집 기간
: 2018 ~ 2022년 (1시간 단위)
- 3) 데이터 종류
: 수온 (°C), TN (mg/L), TP (mg/L), 클로로필 *a* 농도 (mg/m³), 수위 (m), 강수량 (mm)
: 강수량은 대전 지점의 자료를 활용
(출처: 국립환경과학원, 금강홍수통제소, 기상자료개방포털)

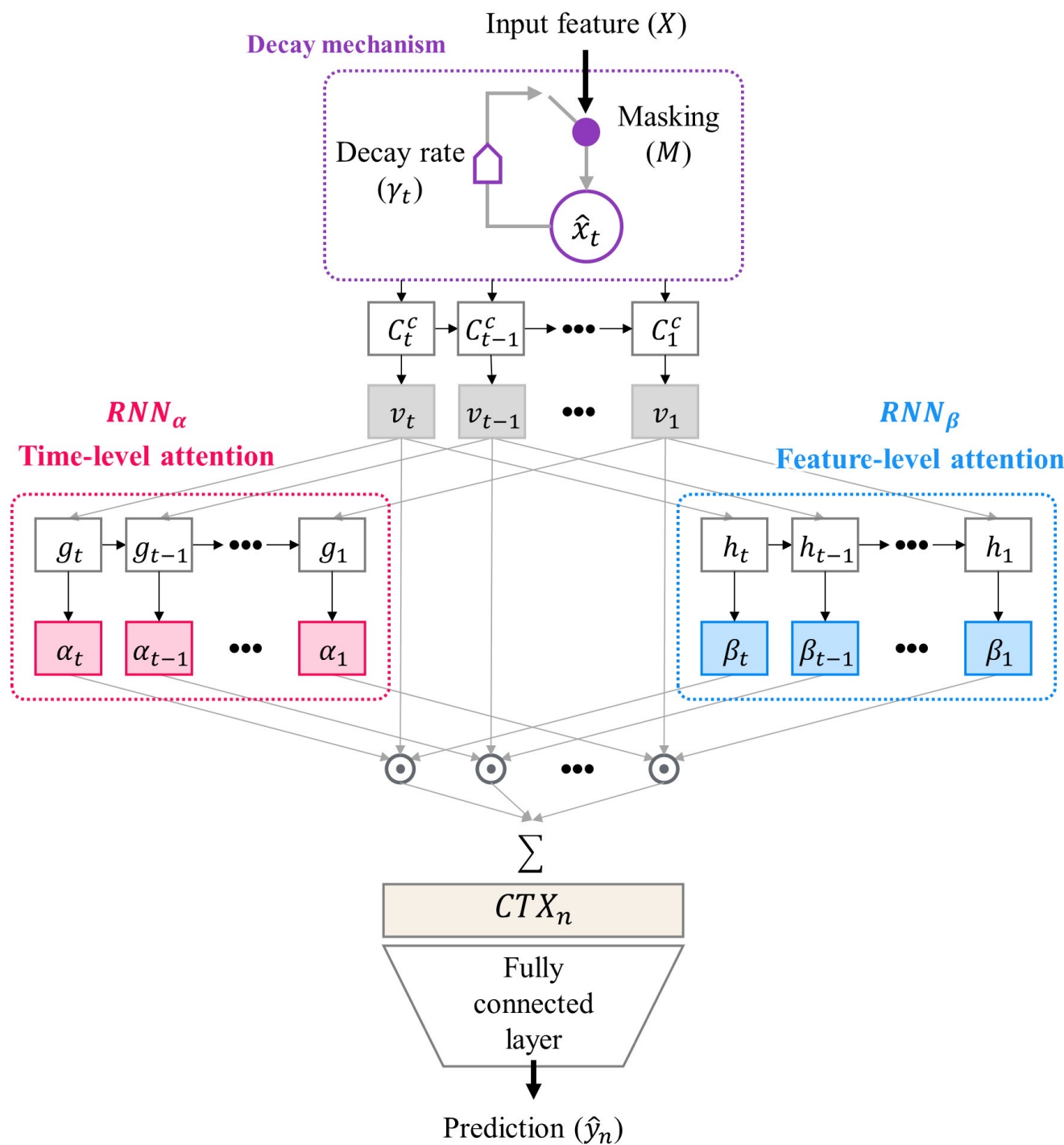


[Fig. 1. 연구 지점]

Interpretable Deep Learning

RETAIN-D (Reverse Time Attention Model with a Decay mechanism)

- 본 연구에서는 하이브리드 딥러닝 모델 중 별도의 조작 없이 확장된 시간해상도로 수질을 예측함과 동시에, 해석을 제공할 수 있는 RETAIN-D를 활용하여 상류 및 지류 지점의 환경인자가 하류 지점의 클로로필 *a* 농도에 미치는 영향을 분석함



[Fig. 2. RETAIN-D 구조]

- RETAIN-D는 1) 학습가능한 감쇄율 (Trainable decay)을 활용하여 입력변수의 결측 패턴을 학습, 결측값에 대한 모델링을 수행하는 감쇄 메커니즘(Decay mechanism)과, 2) 병렬 구조의 어텐션 레이어를 통해 변수 및 시점 각각에 대한 해석 정보를 보존, 예측과 동시에 해석결과를 제공하는 RETAIN(Reverse time attention)이 결합된 형태임

Model implementation

- RETAIN-D의 입력변수로 남면 지점의 수온, TN, TP, 수위, 현도 지점의 수온, TN, TP, 클로로필 *a*, 미호강 및 갑천 지점의 수온 및 클로로필 *a*, 남면 지점의 수위, 그리고 강수량이 활용됨
- RETAIN-D의 종속변수는 수질목표 등급별 남면 지점의 클로로필 *a* 농도 기준 초과/준수 여부임
 - ✓ 현재 클로로필 *a*에 대한 중권역 수질기준이 부재하기 때문에 중권역별 물환경 목표기준을 본 연구에서 클로로필 *a* 기준으로 활용함
 - ✓ 본 연구에서 타겟 지점에 해당하는 남면 지점이 속한 금강공주 중권역의 목표기준은 II 등급(<14 mg/m³)임
 - ✓ 추가로, 연구 대상 지역에서 가장 엄격한 목표기준을 가지고 있는 현도 지점이 속한 대청댐 중권역의 목표기준인 Ia 등급(<14 mg/m³)임
 - ✓ II 등급과 Ia 등급 각각에 대한 모델링을 수행함
- 2018년 1월 1일 00시부터 2021년 12월 31일 23시까지의 데이터를 모델 학습에 활용하였으며, 2022년 1월 1일 00시부터 2022년 12월 31일 23시까지의 데이터를 모델 검증에 활용하였음
- 모델에 입력되는 시계열의 길이는 금강에서의 유하시간을 고려하여 15로 설정하였으며, 결과적으로 RETAIN-D 모델은 현재 시점을 포함 과거 15시간의 정보를 학습하여 예측을 수행하게 됨

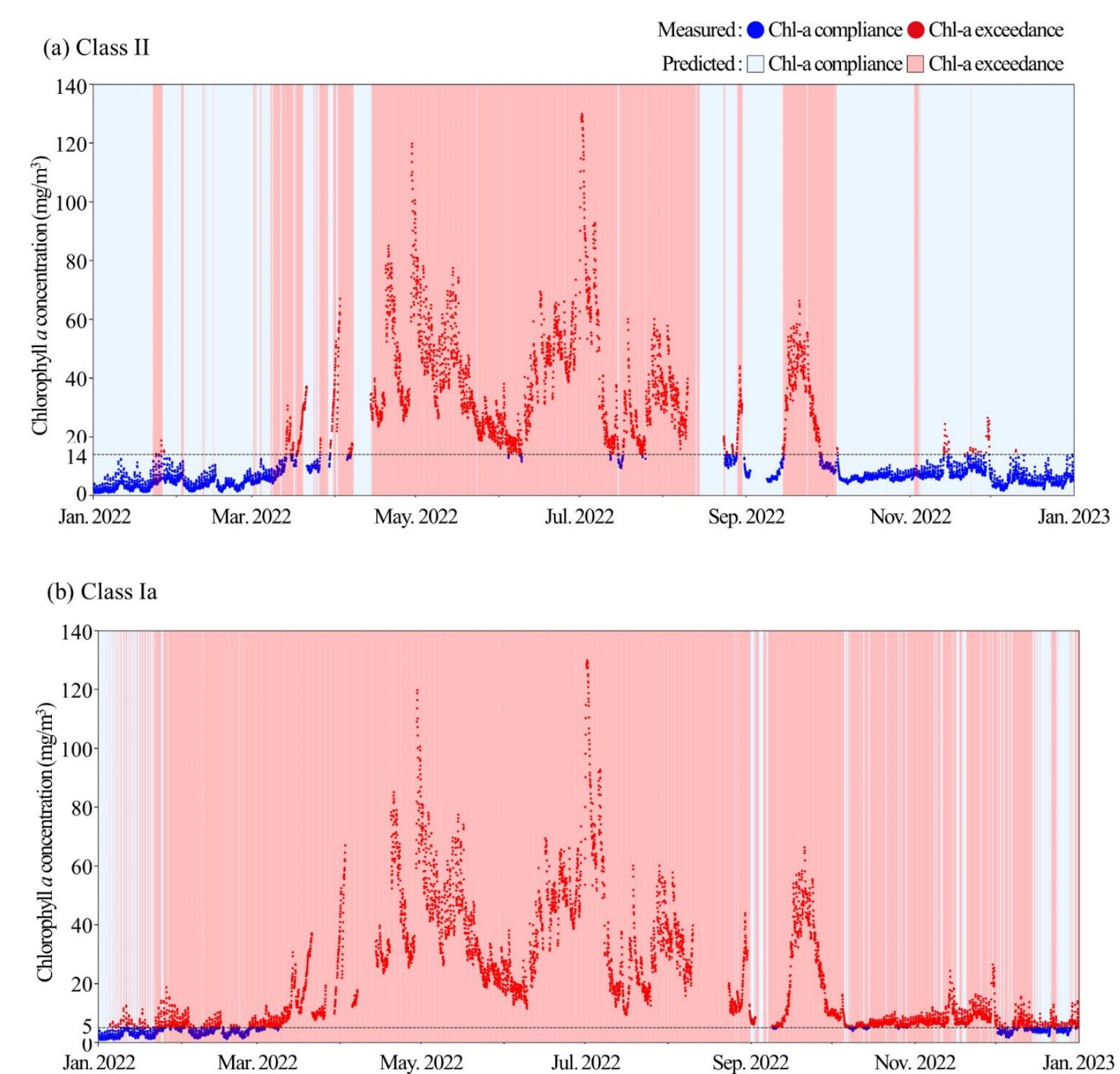
Model Evaluation

- RETAIN-D의 성능을 평가하기 위해 비교모델로 GRU (Gated Recurrent Unit)를 활용함
- GRU 모델링 시에는 시계열 결측 보안을 위해 널리 사용되는 기법 중의 하나인 Kalman Filter를 활용함
- 모델 성능 평가 지표로 Accuracy, AUC (Area Under the ROC Curve), F-measure를 활용함

Results and Discussion

Model performance

- RETAIN-D는 양호한 클로로필 수질 기준 초과 예측 성능을 보이는 것으로 나타남 (Fig. 3)
- RETAIN-D와 GRU 모델의 분류 성능을 비교한 결과, 두 모델 모두 양호한 성능을 보이는 것으로 나타남 (Table 1.)
- 모델 검증 성능의 경우 RETAIN-D와 GRU 모두 기준 II에서 상대적으로 높은 성능을 보이는 것으로 나타남
- Kalman Filter를 통해 사전 결측 보안을 수행한 후 시계열 예측을 수행하는 GRU 모델과 비교했을 때, RETAIN-D는 단일 구조 내에서 결측 보장이 가능하면서도 우수한 성능을 보이는 것으로 나타남



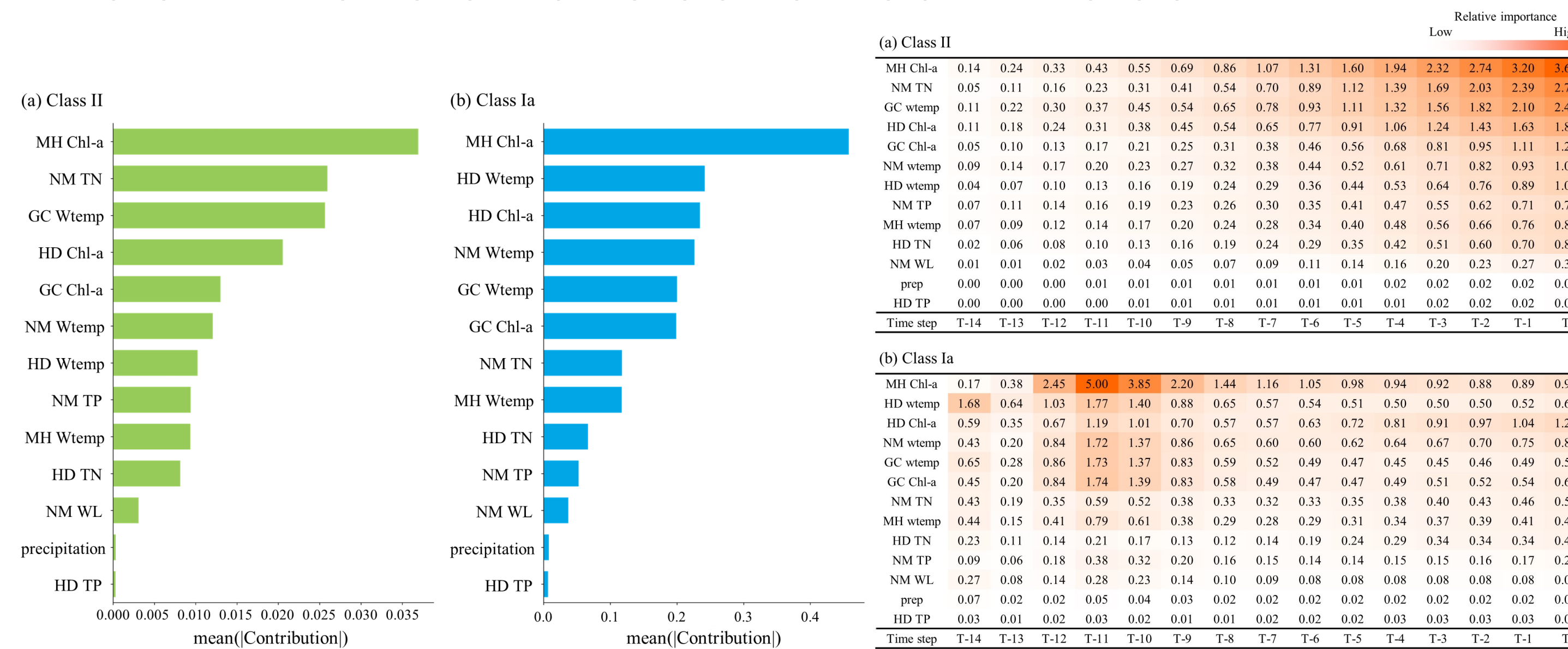
[Fig. 3. RETAIN-D를 활용한 클로로필 *a* 목표기준 예측 결과]

Model	RETAIN-D				GRU			
	Class II		Class Ia		Class II		Class Ia	
Criteria	Training	Test	Training	Test	Training	Test	Training	Test
Accuracy	0.85	0.90	0.98	0.84	0.86	0.91	0.90	0.83
AUC	0.85	0.91	0.97	0.69	0.85	0.90	0.83	0.70
F-measure	0.83	0.89	0.99	0.90	0.83	0.88	0.94	0.89

[Table 1. RETAIN-D 성능 평가 및 비교 결과]

Model interpretation

- RETAIN-D 모델링 결과로부터 종속변수 예측에 대한 입력변수의 변수별, 시점별 기여도를 확인함 (Fig. 4, 5)
- 입력 변수의 변수별 기여도는 두 수질기준 모두에서 미호강 지점의 클로로필 *a* 농도가 가장 높은 것으로 나타났으며, 현도 지점의 TP와 강수량은 기여도가 상대적으로 낮은 것으로 나타남
- 전반적으로 클로로필 *a*, 수온, TN, TP 등의 수질항목의 상대적 기여도가 강수량, 수위의 기여도에 비해 높은 경향이 있는 것으로 나타남
- 입력 변수의 시점별 기여도는 적용된 수질기준에 따라 다르게 나타났는데, 기준 II를 적용한 경우에는 최근 시점의 기여도가 상대적으로 높게 나타났고, 기준 Ia를 적용한 경우에는 T-11 시점과 최근 시점의 기여도가 상대적으로 높게 나타남



[Fig. 4. 변수의 평균 상대적 기여도 분석 결과]

[Fig. 5. 변수별 시점에 따른 상대적 기여도 분석 결과]

Conclusion

- 본 연구에서는 하천 부영양화의 중요한 지표 중의 하나인 클로로필 *a*의 기준 초과 예측에 대한 해석가능한 딥러닝 모델인 RETAIN-D의 적용 가능성을 평가함
- 수집된 고해상도 데이터를 활용한 모델링 결과 RETAIN-D는 양호한 예측 성능을 보였으며, 클로로필 *a*, 수온, TN, TP 등의 수질항목의 상대적 기여도가 높은 것으로 나타남
- RETAIN-D는 단일 구조 내에서 결측 보안, 예측, 변수별/시점별 기여도 해석이 모두 가능하기 때문에 환경 모델링 분야에 높은 활용성이 있을 것으로 기대됨

Reference

1. Kim, T., Shin, J., Lee, D., Kim, Y., Na, E., Park, J. H., ... & Cha, Y. (2022). Simultaneous feature engineering and interpretation: Forecasting harmful algal blooms using a deep learning approach. *Water Research*, 215, 118289.
2. Zhang, Y., Cheng, D., Song, J., Pang, R., & Zhang, H. (2023). How does anthropogenic activity influence the spatial distribution of dissolved organic matter in rivers of a typical basin located in the Loess Plateau, China. *Journal of Environmental Management*, 340, 117984.
3. Hu, R. Y. (2021). Sources and routes from terrestrial exogenous pollutants affect phytoplankton biomass in reservoir bays. *Water Supply*, 21(7), 3913-3931.
4. Meng, C., Liu, H., Li, Y., Shen, J., Li, X., & Wu, J. (2022). Effects of environmental and agronomic factors on pond water quality within an intensive agricultural landscape in subtropical southern China. *Agricultural Water Management*, 274, 107953.

Acknowledgments

이 논문은 2023년도 서울시립대학교 기초·보호학문 및 융복합 분야 R&D기반조성사업에 의하여 지원되었음.