

Research Paper

설명 가능한 그래프 신경망과 SHAP를 활용한 팔당호 수질 예측 및 강우 시 비점오염 우심지역 규명

장유경 · 최이송 · 오종민

경희대학교 환경학 및 환경공학과

Interpretable Water-Quality Prediction of Paldang Lake Using Graph WaveNet and SHAP for Identifying Priority Nonpoint Source Areas under Rainfall

Yukyeong Jang · I Song Choi · Jong-Min Oh

Department of Environmental Science and Engineering, Kyung Hee University

요약: 본 연구는 팔당호의 수질을 예측하고 그 근거를 해석하여 유역의 비점오염 우심지역을 규명하고자 하였다. 강우 시 유입되는 비점오염원은 발생원이 광범위하여 관리 자원의 배분 근거가 요구된다. 이에 남한강·북한강·경안천 수계 63개 지점을 노드로 구성하고, Graph WaveNet으로 팔당호 지점의 BOD, COD, T-N, T-P를 1~10일 후까지 예측하였다. 또한 SHAP로 항목별·지점별 기여도를 산출하고 강우 발생 여부에 따라 비교하였다. 1일 예측에서 전 항목 설명력(R^2) 0.89 이상(T-N 0.94)으로 LSTM·TCN·Transformer와 대등하거나 높았고, 10일 후에도 T-N은 0.37을 유지하였다. 조건부 SHAP 결과, 원거리인 이천 북한천 수계의 영양염류 기여도가 강우 시 뚜렷이 증가하였으며, T-P는 250회 학습 모두에서 강우 시 기여가 높아 해당 지역이 강우에 민감한 비점오염원임을 뒷받침하였다. 경안천 유역에서도 같은 경향이 나타났다. 이러한 결과는 유역 비점오염 저감 사업의 우선순위 설정 근거로 활용될 수 있다.

주요어: Graph WaveNet, XAI, 비점오염, 팔당호, 수질 예측

Abstract: This study forecasts water quality at Paldang Lake and interprets the basis of its predictions to identify priority nonpoint source (NPS) areas in the watershed. Because rainfall-driven NPS pollution originates from widely dispersed sources, evidence is needed to allocate limited management resources. To this end, 63 monitoring stations across the Namhan River, Bukhan River, and Gyeongang Stream were represented as graph nodes, and Graph WaveNet was used to forecast BOD, COD, T-N, and T-P at the Paldang Lake station from 1 to 10 days ahead. SHapley Additive exPlanations (SHAP) was then applied to quantify variable- and station-wise contributions, which were compared according to whether rainfall occurred. For 1-day-ahead forecasts, the explanatory power (R^2) was at least 0.89

First Author: Yukyeong Jang, Tel: +82-31-201-2125, E-mail: ukjang@khu.ac.kr, ORCID: 0009-0007-1557-2588

Corresponding Author: Jong-Min Oh, Tel: +82-31-201-2461, E-mail: jmoh@khu.ac.kr, ORCID: 0000-0002-1104-5867

Co-Author: I Song Choi, Tel: +82-31-201-2125, E-mail: isongchoi67@daum.net, ORCID: 0000-0001-8809-802X

Received: 20 August 2026. Revised: 1 September 2026. Accepted: 21 September 2026.

for all variables (0.94 for T-N), equal to or higher than the LSTM, TCN, and Transformer baselines, and T-N maintained an R^2 of 0.37 even at the 10-day horizon. Conditional SHAP analysis showed that nutrient contributions from the remote Bokha Stream system in Icheon increased markedly under rainfall; for T-P, rainfall contributions exceeded dry-condition contributions in all 250 training runs, indicating a rainfall-sensitive NPS source. The same tendency appeared in the Gyeongan Stream watershed. These results can serve as a basis for prioritizing watershed-scale NPS reduction projects.

Keywords: *Graph WaveNet, Explainable Artificial Intelligence, Nonpoint Source Pollution, Paldang Lake, Water Quality Forecasting*

I. 서론

팔당호는 한강 수계의 남한강과 북한강, 그리고 경안천이 합류하는 지점에 위치한 상수원 호소로서, 수도권 주민에게 용수를 공급하는 핵심 상수원 역할을 담당하고 있다. 이러한 중요성으로 인해 팔당호와 그 유역의 수질은 오랜 기간에 걸쳐 집중적으로 관리되어 왔으며, 하수처리장 방류수 관리와 배출허용기준 강화 등 점오염원에 대한 대책이 상당한 성과를 거두어 왔다. 그러나 점오염원 관리가 성숙 단계에 이른 현재에는, 강우 시 도시·농지·축산 지역 등 유역 전반에서 지표 유출과 함께 유입되는 비점오염원이 상대적으로 더 큰 비중을 차지하며 수질 변동의 주요 원인으로 부각되고 있다 (Park et al., 2021). 비점오염은 그 발생원이 특정 지점에 국한되지 않고 유역 전반에 광범위하게 분포하며 (Mainali & Chang, 2018; Mamun et al., 2022), 강우 사상의 간헐성과 불규칙성으로 인해 유출 부하와 오염 농도가 급격히 변동하는 특성을 지닌다 (Reza et al., 2016; Reza et al., 2019). 따라서 모든 지역을 동일한 강도로 관리하는 것은 비효율적이며, 한정된 예산과 인력을 효과적으로 배분하기 위해서는 강우 시 오염 부하의 기여가 집중적으로 나타나는 우심지역을 사전에 정량적으로 규명하고 이에 관리 역량을 우선 투입하는 접근이 필요하다 (Cho & Lee, 2017; Kim et al., 2022).

한편 수질 예측 분야에서는 최근 인공지능 기반의 딥러닝 기법이 활발하게 적용되어 왔다. 특히 순환신경망(RNN)과 장단기기억(LSTM) 계열의 모델은 수질 시계열의 비선형적·시간적 의존 구조를 효과적으로 학습하여 전통적인 통계 모델 대비 향상된 예측 성능을 보고해 왔으며 (Hochreiter & Schmidhuber, 1997; Jung

et al., 2018), 한강 수계를 비롯한 국내 하천·호소의 조류 및 수질 예측에서도 다수의 적용 사례가 보고되었다 (Ly et al., 2021; Jeong et al., 2022; Kim & Ahn, 2022). 그러나 이들 모델은 대체로 단일 지점의 시계열만을 독립적으로 다루기 때문에, 유역 내 상류와 하류 지점 사이에 존재하는 공간적 상호작용과 오염 신호의 전파 과정을 명시적으로 반영하기 어렵다는 한계가 있다. 실제 하천 수계는 상류 지점의 수질 변화가 하류로 이동·확산되는 방향성을 가진 네트워크 구조를 이루고 있으므로, 이러한 연결 구조를 직접적으로 학습에 반영할 수 있는 접근이 보다 적합하다. 이와 같은 배경에서 그래프 신경망(GNN)은 측정 지점을 노드로, 지점 간 연결 관계를 간선으로 표현함으로써 공간 구조를 모델에 직접 통합할 수 있는 대안으로 주목받고 있으며 (Kipf & Welling, 2016; Li et al., 2017; Yu et al., 2017), 최근 수문·수질 분야에서도 그 적용이 확대되고 있다 (Sun et al., 2021). 그중에서도 Graph WaveNet은 팽창 인과 시간 합성곱과 확산 그래프 합성곱을 결합하고, 사전에 정의된 인접 관계뿐 아니라 관측 자료로부터 지점 간 잠재적 연결 관계를 스스로 학습하는 자가적응 인접행렬 (self-adaptive adjacency matrix)을 도입한 대표적인 시공간 그래프 신경망이다. 특히 교통량 등 대규모 시공간 예측에서 우수한 성능이 검증되었으며, 방향성을 갖는 하천 네트워크와 다변량 수질 자료의 복잡한 시공간 의존성을 효과적으로 포착할 수 있어 본 연구에 적합하다 (Wu et al., 2019).

다만 딥러닝 모델은 높은 예측 설명력을 제공하는 반면, 내부 연산이 복잡하여 예측의 근거를 직관적으로 파악하기 어렵다는 이른바 블랙박스 문제가 지적되어 왔다. 이는 수질 관리와 같이 예측 결과가 실제 정책

과 예산 배분의 근거로 활용되어야 하는 분야에서 특히 중요한 제약으로 작용한다. 이러한 한계를 보완하기 위한 설명 가능 인공지능(XAI) 기법 중 SHAP(Shapley Additive exPlanations)는 협력 게임이론의 샐플리 값(Shapley value)에 이론적 기반을 두고 있어, 개별 예측에 대한 각 입력 변수의 기여도를 가법적(additive)이고 일관성 있게 분해할 수 있다는 장점이 있으며(Dwivedi et al., 2023; Ponce-Bobadilla et al., 2024), 최근 여러 분야에서 블랙박스 모델의 해석 도구로 널리 활용되고 있다(Yang et al., 2022; Liu et al., 2023; D. Ma et al., 2023; Song & Estes, 2023). SHAP을 시공간 그래프 모델에 적용하면 어떤 시점의 어떤 지점에서 어떤 수질 항목이 예측 결과에 기여하였는지를 정량적으로 추적할 수 있으므로, 단순한 예측을 넘어 오염원의 공간적 분포를 해석하는 도구로 활용할 수 있다.

이에 본 연구에서는 다음의 세 가지를 목적으로 하였다. 첫째, 팔당호 유역의 63개 수질측정 지점을 그래프로 구성하여 Graph WaveNet 기반의 팔당호 수질 예측 모델을 구축하고, LSTM·TCN·Transformer 등 대표적인 딥러닝 모델과의 비교를 통해 그 성능을 검증한다. 둘째, 학습된 모델에 SHAP 기법을 적용하여 예측에 대한 입력 항목별·지점별 기여도를 정량화한다. 셋째, 산출된 기여도를 대상 시점의 강우 발생 여부에 따라 강우 시와 비강우 시로 구분·비교함으로써, 강우에 민감하게 반응하여 기여도가 증가하는 비점오염 우심지역을 규명한다. 이를 통해 데이터 기반의 수질 예측과 유역 관리 의사결정을 정량적으로 연결하는 방법론을 제시하고자 한다.

II. 연구방법

1. 연구 대상 유역 및 자료

본 연구의 대상 지역은 팔당호와 이에 유입되는 남한강·북한강·경안천 수계이다(Mainali & Chang, 2018; Cheng et al., 2021). 팔당호로 합류하는 주요 본류와 지류를 따라 배치된 환경부 수질측정망의 지점을 대상으로, 팔당댐 지점을 예측 대상 노드로 설정하고 상류의 남한강 계열 33개 지점, 북한강 계열 17개 지점, 경안천 계열 12개 지점 등 62개 지점과 팔당댐 지점을 합하여

총 63개 지점을 하나의 그래프 노드 집합으로 구성하였다. 남한강 계열에는 여주 본류 구간과 함께 북하천, 양화천, 청미천 등 이천·여주·장호원 지역을 관류하는 주요 지류가 포함되며, 북한강 계열에는 의암호에서 가평에 이르는 본류와 지류가, 경안천 계열에는 광주 일대를 흐르는 경안천 본류와 지류가 포함된다. 이처럼 팔당호로 유입되는 상류 수계 전반을 하나의 그래프로 통합함으로써, 특정 지류에서 발생한 오염 신호가 본류를 따라 팔당호로 전파되는 과정을 모델이 학습할 수 있도록 하였다. 대상 지역의 공간적 범위와 63개 지점의 분포는 Figure 1에 제시하였다. Figure 1에서 파란 원은 62개 수질측정 지점을, 붉은색 별은 예측 대상인 팔당댐 지점을 나타내며, 화살표는 각 지점을 잇는 하천망과 팔당호를 향한 흐름 방향을 보여준다. 이를 통해 남한강·북한강·경안천 세 수계가 팔당호로 합류하는 방향성 있는 네트워크 구조를 확인할 수 있다. 수계별 노드 구성은 Table 1에 정리하였다.

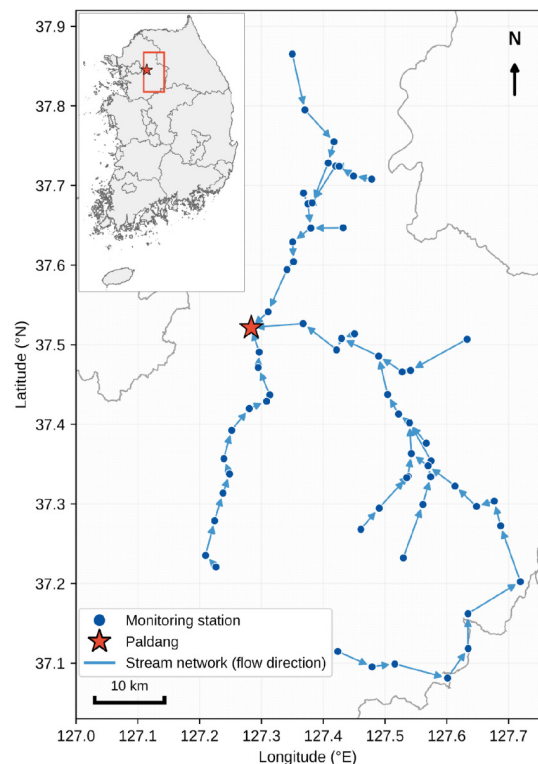


Figure 1. Study area and the 63 water-quality monitoring stations in the Paldang watershed (Namhan River, Bukhan River, Gyeongang Stream)

Table 1. Composition of the 63 graph nodes by sub-watershed and representative tributaries

Sub-watershed	No. of nodes	Representative tributaries / reaches
Namhan River (S1-S33)	33	Yeoju main stem, Bokha Stream, Yanghwa Stream, Cheongmi Stream (Icheon·Janghowon), Eung Stream, etc.
Bukhan River (N1-N17)	17	Uiam Lake-Gapyeong main stem and tributaries
Gyeongang Stream (G1-G12)	12	Gyeongang Stream main stem and tributaries (Gwangju)
Paldang Lake (target)	1	Paldang Lake inflow/intake reach (target)
Total	63	-

Water-quality variables (7): water temperature (Temp), dissolved oxygen (DO), BOD, COD, total nitrogen (T-N), total phosphorus (T-P), pH. Meteorological inputs (2): air temperature (Air temp), rainfall (Rainfall). Study period: January 2020 - January 2025.

Table 2. Descriptive statistics of the daily water-quality variables at the Paldang Lake (target) station (2020-2025, $n = 1,849$)

Variable	Unit	Mean	SD	Min	Median	Max	CV (%)
Water temp	°C	15.15	6.70	2.00	16.32	30.70	44.2
DO	mg/L	10.26	2.19	1.40	10.53	15.70	21.4
BOD	mg/L	1.23	0.46	0.40	1.19	4.30	37.5
COD	mg/L	3.67	0.54	2.50	3.60	6.70	14.6
T-N	mg/L	2.33	0.42	1.39	2.28	3.42	18.2
T-P	mg/L	0.032	0.021	0.008	0.027	0.196	65.7
pH	-	7.97	0.50	6.70	7.97	9.50	6.2

팔당호로 유입되는 지점들은 측정 주기와 자료 가용성이 서로 달라 일부 지점·기간에 결측이 존재하였다. 이를 그래프 신경망 학습에 필요한 일 단위의 정합된 시공간 자료로 구성하기 위해, 각 지점·항목별 시계열의 결측 구간은 시간 축 선형 보간(linear interpolation)으로 보간하여 63개 지점이 동일한 날짜 축을 공유하는 완전한 일 단위 패널(지점당 1,849일)을 구축하였다. 기상 자료 중 강수량은 결측을 무강수(0 mm)로 처리하고 값의 동적 범위가 큰 점을 고려하여 $\log(1+x)$ 변환을 적용한 뒤, 기온과 함께 수질 자료와 동일한 방식으로 정규화하여 날짜 기준으로 모든 지점에 결합하였다. 수질 자료는 위와 같이 완전 패널로 구성되었으므로 보간값(interpolated value)을 학습과 평가에 모두 사용하였고 별도의 결측 마스킹은 적용하지 않았으며, 다만 과거 10일의 입력 창과 예측 시점으로 시계열 표본을 구성할 때 기상 입력에 잔여 결측이 남아 있는 표본은 학습·평가에서 제외하였다. 예측 대상인 팔당댐 지점의 주요 수질 항목별 기초 통계량을 Table 2에 정리하였다.

각 지점·항목별 시계열은 결측과 이상치의 영향을

완화하기 위해 평균과 표준편차 대신 중앙값과 사분위 범위(IQR)를 이용한 강건 정규화를 적용하였다. 이는 강우 시 발생하는 극단적 오염 농도 값이 정규화 척도를 왜곡하는 것을 방지하기 위한 것으로, 임의의 관측 값 x 에 대한 정규화 값 $\tilde{x}$ 는 식 (1)과 같이 정의된다.

$$\tilde{x} = \frac{x - \text{median}(x)}{IQR(x)} \quad (1)$$

여기서 $\text{median}(x)$ 와 $IQR(x)$ 는 각각 해당 지점·항목 시계열의 중앙값과 사분위범위(제3사분위수와 제1사분위수의 차)를 의미한다. 사분위범위가 0인 지점·항목은 1×10^{-12} 로 대체하여 0으로 나누는 것을 방지하였다. 정규화 척도는 정보 누출을 방지하기 위해 학습 구간의 자료만을 사용하여 산정한 뒤 검증·시험 구간에 동일하게 적용하였다.

2. 예측 모델: Graph WaveNet

수질 예측 모델로는 시공간 그래프 신경망인 Graph WaveNet을 채택하였다(Wu et al., 2019). 이 모델은 크게 시간 축의 특성을 추출하는 시간 합성곱 계층과 공간 축의 상호작용을 반영하는 그래프 합성곱 계층으로

구성되며, 두 계층을 교대로 쌓아 시공간 의존성을 동시에 학습한다. 시간 축에서는 팽창 인과 합성곱을 사용하여 적은 수의 계층만으로도 넓은 시간적 수용 역할을 확보한다. 과거의 입력 시계열 X 에 대하여 팽창 계수 d 와 필터 f 를 갖는 팽창 인과 합성곱은 식 (2)와 같이 표현되며, 팽창 계수를 계층마다 지수적으로 증가 시킴으로써 장기 의존성을 효율적으로 포착한다. 여기서 연산자 $*$ 는 합성곱을, K 는 필터의 크기를 의미한다.

$$X *_d f(t) = \sum_{s=0}^{K-1} f(s)X(t-d \cdot s) \quad (2)$$

각 시간 합성곱 계층에서는 정보의 흐름을 선택적으로 조절하기 위해 게이트 활성화(gated activation) 구조를 적용한다. 이는 식 (3)과 같이 하이퍼볼릭 탄젠트(tanh) 활성화를 통과한 성분과 시그모이드(σ) 게이트를 통과한 성분의 원소별 곱($\odot$)으로 정의되며, tanh 성분이 후보 특성을, σ 성분이 해당 특성을 얼마나 통과시킬지를 조절하는 역할을 한다.

$$h = \tanh(W_f * X) \odot \sigma(W_g * X) \quad (3)$$

공간 축에서는 인접한 지점의 정보를 노드로 전파하기 위해 확산 그래프 합성곱을 사용한다(Li et al., 2017). 이는 하천의 흐름 방향을 반영하는 순방향 전이행렬 P_f 와 역방향 전이행렬 P_b , 그리고 학습을 통해 얻어지는 자가적응 인접행렬 $\hat{A}_{apt}$ 에 대한 K 차 확산 과정으로 정의되며 식 (4)와 같다. 여기서 X 는 입력 특성, W 는 학습 대상 가중치, k 는 확산 차수를 의미한다. 물리 인접행렬은 각 지점의 하류 연결(next node) 정보로부터 정의한 뒤 행 방향으로 정규화하여 순방향 전이행렬 P_f 를 구성하였으며, 그 전치행렬을 역방향 전이행렬 P_b 로 사용하였다.

$$Z = \sum_{k=0}^K (P_f^k X W_{k,1} + P_b^k X W_{k,2} + \hat{A}_{apt}^k X W_{k,3}) \quad (4)$$

Graph WaveNet의 핵심적인 특징은 사전에 정의된 하천 연결 관계에만 의존하지 않고, 자료로부터 지점 간의 잠재적 연결 관계를 스스로 학습한다는 점이다. 이처럼 인접 관계를 자료로부터 적응적으로 학습하려는 시도는 여러 시공간 그래프 모델에서 제안되어 왔다(Q. Ma et al., 2023; Liu et al., 2024; Pu et al., 2024). 이를 위해 각 노드에 학습 가능한 임베딩 벡터를 부여하고, 두 노드 임베딩 행렬 E_1, E_2 의 곱에 ReLU와 행 방향 SoftMax를 적용하여 식 (5)와 같이 자가적응 인접행렬을 구성한다. 이 행렬을 통해 관측 자료에 내재된, 물리적 하천 연결로는 설명되지 않는 지점 간 상관 구조까지 모델에 반영할 수 있다. 모델의 전체 구조는 Figure 2에 도시하였다. Figure 2와 같이 63개 지점의 수질·기상 입력은 선형 계층을 거쳐 k 개의 시공간 블록을 통과하며, 각 블록에서 게이트 시간 합성곱과 자가적응 인접행렬 기반의 그래프 합성곱이 잔차·건너뛰 연결로 결합된다. 누적된 건너뛰 출력은 두 개의 선형 계층을 거쳐 최종적으로 BOD, COD, T-N, T-P를 산출한다.

$$\hat{A}_{apt} = \text{SoftMax}(\text{ReLU}(E_1 E_2^T)) \quad (5)$$

입력은 과거 10일의 시간 창으로 구성하였으며, 예측 시계는 1일부터 10일 후까지 설정하여 예측 지점이 멀어짐에 따른 성능 변화를 함께 분석하였다. 예측 대상 항목은 유역 수질 관리에서 중요도가 높은 BOD, COD, T-N, T-P의 4개 항목으로 하였다. 모델의 상대적 성능을 평가하기 위한 대조군으로는 단일 지점 시계열 기반의 LSTM (Hochreiter & Schmidhuber, 1997), 시간 합성곱 기반의 TCN (Bai et al., 2018), 그리고 어텐션 기

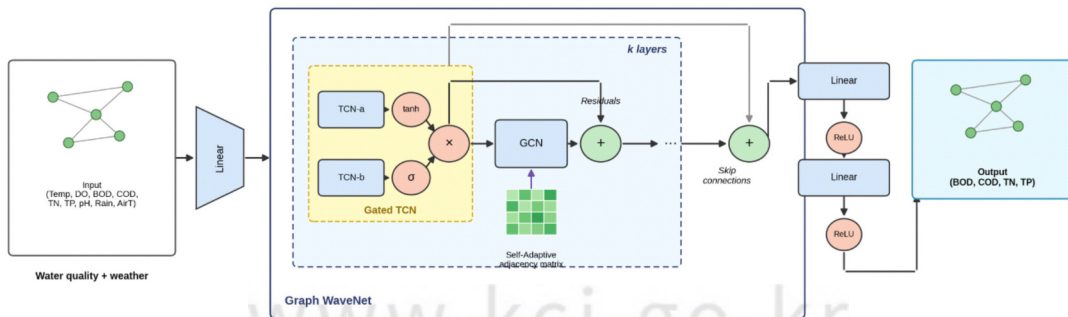


Figure 2. Architecture of the Graph WaveNet model

반의 Transformer (Vaswani et al., 2017) 모델을 동일한 자료와 검증 조건에서 함께 학습하였다.

3. 학습 및 검증

수질 시계열은 시간적 순서를 갖는 자료이므로, 무작위 분할을 사용할 경우 미래의 정보가 학습에 유입되는 시간적 정보 누출이 발생할 수 있다. 이를 완화하기 위해 무작위 분할 대신 시간 구조를 보존하는 시간 블록 교차검증(time-block cross-validation)을 수행하였다. 구체적으로 전체 시계열 표본을 시간 순서대로 10개(2×5)의 연속 블록으로 분할한 뒤, 각 겹 k 에 대하여 $2k$ 번째 블록을 검증(validation), $2k+1$ 번째 블록을 시험(test) 자료로, 나머지 구간 전체를 학습(training) 자료로 배정하였다. 이에 각 겹은 학습 80%·검증 10%·시험 10%로 구성되며, 시험 구간이 전체 기간에 고르게 분포하도록 하였다. 이 방식은 각 예측 표본이 예측 시점 이전의 관측만을 입력으로 사용한다는 점에서 표본 수준의 시간 인과성이 유지된다. 또한 학습의 무작위성을 고려하여 조건마다 5개의 난수 시드로 반복 학습하고, 검증 성능 기준 조기 종료를 적용하였다. 그 결과 4개 항목×10개 예측 시계×5개 겹×5개 시드로 총 1,000회를 학습·평가하였으며, 성능 지표는 조건별 25회의 평균과 표준편차로 제시하였다.

모델 학습에는 Adam 최적화기를 사용하였으며, 학습률 1×10^{-3} , 가중치 감쇠 1×10^{-4} , 배치 크기 32, 드롭아웃 0.3으로 설정하였다. 각 모델은 최대 100 epoch까지 학습하되 검증 성능이 10 epoch 동안 개선되지 않으면 학습을 조기 종료하였다. Graph WaveNet은 잔차 연결과 건너뛰기 연결의 채널 수를 각각 32와 64로 두고, 팽창 계수를 (1, 2, 1, 2)로 하는 시간 합성곱 블록을 쌓았으며, 확산 그래프 합성곱의 확산 차수는 $K=2$, 자가적응 인접행렬의 노드 임베딩 차원은 10으로 구성하였다. 대조 모델(LSTM, TCN, Transformer)은 동일한 입력 창과 교차검증 분할을 적용하여 학습하였으며, 다만 계산 비용을 고려하여 겹당 단일 시드로 수행하였다. 예측 성능은 설명력(R^2), 평균 제곱근오차(RMSE), 평균 절대오차(MAE)의 세 가지 지표로 평가하였으며, R^2 는 모델이 관측값의 분산을 설명하는 정도를 나타내어 1에 가까울수록 우수하고, RMSE와 MAE는 예측 오차

의 크기를 나타내며 0에 가까울수록 우수하다. 각 지표는 관측값 y_i , 예측값 $\hat{y}_i$, 관측값의 평균 $\bar{y}$, 표본 수 n 에 대하여 식 (6)~(8)과 같이 정의된다.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

4. SHAP 기반 기여도 분석 및 강우 조건 구분

학습된 Graph WaveNet의 예측 근거를 해석하기 위해 SHAP을 적용하였다(Ponce-Bobadilla et al., 2024). SHAP은 하나의 예측값 $f(x)$ 를 기준값 ϕ_0 와 각 입력 특성 i 의 기여도 ϕ_i 의 합으로 분해하는 가법적 설명 모델로, 식 (9)와 같이 표현된다. 여기서 M 은 입력 특성의 개수이며, 각 기여도 ϕ_i 는 게임이론의 샐플리 값으로서 식 (10)과 같이 해당 특성이 존재할 때와 존재하지 않을 때의 예측값 차이를 가능한 모든 특성 부분집합 S 에 대해 가중 평균하여 산정한다.

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i \quad (9)$$

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M-|S|-1)!}{M!} [f(S \cup \{i\}) - f(S)] \quad (10)$$

그러나 식 (10)의 엄밀한 계산은 특성 수가 많은 딥러닝 모델에서는 현실적으로 수행하기 어려우므로, 본 연구에서는 신경망의 기울기 정보를 이용하여 이를 효율적으로 근사하는 GradientExplainer를 사용하였다(Dwivedi et al., 2023). GradientExplainer는 기댓값 기울기 방식에 기반하여, 배경 표본 x' 로부터 실제 입력 x 까지의 경로를 따라 모델 출력의 입력에 대한 편미분을 적분한 값으로 각 특성의 기여도를 근사하며, 이는 식 (11)과 같다.

$$\phi_i \approx (x_i - x'_i) \cdot \int_0^1 \frac{\partial f(x' + \alpha(x - x'))}{\partial x_i} d\alpha \quad (11)$$

각 예측에 대하여 산출되는 SHAP 값은 입력 텐서와 동일하게 (시간 창×지점×항목)의 3차원 구조를 가진다. 본 연구에서는 이를 두 가지 방식으로 집계하였다.

Table 3. Test R^2 of Graph WaveNet by target variable and forecast horizon (mean $\pm$ standard deviation over 5 folds $\times$ 5 seeds)

Horizon (day)	BOD	COD	T-N	T-P
1	0.89 ± 0.02	0.90 ± 0.05	0.94 ± 0.02	0.90 ± 0.07
2	0.74 ± 0.04	0.78 ± 0.09	0.86 ± 0.05	0.77 ± 0.14
3	0.50 ± 0.12	0.59 ± 0.16	0.78 ± 0.06	0.56 ± 0.28
4	0.25 ± 0.24	0.38 ± 0.25	0.71 ± 0.07	0.39 ± 0.40
5	-0.04 ± 0.31	0.16 ± 0.43	0.64 ± 0.11	0.30 ± 0.30
6	-0.33 ± 0.36	0.07 ± 0.37	0.57 ± 0.13	-0.02 ± 0.47
7	-0.36 ± 0.30	0.01 ± 0.34	0.52 ± 0.14	0.09 ± 0.35
8	-0.47 ± 0.52	-0.01 ± 0.32	0.44 ± 0.16	-0.03 ± 0.37
9	-0.28 ± 0.33	0.01 ± 0.34	0.39 ± 0.14	-0.04 ± 0.34
10	-0.18 ± 0.21	0.02 ± 0.32	0.37 ± 0.20	-0.13 ± 0.43

먼저 시간과 지점 축으로 합산하여 각 입력 항목 c 의 전체 기여도 Φ_c 를 산출함으로써 식 (12)와 같이 어떤 수질·기상 항목이 예측에 기여하는지를 분석하였으며, 이후 식 (13)과 같이 시간과 항목 축으로 합산하여 각 지점 n 의 기여도 Φ_n 를 산출함으로써 어떤 지점이 예측에 기여하는지를 분석하였다.

$$\Phi_c = \sum_i \sum_n \varphi_{i,n,c} \quad (12)$$

$$\Phi_n = \sum_i \sum_c \varphi_{i,n,c} \quad (13)$$

비점오염의 특성상 오염 부하는 강우 사상에 의해 크게 좌우되므로, 각 예측 시점을 대상 시점의 강수 발생 여부에 따라 강우 시와 비강우 시로 구분하고 지점별 기여도를 조건부로 산출하였다. 강우 시는 대상 시점의 유역 평균 일 강수량이 5 mm를 초과하는 경우로, 비강우 시는 그 이하로 정의하였다. 5 mm는 국내 비점오염 관리에서 지표 유출을 유발하는 최소 강수량으로 통용되는 값을 고려한 것이다(Kim et al., 2022). 특정 지점 n 의 강우 민감도는 강우 시 평균 기여도를 비강우 시 평균 기여도로 나눈 강우/비강우 비(RDR, Rain/Dry ratio)로 정량화하였으며, 식 (14)와 같이 정의된다. 이 값이 1보다 클수록 해당 지점의 기여가 강우 시에 상대적으로 증가함을, 즉 비점오염원으로서의 성격이 강함을 의미한다. 본 연구에서는 이천 지역의 복하천 수계를 중심으로 RDR를 산출하고, 이천 지역 전체와 경안천 유역에 대해서도 함께 검토하였으며, 항목당 250회의 학습 결과 전반에서 그 경향이 얼마나 일관되게 나타나는지를 확인하였다.

$$RDR_n = \frac{mean_{rain}(\Phi_n)}{mean_{dry}(\Phi_n)} \quad (14)$$

III. 결과 및 고찰

1. 수질 예측 성능

Graph WaveNet 모델의 예측 시계별 성능을 4개 항목에 대하여 정리한 결과는 Table 3과 같다. 단기(1일 후) 예측에서는 모든 항목이 R^2 0.89 이상의 설명력을 보였으며, T-N은 0.94, T-P와 COD는 0.90, BOD는 0.89로 나타나, 63개 지점의 시공간 정보를 활용한 그래프 기반 예측이 팔당호 수질을 단기적으로 높은 설명력으로 재현함을 확인하였다. 예측 시계가 길어질수록 성능은 항목별로 뚜렷하게 다른 감소 양상을 보였다. T-N은 예측 시계가 늘어남에 따라 완만하게 감소하여 3일 후 0.78, 5일 후 0.64, 7일 후 0.52의 설명력을 유지하였고, 10일 후에도 R^2 0.37로 네 항목 중 유일하게 전체 예측 구간에서 양의 설명력을 유지하였다. 이는 T-N이 강우 사상에 급격히 반응하기보다 유역의 배경 부하와 계절적 변동에 의해 완만하게 변화하는, 상대적으로 자기상관이 강한 항목이기 때문으로 추정된다. 반면 COD는 3일 후 0.59, 4일 후 0.38, 5일 후 0.16으로 비교적 빠르게 감소한 뒤 6일 후부터는 0 부근에서 등락하여 8일 후에는 -0.01을 기록하였고, BOD와 T-P는 3일 후까지 실용적인 수준의 예측력을 보였으나 이후 BOD는 5일 후부터 0 이하로, T-P는 6일 후부터 0 부근으로 낮아졌으며 반복 실험 간 표준편차도 크게 증가하였다. 이러

Table 4. RMSE and MAE of Graph WaveNet by target variable at forecast horizons of 1, 3, 5, 7, and 10 days (mg/L; mean over 5 folds × 5 seeds)

Variable	Metric	H1	H3	H5	H7	H10
BOD	RMSE	0.145	0.292	0.418	0.477	0.454
	MAE	0.093	0.200	0.296	0.347	0.342
COD	RMSE	0.151	0.305	0.437	0.474	0.473
	MAE	0.105	0.220	0.335	0.365	0.370
T-N	RMSE	0.096	0.180	0.229	0.266	0.306
	MAE	0.071	0.135	0.174	0.206	0.235
T-P	RMSE	0.006	0.013	0.017	0.020	0.022
	MAE	0.004	0.008	0.011	0.013	0.014

한 결과는 유기물(BOD, COD)과 인(T-P)이 강우 유출에 의해 간헐적으로 급증하는 특성을 지녀 장기 예측이 본질적으로 어렵다는 점을 반영한다(Reza et al., 2019; Mamun et al., 2022).

R^2 0.5를 실용적 기준으로 보면 BOD·COD·T-P는 약 3일, T-N은 약 7일 이내의 예측에서 신뢰할 수 있는 설명력을 제공하는 것으로 사료되며, 동일한 조건에서 산출한 원 단위 오차(RMSE, MAE)는 Table 4에 제시

하였다. 1일 후 RMSE는 BOD 0.145 mg/L, COD 0.151 mg/L, T-N 0.096 mg/L, T-P 0.006 mg/L 수준이었으며, 10일 후에는 각각 0.454, 0.473, 0.306, 0.022 mg/L로 증가하여 설명력의 감소 경향과 일관되었다. 관측값과 예측값의 시계열 비교 및 잔차 분포는 Figure 3에 제시하였다. Figure 3의 (a)부터 (d)는 각각 BOD, COD, T-N, T-P에 대한 결과로, 항목별 상단에는 관측값(검은 실선)과 예측 시계열 예측값(색 점선)을, 하단에는 잔차

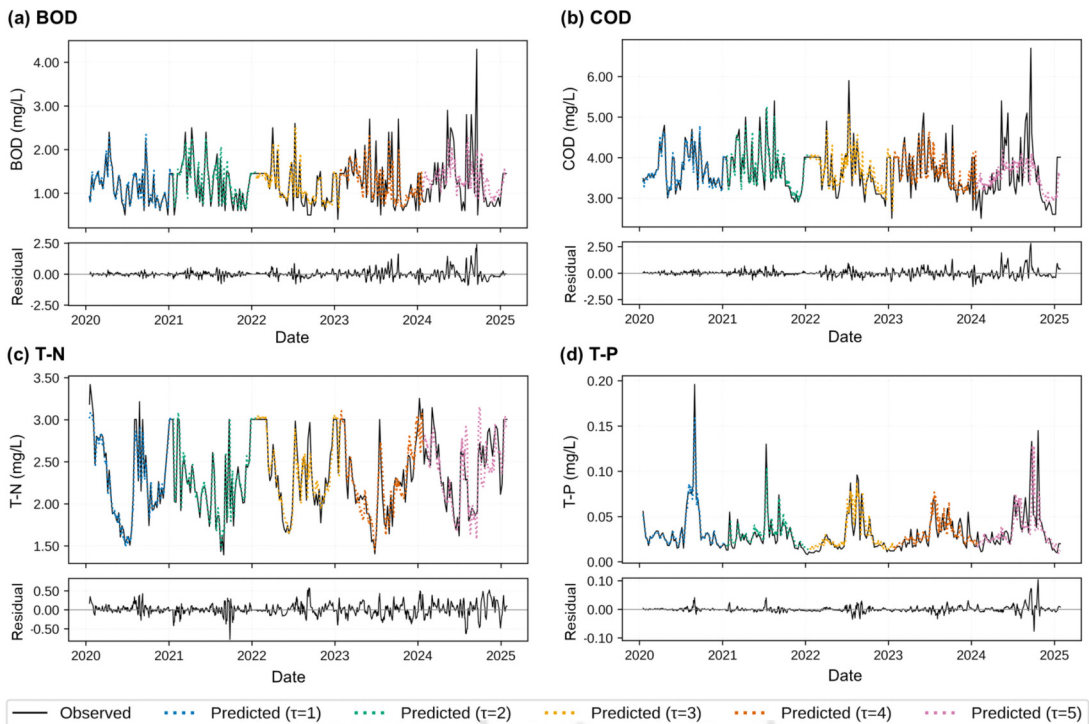


Figure 3. Observed and predicted daily time series (upper) with residuals (lower) at the Paldang Lake station for (a) BOD, (b) COD, (c) T-N, and (d) T-P

Table 5. Test R2 of Graph WaveNet and baseline models for forecast horizons of 1-5 days (mean $\pm$ standard deviation)

Variable	Model	H1	H2	H3	H4	H5
BOD	GWNet	0.89 $\pm$ 0.02	0.74 $\pm$ 0.04	0.50 $\pm$ 0.12	0.25 $\pm$ 0.24	-0.04 $\pm$ 0.31
	LSTM	0.87 $\pm$ 0.02	0.65 $\pm$ 0.04	0.43 $\pm$ 0.12	0.07 $\pm$ 0.13	-0.30 $\pm$ 0.40
	TCN	0.78 $\pm$ 0.04	0.64 $\pm$ 0.09	0.44 $\pm$ 0.08	0.19 $\pm$ 0.11	-0.09 $\pm$ 0.25
	Transformer	0.82 $\pm$ 0.04	0.63 $\pm$ 0.04	0.32 $\pm$ 0.13	0.07 $\pm$ 0.26	-0.13 $\pm$ 0.39
COD	GWNet	0.90 $\pm$ 0.05	0.78 $\pm$ 0.09	0.59 $\pm$ 0.16	0.38 $\pm$ 0.25	0.16 $\pm$ 0.43
	LSTM	0.90 $\pm$ 0.06	0.76 $\pm$ 0.14	0.57 $\pm$ 0.24	0.31 $\pm$ 0.33	0.12 $\pm$ 0.36
	TCN	0.83 $\pm$ 0.06	0.71 $\pm$ 0.11	0.53 $\pm$ 0.18	0.34 $\pm$ 0.27	0.17 $\pm$ 0.40
	Transformer	0.83 $\pm$ 0.07	0.66 $\pm$ 0.15	0.37 $\pm$ 0.29	0.22 $\pm$ 0.39	0.05 $\pm$ 0.50
T-N	GWNet	0.94 $\pm$ 0.02	0.86 $\pm$ 0.05	0.78 $\pm$ 0.06	0.71 $\pm$ 0.07	0.64 $\pm$ 0.11
	LSTM	0.94 $\pm$ 0.02	0.86 $\pm$ 0.05	0.74 $\pm$ 0.10	0.66 $\pm$ 0.06	0.54 $\pm$ 0.25
	TCN	0.84 $\pm$ 0.05	0.77 $\pm$ 0.06	0.72 $\pm$ 0.05	0.64 $\pm$ 0.09	0.62 $\pm$ 0.10
	Transformer	0.84 $\pm$ 0.10	0.80 $\pm$ 0.10	0.72 $\pm$ 0.10	0.64 $\pm$ 0.11	0.62 $\pm$ 0.10
T-P	GWNet	0.90 $\pm$ 0.07	0.77 $\pm$ 0.14	0.56 $\pm$ 0.28	0.39 $\pm$ 0.40	0.30 $\pm$ 0.30
	LSTM	0.86 $\pm$ 0.17	0.79 $\pm$ 0.09	0.53 $\pm$ 0.16	0.29 $\pm$ 0.22	0.21 $\pm$ 0.24
	TCN	0.78 $\pm$ 0.07	0.68 $\pm$ 0.11	0.55 $\pm$ 0.15	0.38 $\pm$ 0.25	0.28 $\pm$ 0.22
	Transformer	0.75 $\pm$ 0.22	0.62 $\pm$ 0.22	0.51 $\pm$ 0.20	0.30 $\pm$ 0.27	0.19 $\pm$ 0.31

(관측값-예측값)를 함께 나타내었다. T-N은 전 예측 시계에 걸쳐, T-P는 단기 예측 시계에서 강우기의 급변 구간까지 비교적 잘 재현된 반면, BOD와 COD는 여름철 고농도 침두에서 잔차가 크게 확대되어 앞서 제시한 정량적 결과와 부합하였다.

다음으로 Graph WaveNet의 상대적 성능을 확인하기 위해 동일한 자료와 교차검증 분할에서 LSTM, TCN, Transformer 모델과 비교하였으며, 그 결과를 Table 5에 정리하였다. 전반적으로 Graph WaveNet은 대부분의 항목과 예측 시계에서 대조 모델과 대등하거나 더 높은 R² 값을 나타냈다. 단기(1일 후) 예측에서 Graph WaveNet은 BOD 0.89, COD 0.90, T-N 0.94, T-P 0.90을 기록하였으며, 단일 지점 정보만을 사용하는 LSTM과 비교하면 T-P는 0.86에서 0.90으로, BOD는 0.87에서 0.89로 향상된 반면 COD와 T-N에서는 두 모델이 대등한 수준을 보였다. 이는 공간 정보의 활용에 따른 개선 효과가 항목별로 상이하며, 강우 유출의 영향을 크게 받는 항목에서 상대적으로 뚜렷하게 나타남을 시사한다. 다만 예측 시계가 5일 이상으로 길어지는 경우 유기물 항목에서는 모델 간 성능이 모두 낮아지면서 순위가 뒤바뀌는 구간도 관찰되었는데, 이는 장기 예측에서는 어느 모델이든 강우에 기인한 변동을 안정적으

로 재현하기 어렵다는 점을 시사한다. 종합하면 Graph WaveNet은 실무적으로 의미 있는 단기 및 중기 예측 구간에서 대조 모델 대비 대등하거나 우수한 성능을 제공하며, 공간 구조를 명시적으로 학습하는 접근의 타당성을 뒷받침한다. 그래프 기반 접근의 유용성은 최근 여러 하천 수질 예측 연구에서도 보고된 바 있으며 (Huan et al., 2023; Li et al., 2024; Liu et al., 2025; Wan et al., 2025; M. Yuan et al., 2025; Li, 2026), 딥러닝을 이용한 수질·영양염류 예측의 유효성 또한 다수 확인된 바 있다(Baek et al., 2020; He et al., 2022; Perumal et al., 2023).

2. 입력 항목별 기여도

학습된 모델의 예측 근거를 항목 수준에서 분석하기 위해, 식 (12)에 따라 SHAP 값을 시간과 지점 축으로 합산하여 각 입력 항목의 기여도를 산출하고 그 분포를 Figure 4에 제시하였다. Figure 4의 (a)-(d)는 각각 BOD, COD, T-N, T-P의 예측에 대한 결과이며, 각 점은 하나의 표본을, 가로 위치는 SHAP 기여도를, 색은 해당 입력값의 정규화된 크기를 나타낸다. 이때 각 예측 대상 항목에 대해서는 자기 자신의 항목 기여를 제외한 상태에서 나머지 입력 항목의 기여도 크기와 방향성을 나타내었다. 분석 결과, 각 수질 항목의 예측에는 해당 항

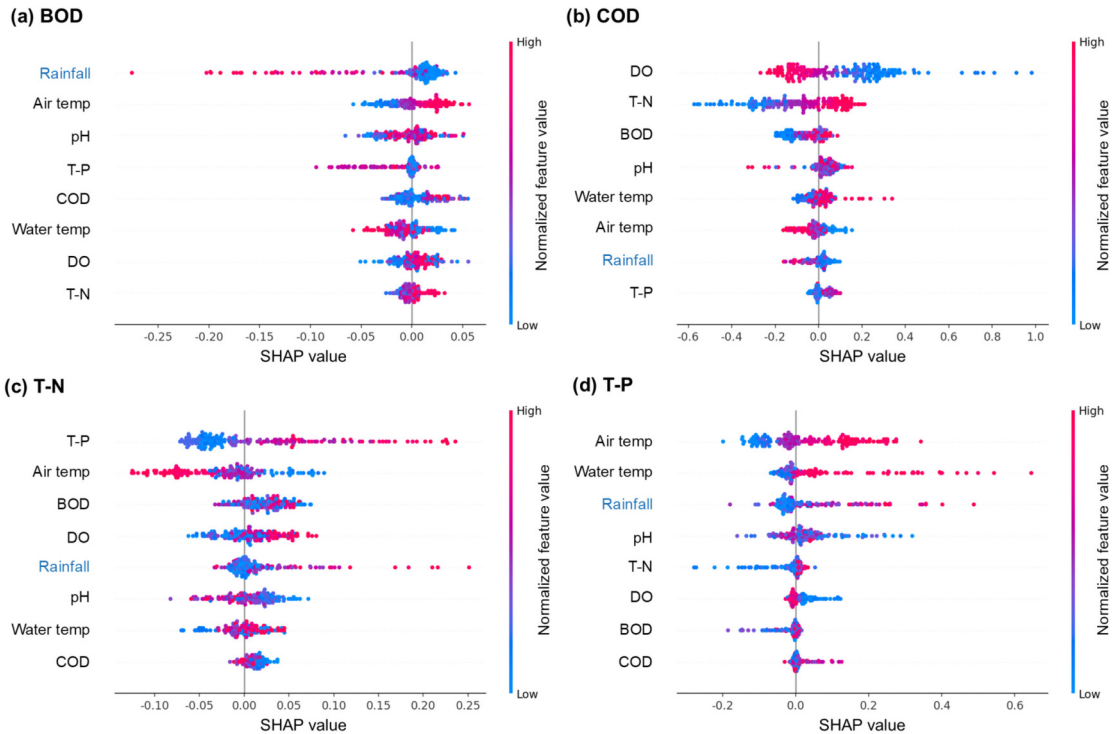


Figure 4. SHAP summary plots for (a) BOD, (b) COD, (c) T-N, and (d) T-P (each target's own variable excluded)

목 자체의 과거 수치뿐 아니라 물리적으로 연관된 다른 수질 항목과 기상 변수가 뚜렷한 크기로 기여하는 것으로 나타났다. 특히 강수량(Rainfall)은 여러 항목의 예측에서 큰 기여를 보였으며, 강수량이 큰 시점일수록 SHAP 값의 부호와 크기가 체계적으로 변화하는 경향이 관찰되었다. 이는 강우가 팔당호 수질 변동과 밀접하게 연관된 인자이며, 모델이 이러한 연관 구조를 자료로부터 포착하였음을 시사한다. 영양염류 항목(T-N, T-P)의 예측에서는 유기물 항목과 수온·용존산소 등이 함께 기여하여, 유역 내 오염물질 간의 동반 거동이 예측에 반영되고 있음이 확인되었다. SHAP를 이용한 이러한 입력 변수 기여도 분석은 최근 수질 예측 모델의 해석 가능성을 높이는 방법으로 활발히 활용되고 있다(Wang et al., 2022; Makumbura et al., 2024; Choudhary et al., 2025; Karahan et al., 2025; Y. Yuan et al., 2025; Alwateer, 2026).

3. 강우 시 비점오염 우심지역

지점 수준의 기여도를 대상 지점의 강우 발생 여부

에 따라 구분하여 비교한 결과, 팔당호와 지리적으로 상당한 거리가 있음에도 불구하고 이천 지역의 복하천 수계에서 영양염류 항목의 강우 시 기여도가 비강우 시 대비 뚜렷하게 증가하는 특징이 확인되었다. 복하천 수계와 이천 지역 전체, 그리고 비교를 위해 함께 검토한 경안천 유역의 항목별 강우/비강우 기여도 비(RDR)와, 항목당 250회 학습 중 강우 시 기여도가 비강우 시보다 높게 나타난 회차의 비율을 Table 6에 정리하였다. 영양염류인 T-P는 복하천에서 RDR 중앙값 1.42로 강우 시 기여도가 비강우 시의 약 1.4배 수준이었으며, 전체 학습 회차의 100%에서 강우 시 기여도가 비강우 시보다 높게 나타나 일관된 강우 민감성을 보였다. T-N 또한 RDR 1.15, 일치 비율 86%로 강우 민감성이 뚜렷하였다. 복하천을 포함한 이천 지역 전체를 대상으로 한 분석에서도 T-P는 RDR 1.39·일치 비율 99%, T-N은 RDR 1.12·일치 비율 84%로 동일한 경향이 확인되었다.

한편 팔당호에 직접 유입되는 경안천 유역에서도 T-P 1.57(일치 비율 100%), T-N 1.25(일치 비율 90%)로 같은 방향의 경향이 나타났으나, 경안천은 팔당호에

Table 6. Median Rain/Dry ratio (*RDR*) and, in parentheses, the percentage of runs with rain > dry contribution ($n = 250$ runs per variable)

Variable	Bokha Stream	Icheon area	Gyeongang Stream
BOD	0.95 (36%)	0.92 (30%)	1.04 (60%)
COD	1.01 (55%)	0.97 (43%)	1.09 (72%)
T-N	1.15 (86%)	1.12 (84%)	1.25 (90%)
T-P	1.42 (100%)	1.39 (99%)	1.57 (100%)

인접한 직유입 지류로서 기존 유역 관리에서 이미 주요 대상으로 다루어져 온 지역이다. 본 연구가 주목하는 대상은 팔당호로부터 원거리에 위치하여 상대적으로 주목받지 못하였음에도 이에 준하는 강우 민감성을 보인 복하천 수계이다. 이러한 결과는 강우 유출 시 이천 일대 농경지·시가지의 비점오염원에서 유출되는 영양염류 부하가 하류 팔당호의 예측에 실질적으로 반영되고 있음을 시사하며(Reza et al., 2016; Reza et al., 2019; Lee et al., 2025), 해당 지역이 강우에 민감한 비점오염 우심지역임을 정량적으로 뒷받침한다(Kim et al., 2022; Shim & Choi, 2024).

한편 유기물 항목인 BOD와 COD는 복하천에서 *RDR*이 각각 0.95, 1.01로 1에 근접하였고 강우 시 기여도가 비강우 시보다 높게 나타난 회차의 비율도 36~55% 수준에 머물러, 평균적으로는 뚜렷한 강우 강화 효과가 관찰되지 않았다. 경안천 유역에서도 BOD 1.04, COD 1.09로 1을 다소 상회하는 데 그쳐 영양염류 항목에 비해서는 약한 수준이었다. 이는 유기물의 경우 강우 유출뿐 아니라 하천 내 자정작용과 조류 생산 등 다양한 과정이 복합적으로 작용하기 때문으로 판단되며, 강우 시 비점오염의 신호가 상대적으로 명확하게 드러나는 항목은 영양염류, 특히 인(T-P)임을 시사한다. 이러한 항목별 차이는 비점오염 관리의 우선 대상 항목을 설정하는 데 실무적 시사점을 제공한다.

지점 단위의 세부 분석에서는 팔당호 인접 분류 합류부에 위치한 지점들의 절대적 기여도가 가장 크게 나타났다. 이는 하류에 가까울수록 상류의 누적 신호가 집약되는 위치적 특성에 기인한다. 그러나 팔당호와 거리가 먼 상류 지류 중에서도 이천 일대의 복하천 지점들은 영양염류 항목에서 강우 시 기여가 높은 회차

의 비율이 크게 나타났으며, T-P의 경우 S13-S17 각 지점에서 93~100%로 수계 전체 집계값(100%)과 일관된 수준을 보였다. 반면 BOD·COD 등 유기물 항목에서는 같은 지점들의 일차 비율이 50~60% 수준에 머물러 영양염류와 뚜렷한 대조를 이루었다. 항목별 강우 시 우심 지역의 공간적 분포는 Figure 5에 지도 형태로 제시하였다. Figure 5의 (a)~(d)는 각각 BOD, COD, T-N, T-P에 대한 결과로, 원의 크기는 강우 시 기여도의 크기를, 색은 강우/비강우 기여도 비(*RDR*)를 나타내며, 파란 타원은 본 연구에서 규명한 원거리 비점오염 우심지역(Priority NPS area), 즉 이천 일대의 복하천을 포함한 수계를 표시한 것이다. 팔당호에 직접 유입되는 경안천 유역은 표시하지 않았다.

종합하면, 본 연구에서 규명된 이천 지역의 복하천 수계는 강우 시 팔당호 수질, 특히 영양염류에 대한 기여가 크게 증가하는 비점오염 우심지역으로 판단된다. 복하천 수계는 팔당호로부터 원거리에 위치함에도 근거리 지류인 경안천에 준하는 강우 민감성을 보였다는 점에서 관리상 주목할 필요가 있다. 이러한 해석 결과는 이천 일대가 강우 시 비점오염 기여가 큰 지역으로 보고되어 온 기존 유역 조사와도 정성적으로 부합하며(Reza et al., 2016; Reza et al., 2019; Mamun et al., 2022), 설명 가능한 그래프 신경망의 해석이 실제 유역 특성을 반영할 수 있음을 시사한다. 설명 가능 인공지능을 유역 수질 관리 의사결정에 연결하려는 이러한 시도는 국내 수질 관리 연구에서도 점차 확대되고 있다(Jeong et al., 2026). 다만 유기물 항목의 경우 평균적으로는 강우 효과가 뚜렷하지 않은 경우도 관찰되었으므로, 항목별 특성과 하천 내 반응 과정을 함께 고려한 신중한 해석이 요구된다.

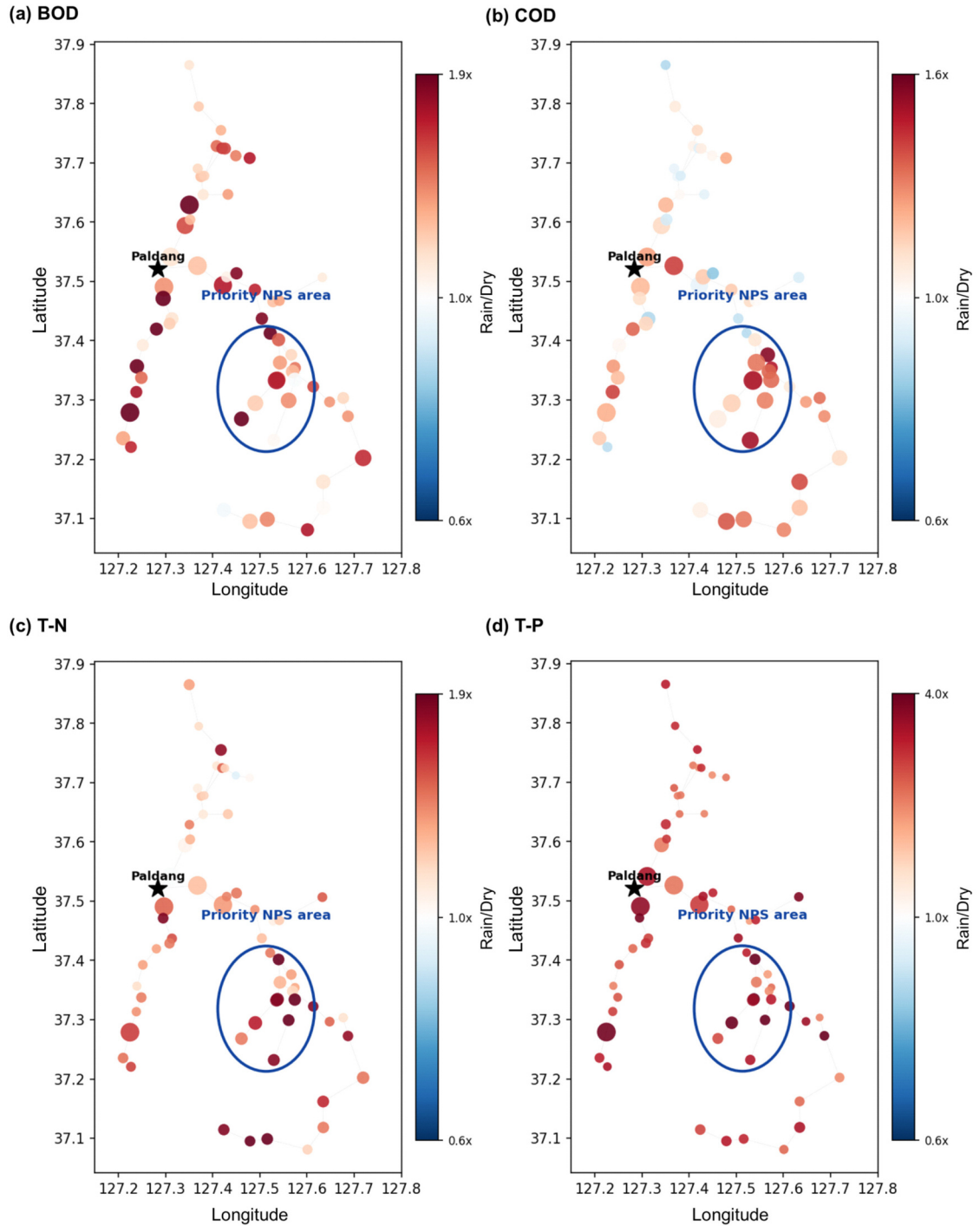


Figure 5. Spatial distribution of rainfall-sensitive contributions for (a) BOD, (b) COD, (c) T-N, and (d) T-P. Circle size: contribution under rainfall; color: Rain/Dry ratio. The blue ellipse marks the remote priority NPS area (Bokha Stream, Icheon area).

4. 연구의 한계 및 향후 과제

본 연구의 해석 결과를 활용함에 있어서는 몇 가지 한계를 함께 고려할 필요가 있다. 첫째, SHAP가 제공하는 기여도는 관측된 오염 부하 그 자체가 아니라 학습된 모델이 예측에 이르는 판단 근거를 설명하는 값이라는 점이다(Ponce-Bobadilla et al., 2024). 따라서 SHAP 기여도는 학습에 사용된 자료의 분포와 기간, 그리고 모델이 학습한 입력 변수 간 상관 구조에 종속되며, 서로 강하게 상관된 항목이나 지점 사이에서는 기여도가 분산되거나 특정 변수로 치우쳐 나타날 수 있다(Yang et al., 2022). 둘째, 본 연구에서 활용한 자가적응 인접행렬(self-adaptive adjacency matrix)은 물리적 하천 연결이 아니라 자료로부터 학습된 통계적 연결이므로, 이를 실제 오염물질의 이동 경로로 직접 해석하는 데에는 신중을 기해야 한다(Sun et al., 2021). 셋째, 자료가 일 단위이고 예측이 보간된 자료로 구성되어 강우 사상 내에서 수 시간 단위로 진행되는 초기 유출 등 단기 유출 동역학을 충분히 포착하지 못할 가능성이 있으며, 측정 지점의 공간적 불균형 또한 기여도 해석에 영향을 줄 수 있다. 마지막으로 강우 시와 비강우 시의 구분에 적용한 5 mm 기준은 국내 비점오염 관리 관행을 참고한 값으로, 임계값의 설정에 따라 조건부 기여도의 크기가 달라질 수 있다.

특히 본 연구에서 규명한 강우 시 기여도의 증가는 강우와 오염 신호 사이의 통계적 연관성을 의미하는 것으로, 이를 실제 비점오염 부하의 인과적 기여로 확정하기 위해서는 추가적인 정량 분석이 요구된다. 구체적으로는 대상 지역에 대한 강우 사상별 현장 모니터링을 통해 유량·중량·농도와 오염 부하량을 직접 산정하고(Reza et al., 2016; Reza et al., 2019), SWAT·HSPF 등 유역 기작 모델과의 연계성을 통해 데이터 기반 해석 결과를 물리적으로 교차 검증하는 접근이 필요할 수 있다. 아울러 추후 시간 단위 자동측정망(TMS) 등 고빈도 자료를 결합하면 강우 사상 내 단기 변동까지 반영한 보다 정밀한 기여도 분석이 가능할 것으로 기대되며, 다른 유역과 기간으로의 확장 적용을 통해 본 방법론의 일반화 가능성을 검증하는 것도 향후 과제로 남는다. 이러한 한계에도 불구하고, 본 연구는 예측 설명력과 공간적 해석을 동시에 제공함으로써 비점오염 우

심지역에 대한 우선순위 설정의 정량적 출발점을 제시한다는 점에서 의의가 있다.

IV. 결론

본 연구에서는 팔당호 유역의 63개 수질측정 지점을 그래프로 구성하고 시공간 그래프 신경망인 Graph WaveNet과 설명 가능 인공지능 기법인 SHAP를 결합함으로써, 팔당호 수질을 예측하는 동시에 그 예측의 근거를 공간적으로 해석할 수 있는 설명 가능한 수질 예측 체계를 제시하였다. 예측 성능 측면에서 Graph WaveNet은 단기 예측에서 전 항목 R^2 0.89 이상의 설명력을 보였으며, LSTM·TCN·Transformer 등 대조 모델과 대부분의 항목 및 예측 시계에서 같거나 높은 성능을 나타내어 공간 구조를 명시적으로 학습하는 접근의 타당성을 확인하였다. 항목별로는 T-N의 장기 예측 안정성이 가장 우수하여 10일 후에도 유의한 설명력을 유지한 반면, 강우 유출에 의해 단속적으로 변동하는 유기물과 인 항목은 예측 시계가 길어질수록 설명력이 빠르게 감소하였다. SHAP 기반의 항목별 기여도 분석에서는 강수량이 여러 수질 항목의 예측에 뚜렷하게 기여하여 강우가 팔당호 수질 변동의 핵심 구동 인자임을 확인하였다.

특히 SHAP 기여도를 강우 시와 비강우 시로 구분하여 비교한 결과, 팔당호와 거리가 먼 이천 지역의 복하천 수계에서 영양염류 항목의 강우 시 기여도가 비강우 시 대비 뚜렷하게 증가함을 확인하였다. T-P는 항목당 250회 학습 회차의 100%에서 복하천의 강우 시 기여도가 비강우 시보다 높게 나타나 일관된 강우 민감성을 보였으며, 이는 해당 지역이 강우에 민감한 원거리 비점오염원임을 정량적으로 뒷받침한다. 팔당호 직유입 지류인 경안천 유역에서도 같은 방향의 경향이 확인되었다. 따라서 팔당호의 수질 관리를 위해서는 이천 지역의 복하천 수계와 같이 강우 시 오염 기여가 집중되는 원거리 지류에 대한 비점오염 조사와 저감 대책의 우선 적용이 필요하다. 본 연구에서 제시한 설명 가능 그래프 신경망 기반 접근은 향후 유역 단위 비점오염 관리의 우선순위를 설정하고 한정된 자원을 효율적으로 배분하기 위한 정량적 의사결정 근거로 활용될

수 있을 것으로 기대된다.

사사

본 연구는 경기녹색환경지원센터 「2026년도 연구 개발사업(No. 26-1-30-33)」의 지원을 받아 수행되었습니다.

References

- Alwateer, M. (2026). Water Quality prediction using explainable machine learning models. *Sustainability*, 18(4), 1721. <https://doi.org/10.3390/su18041721>
- Baek, S.-S., Pyo, J., & Chun, J. A. (2020). Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. *Water*, 12(12), 3399. <https://doi.org/10.3390/w12123399>
- Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.
- Cheng, B., Zhang, Y., Xia, R., Wang, L., Zhang, N., & Zhang, X. (2021). Spatiotemporal analysis and prediction of water quality in the Han River by an integrated nonparametric diagnosis approach. *Journal of Cleaner Production*, 328, 129583. <https://doi.org/10.1016/j.jclepro.2021.129583>
- Cho, J., & Lee, J. (2017). Stormwater runoff characteristics and effective management of nonpoint source pollutants from a highland agricultural region in the Lake Soyang watershed. *Water*, 9(10), 784. <https://doi.org/10.3390/w9100784>
- Choudhary, R., Kumar, A., C., P., Naik, M. M., Choudhury, M., & Khan, N. A. (2025). Predicting water quality index using stacked ensemble regression and SHAP based explainable artificial intelligence. *Scientific Reports*, 15, 31139. <https://doi.org/10.1038/s41598-025-09463-4>
- Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., & Ranjan, R. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, 55(9), 1-33. <https://doi.org/10.1145/3561048>
- He, M., Wu, S., Huang, B., Kang, C., & Gui, F. (2022). Prediction of total nitrogen and phosphorus in surface water by deep learning methods based on multi-scale feature extraction. *Water*, 14(10), 1643. <https://doi.org/10.3390/w14101643>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Huan, J., Liao, W., Zheng, Y., Xu, X., Zhang, H., & Shi, B. (2023). A deep learning model with spatio-temporal graph convolutional networks for river water quality prediction. *Water Supply*, 23(7), 2940-2957. <https://doi.org/10.2166/ws.2023.164>
- Jeong, B., Chapeta, M. R., Kim, M., Kim, J., Shin, J., & Cha, Y. (2022). Machine learning-based prediction of harmful algal blooms in water supply reservoirs. *Water Quality Research Journal*, 57(4), 304-318. <https://doi.org/10.2166/wqrj.2022.019>
- Jeong, J. G., Ryu, Y. M., & Lee, E. H. (2026). Development of XAI-Based explainable planning management for Chl-a reduction. *Water*, 18(1), 7. <https://doi.org/10.3390/w18010007>
- 정성호, 조효섭, 김정엽, 이기하. (2018). 딥러닝 기반 LSTM 모델을 이용한 감조하천 수위 예측. 한국수자원학회논문집, 51(12), 1207-1216.
- Jung, S., Cho, H., Kim, J., & Lee, G. (2018). Prediction of water level in a tidal river using a deep-learning based LSTM model. *Journal of Korea Water Resources Association*, 51(12), 1207-1216. <https://doi.org/10.3741/jkwra.2018.51.12.1207>
- Karahan, S. M., Vandenbruwaene, W., Castano, J. A., & Verwaeren, J. (2025). Explainable AI for aquatic environmental intelligence: A SHAP-enhanced LSTM approach using high-frequency water quality

- data in a river system. *Journal of Hydroinformatics*, 27(12), 1918-1928. <https://doi.org/10.2166/hydro.2025.118>
- Kim, J., Choi, J., Park, M., Min, J.-H., Lee, J. M., Lee, J., Na, E. H., & Jang, H. (2022). A study on identifying priority management areas and implementing best management practice for effective management of nonpoint source pollution in a rural watershed, Korea. *Sustainability*, 14(21), 13999. <https://doi.org/10.3390/su142113999>
- Kim, K.-M., & Ahn, J.-H. (2022). Machine learning predictions of chlorophyll-a in the Han river basin, Korea. *Journal of Environmental Management*, 318, 115636. <https://doi.org/10.1016/j.jenvman.2022.115636>
- Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
- Lee, S., Kim, Y.-S., Bak, M., & Hong, E. (2025). Effects of straw mulching on nonpoint source pollutant runoff during snowmelt in Korean highland agricultural areas. *Water*, 17(18), 2675. <https://doi.org/10.3390/w17182675>
- Li, L. (2026). Spatiotemporal prediction of water quality and ecological risk assessment in a river basin using T-GCN based on river network topology constraints. *Scientific Reports*, 16, 4553. <https://doi.org/10.1038/s41598-025-34722-9>
- Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2017). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. arXiv preprint arXiv:1707.01926.
- Li, Z., Liu, H., Zhang, C., & Fu, G. (2024). Real-time water quality prediction in water distribution networks using graph neural networks with sparse monitoring data. *Water Research*, 250, 121018. <https://doi.org/10.1016/j.watres.2023.121018>
- Liu, J., Li, C., Ouyang, P., Liu, J., & Wu, C. (2023). Interpreting the prediction results of the tree-based gradient boosting models for financial distress prediction with an explainable machine learning approach. *Journal of Forecasting*, 42(5), 1112-1137. <https://doi.org/10.1002/for.2931>
- Liu, L., Wang, F., Liu, H., Zhu, S., & Wang, Y. (2024). HD-Net: A hybrid dynamic spatio-temporal network for traffic flow prediction. *IET Intelligent Transport Systems*, 18(4), 672-690. <https://doi.org/10.1049/itr2.12462>
- Liu, Y., Zheng, H., & Zhao, J. (2025). Enhanced water quality prediction by LSTM and graph attention network (L-GAT): An analytical study of the Pearl River Basin. *Water Research X*, 28, 100383. <https://doi.org/10.1016/j.wroa.2025.100383>
- Ly, Q. V., Nguyen, X. C., Lê, N. C., Truong, T.-D., Hoang, T.-H. T., Park, T. J., Maqbool, T., Pyo, J., Cho, K. H., Lee, K.-S., & Hur, J. (2021). Application of Machine Learning for eutrophication analysis and algal bloom prediction in an urban river: A 10-year study of the Han River, South Korea. *Science of The Total Environment*, 797, 149040. <https://doi.org/10.1016/j.scitotenv.2021.149040>
- Ma, D., Bortnik, J., Chu, X., Claudepierre, S. G., Ma, Q., & Kellerman, A. (2023). Opening the Black Box of the Radiation Belt Machine Learning Model. *Space Weather*, 21(4). <https://doi.org/10.1029/2022sw003339>
- Ma, Q., Sun, W., Gao, J., Ma, P., & Shi, M. (2023). Spatio-temporal adaptive graph convolutional networks for traffic flow forecasting. *IET Intelligent Transport Systems*, 17(4), 691-703. <https://doi.org/10.1049/itr2.12296>
- Mainali, J., & Chang, H. (2018). Landscape and anthropogenic factors affecting spatial patterns of water quality trends in a large river basin, South Korea. *Journal of Hydrology*, 564, 26-40. <https://doi.org/10.1016/j.jhydrol.2018.06.074>
- Makumbura, R. K., Mampitiya, L., Rathnayake, N., Meddage, D. P. P., Henna, S., Dang, T. L., Hoshino, Y., &

- Rathnayake, U. (2024). Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature. *Results in Engineering*, 23, 102831. <https://doi.org/10.1016/j.rineng.2024.102831>
- Mamun, M., Kim, J. Y., Kim, J.-E., & An, K.-G. (2022). Longitudinal chemical gradients and the functional responses of nutrients, organic matter, and other parameters to the land use pattern and monsoon intensity. *Water*, 14(2), 237. <https://doi.org/10.3390/w14020237>
- Park, M., Cho, Y., Shin, K., Shin, H., Kim, S., & Yu, S. (2021). Analysis of water quality characteristics in unit watersheds in the Hangang basin with respect to TMDL implementation. *Sustainability*, 13(18), 9999. <https://doi.org/10.3390/su13189999>
- Perumal, B., Rajarethinam, N., Velusamy, A. D., & Sundramurthy, V. P. (2023). Water quality prediction based on hybrid deep learning algorithm. *Advances in Civil Engineering*, 2023, 1-10. <https://doi.org/10.1155/2023/6644681>
- Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and Translational Science*, 17(11), e70056. <https://doi.org/10.1111/cts.70056>
- Pu, X., Zhu, J., Wu, Y., Leng, C., Bo, Z., & Wang, H. (2024). Dynamic adaptive spatio-temporal graph network for COVID-19 forecasting. *CAAI Transactions on Intelligence Technology*, 9(3), 769-786. <https://doi.org/10.1049/cit2.12238>
- Reza, A., Eum, J., Jung, S., Choi, Y., Jang, C., Kim, K., Owen, J. S., & Kim, B. (2019). Phosphorus Budget for a Forested-Agricultural Watershed in Korea. *Water*, 11(1), 4. <https://doi.org/10.3390/w11010004>
- Reza, A., Eum, J., Jung, S., Choi, Y., Owen, J. S., & Kim, B. (2016). Export of non-point source suspended sediment, nitrogen, and phosphorus from sloping highland agricultural fields in the East Asian monsoon region. *Environmental Monitoring and Assessment*, 188(12), 692. <https://doi.org/10.1007/s10661-016-5681-9>
- Shim, S. H., & Choi, J. H. (2024). Building an XGBoost model based on landscape metrics and meteorological data for nonpoint source pollution management in the Nakdong river watershed. *Ecological Indicators*, 165, 112156. <https://doi.org/10.1016/j.ecolind.2024.112156>
- Song, L., & Estes, L. (2023). itsdm: Isolation forest-based presence-only species distribution modelling and explanation in r. *Methods in Ecology and Evolution*, 14(3), 831-840. <https://doi.org/10.1111/2041-210x.14067>
- Sun, A. Y., Jiang, P., Mudunuru, M. K., & Chen, X. (2021). Explore spatio-temporal learning of large sample hydrology using graph neural networks. *Water Resources Research*, 57(12). <https://doi.org/10.1029/2021wr030394>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Wan, H., Xiang, L., Cai, Y., Xie, Y., & Xu, R. (2025). Temporal and spatial feature extraction using graph neural networks for multi-point water quality prediction in river network areas. *Water Research*, 281, 123561. <https://doi.org/10.1016/j.watres.2025.123561>
- Wang, S., Peng, H., & Liang, S. (2022). Prediction of estuarine water quality using interpretable machine learning approach. *Journal of Hydrology*, 605, 127320. <https://doi.org/10.1016/j.jhydrol.2021.127320>
- Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph WaveNet for deep spatial-temporal graph modeling. arXiv preprint arXiv:1906.00121.

- Yang, Y., Yuan, Y., Han, Z., & Liu, G. (2022). Interpretability analysis for thermal sensation machine learning models: An exploration based on the SHAP approach. *Indoor Air*, 32(2). <https://doi.org/10.1111/ina.12984>
- Yu, B., Yin, H., & Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875.
- Yuan, M., Li, Y., Zhang, L., Zhao, W., Zhang, X., & Li, J. (2025). Rapid prediction approach for water quality in plain river networks: A data-driven water quality prediction model based on graph neural networks. *Water*, 17(17), 2543. <https://doi.org/10.3390/w17172543>
- Yuan, Y., Zhou, C., Wu, J., Deng, F., Liu, W., Sun, M., & Li, L. (2025). An interpretable deep learning framework for river water quality prediction-A case study of the Poyang Lake Basin. *Water*, 17(16), 2496. <https://doi.org/10.3390/w17162496>