

GAN 데이터 기반의 머신러닝 모델을 통한 미계측 하천에서의 수생태계 건강성 등급 예측 방안 연구

A study on the prediction of aquatic ecosystem health grade in
ungauged rivers through the machine learning model based on GAN data

이서로*, 이지민**, 이관재***, 김종건****, 임경재*****

Seoro Lee, Jimin Lee, Gwanjae Lee, Jonggun Kim, Kyoung Jae Lim

요 지

최근 급격한 기후변화와 도시화 및 산업화로 인한 지류하천에서의 수량과 수질의 변동은 생물 다양성 감소와 수생태계 건강성 저하에 큰 영향을 미치고 있다. 효율적인 수생태 관리를 위해서는 지속적인 유량, 수질, 그리고 수생태 모니터링을 통한 데이터 축적과 더불어 면밀한 상관 분석을 통해 수생태계 건강성의 악화 원인을 규명해야 할 필요가 있다. 그러나 수많은 지류하천을 대상으로 한 지속적인 모니터링은 현실적으로 어려움이 있으며, 수생태계의 특성 상 단일 영향 인자만으로 수생태계의 건강성 변화와의 관계를 정확히 파악하는데 한계가 있다. 따라서 지류하천에서의 유량 및 수질의 시공간적인 변동성과 다양한 영향 인자를 고려하여 수생태계의 건강성을 효율적으로 예측할 수 있는 기술이 필요하다. 이에 본 연구에서는 경험적 데이터 기반의 머신러닝 모델 구축을 통해 미계측 하천에서의 수생태계 건강성 지수(BMI, TDI, FAI)의 등급(A to E)을 예측하고자 하였다. 머신러닝 모델은 학습 데이터셋의 양과 질에 따라 성능이 크게 달라질 수 있으며, 학습 데이터셋의 분포가 불균형적일 경우 과적합 또는 과소적합 문제가 발생할 수 있다. 이를 보완하고자 본 연구에서는 실제 측정망 데이터셋을 바탕으로 생성적 적대 신경망 GAN(Generative Adversarial Network) 알고리즘을 통해 머신러닝 모델 학습에 필요한 추가 데이터셋(유량, 수질, 기상, 수생태 등급)을 확보하였다. 머신러닝 모델의 성능은 5차 교차검증 과정을 통해 평가하였으며, GAN 데이터셋의 정확도는 실제 측정망 데이터셋의 정규분포와의 비교 분석을 통해 평가하였다. 최종적으로 SWAT(Soil and Water Assessment Tool) 모형을 통해 예측된 미계측 하천에서의 데이터셋을 머신러닝 모델의 검증 자료로 사용하여 수생태계 건강성 등급 예측 정확도를 평가하였다. 본 연구에서의 GAN에 의해 강화된 머신러닝 모델은 수질 및 수생태 관리가 필요한 우심 지류하천 선정과 구조적/비구조적 최적관리기법에 따른 수생태계 건강성 개선 효과를 평가하는데 활용될 수 있을 것이다. 또한 이를 통해 예측된 미계측 하천에서의 수생태계 건강성 등급 자료는 수량-수질-수생태를 유기적으로 연계한 통합 물관리 정책을 수립하는데 기초자료로 활용될 수 있을 것이라 사료된다.

핵심용어 : GAN, 수생태계 건강성, SWAT, 미계측 하천, 머신러닝

감사의 글

본 결과물은 환경부의 재원으로 한국환경산업기술원의 수생태계 건강성 확보 기술개발사업의 지원을 받아 연구되었습니다.(2020003030004)

* 정회원 · 강원대학교 농업생명과학대학 지역건설공학과 박사과정 · E-mail : seorolee91@gmail.com

** 정회원 · 강원대학교 농업생명과학대학 지역건설공학과 박사과정 · E-mail : jiminlee217@gmail.com

*** 정회원 · 강원대학교 농업생명과학대학 지역건설공학과 박사과정 · E-mail : lkj8151@kangwon.ac.kr

**** 정회원 · 강원대학교 농업생명과학대학 지역건설공학과 교수 · E-mail : kjmjg23@gmail.com

***** 정회원 · 강원대학교 농업생명과학대학 지역건설공학과 교수 · E-mail : kjlim@kangwon.ac.kr