**3.1 데이터**

본 연구에서는 스마트 부표 디바이스를 실제 운영 중인 기업의 스마트 부표를 통해 수집한 해양데이터를 바탕으로 바이오파울링 예측모델을 수립하였다. 본 연구는 부표에 탑재된 센서 중 바이오파울링에 민감한 광학센서인 용존산소 센서 데이터를 주로 활용하였다. 실제 데이터는 동일한 지역에 설치된 세 개의 부표(Wando01, Wando01b, Wando02)에서 수집된 자료를 사용하였으며, Wando01은 5월 23일부터 11월 1일까지, Wando01b는 10월 24일부터 11월 1일까지, Wando02는 8월 15일부터 11월 1일까지의 데이터를 수집하여 분석하였다.

원시 데이터에서 Fig. 1과 같이 시계열 예측 및 기기 구분을 위해 기기 ID, 측정 시간, 용존산소 수치만을 추출하여 전처리하였다.

Fig. 1. 기기별 용존산소 시계열 그래프

그래프, 텍스트, 도표, 라인이(가) 표시된 사진

자동 생성된 설명

해양 생물 다양성과 저산소상태에 관한 선행 연구[1]에 따르면, <Table 1>처럼 용존산소 농도가 3.0 mg/L 이하일 때 저산소 상태로 정의된다. 본 연구에서는 무산소 상태인 0 mg/L를 제외한 평균 용존산소 농도가 5.272 mg/L로 나타났다. 해양 환경에서 용존산소 농도가 단기간에 5.0 mg/L에서 3.0 mg/L로 급격히 감소하는 현상은 비정상적인 상태를 시사하며, 바이오파울링의 시작 기준으로 3.0 mg/L를 설정하였다. 비정상 데이터는 3.0 mg/L로 감소하는 시점의 719개 이전 데이터를 포함하며, 그 외의 데이터를 정상 데이터로 정의하였다.

**<Talbe1> 용존 산소에 따른 해양 상태**

|  |  |
| --- | --- |
| 용존 산소(mg/L) | 상태 |
| DO ≥ 5.0 mg/L | 정상 |
| 3.0 mg/L ≤ DO < 5.0 mg/L | 경계저산소 |
| 2.0 mg/L ≤ DO < 3.0 mg/L | 저산소 |
| DO < 2.0 mg/L | 심각한 저산소 |
| DO ≈ 0 mg/L | 무산소 |

원시 데이터는 1분 이내의 불규칙한 시간 간격으로 수집되었기 때문에, 시계열 데이터 예측에 적합하도록 <Table 1>과 같은 칼럼을 추출한 뒤 2분 간격으로 샘플링하고, 결측치는 보간법으로 처리하여 최종 데이터셋을 구성하였다. 이 과정에서 비정상 데이터는 72,682개, 정상 데이터는 116,570개로 총 189,152개의 데이터를 확보하였다.

**<Table 2> 원시 데이터 개수**

|  |  |  |
| --- | --- | --- |
| 상태 | 개수 | 샘플링 후 개수 |
| 전체 데이터 | 229,287 | - |
| 정상 데이터 | 161,395 | 116,570 |
| 비정상 데이터 | 67,892 | 72,582 |

본 연구는 해양데이터가 일반적으로 하루를 주기로 특정 패턴이 반복되는 것을 고려하여720개의 연속적인 데이터를 하나의 시퀀스로 정의하였으며, 슬라이딩 윈도우 기법을 사용하여 정상 데이터 시퀀스 159,238개와 비정상 데이터 시퀀스 66,454개를 구축하였다. 전체 데이터는 <Table 3>처럼 학습(train), 검증 (validation), 테스트(test) 용도로 70:15:15의 비율로 무작위 분할하여 구성하였다.

**<Table 3> 데이터셋 분할 개수**

|  |  |
| --- | --- |
| 데이터 구분 | 개수 |
| Train | 157,984 |
| Validation | 33,853 |
| Test | 33,855 |

**참고 문헌**

[1] Vaquer-Sunyer, R., & Duarte, C. M. (2008). "Thresholds of hypoxia for marine biodiversity." *Proceedings of the National Academy of Sciences*, 105(40), 15452-15457.1