**스마트부표 센서 데이터의 바이오파울링 예측을 위한 AI 개발: 용존산소 시계열 데이터 기반의 순환신경망 예측 모델**

임주혁[[1]](#footnote-1), 박영찬[[2]](#footnote-2), 김민석[[3]](#footnote-3), 김민수[[4]](#footnote-4), 정설영[[5]](#footnote-5), 이현상[[6]](#footnote-6)

**1. Introduction**

최근 이상 기후 현상으로 해양 환경이 급격하게 변화하면서[1][2], 해양 환경 관측 기기의 중요성이 더욱 부각되고 있다[3]. 실시간으로 수집되는 해양 데이터는 생태계 보호와 해양 자원의 효율적 관리에 가장 기본적이고 핵심적인 요소이다[3]. 그러나 부표에 설치한 센서와 같이 해상에 직접 배치된 관측 기기는 해양에서 발생하는 외부 요인에 의해 오작동하기도 한다[4]. 특히 지속적, 반복적으로 발생하는 바이오파울링(Biofouling)에 의해 센서의 성능이 저하되고 수집 중인 데이터의 정확도가 크게 떨어질 수 있다[5].

바이오파울링이란 따개비나 해조류 등 해양 생물이 선박이나 부표 등 구조물에 부착하여 본래의 기능 수행을 어렵게 만드는 현상을 말한다[6]. 현재는 주기적인 현장 점검 및 부착된 유기물 제거를 통해 바이오파울링 발생에 대처하고 있다[5]. 그러나 최근에는 해양 환경의 급격한 변화로 인해 바이오파울링 발생 주기의 예측이 점점 어려워지고 있다[7]. 이러한 상황에서, 해양 센서에 발생하는 바이오파울링에 대해서는 측정 데이터의 이상 징후를 통해 예측해보는 방식을 고려해볼 필요가 있다.

기존 해양 환경 관측 기기의 데이터 품질 관리 연구는 주로 통계적인 접근을 통해 이루어졌고, 수작업 기반의 작업을 자동화하는 것에 집중하였다[8]. 최근에는 인공지능 기술을 활용하여 해양 관측 데이터에서 이상 데이터를 탐지하는 기법에 대한 연구가 진행되었다[9]. 그러나 일반적인 데이터 이상 징후 탐지를 넘어서, 구체적으로 센서 바이오파울링을 예측하는 접근에 대한 연구는 아직 제한적이다.

본 연구에서는 부표에 설치된 해양 센서가 측정한 데이터에서 시계열 패턴을 학습하고 바이오파울링 발생을 예측하는 AI 모델을 개발하였다. 데이터의 흐름을 추적하면서 바이오파울링을 실시간으로 예측할 수 있다면, 보다 합리적이고 체계적인 센서 바이오파울링 대응이 가능해질 것이다. 또한 수집된 해양 데이터의 신뢰성을 보장하고, 데이터 관리의 자동화를 지원하여 효율적인 시스템 관리 및 운영에 기여할 수 있을 것으로 기대한다.

**2. Related Work**

**2.1 기존 해양 환경 이상 탐지 연구**

과거 연구에서는, 해양 측정 기기의 데이터 품질을 전문가가 주로 수작업으로 처리하고, 이상 탐지를 전문가의 경험에 기반한 방법을 통해 이루어졌다. 이로 인해 자동화된 이상 탐지의 필요성이 대두된 배경이 포함되어 있다. 특히, 기존 수작업 기반의 한계를 지적하고, 과거 연구에서는 자동 품질 관리 시스템(QC)을 통해 데이터 품질을 높이려는 방향이 필요하다는 점을 강조한다[10].

최근에는 Argo profile float라는 해양 관측 장치를 통해 해양의 여러 정보를 수집하고 있다. Argo profile float는 일정 기간 동안 수심이 1000m에서 2000m 사이의 해양 깊이를 수직으로 이동하면서, 일반적으로 10일 동안 해양의 온도, 염도, 깊이에 대한 정보를 총체적으로 수집한다. 하지만 이 관측 장치도 갑작스러운 수괴층을 만나거나 해양 오염에 노출될 경우 데이터가 왜곡될 가능성이 있다. 이 문제를 해결하기 위해 trajectory clustering 방법을 활용한 이상 탐지 연구가 진행되고 있다. 기존의 LOF[11]와 DBSCAN[12] 방식보다 우수한 성능을 보여, 해양 데이터의 신뢰성을 높이는 데 중요한 기여를 하고 있다고 보고되었다[13].

**2.2 ai 기반 센서 이상 탐지**

과거 연구에서는 센서 데이터의 이상 탐지는 전통적으로 통계적 방법이나 임곗값 기반 접근법을 통해 이루어졌다. 평균, 분산, 상관계수 등의 통계 지표를 활용하여 데이터의 이상 여부를 판단하고, 사전에 정의된 임곗값을 초과하는 경우 이상으로 간주하는 방식이다. 이러한 방법들은 구현이 간단하고 이해하기 쉽지만, 복잡한 패턴이나 다차원 데이터에서의 이상 탐지에는 한계가 있다[14].

최근에는 인공지능의 발전으로 비지도 학습과 딥러닝을 활용한 이상 탐지 기법이 등장하였다. 오토인코더(Autoencoder)를 이용하여 정상 데이터의 패턴을 학습하고, 재구성 오차를 기반으로 이상치를 감지하는 방법이 대표적이다. 이는 사전에 이상 데이터를 레이블링하지 않아도 되어 실용적이다[15]. 또한, LSTM과 같은 시계열 모델을 이용한 이상 탐지와, PyOD를 활용한 다양한 이상 탐지 기법에 대해 다루었다. 예를 들어, LSTM을 통해 시간에 따른 데이터 변화를 효과적으로 모델링하고[16], PyOD 툴 박스를 이용해 다양한 이상 탐지 모델을 테스트하여 센서 데이터의 이상 탐지를 수행하는 연구가 진행되었다[17].

위의 연구들을 통해 센서 데이터 이상 탐지 방법이 전통적인 통계적 수치를 기반으로 전문가의 경험에 의존하여 판단하던 방식에서 AI 기반의 지능형 시스템으로 발전했음을 확인할 수 있다. 본 연구에서는 이러한 최신 기법들을 적용하여 해양 환경에서 바이오파울링 발생 여부를 보다 정확하게 예측하고자 한다.

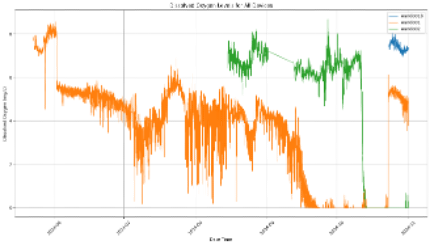
**3. Methods**

**3.1 Data**

본 연구에서는 (주)오든(ODN)기에서 제조 및 설치한 스마트 부표를 통해 수집한 해양 데이터를 바탕으로 바이오파울링 예측모델을 수립하였다. 본 연구는 부표에 탑재된 센서 중 바이오파울링에 민감한 광학센서인 용존산소 센서 데이터를 사용하였다. 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



해양 생물 다양성과 저산소상태에 관한 선행 연구[18]에 따르면, <Table 1>과 같이 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.

**<Table 1> 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 하루치 데이터를 예측하기 위해 720개의 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 |

**3.2 Modeling**

본 연구에서는 해양 환경에서의 바이오파울링 발생 여부를 예측하기 위해 순환신경망 계열의 모델인 RNN, GRU, LSTM을 활용하였다. 각 모델은 시계열 데이터 처리에 있어 고유한 특성을 지니고 있다. 기본적인 RNN은 순차적 데이터를 처리하는 데 효과적이지만, 긴 시퀀스에서 기울기 소실 문제가 발생할 수 있다는 한계를 가진다[19]. GRU는 Update Gate와 Reset Gate를 통해 이러한 문제를 개선하였으며, 비교적 단순한 구조로 효율적인 학습이 가능하다[20]. LSTM은 Input Gate, Forget Gate, Output Gate와 Memory Cell를 활용하여 장기 의존성 문제를 효과적으로 해결할 수 있으며, 복잡한 시계열 데이터에서 정보의 저장과 전달을 선택적으로 제어함으로써 안정적인 학습이 가능하다[20].

바이오파울링 예측 문제는 시간에 따른 용존산소량의 변화 패턴을 분석해야 하는 특성을 가진다. 특히, 센서 데이터의 미세한 변동성과 장기적인 추세를 모두 고려해야 하므로, 단순한 통계적 방법이나 전통적인 시계열 분석 기법으로는 한계가 있다. 이에 순환신경망 계열의 모델들을 적용함으로써, 시간적 의존성을 고려한 효과적인 패턴 학습이 가능할 것으로 기대하였다. 또한, 과적합을 방지하고 안정적인 학습을 보장하기 위해 학습률 조정(learning rate scheduling), 조기 종료(early stopping) 기법을 적용하였다. 특히, 검증 손실이 개선되지 않을 경우 학습률을 자동으로 조정하는 `ReduceLROnPlateau` 스케줄러를 사용하여 학습의 안정성을 확보하였다[21].

모델의 학습 과정에서는 Cross-Entropy Loss를 손실 함수로 사용하여 이진 분류 문제에 적합한 학습이 이루어지도록 하였다. Adam 옵티마이저를 통해 효율적인 파라미터 업데이트를 수행하였으며[22], 초기 배치 크기는 32로 설정하였다. 이는 선행 연구[23]에서 배치 크기가 32 이하일 때 일반화 성능이 가장 우수하다는 실험 결과를 반영한 것이다.

**3.2.1 Metrics**

본 연구에서는 바이오파울링 예측 모델의 성능을 평가하기 위해 이진 분류에서 일반적으로 사용되는 4가지 평가 지표를 활용하였다. 정확도(Accuracy)는 전체 예측 중 정확히 분류된 비율을 나타내는 지표로, 전반적인 모델의 성능을 평가하는 데 사용된다. 그러나 클래스 불균형이 있는 경우 이 지표만으로는 모델의 성능을 정확히 평가하기 어렵다[24]. 정밀도(Precision)는 모델이 비정상(바이오파울링 발생)이라고 예측한 케이스 중 실제로 비정상인 비율을 나타낸다. 이는 오탐지(false positive)를 최소화하는 것이 중요한 상황에서 유용한 지표이다. 재현율(Recall)은 실제 비정상 케이스 중 모델이 정확히 비정상이라고 예측한 비율을 의미한다. F1 Score는 Precision와 Recall의 조화평균으로, 두 지표 간의 균형을 평가하는 데 사용된다. 이는 정밀도와 재현율 중 어느 한쪽으로 치우치지 않고 두 성능이 모두 우수한 모델을 선별하는 데 효과적이다.

**3.2.2 Experimental Results**

본 연구에서는 RNN, GRU, LSTM 모델의 성능을 두 가지 관점에서 평가하였다. 첫째로 동일한 하이퍼파라미터 환경에서의 기본 성능을 비교하였고, 둘째로 각 모델별 최적화된 파라미터에서의 성능을 비교하였다. 이러한 이원화된 평가 방식을 채택한 이유는 다음과 같다. 동일 환경에서의 성능 비교는 각 모델의 구조적 특성이 성능에 미치는 순수한 영향을 분석할 수 있게 하며, 최적화된 환경에서의 비교는 각 모델이 실제 현장에서 달성할 수 있는 최대 성능을 평가할 수 있게 한다. 순환신경망의 경우, 모델 구조의 차이가 성능에 미치는 영향과 하이퍼파라미터 최적화를 통한 성능 개선 가능성을 구분하여 분석하는 것이 모델 선택에 있어 중요한 기준이 될 수 있다.

먼저, 동일한 하이퍼파라미터 환경에서의 성능 비교 결과는 <Table 4>와 같다. 각 실험은 무작위 초기화 및 데이터 셔플에 따른 변동성을 고려하기 위해 10회 반복 수행되었으며, 결과는 평균값으로나타내었다. 기본 하이퍼파라미터는 hidden size 32, learning rate 0.001, batch size 64로 설정하였다.

**<Table 4> 동일 환경에서의 모델별 성능 비교**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1 Score |
| RNN | 0.9441 | 0.9343 | 0.9340 | 0.9333 |
| GRU | 0.9838 | 0.9878 | 0.9732 | 0.9801 |
| LSTM | 0.9619 | 0.9610 | 0.9474 | 0.9535 |

동일 환경에서의 실험 결과, GRU가 모든 평가 지표에서 가장 우수한 성능을 보였다. 이는 GRU의 Update Gate와 Reset Gate 구조가 바이오파울링으로 인한 용존 산소량의 시계열 패턴을 가장 효과적으로 포착할 수 있음을 시사한다.

다음으로, 각 모델의 최적 성능을 도출하기 위해 Optuna 라이브러리를 활용한 하이퍼파라미터 최적화를 수행하였다[25]. 참고 문헌[21]에 따르면 hidden size, learning rate, batch size가 순환신경망 모델의 성능에 큰 영향을 미치는 것으로 알려져 있다. 각 모델별로 도출된 최적의 하이퍼파라미터는 <Table 5>과 같다.

**<Table 5> 모델별 최적 하이퍼파라미터**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | hidden\_size | learning\_rate | batch\_size |
| RNN | 32 | 0.001 | 16 |
| GRU | 32 | 0.006 | 64 |
| LSTM | 64 | 0.001 | 32 |

최적화된 환경에서 각 모델의 성능을 비교한 결과는 <Table 6>과 같다. 이 또한 각 실험을 10회 반복하여 평균값으로 성능을 제시하였다.

**<Table 6> 최적화된 환경에서의 모델별 성능 비교**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1 Score |
| RNN | 0.95 | 0.94 | 0.93 | 0.94 |
| GRU | 0.99 | 0.99 | 0.98 | 0.99 |
| LSTM | 0.98 | 0.98 | 0.96 | 0.97 |

하이퍼파라미터 최적화 결과, 모든 모델에서 성능이 향상되었으며 특히 GRU는 F1 점수 0.99를 달성하며 가장 우수한 성능을 보였다. 주목할 만한 점은 GRU가 비교적 작은 은닉층 크기(32)에서도 최고의 성능을 달성했다는 것이다. 또한 GRU는 정상과 비정상 상태 모두에서 매우 균형 잡힌 성능을 보여주었다. 이는 GRU가 계산 효율성과 예측 정확도 측면에서 바이오파울링 예측에 가장 적합한 모델임을 입증한다.

**4. Conclusion**

본 연구에서는 해양 센서의 바이오파울링을 예측하기 위한 AI 기반 접근 방식을 제안하였다. 완도 지역의 3개 부표에서 수집한 실제 데이터를 활용하여 RNN, GRU, LSTM 모델의 성능을 비교 분석하였으며, 특히 GRU 모델이 F1 Score 0.99를 달성하며 가장 우수한 성능을 보였다. GRU 모델은 작은 hidden size(32)로도 높은 성능을 달성했다는 점에서 계산 효율성과 예측 정확도를 동시에 확보했다는 의의가 있다.

추가적으로, 시퀀스 길이(180, 360, 720, 1440, 2880 타임스텝)에 따른 성능 실험을 수행한 결과, 더 긴 시퀀스 길이에서 더 높은 예측 정확도를 보였으며, 특히 2880   
(이틀)에서 F1 Score 0.993으로 가장 우수한 성능을 달성하였다. 그러나 실제 시스템 구현에서는 하루 전 예측이라는 실용적 목적을 고려하여 720 시퀀스를 사용하였다. 이는 향후 예측 기간 확장이 필요한 경우 시퀀스 길이 조정을 통한 성능 개선이 가능함을 시사한다.

그러나 본 연구는 몇 가지 중요한 한계점을 가지고 있다. 가장 핵심적인 한계점은 바이오파울링의 실제 발생 시점에 대한 정확한 레이블링 데이터가 부족한 상태에서 연구가 진행되었다는 점이다. 본 연구에서는 용존산소량이 3.0 mg/L 이하로 떨어지는 시점을 바이오파울링 발생의 기준으로 설정하였으나, 이는 간접적인 지표에 불과하며 실제 바이오파울링 발생과의 상관관계에 대한 실증적 검증이 부족하다. 또한, 단일 지역(완도)의 데이터만을 활용하여 지역적 일반화 가능성에 대한 검증이 부족하며, 5월부터 11월까지의 데이터만을 사용하여 계절적 영향을 포괄하지 못했다. 수온, 염도 등 다양한 해양 환경 변수의 통합적 고려가 이루어지지 않았다.

이러한 한계점을 보완하기 위한 향후 연구 방향은 다음과 같다. 우선, 바이오파울링 발생에 대한 정확한 레이블링 데이터 확보가 필수적이다. 이를 위해 센서 표면의 이미지 데이터나 정기적인 현장 검사 결과 등과 같은 직접적인 바이오파울링 관측 데이터를 수집하고, 이를 용존산소량 등 간접적 지표들과 연계하여 분석할 필요가 있다. 데이터 측면에서는 연안/근해/원해 등 다양한 해역의 데이터를 확보하고, 장기간 데이터 수집을 통해 계절적 영향을 분석할 필요가 있다. 또한, 수온, 염도, pH 등 복합 환경 변수를 통합적으로 고려하는 연구가 필요하다.

실용화 측면에서는 바이오파울링 예측 모델을 기반으로 한 자동 알림 시스템과 유지보수 일정 최적화 등을 포함하는 통합 관리 플랫폼 개발을 진행하고 있다. 이를 통해 해양 센서의 효율적인 운영 및 관리가 가능할 것으로 기대된다.

결론적으로 본 연구는 해양 센서의 바이오파울링 예측이라는 실제적 문제에 대해 AI 기반의 해결 방안을 제시했다는 점에서 의의가 있다. 특히 실제 현장에서 즉시 활용 가능한 수준의 예측 정확도와 계산 효율성을 달성했다는 점에서 실용적 가치가 높다. 비록 실제 바이오파울링 발생 데이터의 부족이라는 근본적 한계를 가지고 있으나, 향후 제시된 보완 연구들을 통해 이러한 한계점들을 극복하고, 더욱 견고하고 실용적인 시스템으로 발전시킬 수 있을 것으로 기대된다.

**5. References**

[1] IPCC. (2019). IPCC Special Report on the Ocean and Cryosphere in a Changing Climate.

[2] Hoegh-Guldberg, O., & Bruno, J. F. (2010). The impact of climate change on the world’s marine ecosystems. Science, 328(5985), 1523-1528.

[3] Lindstrom, E., et al. (2012). A framework for ocean observing. UNESCO.

[4] Venkatesan, R., Tandon, A., D’Asaro, E., & Atmanand, M. A. (Eds.). (2018). Observing the oceans in real time. Springer. ISBN 978-3-319-66492-7.

[5] Delauney, L., Compère, C., and Lehaitre, M.: Biofouling protection for marine environmental sensors, Ocean Sci., 6, 503–511, <https://doi.org/10.5194/os-6-503-2010>, 2010.

[6] Growcott, A., Kluza, D., & Georgiades, E. (2017). In-water systems to reactively manage biofouling in sea chests and internal pipework. Marine Technology Society Journal, 51(2), 89-104.

[7] Poloczanska, E. S., & Butler, A. J. (2010). Biofouling and climate change. In S. Dürr & J. C. Thomason (Eds.), Biofouling (pp. 333-342). Wiley-Blackwell. ISBN 978-1-4051-6926-4.

[8] Wong, A., Keeley, R., & Carval, T. (2018). Argo Quality Control Manual for CTD and Trajectory Data, Version 3.1.

[9] Wang, Y., Han, L., Liu, W., Yang, S., & Gao, Y. (2019). Study on wavelet neural network based anomaly detection in ocean observing data series. Ocean Engineering, 186, 106129.

[10] 민용침, 정진용, 장찬주, 이재익, 정종민, 민인기, 심재설, 김용선, 해양과학기지 시계열 관측 자료 품질관리 시스템 구축 : 국제 관측자료 품질관리 방안 수온 관측 자료 시범적용과 문제점, 2020.

[11] Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. KDD 96, 226–231, 1996.

[12] Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J. LOF: identifying density-based local outliers. In: Proceedings of the SIGMOD Conference, pp. 93–104 , 2000.

[13] Cai, W.-Y., Liu, Z.-Q., & Zhang, M.-Y. . Trajectory clustering based oceanic anomaly detection using Argo profile floats. In ChinaCom 2019: International Conference on Communications in China (pp. 498–508). Springer, Cham.<https://doi.org/10.1007/978-3-030-41114-5_37>, 2020.

[14] Hodge, V. J., & Austin, J. A survey of outlier detection methodologies. Artificial Intelligence Review, 22(2), 85-126, 2004.

[15] Chalapathy, R., & Chawla, S. Deep Learning for Anomaly Detection: A Survey. arXiv preprint arXiv:1901.03407, 2019.

[16] Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 387-395, 2018.

[17] Zhao, Y., Nasrullah, Z., & Li, Z. PyOD: A Python Toolbox for Scalable Outlier Detection. Journal of Machine Learning Research, 20(96), 1-7, 2019.

[18] Vaquer-Sunyer, R., & Duarte, C. M. (2008). "Thresholds of hypoxia for marine biodiversity." *Proceedings of the National Academy of Sciences*, 105(40), 15452-15457.

[19] Bengio, Y., Simard, P., & Frasconi, P. (1994). "Learning Long-Term Dependencies with Gradient Descent is Difficult." IEEE Transactions on Neural Networks, 5(2), 157-166.

[20] Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). "On the Properties of Neural Machine Translation: Encoder–Decoder Approaches." Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8), 103-111.

[21] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press..

[22] Kingma, D. P., & Ba, J. (2015). "Adam: A Method for Stochastic Optimization." International Conference on Learning Representations (ICLR).

[23] Masters, D., & Luschi, C. (2018). "Revisiting Small Batch Training for Deep Neural Networks." arXiv preprint arXiv:1804.07612.

[24] Saito, T., & Rehmsmeier, M. (2015). "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets." PLOS ONE, 10(3), e0118432.

[25] Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). "Optuna: A Next-generation Hyperparameter Optimization Framework." Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2623-2631.

1. 경북대학교 컴퓨터학부 학사과정, E-mail: wngur1104@knu.ac.kr [↑](#footnote-ref-1)
2. 경북대학교 컴퓨터학부 학사과정, E-mail: [lgihtyear@gmail.com](mailto:lgihtyear@gmail.com) [↑](#footnote-ref-2)
3. 경북대학교 컴퓨터학부 학사과정, E-mail: [cpyy5526@knu.ac.kr](mailto:cpyy5526@knu.ac.kr) [↑](#footnote-ref-3)
4. 경북대학교 컴퓨터학부 학사과정, E-mail: [kimhyonjin10@gmail.com](mailto:kimhyonjin10@gmail.com) [↑](#footnote-ref-4)
5. 경북대학교 컴퓨터학부 교수, E-mail: [snowflower@knu.ac.kr](mailto:snowflower@knu.ac.kr) [↑](#footnote-ref-5)
6. 경북대학교 경영학부 박사과정, E-mail: [coolwin200@gmail.com](mailto:coolwin200@gmail.com)

   **Development of AI for Biofouling Prediction Using Smart Buoy Sensor Data: Recurrent Neural Network Prediction Model Based on Dissolved Oxygen Time Series Data**

   Ju-Hyeok Lim1, Yung-Chan Park1, Min-Seok Kim1, Min-Su Kim1

   Seol-Young Jung2, Hyeon-Sang Lee3

   1Bachelor of Science in, Computer Science and Engineering, Kyungpook National University

   2 Professor, School of Computer Science and Engineering, Kyungpook National University

   3 Ph.D. Student, School of Business Administration, Kyungpook National University

   **Abstract**

   최근 기후 변화로 해양 환경이 급격히 변하면서, 부표에 설치된 센서에 바이오파울링이 발생하여 데이터 정확도가 저하되고 있다(서론을 요약하다보니 논리적 비약 발생, 기후 변화로 인해 바이오파울링이 발생하는 것은 아님. 수정 필요). 바이오파울링은 해양 생물이 구조물에 부착하여 기능 수행을 방해하는 현상으로,

   발생 주기의 예측이 어려워(예측이 어렵다기 보다 사전적 예측이 필요하다로 제시) 데이터 이상 징후를 통한 예측이 필요하다.

   기존의 해양 환경 관측 데이터 품질 관리는 전문가의 수작업과 통계적 접근에 의존해왔으며,

   AI 기반의 이상 탐지 연구는 제한적으로 이루어져 왔다(어떤 제한점이 있는지 구체적으로 제시).

   본 연구는 완도 지역 3개 부표의 용존산소 데이터를 수집하여 바이오파울링 발생 시점(DO ≤ 3.0mg/L)을 기준으로 데이터를 레이블링하고, 2분 간격 시계열 데이터셋을 구축하였다.

   시계열 패턴 학습을 위해 RNN, GRU, LSTM 모델을 구현하고 다양한 시퀀스 길이(180-2880)에서 성능을 비교한 결과, GRU 모델이 작은 hidden size(32)로도 F1 Score 0.99를 달성하며 가장 우수한 성능을 보였다(논의 필요).

   본 연구는 해양 센서의 바이오파울링 예측에 있어 실용적이고 즉시 활용 가능한 AI 기반 접근법을 제시함으로써 해양 관측 시스템의 효율적 운영과 관리를 지원할 수 있음을 보여준다. [↑](#footnote-ref-6)