**\*\*abstract\*\***

**바다숲은 해조류 및 해초류가 바닷속에서 무성하게 이룬 숲으로, 갯녹음 현상을 막아 어획량을 늘리고 해양생태계를 유지하는 효과가 있으며 블루카본의 핵심 자원으로 주목받고 있다. 하지만 현재, 해역별 특수성을 고려하지 않은 바다숲 조성 사업으로 인해 일부 해역에서는 조성 전보다 해조류가 감소하는 현상이 나타나고 있다. 따라서 본 연구에서는 해역별 특성을 고려한 바다숲 조성이라는 목표를 달성하기 위해, 해양 데이터를 활용한 갯녹음 발생률 예측 모델을 생성하였다. 95%의 설명력을 보인 RandomForest 모델을 통해 갯녹음 발생률이 높은 상위 20곳을 선정하고 바다숲 조성을 통한 이산화탄소 감축량을 계산하였다. 계산 결과, 해당 지역에 바다숲을 조성했을 시 약 114만 톤의 이산화탄소가 흡수될 것으로 기대되며, 이는 기존 8만여 톤이었던 이산화탄소 감축량에 비해 약 14배에 달한다.**

**1.서론**

**최근 들어 자주 발생되는 이상기후와 지구온난화로 인해 탄소중립의 중요성이 강조되고 있다. 탄소 중립을 달성하기 위해 탄소 배출량만큼 탄소 흡수량도 주목받고 있다. 탄소 흡수원으로 부상하고 있는 블루카본은 해양생태계에서 흡수하는 탄소를 말하며, 이는** 육상 생태계보다 탄소를 최대 50배 이상 빠르게 흡수할 수 있다고 알려져 있어 잠재 가치를 높게 평가받고 있다.

블루카본의 핵심 자원인 바다숲은 해조류와 해초류로 구성되어 있어 바다 생태계를 유지하는 역할을 한다. 2050 탄소중립의 수단으로 바다숲 조성이 주목받으면서, 2021년까지 211개소에 총면적 26,644ha가 조성되었다. 하지만 해역별 특수성을 고려하지 않아 이에 대한 효과는 그다지 크지 않은 것으로 알려졌다.

다양한 원인 중 바다 사막화 현상인 갯녹음이 가장 큰 원인으로 꼽힌다. 갯녹음은 연안 암반 지역에서 해조류가 사라지고 흰색의 석회 조류가 달라붙어 암반 지역이 흰색으로 변하는 현상으로 해양 생태계와 어업에 큰 영향을 준다. 우리나라의 전 해역에 발생되고 있으며, 그 면적은 점차 증가하고 있다. 이 현상에 대한 원인은 아직 정확하게 밝혀지지 않았으나, 물리적, 화학적, 생물학적인 복합적인 요인으로 인해 발생하는 것으로 알려져 있다.

이에 본 논문은 갯녹음 발생에 영향을 줄 수 있는 해양생태학적 요인을 파악하여 해당 요인들에 대한 시계열 분석으로 갯녹음 면적이 가장 클 것으로 보이는 상위 20 지역을 선정하는 모델을 생성하여 해역별 특성을 고려한 최적지를 선정한 뒤 바다숲 조성으로 예상되는 연간 이산화 탄소 감축량을 계산하고자 한다.

**2. 관련연구**

**2.1 갯녹음 발생 요인**

앞서 언급한 듯이 갯녹음 현상의 원인은 복합적이다. 때문에 분석을 수행하기 전 변수를 선정하기 위해서 갯녹음의 요인을 살펴볼 필요가 있다.

갯녹음 현상이 발생하는 원인으로는 크게 물리적 원인, 화학적 원인, 생태학적 원인으로 구분할 수 있다(유혜민 외, 2016). 물리적 원인으로는 환경오염으로 인한 수온의 상승과 이로 인한 바다속 용존산소량의 감소가 대표적이다. 이외에도 해조류를 섭식하는 조식동물들의 지나친 증식과 오염된 담수의 유입을 꼽을 수 있다. 화학적 원인으로는 해안 개발과 양식업장으로 인한 해양오염과 해수의 저염수화, 해양의 영양 결핍이 있다. 생태학적인 원인 중 가장 유력한 식해설에 따르면 무절석회조류의 과포화가 원인이다. 국내 성게 어획량이 감소하여 개체수가 증가하였고 이의 먹이가 되는 해조류가 줄어들어 그 자리에 무절석회조류의 분포가 확산된 것이다. 이외에도 인공적인 원인이 존재하는데 해양오염으로 인한 해수의 불투명도 증가와 수질 부유물질의 영향이 대표적이다(권혜옥 외, 2007). 이를 통해서 갯녹음이 발생하는 요인으로 물리적 요인과 화학적 요인 중 일부를 택하여 갯녹음 현상을 설명하는 변수로 꼽았다.

**2.2 바다숲**

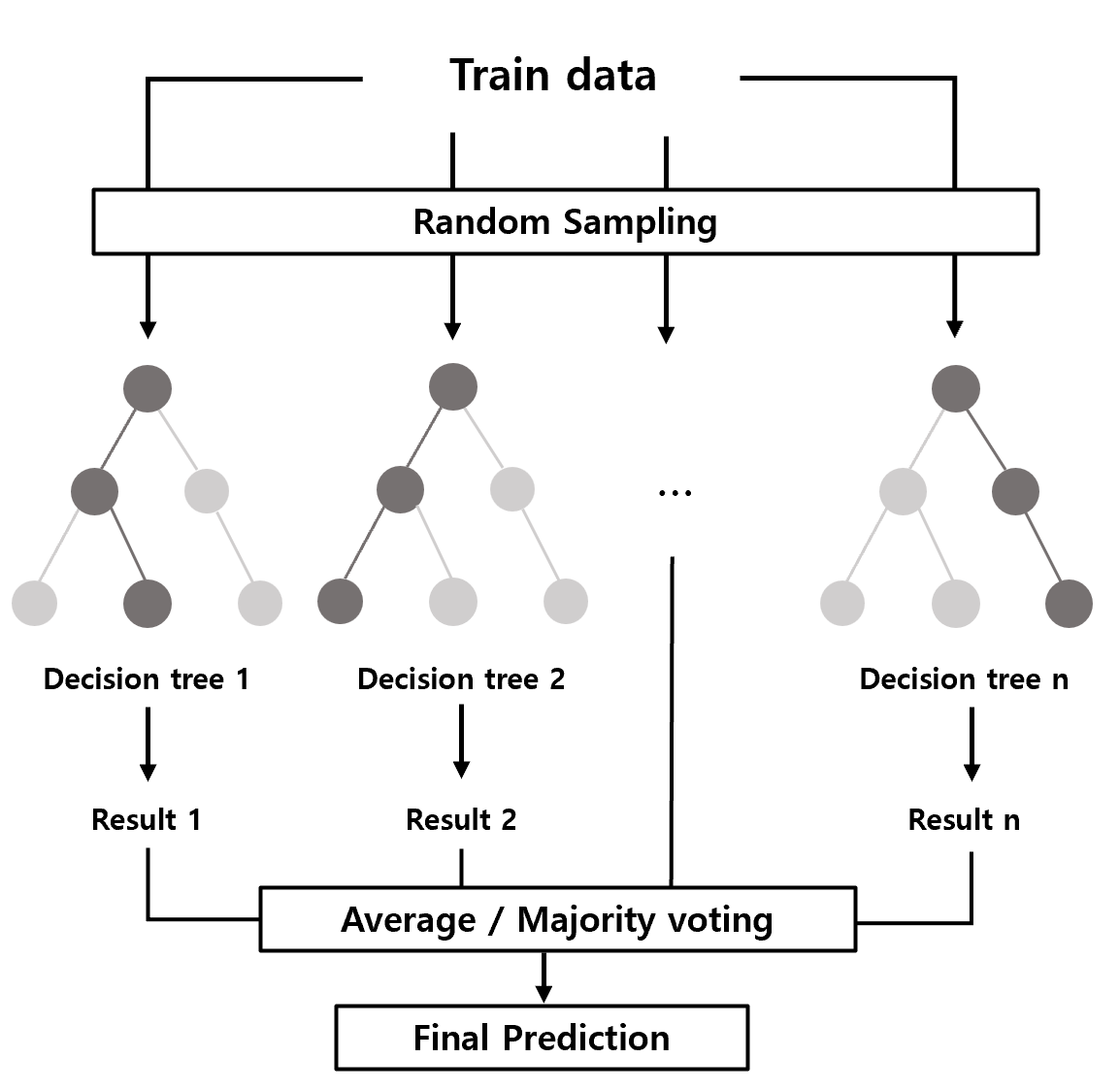
바다숲은 생태계의 1차 생산자 역할을 담당하고, 해양생물의 서식처 및 산란처로 활용되며, 해양생태계의 복원 및 균형유지와 같은 환경보전기능을 수행한다(김대권 등,2011). 바다숲 조성면적을 확대함으로 갯녹음 확산을 방지 및 연안생태계 복원 효과를 기대하며 2030년까지 총260개소(5만4000ha)를 조성할 계획이다(해양수산부, 2016). 바다숲 조성해역의 물리적 환경을 고려한 최적적지선정(오태건, 2010)은 바다숲 조성에 영향을 주는 물리적 환경인 경사도, 퇴적물, 보상심도, 유속 및 파랑 등의 해양물리항목을 고려하여 최적적지를 도출하였다. 그러나 물리환경과 해조류의 생리 및 생태에 대해서 보다 면밀하게 고려하여 물리와 생물환경을 종합적으로 고찰하여 적지를 도출할 경우 보다 정확한 해양환경 분석이 가능할 것이라 판단하여, 본 연구에서는 해역, 수온 같이 물리적 환경과 갯녹음 영향 요인을 고려하여 변수를 선정하였다.

**3. 알고리즘**

**3.1 랜덤포레스트 모델**

랜덤포레스트 모델은 앙상블 학습 방법 중 하나로 분류, 회귀 분석 등에서 사용되고 있는 머신러닝 알고리즘 기법이다. 랜덤포레스트 모델은 부트스트랩 방식으로 데이터를 샘플링하여 의사결정나무를 생성한다. 이러한 의사결정나무를 복수 개 생성하고 각각의 의사결정나무가 도출한 결과들을 통해 최종 값을 도출한다. 분류 모형일 경우에는 투표를 통한 다수결로, 회귀 모형일 경우에는 결과들의 평균으로 최종 결과를 나타낸다.

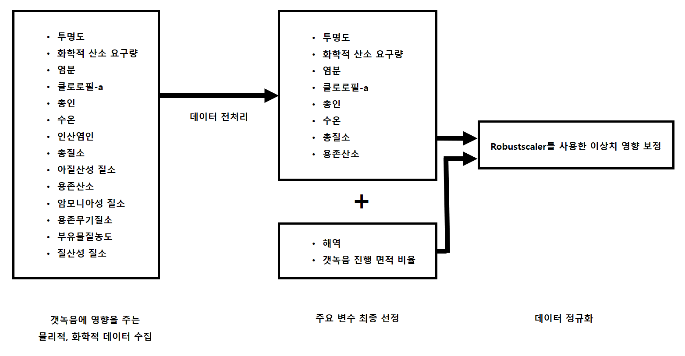
본 연구에서는 랜덤포레스트를 통해 예측을 진행하고자 한다. 예측변수의 수가 많아지면 예측 변수의 외생성이나 오차항 분포에 대한 제약이 있는 기존 전통적인 회귀분석이나 로지스틱 회귀분석 모형의 한계를 보완하여 랜덤포레스트 모델은 설명변수의 수가 많아지면 예측력이 높아지는 장점을 가지고 있다. 이러한 점은 많은 설명 변수를 가진 갯녹음 현상을 분석하는 데 적합하다고 판단하여 해당 모델을 선택하였다.



<Figure 1> Random Forest Simplified

**4. 머신러닝을 활용한 갯녹음 위험 지역 예측 방법**

**4.1 해양 데이터 전처리 및 분석을 위한 주요 변수 선정**



<Figure 2> Structural Diagram of the Selection of Major Variables for Preprocessing and Analysis of Ocean Data for Prediction of Tidal Whitening Event Risk Areas

본 논문에서는 머신러닝을 활용한 갯녹음 위험 지역 예측 방법을 제안하고자 한다. 이를 위해 먼저 해양 데이터 전처리 및 분석을 위한 주요 변수 선정 과정을 거친다. <Figure 1>은 수집한 해양 데이터를 전처리하는 과정에 대한 전체 구조도이다.

우선, 갯녹음 발생 예상 지역을 예측하기 위해 갯녹음을 유발하는 물리적, 화학적 데이터를 수집한다. 수집한 갯녹음 발생 요인 데이터를 분석을 위해 2002년부터 2020년까지의 해양환경측정망 데이터를 병합한 후 다량의 결측치가 발생한 행(지점) 또는 열(변수)를 제거한다. 이후 갯녹음과 관계가 있을 것으로 예상되는 변수들을 선별하고, 지점별로 구분하여 결측값이 발생한 경우 열의 평균으로 대치한다. 이후 Station 별로 해역을 구분하기 위해 동해=1, 남해=2, 서해=3으로 ‘해역’ 변수를 추가한다. 또한, 전국 갯녹음 발생 현황 데이터로부터 갯녹음 진행면적 비율 데이터를 추출하여 분석 데이터에 ‘getnok’ 변수를 추가한다.

이후 수집한 변수들이 실제로 갯녹음 심화에 영향을 미치는지 알아보기 위해 상관관계 확인 과정을 거친다. 상관계수는 한 변수의 변화가 다른 변수의 변화에 따라 어떤 변화가 일어나는지 보여주는 지표 역할을 하며, 일반적으로 0.3이상의 값에서 뚜렷한 양의 상관관계를 가진다. 따라서 상관계수가 0.3이상인 변수를 분석 변수로 선택하고자 하였다. 또한, 초기 수집 데이터 중 인산염인과 아질산성 질소의 경우 상관계수와 무관하게 총인과 총질소가 변수로 선정되었으므로 변수가 중복되지 않기 위해 제외한다.

최종적으로 선정한 주요 변수 중 갯녹음 진행 면적 비율을 제외한 9개의 변수에 대해Robustscaler를 사용하여 데이터 정규화를 진행한다. 머신러닝은 데이터가 가진 feature들을 비교하여 데이터의 패턴을 찾는데, 이때 머신러닝 모델에 투입되는 모든 데이터가 같은 scale로 반영되어 이상치에 큰 영향을 받지 않도록 정규화 과정을 거친다. 이를 위해 Robustscaler를 사용하여 각 feature의 median(Q2)에 해당하는 데이터를 0으로 설정한 후, Q1, Q3 사분위수의 차이인 IQR(Q3-Q1)로 나누어 정규화를 진행한다. 이와 같은 데이터 전처리 과정을 통해 이상치의 영향을 최소화하며 이후 적용하게 될 머신러닝 모델의 학습 정확도를 높일 수 있다.

**4.2 시계열 데이터 분석을 통한 미래 해양 데이터 예측**

앞서 전처리 과정을 거친 해양환경측정망 데이터를 이용하여 fbprophet을 기반으로 미래 해양 데이터 내 갯녹음을 유발하는 주요 변수 값을 예측한다. Facebook에서 공개한 시계열 데이터 분석 라이브러리인 fbprophet은 정확도가 높고 빠르며 직관적인 하이퍼 파라미터로 모델 훈련이 용이하다는 장점을 가진다. 본 논문에서는 2002년부터 2020년까지의 해양환경측정망 데이터를 이용하여 2021년의 데이터를 예측하기 위해, fbprophet을 사용하여 분석에 사용할 9개의 주요 변수 중 ‘해역’을 제외한 8개의 변수에 대한 시계열 예측을 진행한다. 하이퍼 파라미터 튜닝을 통해 각 feature 별 Root Mean Square Error(RMSE) 값이 가장 작은 모델로 예측한다. RMSE는 모델의 예측값과 실제 값의 차이를 다룰 때 사용하는 측도로 0에 가까울수록 예측값과 실제값의 차이가 작음을 수치적으로 확인할 수 있다. <Figure 2>는 최종 선정된 시계열 예측 모델의 각 feature 별 평균 RMSE를 나타낸다.



<Figure 3> Average RMSE for each feature of the finally selected time series prediction model

**4.3 머신러닝을 활용한 갯녹음 비율 예측 모델의 바다숲 조성 위치 선정**

본 논문에서는 머신러닝을 활용한 갯녹음 발생 예상 면적 비율을 예측하기 위해 랜덤 포레스트 모델을 사용하고자 한다. 랜덤포레스트 모델에는 n\_estimator, max\_depth, gamma 등의 파라미터가 존재한다. n\_estimator는 랜덤포레스트 안의 의사결정나무의 개수이며 이 값이 클수록 과대적합을 막아주지만 메모리와 훈련시간이 증가하기 때문에 적절한 수준으로 설정해 줄 필요가 있다. max\_depth는 트리의 깊이를 뜻하며 이 값이 클수록 복잡한 트리가 생성된다. 이 파라미터를 활용해 모델이 복잡해져 과적합이 나타나는 상황을 사전에 대처할 수 있다. gamma는 매개변수가 하나의 훈련 샘플이 미치는 영향의 범위를 결정한다. Gamma 값이 클수록 영향을 미치는 범위가 좁다는 것을 의미하며, 모델이 복잡해진다. 해당 연구에서는 gamma = 1, max\_depth = 4, n\_estimator = 300으로 설정하였다.

최종적으로 Random Forest를 통해 0.9580의 R\_squared 값을 확인할 수 있다. R\_squared는 회귀 모델의 성능에 대한 평가 지표로, 독립변수가 종속변수를 얼마만큼 설명해주는지를 보여주며 1에 가까울수록 설명력이 높다. 이를 바탕으로 갯녹음 비율 증감량이 높은 20개 지점을 바다숲 조성 위치로 우선 선정한다. 이때 갯녹음의 비율 증감량은 2021년 갯녹음 예상 면적 비율에서 과거 갯녹음 면적 비율을 뺀 값으로 설정하였다.

**5. 결론**

|  |  |  |  |
| --- | --- | --- | --- |
| 삼척연안2 | 20.38814 | 삼척연안1 | 16.36714 |
| 고성자란만3 | 14.61409 | 기장연안2 | 12.41957 |
| 고흥연안2 | 12.12102 | 고성자란만2 | 10.99447 |
| 김포연안1 | 10.80200 | 주문진연안1 | 8.69529 |
| 사천연안3 | 8.41482 | 진주만1 | 8.32828 |
| 사천연안2 | 8.12024 | 가막만3 | 8.05346 |
| 전주포연안1 | 6.90004 | 고성자란만1 | 6.65934 |
| 군산연안4 | 6.56848 | 전주포연안2 | 6.55727 |
| 김포연안2 | 6.34936 | 군산연안3 | 6.33599 |
| 군산연안1 | 6.25741 | 온산연안2 | 6.13800 |

<Figure 4 갯녹음 비율 증감량>

<Figure 4>는 앞서 사용한 사용한 random forest 통해 2021년 갯녹음 예상 면적 비율에서 과거 면적 비율을 제거한 수치가 가장 높은 지역들을 나열한 표이다.

우선, 한국수산자원공단에서 제공한 자료를 바탕으로 사업 해역 1개소 당 약 126.275ha 만큼의 바다숲을 위 지역들에 설치한다. 연간 예상 이산화탄소 감축량 (Daisuke Muraoka, 2004)을 이용하여 계산하기 위해 각 해역별로 이산화탄소 감축량이 가장 큰 해조류를 선택한다. 선정된 20곳 중 7개의 동해 지역은 다시마, 서해 5곳은 해초, 8개의 남해지역에도 이산화탄소 감축량이 큰 다시마를 바다숲에 심는다고 가정하면 연간 예상 이산화탄소 감축량은 약 1,146,855tCO2이다. 이는 서울과 뉴욕을 왕복 18,024번을 이동할 때 발생하는 이산화탄소를 흡수하며 같은 면적을 가진 산림이 흡수하는 이산화탄소 양의 약 45배이다. 이로써 2050 탄소중립의 수단으로 바다숲 조성은 타당하며 바다숲을 통해 해양생태계 복원, 갯녹음 확산예방, 어업 이익 증대와 같은 추가적인 기대효과 또한 기대해 볼 수 있다.

**참고문헌**

김대권·황선재·최옥인·최임호·한명일·신영재, 2011, “제주연안에서 기후변화가 갯녹음 확산에 미치는 영향,” 수산자원관리 1(1), pp.1-17

오태건, 김대권, 김창길, 이문옥, 조재권. (2010). 바다숲 조성해역의 물리적 환경을 고려한 최적적지선정. 한국마린엔지니어링학회지, 34(1), 183-194.

유혜민, 박지연, 송예슬, 이은정, 강송이, 김주효, 박민지, 변유진, 성희원, 오현경, 윤화영, 이수현, 정소영. (2016). 한국의 갯녹음 실태 연구. 응용지리, (33), 86-89.

권혜옥, 김진만, 이식모. (2007). 갯녹음 해역의 해조장 복원 모듈. 한국해양환경·에너지학회 학술대회논문집, (), 72.

Daisuke Muraoka. (2004). Seaweed resources as a source of carbon fixation., 3p