

# SHAP 분석 기반의 넙치 질병 분류 입력 파라미터 최적화

조경원\* · 백란\*\*

## Optimizing Input Parameters of *Paralichthys olivaceus* Disease Classification based on SHAP Analysis

Kyung-Won Cho\* · Ran Baik\*\*

### 요약

머신러닝을 이용한 텍스트 기반 어류 질병 분류에서 머신러닝 모델의 입력 파라미터가 너무 많은 문제가 존재하지만, 성능의 문제로 임의로 입력 파라미터를 줄일 수 없다. 본 논문에서는 이 문제를 해결하고자 SHAP 분석 기법을 활용해 넙치 질병 분류에 특화된 입력 파라미터 최적화 방안을 제시한다. 제안한 방법은 SHAP 분석 기법을 적용하여 넙치 질병 문진표에서 추출한 질병 정보의 데이터 전처리와 AutoML을 활용한 머신러닝 모델 평가 과정을 포함한다. 이를 통해 AutoML의 입력 파라미터의 성능을 평가하고, 최적의 입력 파라미터 조합을 도출한다. 본 연구에서 제안 방법은 필요한 입력 파라미터 수를 감소시키면서도 기존의 성능을 유지할 수 있을 것으로 기대되며, 이는 텍스트 기반 넙치 질병 분류의 효율성 및 실용성을 높이는 데 기여할 것이다.

### ABSTRACT

In text-based fish disease classification using machine learning, there is a problem that the input parameters of the machine learning model are too many, but due to performance problems, the input parameters cannot be arbitrarily reduced. This paper proposes a method of optimizing input parameters specialized for *Paralichthys olivaceus* disease classification using SHAP analysis techniques to solve this problem. The proposed method includes data preprocessing of disease information extracted from the halibut disease questionnaire by applying the SHAP analysis technique and evaluating a machine learning model using AutoML. Through this, the performance of the input parameters of AutoML is evaluated and the optimal input parameter combination is derived. In this study, the proposed method is expected to be able to maintain the existing performance while reducing the number of input parameters required, which will contribute to enhancing the efficiency and practicality of text-based *Paralichthys olivaceus* disease classification.

### 키워드

*Paralichthys olivaceus* disease, Text-based classification, SHAP Analysis, Machine Learning  
넙치 질병, 텍스트기반 분류, SHAP 분석, 기계학습

\* 호남대학교 박사과정(kwcho90@naver.com)

\*\* 교신저자 : 호남대학교 컴퓨터공학과

• 접수일 : 2023. 10. 28

• 수정완료일 : 2023. 11. 19

• 게재확정일 : 2023. 12. 27

• Received : Oct. 28, 2023, Revised : Nov. 19, 2023, Accepted : Dec. 27, 2023

• Corresponding Author : Ran Baik

Dept. Computer Engineering, Honam University

Email : baik@honam.ac.kr

## I. 서론

FAO의 농업 전망 보고서[1]에 따르면 세계시장에서 수산물은 농축 수산물 중에서 가장 빠르게 성장하는 단백질 공급원 보고 있고 1인당 연간 소비량도 계속 증가하고 있다. 이에 따라, 한국 수산업은 어획 어업생산량은 갈수록 줄어들고 있지만 양식 어업 생산량은 계속 증가하고 있다[2].

양식 어업에 가장 큰 어려움은 어류의 질병 발생 및 질병에 따른 폐사로 인한 생산성 저하이다. 특히, 최근에는 발생 시기와 관계없이 어류의 질병이 발생하고 있으며 질병에 따른 집단 폐사율이 총사육량의 25-30% 정도가 된다. 또한 이에 따른 연간 피해액은 2014년 기준 약 2,500억 원으로 추정된다[3].

국립수산물품질관리원 국가수산물질병정보 질병발생통계에 따르면 어류의 질병 종류별 발생 현황을 보면, 세균성, 기생충성, 세균혼합감염증, 세균 및 기생충 혼합감염증, 바이러스성 질병 등 순으로 다양한 질병이 발생하고 있다[4]. 이러한 수산 양식에서 질병 감염 확산을 막기 위해서는 어류의 질병을 빠르게 예측하는 기술이 필요하다. 이에 따라 최근 인공지능 기술인 딥러닝 모델을 활용한 어류 질병 진단 및 예측에 관한 연구가 진행되고 있다[5-7]. 그러나 대부분 이미지 데이터를 기반으로 딥러닝 객체 탐지 모델을 사용하여 질병 분류 연구를 하고 있다.

일반적인 텍스트 기반 분류 모델에서는 머신러닝 모델의 입력 파라미터의 개수가 많이 존재할 수가 있다. 이것은 데이터를 수집하는 과정에서 질병 분류에 영향을 미치는 가능한 속성들을 모두 포함하는 데이터를 추가하면서 발생한다. 넓치 질병 분류 모델에서는 이미지 데이터도 수집하지만, 질병 검사 대상 문진표를 이용하여 넓치의 상태(크기, 전장, 중량 등)와 질병 정보(피사, 발적, 궤양, 출혈 등) 등을 수집할 수 있다. 이러한 문진표 데이터를 이용하면, 텍스트 기반의 머신러닝 모델을 필요하여 넓치 질병 분류를 수행할 수 있다. 그러나 이 경우에도 전처리 과정을 거친 후에도 총 51개의 파라미터가 입력으로 들어가고 학습을 통해서 하나의 질병 분류 결과를 출력하게 된다. 입력 파라미터의 개수가 많아지면 많아질수록 넓치 질병 분류를 위한 서비스 개발에 어려움이 발생하지만, 입력 속성 파라미터 개수가 많다고 해서 성능에 영향을 주

기 때문에 임의로 파라미터를 줄일 수도 없다.

본 연구에서는 머신러닝을 이용한 텍스트 기반 어류 질병 분류 모델의 입력 파라미터 과다함의 문제를 해결하고자 SHAP(: SHapley Additive exPlanations)[8] 분석을 기반한 입력 파라미터의 개수 최적화 방안을 제시한다. 제안한 방법은 질병 검사 대상 문진표에서 추출한 질병 정보의 데이터 전처리, AutoML(: Automated Machine Learning)[9]을 활용한 모델의 성능 평가, SHAP 분석 과정을 포함한다. 이를 통해 각 파라미터의 성능을 평가하고, 최적의 입력 파라미터 조합을 도출한다. 본 연구의 제안 방법은 필요한 파라미터 수를 감소시키면서도 기존의 성능을 유지할 수 있을 것으로 기대되며, 이는 텍스트 기반 넓치 질병 분류의 효율성 및 실용성을 높이는 데 기여할 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로서 AutoML 기계학습 모델을 소개한다. 3장에서는 제안한 딥러닝 기반의 SHAP 분석을 기반한 입력 파라미터 최적화 방법을 제안한다. 마지막으로 4장에서는 결론 및 향후 연구에 관해서 기술한다.

## II. 관련연구

AutoML은 기계학습 모델의 선택, 구성, 그리고 최적화 과정을 자동화하는 기술로 데이터 전처리, 특징 추출, 모델 선택, 하이퍼파라미터 조정 등을 자동화하여, 기계학습을 쉽고 빠르게 개발할 수 있도록 도와준다. 본 논문에서는 AutoML의 여러 모델 중에서 성능이 좋은 ExtraTree[10], RandomForest[11], XGB[12], LGBM[13], Bagging[14], SVC[15]를 사용한다.

ExtraTrees는 결정트리 모델을 이용한 Bagging 학습을 하는 앙상블 학습 모델이다. 이 모델은 결정트리 모델이 노드를 분리할 때 특성을 임의로 선택할 뿐만 아니라 무작위로 선택하여 학습하는 모델이다. 훈련에 사용한 특성을 무작위 선택 후 그 중에서 최적을 선택하고 임계 값을 무작위로 선택 후에 그 중 최적을 선택한다.

RandomForest는 ExtraTrees 모델처럼 결정트리를 이용한 배깅 학습을 하는 앙상블 학습 모델이다. 이 모델은 데이터셋의 다양한 하위 샘플에 의사결정 트리 분류기를 적합시키고 평균화를 사용하여 예측 정확도를 향상시키고 과적합을 제어한다.

XGB( eXtreme Gradient Boosting)는 결정 트리를 앙상블하여 분류기를 만드는 기법으로, 각 트리가 이전 트리의 오류를 보정하면서 학습한다. 이 모델은 과적합을 방지할 수 있는 다양한 정규화 기법을 제공하고 누락된 데이터를 자동으로 처리할 수 있고, 가지치기를 통해 복잡도를 관리한다.

LGBM( Light Gradient Boosting Model)는 결정 트리 기반의 학습 알고리즘으로, 큰 규모의 데이터와 다양한 종류의 예측 모델링 문제에 적합하게 설계되었다. 이 모델은 효율적인 트리 구조인 리프 중심 트리 분할 방식을 사용하여, 더 낮은 손실로 모델을 빠르게 학습시키고, 고차원 대규모 데이터셋에서도 우수한 성능을 발휘한다. 또한, 메모리 사용량이 적고 실행 속도가 빠른 것이 특징이며, 결측치 처리 및 범주형 특성을 기본적으로 지원하여 데이터 전처리 작업을 간소화한다.

Bagging은 앙상블 학습 기법 중 하나인 배깅을 구현한 것이다. 이 모델은 여러 개의 결정 트리를 병렬적으로 훈련시키고, 이들의 결과를 종합하여 최종 예측을 결정한다. 원본 훈련 데이터셋에서 중복을 허용한 랜덤 샘플링을 통해 생성된 샘플을 사용하여 독립적으로 훈련되고 과적합을 줄이고 일반화 성능을 향상시킨다.

SVC( Support Vector Classification)는 데이터 포인트를 고차원 공간으로 매핑하여, 서로 다른 클래스를 최대한 멀리 떨어뜨리는 결정 경계를 찾는다. 이 모델은 클래스 간의 경계가 불분명한 경우에도 효과적이며, 비선형 분류를 위해 커널 트릭을 사용하여 선형이 아닌 데이터를 효율적으로 처리할 수 있다. 마진(margin)이라고 하는 클래스 사이의 간격을 최대화하는 방식으로 작동하며, 이를 통해 모델의 일반화 능력을 강화한다..

### III. 입력 파라미터 최적화 방법

본 연구에서는 넙치 질병 증상을 기록하는 문진표의 텍스트 데이터를 이용하여 그림 1과 같은 머신러닝 모델 기반의 넙치 질병 분류 모델 입력 파라미터 최적화 방법을 제안한다. 본 연구에서 제안하고 있는 입력 파라미터 최적화 방법은 그림 1과 같이 넙치 양식장에서 데이터 수집한 후 전처리, AutoML을 이용한 데이터 학습, SHAP 분석, 입력 파라미터 조정하면서 훈련하는 총 5단계로 구성된다.

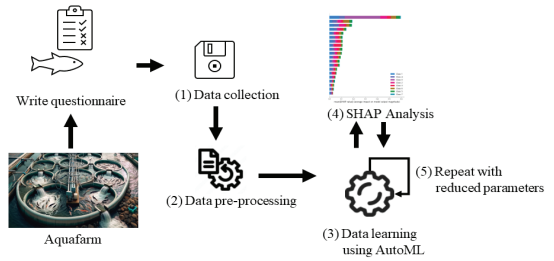


그림 1. SHAP를 이용한 입력 파라미터 최적화 방법  
Fig. 1 The method of optimizing input parameters using SHAP

### 3.1 데이터 수집

넙치에 대한 데이터 수집은 2022년 10월 부터 2023년 06월까지 제주 11곳의 양식장과 완도 지역에 13곳의 양식장에서 한 달에 한 번씩 한 양식장 당 10마리씩 선정해서 수집한 것이다. 양식장에서 넙치를 가져오면 검사자가 질병 검사 대상의 양식에 넙치에 대한 일반적인 정보 및 질병에 대한 내용을 작성한다.

#### 질병 검사 대장

검사일 : 2023-05-13  
검사자 :

■ 어장 정보

양식장: 칠진수산	주소: 서귀포시 성산읍 온평리 1297-5	대표: 김병욱	연락처: 010-9723-1026
수면적(㎡): 4,703	수온(℃): 15.6	환수량(평일/일): 31	사육수: 자연수/지하수
사료: MP			

■ 시료 정보

시료 번호: JCN230506	입식일: 22.06	전장(cm): 37.7	체중(cm): 14.4	중량(g): 660
사육 밀도(kg/㎡): 18	토양 이력: 0/2년 5일간 장두묘어	기타:		

■ 외부 증상

기관	증상	기관	증상	기관	증상
체표	<input type="checkbox"/> 발적 <input type="checkbox"/> 울혈 <input type="checkbox"/> 궤양 <input type="checkbox"/> 근육출혈 <input type="checkbox"/> 피사	복부	<input type="checkbox"/> 팽만	입	<input type="checkbox"/> 주둥이 괴양
지느러미	<input type="checkbox"/> 여혈 <input type="checkbox"/> 반열피사 <input type="checkbox"/> 주부출혈 <input type="checkbox"/> 비병부출혈 <input type="checkbox"/> 출혈	눈	<input type="checkbox"/> 들음 <input type="checkbox"/> 출혈	형통	<input type="checkbox"/> 염종 <input type="checkbox"/> 탈양
아가미뚜껑	<input type="checkbox"/> 안쪽 염종, 출혈	기타			

■ 내부 증상

기관	증상	기관	증상	기관	증상	기관	증상
아가미	<input checked="" type="checkbox"/> 변형 <input type="checkbox"/> 피사 <input type="checkbox"/> 결손	복강	<input type="checkbox"/> 출혈성 복수 (복수)	신장	<input type="checkbox"/> 비대 <input type="checkbox"/> 염종	비장	<input type="checkbox"/> 비대
간	<input type="checkbox"/> 비대 <input type="checkbox"/> 울혈 <input type="checkbox"/> 염종	장	<input type="checkbox"/> 염증 <input type="checkbox"/> 출혈	생식소	<input type="checkbox"/> 출혈	기타	
	<input type="checkbox"/> 농양 <input type="checkbox"/> 섬유화 <input type="checkbox"/> 출혈		<input type="checkbox"/> 국소적 출혈		<input type="checkbox"/> 국소적 출혈		

■ 검사 결과

○ 기생충

외부 기생충	
내부 기생충	장내 엔테로톡소증 소양

○ 세균 (배양 배지: ■ BHEA, □ TSA, ■ SS, ■ TCBS, □ MA / 배양 온도: 28 °C)

①	검출 부위	간	신장	비장	기타:
②	검출 부위	간	신장	비장	기타:
③	검출 부위	간	신장	비장	기타:
④	검출 부위	간	신장	비장	기타:

○ 바이러스: (생물형 조직 ■ 심장 ■ 비장 □ 기타: )

○ 기타

● 종합결과: 엔테로톡소증

■ 검사자 소견  
장내 엔테로톡소증 소양 감염됨

그림 2. 넙치 질병 정보를 위한 문진표 예시  
Fig. 2 Example of a questionnaire for *Paralichthys olivaceus* disease information

그림 2와 같이 문진표에는 각 양식장 별로 수집된 어장 정보와, 낚치의 크기와 중량 등과 같은 일반적인 정보, 외부 및 내부 증상, 검사결과, 검사소견 정보가 있다. 총 컬럼의 수는 73개이고 전체 1,930개의 문진표에 입력된 정보를 엑셀파일로 만들고 이를 CSV 파일로 변환한다.

### 3.2 데이터 전처리

문진표를 그대로 CSV 파일로 변환하여 바로 사용하지 못하므로 필요한 데이터를 선별하고 가공하기 쉬운 상태로 전처리를 수행한다. 데이터 전처리 단계는 불필요한 컬럼 삭제, 결측치가 높은 데이터 삭제 및 결측값 대체, 학습 가능한 형태로 데이터 변형 단계를 통해서 데이터 전처리를 수행한다.

#### (1) 불필요한 속성 삭제

이 단계에서는 데이터 중에서 학습에 필요한 것인지 판단하여 불필요한 속성과 학습하기 어려운 데이터들은 제거한다. 전체 73개의 속성중에서 '순서', '개체 ID', '지역', '양식장명', '수집월', '촬영 날짜', '원본 해상도', '이미지수량', '검사일', '입식일', '투약이력', '검사사소견', '비고'는 학습에 필요한 정보가 아니라고 판단하여 삭제한다.

#### (2) 결측치가 높은 데이터 삭제 및 결측값 대체

이 단계에서는 데이터에 결측 값이 많으면 학습에 악영향을 미치므로 이를 제거한다. 그림 2의 문진표를 분석한 결과 질병 증상 속성을 제외한 수조면적, 수온, Do, pH, 염분농도, 사료, 환수량 등이 결측치가 매우 높은 것을 확인하였고 이 속성들을 제거한다. 또한 크기, 전장, 중량은 결측 값이 1% 정도가 존재하여 결측 값이 존재하는 행을 삭제한다. 질병 증상 데이터가 존재하는 경우는 'O', 증상은 있지만 질병이 아니라면 '△', 증상이 없으면 null로 되어 있다. 그러므로 null은 'X' 값으로 대체 한다.

#### (3) 학습 가능한 형태로 데이터 변형

이 과정에서는 텍스트로 되어 있는 정보를 학습이 가능한 숫자 값으로 변경한다. 수집한 데이터에는 증상 별 값은 'O', '△', 'X'로 되어 있어 학습이 불가능 하기 때문에 'X'는 0, 'O'는 1, '△'는 2로 숫자 값으로 변경

한다. 또한 검사결과(질병)은 낚치의 질병명의 텍스트로 되어 있기 때문에 표 1과 같이 숫자로 변경한다. 텍스트 데이터를 통해 어떤 질병 혹은 정산인지 판별하는 것이 목적이기 때문에 이 검사결과 값이 라벨(정답)이 된다. 그 외의 전장 및 중량 값이 편차가 크기 때문에 Standard Scaler를 이용하여 데이터를 정규화 해준다. 데이터 전처리 결과 전체 73개의 컬럼 중에 51개를 훈련에 사용한다.

표 1. 질병명에 대한 변환 값 리스트  
Table 1. List of conversion values for disease names

Name of the disease	Value
Normal	0
Vibrio	1
Scutica	2
Viral Hemorrhagic Septicemia	3
Streptococcosis	4
Flexibacter maritimus	5
Edward	6
Emaciation	7

### 3.3 파라미터 조정 없는 AutoML 학습

이 단계에서는 입력 파라미터의 조정 없이 AutoML을 이용하여 학습을 수행하고 학습된 결과를 평가한다. 본 논문에서는 AutoML중에 lazypredict<sup>1)</sup>을 사용한다. 표 2는 lazypredict을 사용해서 학습한 후 최상위 6개 모델의 결과를 표시한 것이다.

표 2. AutoML을 이용한 학습 결과  
Table 2. Learning results using AutoML

Model	Accuracy	F1 Score
ExtraTreesClassifier	0.82028	0.82365
RandomForestClassifier	0.78802	0.78791
XGBClassifier	0.78802	0.79145
LGBMClassifier	0.78341	0.78796
BaggingClassifier	0.75576	0.75564
SVC	0.73272	0.72426

1) Lazy Predict, <https://lazypredict.readthedocs.io/>

3.4 최상위 모델에 대한 SHAP 분석

이 단계에서는 이전 단계에서 학습 결과 중에서 가장 좋은 모델을 선택하고 SHAP 분석을 수행한다. SHAP 분석은 여러 모델을 동시에 하지 못하기 때문에 특정 모델을 선택해야 한다. ExtraTreesClassifier가 가장 높은 정확도를 얻어, 이 모델을 SHAP로 분석한다. 그림 3은 ExtraTreesClassifier 모델에 대해서 SHAP를 이용하여 분석한 결과이다. 최상위부터 질병을 분류하는데 가장 많이 영향을 주는 입력 값이 나열된 것이다. 그래프의 색상 정보(Class 0-7)는 표 1의 질병명의 값을 의미한다.

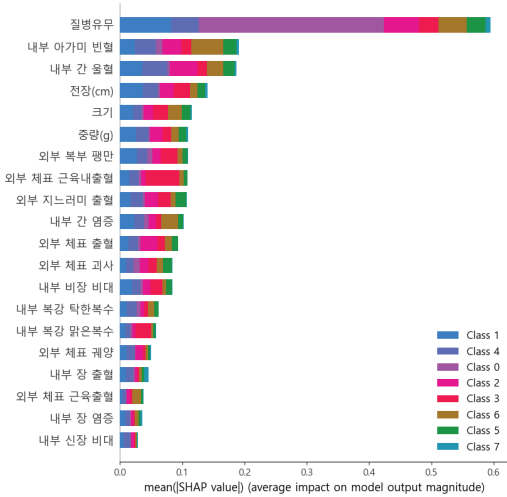


그림 3. SHAP 분석 결과  
Fig. 3 The result of SHAP analysis

3.5 입력 파라미터 조정에 따른 AutoML 학습

이 단계에서는 SHAP 분석으로 출력한 값의 가장 낮은 값부터 정렬하고 입력 값을 줄여가면서 학습을 반복 수행한다. 그림 4는 6개 모델에 대한 입력 파라미터 값을 줄여가면서 혼련한 정확도 측정 결과를 그래프로 표현한 것이다. 이 그래프를 통해서 입력 파라미터를 줄이면서 어느 정도의 정확도를 유지하는 결과를 얻을 수 있다. 우리는 결과적으로 25개의 입력 값을 줄이면서 ExtraTreesClassifier이 정확도를 82%를 유지할 수 있었다.

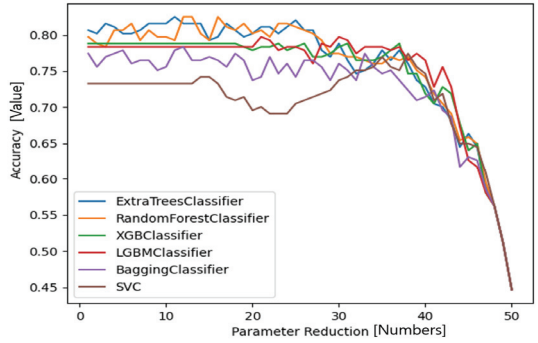


그림 4. 파라미터 감소에 대한 모델들의 정확도 측정 결과  
Fig. 4 The result of measuring the accuracy of models for parameter reduction

IV. 결 론

본 연구에서는 텍스트 기반 분류에서 나타나는 입력 파라미터 과다의 문제를 해결하기 위해, 넓치 질병 분류에 적용된 설문지 데이터를 사용하여 SHAP 분석과 AutoML을 통한 입력 파라미터 수 최적화 방안을 제시하였다. 제안한 방법은 크기, 전장, 중량, 질병 증상 등의 문진표 기반 정보를 포함한 데이터 수집이 많은 입력 파라미터를 가지지만, 많은 수의 입력 데이터가 성능 향상으로 이어지는 않는다는 것을 보여주었다. 그리고 SHAP 분석과 AutoML을 사용하여 입력 파라미터 수를 줄이면서도 성능을 유지하는 최적화된 접근 방식을 제안하였다.

이 방법은 질병 검사 대장 문진표에서 추출된 데이터의 전처리, AutoML을 통한 모델 성능 평가, 각 파라미터의 효과성 평가를 위한 SHAP 분석을 포함하고 입력 파라미터의 최적 조합을 성공적으로 식별하여 텍스트 기반 질병 분류에 대한 보다 효율적이고 실용적인 방법을 제공했다. 그 결과 일정량의 입력 파라미터 수를 줄여도 분류 성능에 부정적인 영향을 미치지 않음을 확인했고 적절한 수치를 그래프로 결정할 수 있도록 하였다. 제안한 방법은 텍스트 기반 분류 시스템, 특히 수산양식 분야의 질병 진단에서 나타나는 입력 파라미터 과다 문제에 대한 실용적인 해결책을 제공한다.

향후 연구로 이 방법론을 보다 광범위한 적용 분야에 맞게 정제하고, 과도한 파라미터 문제에 대한 해결과 성능을 향상시킬 수 있는 방법의 추가 연구가 필요



하다. 또한, 더 많은 수의 데이터를 확보한 상태에서도 동일한 결과가 나타나는지에 대한 확인도 필요하다.

감사의 글

본 논문은 2023년도 호남대학교 학술연구비 지원으로 수행되었음.

References

[1] FAO, "2018 The STATE OF WORLD FISHERIES AND AQUACULTURE," pp. 85, 2018.

[2] Statistics Korea, "2020 Fish Farming Trend Survey Results", <https://eiec.kdi.re.kr/policy/materialView.do?num=211933&topic=O&pp=20>

[3] Fisheries Newspaper, "Increase in Infectious Diseases in Domestic and Imported Seafood in Korea", <https://www.isusanin.com/news/articleView.html?idxno=30689>

[4] National Institute of Fisheries Science, "Fish disease incidence statistics", <https://www.nifs.go.kr/fishguard/disease03>

[5] H. S. Son, H. S. Choi, "Image Augmentation of Paralicthys Olivaceus Disease Using SinGAN Deep Learning Model," *The J. of the Korea Contents Association*, vol. 21, no. 12, 2021, pp. 322-330.

[6] J. Y. Kang, H. S. Son, H. S. Choi, "A Comparison of Pre-Processing Techniques for Enhanced Identification of Paralicthys olivaceus Disease based on Deep Learning," *The J. of the Korea Contents Association*, vol. 22, no. 3, 2022, pp. 71-80.

[7] H. S. Son, H. K. Lim, H. S. Choi, "A Study on Disease Prediction of Paralicthys Olivaceus using Deep Learning Technique," *Smart Media J.*, vol. 11, no. 4, 2022, pp. 62-68.

[8] S. M. Lundberg, S. I. Lee. "A unified approach to interpreting model predictions," In *Proc. Advances in neural information processing systems 30*, 2017.

[9] S. K. Karmaker, M. M. Hassan, M. J. Smith, L. Xu, C. Zhai, K. Veeramachaneni, "Automl to date and beyond: Challenges and opportunities," *ACM Computing Surveys*, vol.

54, no. 8, 2021, pp. 1-36.

[10] P. Geurts, D. Ernst, L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, 2006, pp. 3-42.

[11] L. Breiman, "Random forests," *Machine learning*, vol. 45, 2001, pp. 5-32.

[12] T. Chen, C. Guestrin, "Xgboost: A scalable tree boosting system," In *Proc. the 22nd acm sigkdd int. conf. on knowledge discovery and data mining*, August 2016, pp. 785-794.

[13] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T. Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," In *Proc. Advances in neural information processing systems 30*, 2017.

[14] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, 1996, pp. 123-140.

[15] C. Cortes, V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, 1995, pp. 273-297.

저자 소개

조경원(Kyung-Won Cho)



2018년 뉴캐슬대학교 컴퓨터공학과 졸업(공학사)  
 2021년 광주 과학기술원 지능로봇학과 졸업(공학석사)  
 2023년 호남대학교 컴퓨터공학과 박사과정

※ 관심분야 : 인공지능, 스마트 아쿠아팜 융합 연구

백란(Baik Ran)



1983년 성균관대학교 수학과(이학사)  
 1988년 노스켈로라이나주립대학교 응용수학&컴퓨터과학(이학석사)  
 1996년 노던일리노이대학교 계산수학(이학박사)  
 2002년: 전남대학교 금융공학(박사수료)

1997년 ~ : 호남대학교 컴퓨터공학과 교수  
 ※ 관심분야 : 인공지능, 수치해석(선형), 병렬알고리즘, 데이터분석, 패턴인식