학습된 머신러닝의 표류 현상에 관한 고찰

(A Study on Drift Phenomenon of Trained ML)

신병춘[1], 차윤석[2], 김채윤[3], 차병래[4]

(ByeongChun Shin, YoonSeok Cha, Chaeyun Kim, ByungRae Cha)

요 약

학습된 머신러닝은 시간 경과에 따른 학습 모델과 학습 데이터 측면의 표류 현상이 발생과 동시에 머신러닝의 성능이 퇴화하게 된다. 이를 해결하기 위한 방안으로 머신러닝의 재학습 시기를 결정하기 위한 ML 표류의 개념과 평가 방법을 제안하고자 한다. 딸기와 선명도에 따른 XAI 테스트 및 사과 이미지의 XAI 테스트를 진행하였다. 딸기의 경우 선명도 값에 따른 ML 모델의 XAI 분석의 변화는 미미하였으며 사과 이미지의 XAI의 경우 사과는 정상적으로 객체 분류 및 히트맵 영역을 표시하였으나 사과꽃 및 꽃봉오리의 경우 그 결과가 딸기나 사과에 비해 미미하였다. 이는 사과꽃 및 꽃봉오리의 학습 이미지 수가 부족하기에 발생한 것으로 예상되며 추후 더 많은 사과꽃 및 꽃봉오리 이미지를 학습하여 테스트할 계획이다.

■ 중심어 : 머신러닝 ; 인공지능 ; 표류현상 ; IoU ; XAI

Abstract

In the learned machine learning, the performance of machine learning degrades at the same time as drift occurs in terms of learning models and learning data over time. As a solution to this problem, I would like to propose the concept and evaluation method of ML drift to determine the re-learning period of machine learning. An XAI test and an XAI test of an apple image were performed according to strawberry and clarity. In the case of strawberries, the change in the XAI analysis of ML models according to the clarity value was insignificant, and in the case of XAI of apple image, apples normally classified objects and heat map areas, but in the case of apple flowers and buds, the results were insignificant compared to strawberries and apples. This is expected to be caused by the lack of learning images of apple flowers and buds, and more apple flowers and buds will be studied and tested in the future.

■ keywords : Machine Leaning ; AI ; Drift ; IoU ; XAI

Ⅰ. 서 론

사람은 모르는 문제가 있으면 이를 해결하기 위해 학습과 기억을 하게 되며, 이후 시간이 지나면 모르는 정보는 늘어나게 되며, 그럴 때마다 사람은 이를 해결하기 위해 재학습을 수행하게 된다. 이것은 머신러닝(Machine learning, ML) 모델도 마찬가지이며, 한번 학습된 모델이 최근 테스트 데이터에 관하여 유의미한 성능을 보이더라도 어느 정도의 시간이 경과 후에 테스트 데이터에 대해 동일한 성능이 나온다는 것을 보장하지 못한다. 이것은 ML 모델의 재학습이 필요한 상황이다.

라벨링을 수행하고, ML 모델의 학습 후에 객체 인식 테스트를 수행하게 된다. 그림 1에는 객체 탐지 및 정확히 인식한 결과를 확인할 수 있으나 학습되지 않은 새로운 이미지의 객체에 대해서는 적정한 성능을 보장하지 못할 수도 있다.

[1] 정회원, 전남대학교 수학과 교수 신병춘 [2] 정회원, 제노테크(주) 선임연구원 차윤석 [3] 정회원, 제노테크(주) 연구소장 김채윤
[4] 정회원, 광주과학기술원 AI 대학원 연구부교수 & 제노테크(주) 대표이사 차병래

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (2017R1E1A1A03070059). And this work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry(IPET) through (Advanced Production Technology Development Program), funded by ministry of Agriculture, Food and Rural Affairs(MAFRA)(No.320030-3).

접수일자 : 2022년 07월 15일 게재확정일 : 2022년 08월 22일

교신저자 : 차병래 e-mail : brcha@smartx.kr

그림 1. 자체 개발 모델로 객체인식 된 이미지

머신러닝 모델을 만들기 위해서는 일반적으로 그림 2와 같은 Train-run 절차로 기본적으로 4단계로 진행하게 된다[1,2].

학습된 ML 모델을 검증 및 테스트 데이터를 이용하여 원하는 최적의 성능 달성이 가능하다. 하지만 시간의 흐름에 따른 영구적인 최적의 성능을 달성하지는 못하고 학습 결과가 퇴화하는 현상이 발생하게 된다. 이러한 경우에 ML 모델은 재학습이 필요하며, 성능 퇴화의 원인인 ML 모델의 표류를 탐지하기 위하여 XAI 분석을 적용하여 본다. 본 연구에서는 재학습을 위한 ML 모델의 표류 및 XAI를 이해하고 딸기 분류 모델의 XAI 과정을 살펴본 후 ML Model Factory를 제안한다.

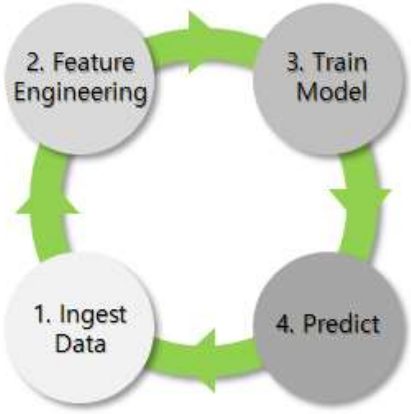

그림 2. ML의 Train-run 절차

Ⅱ. 관련연구

1. ML 모델 최적화와 IoU

ML 모델 최적화를 위한 자동화(Automation) 단계는 그림 3 과 같이 Hand cranking 단계, 하이퍼-파라미터 최적화(Hyper-parameter optimization) 단계, 그리고 AutoML(Automated ML) 단계로 구분한다.

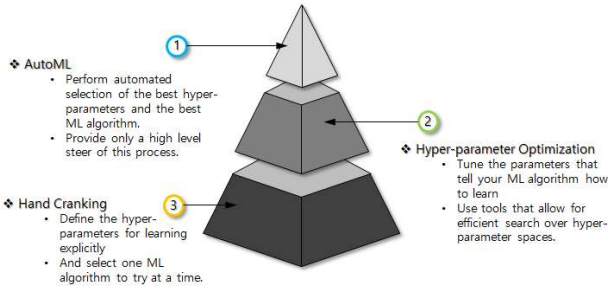

그림 3. ML 모델 최적화의 자동화 단계

Hand cranking 단계는 가장 낮은 수준의 ML 구현에서 훈련 프로세스가 어떻게 진행되기를 원하는지에 대한 세부 정보를 제공하게 되며, 코드에서 훈련 실행에 사용할 정확한 하이퍼-파라미터 세트를 수동으로 정의한다. 하이퍼-파라미터 최적화 단계는 Hand cranking 단계보다 추상화 수준을 한 단계 높여 하이퍼-파라미터에 대한 범위와 경계를 각각에 대한 모델의 성능을 효율적으로 샘플링하고 테스트하도록 설계된 도구를 제공하며, 자동화된 하이퍼-파라미터의 조정(Tuning)이 가능하게 된다. 마지막으로 AutoML 단계는 하이퍼-파라미터 최적화 단계보다 더 높은 수준의 추상화를 제공하며, 제공된 다양한 알고리즘들에 의한 자동화된 ML 최적화가 가능하게 된다[3].

객체 탐지의 IoU((Intersection over Union)는 객체 탐지의 정확도를 측정하는데 이용되는 평가 지표이다. 객체 탐지 알고리즘이 출력한 예측 바운딩 박스는 IoU를 이용해서 평가될 수 있다[4]. IoU를 적용하기 위해서는 그림 4와 같이

Ground-truth bounding boxes(학습을 위한 테스트 데이터셋에서 객체 위치를 라벨링한 것)와 Predicted bounding boxes (ML 모델이 출력한 객체 위치 예측값)의 이 두 가지가 있으면 IoU를 적용할 수 있다.

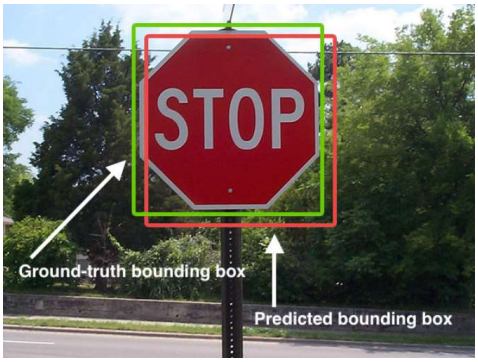

그림 4. IoU의 예제

2. XAI

미국 방위고등연구계획국(DARPA, Defense Advanced Research Projects Agency)는 2016년 XAI(eXplainable Artificial Intelligence) 프로그램을 발표하였다. AI가 일반적인 업무에서 의사결정을 수행하는 등 더 핵심적인 업무로 이동됨에 따라 결정한 결과의 근거와 도출과정의 타당성을 제공하지 못하는 블랙박스 사례가 늘고 있다. 이에 AI가 내린 결정이나 답을 AI 스스로가 사람이 이해하는 형태로 설명하고 제시할 수 있는 화이트박스 기술로 '설명 가능한 AI(XAI)'가 핵심 기술로 대두되고 있다[5,6].

딥러닝을 비롯한 현대 머신러닝 알고리즘은 불투명한 편이다. 그러나 XAI 기법 중 피처 중요도(Feature Importance) 표시나 필터 시각화(Filter Visualization) 기법, LRP(Layer-wise Relevance Propagation)를 이용한 히트맵 분석 등은 불투명한 인공지능의 의사 결정 과정을 설명적으로 개선한다. XAI는 시각화(Visualization)와는 다르다. XAI의 많은 산출물이 시각화 기법에 의존하지만, 머신러닝 모델의 과정을 시각화했다고 해서 모두 XAI라고는 말할 수는 없다. XAI의 핵심은 해석 가능성이다. XAI는 대리 분석(Surrogate Analysis), 부분 의존성 플롯(Partial Dependence Plotsm PDPs), 유사도 분석(Similarity Measure), 피처 중요도(Feature Importance) 등의 기법으로 데이터와 모델을 설명한다[5].

Ⅲ. 머신러닝 드리프트 현상의 고찰

학습된 ML 모델도 최근 테스트 데이터에 관하여 유의미한 결과를 도출한다고 하더라도 이후에 시간이 지난 새로운 테스트 데이터에서도 동일한 결과가 나올지는 미지수이다. 그리고 일정 시간 경과 후에 새로운 테스트 데이터에서 원하는 결과가 나타나지 않는다면 이 모델은 재학습이 필요한 상황이다. ML의 표류 현상은 시간이 지남에 따라 모델의 성능 하락의 다양한 이유를 포괄하는 단어이며 Data drift와 Concept drift의 두 가지 유형으로 나눌 수 있다.

- **Data drift** – Data drift는 사용 중인 변수의 통계적 특성이 변경된 경우 발생한다.
- **Concept drift** – Concept drift는 데이터의 특징과 예측 결과 간의 근본적 관계에 변화가 생길 시 발생한다.

먼저 통계적 속성의 변화에 따른 Data drift의 예를 들면 나이를 특징으로 하여 사용하는 모델이 있는데 학습된 연령대는 10~20대이다. 이후 모델을 사용하는데 10~20대의 청년이 아닌 30대 이상을 수집하게 되면서 Data drift가 발생하게 된다. 그리고 근본적인 관계 변화에 의한 Concept drift의 예를 들면 학습 데이터를 특징과 결과 사이의 선형 관계를 보여주는 데이터 하위 샘플만 사용하여 학습을 진행했는데 학습 후 다른 데이터를 수집해 보니 특징과 결과의 관계가 비선형인 것으로 판명이 되는 것이다. 이를 해결하기 위해선 관계를 더욱 잘 나타내는 새로운 데이터와 모델로 재학습하는 것이다.

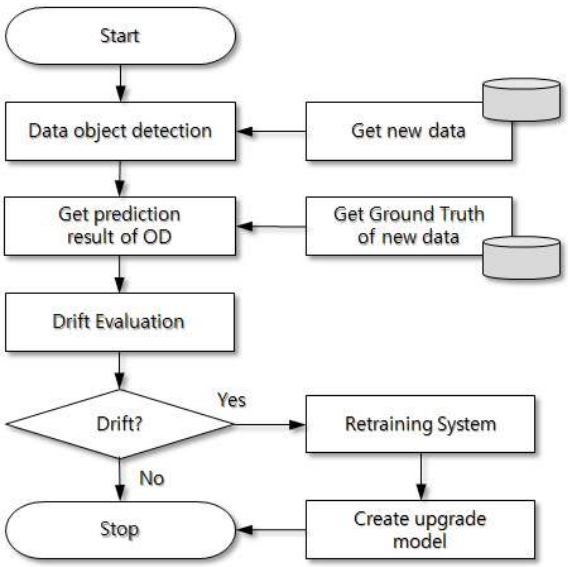

그림 5. ML 표류 탐지와 재훈련 시스템의 플로우-챠트

ML의 재학습을 결정하기 위한 아이디어는 그림 5와 같이 머신러닝의 표류 탐지 과정과 재훈련 시스템이 연계된 플로우-챠트를 나타낸다. Data drift는 통계적 특성에 따른 학습 데이터의 p-value의 계산에 의한 머신러닝의 학습이 표류됨을 예상할 수 있으며, p-value이 낮을수록 관찰된 차이의 통계적 유의성이 커지게 된다. Concept drift는 ML 모델의 표류 문제로 학습 모델과 학습 데이터의 특성 공학(Feature engineering) 측면의 차원 저주(Curse of dimensionality) 문제의 분석이 필요하다. 이러한 ML 표류를 미리 탐지하여 재학습을 할 수 있도록 학습 시스템을 구축한다면 많은 컴퓨팅 자원과 학습 시간의 절약, 그리고 ML 모델의 적정한 성능의 유지가 가능하게 된다.

Ⅳ. 딸기 분류 모델의 XAI 및 ML Model Factory

1. 딸기 분류 모델의 ML Drift 탐지를 위한 XAI 적용

농작물 딸기의 성숙 단계를 분류하기 위하여 먼저 600개의 딸기 사진을 사용하여 딸기 분류 학습 모델을 학습 및 생성을 그림 6과 같이 진행하였다.

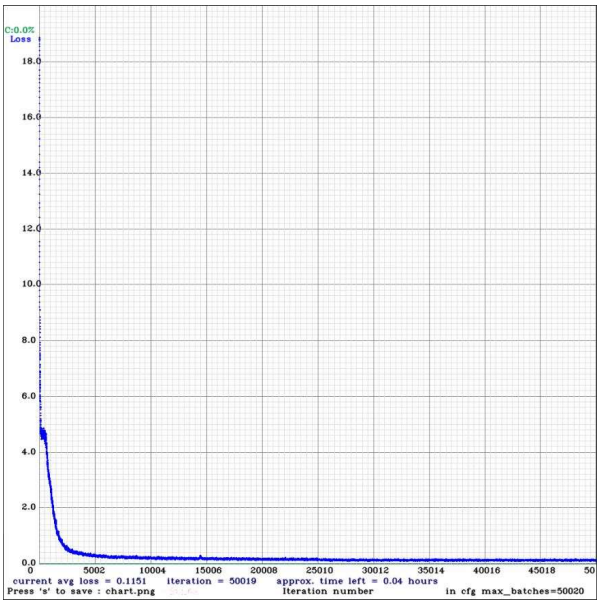

그림 6. 딸기분류 모델의 학습차트

딸기 분류를 위한 학습된 ML 모델을 사용하여 딸기 사진의 객체 인식을 수행하였으며, 그림 7은 객체 인식 및 분류를 위한 딸기 사진의 원본이며, 그림 8는 객체 인식결과를 나타낸 것이다. 객체 인식결과 앞쪽의 딸기만 분류하는 것을 확인하였고 왜 그렇게 분류하였는지 모델이 관심을 가지는 이미지 영역을 XAI 분석의 히트맵으로 그림 9와 같이 표시하였다. XAI 분석의 히트맵의 왼쪽 앞에 있는 딸기를 핵심 영역으로 인식하였으며, 그에 비해 오른쪽 흐릿한 영역은 전혀 관심을 두지 않을 것을 확인하였다.

그림 7. 객체 분류를 위한 딸기 사진 원본

그림 8. 객체 분류의 결과

그림 9. 딸기 분류를 위한 ML 모델의 히트맵 영역

2. ML Factory Model 제안

ML의 단일 모델인 Train-run 절차(그림 1 참조)의 단일 모델에 대해 학습 모델과 학습 데이터 측면의 표류 현상의 극복과 ML의 추상화된 다중 모델을 위한 Model Factory의 Train-persist 절차를 그림 10과 같이 제안한다. Model Factory의 구성 요소들은 재학습 시스템, Model Store, 예측 시스템, 그리고 드리프트 탐지기로 구성된다.

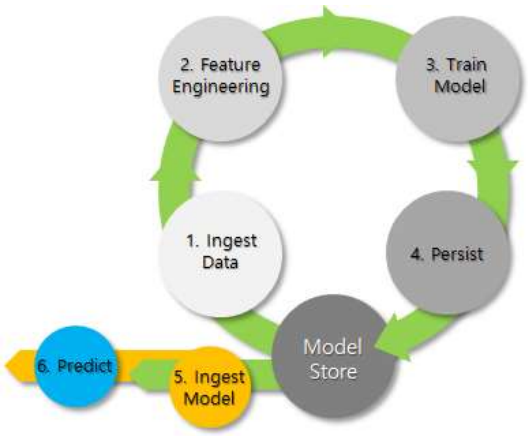

그림 10. ML Model Factory의 Train-persist 절차

V. 딸기와 사과 이미지의 선명도에 따른 XAI 테스트

1. 딸기 분류 모델의 XAI 적용

딸기 이미지의 선명도에 따른 ML 모델의 히트맵 영역의 차이를 확인하기 위하여 선명도를 조정한 딸기 이미지를 객체인식 및 관심을 가지는 이미지 영역을 XAI 분석의 히트맵으로 표현하는 테스트를 진행하였다.

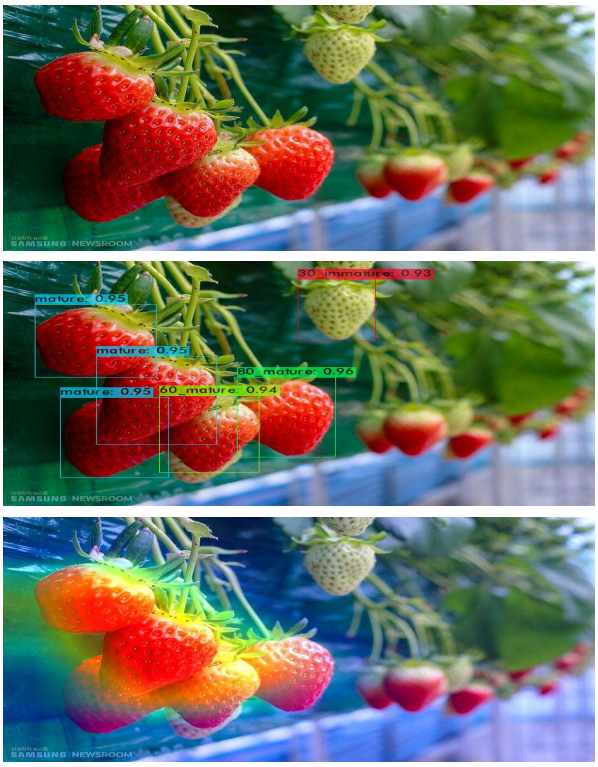

그림 11. 선명도 factor 2의 딸기이미지 원본, 객체 분류 및 히트맵 영역

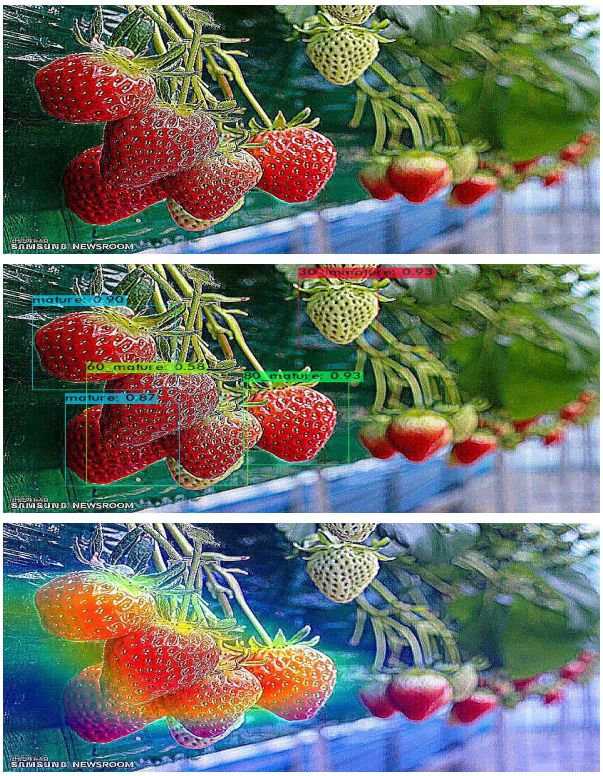

그림 13. 선명도 factor 50의 딸기이미지 원본, 객체 분류 및 히트맵 영역

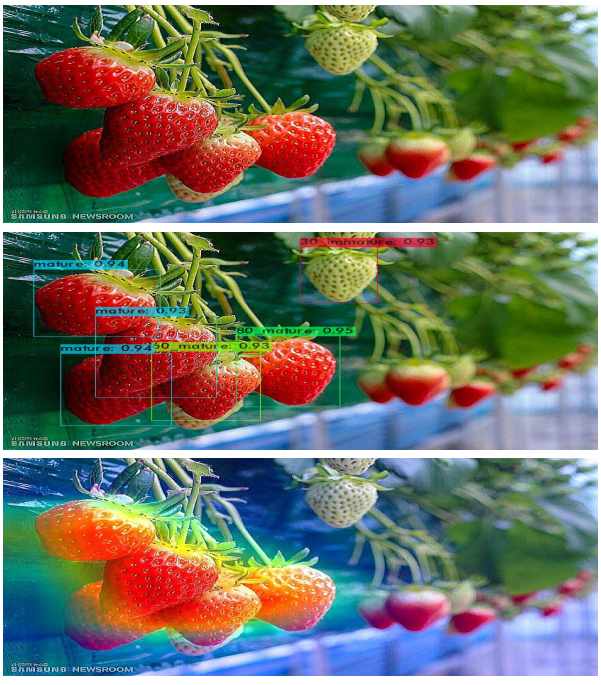

그림 12. 선명도 factor 10의 딸기이미지 원본, 객체 분류 및 히트맵 영역

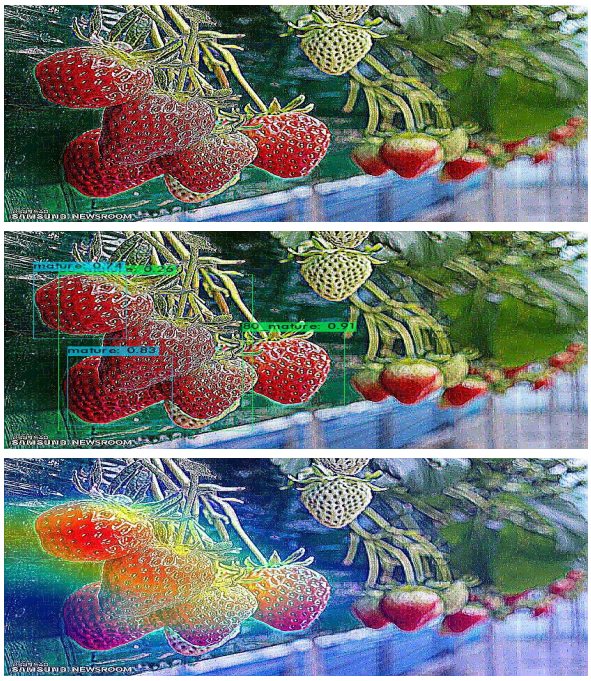

그림 14. 선명도 factor 100의 딸기이미지 원본, 객체 분류 및 히트맵 영역

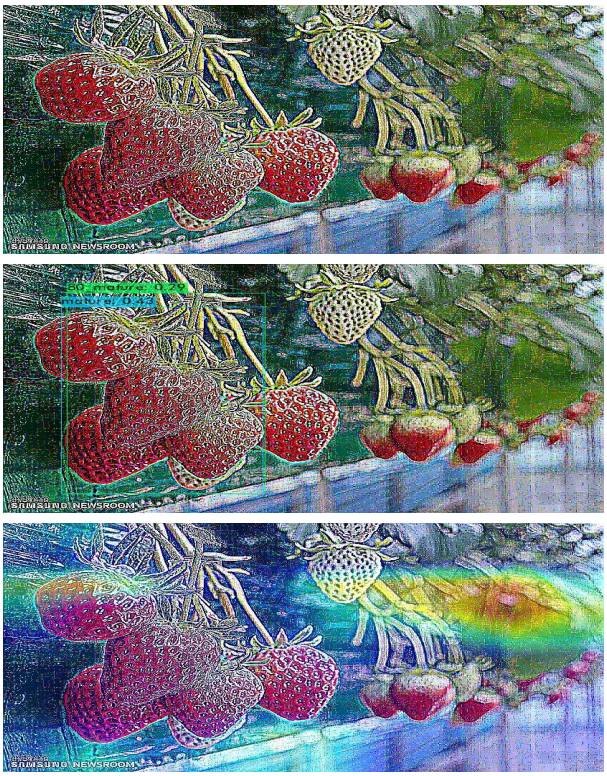

그림 15. 선명도 factor 200의 딸기이미지 원본, 객체 분류 및 히트맵 영역

그림 11부터 그림 15는 차례대로 선명도의 factor를 2, 10, 50, 100, 200으로 증가시킨 딸기이미지의 원본, 객체 분류 및 히트맵 영역이다. factor값이 2~100까지는 히트맵 영역이 비슷하지만 200에서는 모델이 왼쪽영역이 아닌 오른쪽 영역에 관심을 가지고 있는 것을 확인할 수 있다. 그러나 해당 영역의 부분은 열매가 아닌 잎 부분으로 이미지 자체를 잘못 이해하고 있는 것으로 볼 수 있다. 테스트 결과 이미지 선명도의 factor값에 따른 ML 모델의 XAI 분석의 변화는 미미하였으며 factor값이 지나치게 높아지면 ML 모델이 오히려 잘못 인식하게 되는 것을 확인할 수 있었다.

2. 사과 분류 모델의 XAI 적용

이어서 사과 760개, 사과꽃 57개 이미지를 학습한 모델로 객체 인식을 한 사진과 XAI 분석의 히트맵으로 분석하였다. 사과의 경우 학습한 이미지의 수가 많기 때문에 히트맵 표시 부분이 정확하게 사과 쪽을 인식하는 것을 확인할 수 있으나 사과꽃과 사과 꽃 봉우리의 경우 사과 이미지에 비하여 적은 수를 학습해서 그런지 히트맵 부분이 정확하게 표시하지 않는 것을 확인할 수 있다.

그림 16은 사과나무를 객체 분류한 것으로 이미지 앞에 가장 잘 보이는 두 개의 사과가 인식되었으며 히트맵 영역도 두 사과를 중점으로 표시되어있는 것을 확인할 수 있다. 그러나 가지 뒤의 사과나 앞의 두 사과 뒤에 있는 사과의 경우는 인식하지 못하는 것을 확인할 수 있다.

그림 17은 과일상자를 객체 분류한 것으로 일부 가려진 사과를 제외하고는 사과만 정확하게 객체 분류한 것을 확인할 수 있으며 히트맵 영역도 뭉쳐있는 사과를 중심으로 표시되어있는 것을 확인할 수 있다.

그림 18은 사과꽃 및 꽃봉오리 이미지를 객체 분류한 것으로 일부분의 사과꽃 및 꽃봉오리를 객체 분류 및 히트맵 영역을 표시하였다. 그러나 사과꽃의 경우 딸기와 사과와는 다르게 주로 잘 보이는 부분만 제대로 객체 분류하였으며 흐릿한 부분은 객체 분류를 못 한 것을 확인할 수 있었다. 이는 사과꽃이 딸기나 사과에 비해 매우 적은 이미지 수로 학습되었기에 발생한 결과로 예상된다. 추후 좀 더 많은 사과꽃 및 꽃봉오리 이미지를 학습하여 테스트를 진행할 계획이다.

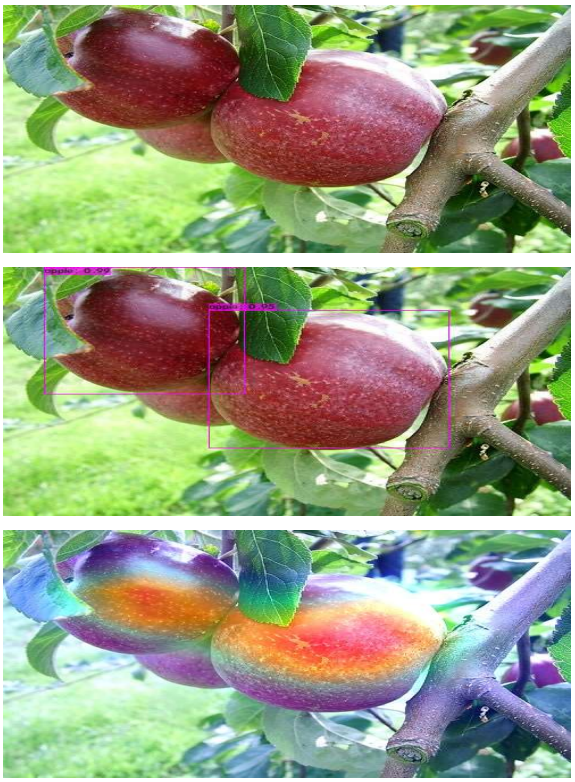

그림 16. 사과나무 이미지 원본, 객체 분류 및 히트맵 영역

그림 17. 과일상자 이미지 원본, 객체 분류 및 히트맵 영역

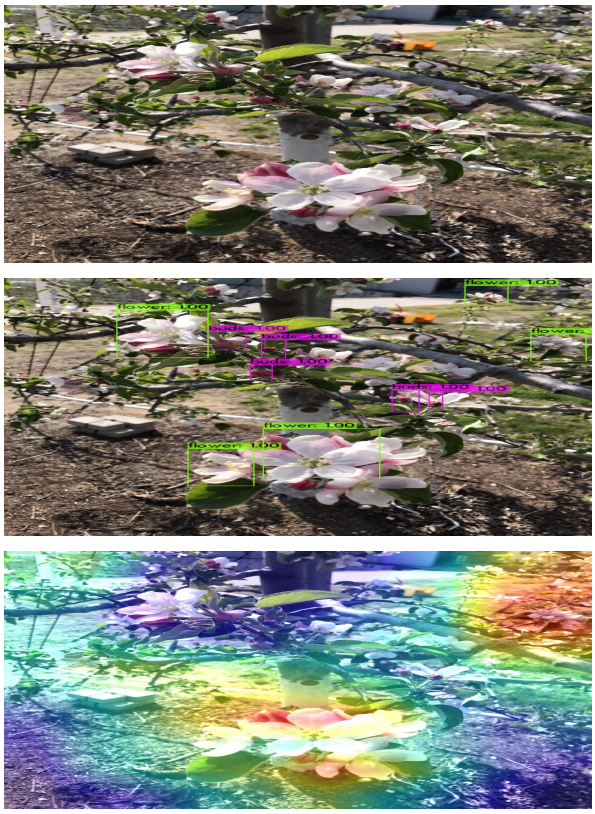

그림 18. 사과꽃 및 꽃봉오리 이미지 원본, 객체 분류 및 히트맵 영역

Ⅵ. 결 론

AI가 일반적인 업무에서 의사 결정을 수행하는 핵심적인 업무로 이동함과 동시에 디지털 전환을 위한 중요한 역할을 수행하고 있다. 그러나 학습된 ML은 시간 경과에 따른 학습 모델과 학습 데이터 측면의 표류 현상에 의하여 ML의 성능이 퇴화하게 된다. 현재 많은 학습 모델은 시간이 경과에 따른 ML drift에 의하여 성능 저하를 피할 수 없다. 이러한 문제 해결을 위한 방법으로 XAI 분석을 적용한다. 더불어, ML의 단일 모델에서 추상화된 다중 모델을 위한 ML Model Factory를 제안하였다.

이어서 딸기와 선명도에 따른 XAI 테스트 및 사과 이미지의 XAI 테스트를 진행하여 딸기의 경우 선명도 값에 따른 ML 모델의 XAI 분석의 변화는 미미하고 사과꽃 및 꽃봉오리의 경우 학습 이미지 수가 부족하면 객체 분류 및 히트맵

영역의 정확도가 줄어드는 것을 확인할 수 있었다. 사과꽃 및 꽃봉오리의 경우 추후 더 많은 이미지를 학습하는 것으로 해결할 수 있을지 테스트할 계획이다.

REFERENCES

[1] Andrew P. McMahon, “Machine Learning Engineering with Python - Manage the production life cycle of machine learning models using MLOps with practical examples,” *Packt Publishing*, 2021.

[2] 차윤석, 박진영, 박선, 김종원, 차병래, "ML 모델의 Drift 탐지를 위한 XAI 분석에 따른 머신러닝 모델 팩토리의 제안," *스마트미디어종합학술대회*, 2022년 6월

[3] Frank Hutter, Lars Kotthoff, Joaquin Vanschroen, “Automated Machine Learning - Methods, Systems, Challenges,” *Springer*, 2019.

[4] IoU, https://deep-learning-study.tistory.com/402 (accessed Jul. 10, 2022)

[5] 안재현, "XAI 설명 가능한 인공지능, 인공지능을 해부하다," *위키북스*, 2020년

[6] NVIDIA KOREA, "설명 가능한 AI란 무엇인가?," https://blogs.nvidia.co.kr/2021/07/27/what-is-explainable-ai/ (accessed Jul. 11, 2022)

저 자 소 개

신병춘

2002년 전남대학교 수학과 조교수
2011년 ~ 현재 전남대학교 수학과 교수
<주관심분야: 수치해석, 인공신경망, 컴퓨터 비전>

차윤석

2014년 고려대학교 컴퓨터정보학과 공학사
2015년 ~ 현재 제노테크(주) 기업부설연구소 선임연구원
<주관심분야: IoT, BigData, Cloud Computing, VoIP, NFC, 대용량 스토리지 기술>

김채윤

2010년 전남대학교 대학원 화학과 졸업(이학박사)
2011년 미국 위스콘신대학교-메디슨 화학과 박사후연구원
2013년 포항공과대학교 미세유체응용화학연구단 박사후연구원
2022년 ~ 현재 제노테크(주) 기업부설연구소 연구소장
<주관심분야: 정보검색, 데이터마이닝, 클라우드 컴퓨팅, IoT, 스토리지 시스템>

차병래

2004년 목포대학교 대학원 컴퓨터공학과 졸업(공학박사)
2005년 호남대학교 컴퓨터공학과 전임강사
2009년 광주과학기술원 전기전자컴퓨터공학부 연구부교수
2020년 ~ 현재 광주과학기술원 AI 대학원 연구부교수
2012년 ~ 현재 제노테크(주) 대표이사
<주관심분야: 정보보안&IDS, ANN, Cloud Computing, VoIP, NFC, SDS 등>