

수경재배 환경에서 머신러닝 기반 불량 모종 진단을 위한 증강 데이터 활용 연구

이 규 진*

A Study of using Augmented Data with Machine Learning for Growth Detection Under Hydroponic Conditions

Kyujin Lee*

이 논문은 ㈜교원프로퍼티의 데이터 사용 허가를 통해 수행된 연구임

요 약

스마트 농업 분야에서 빅데이터 분석 기술 및 인공지능을 활용한 연구의 필요성이 지속적으로 요구되고 있다. 인공지능 연구를 위해서는 대용량의 고품질 데이터 확보 및 라벨링이 필수적이나, 수경재배 영역에서의 데이터가 상대적으로 부족한 현황이다. 이에 따라 본 논문에서는, 향암쌈채 모종의 원본 생육 데이터의 머신러닝 분류 모델 적용 시 효과를 확인한 후, 증강 데이터를 활용한 분류 모델 개발 및 성능개선을 실험하였다. 실험 결과, DCGAN 증강 데이터만으로 학습한 모델은 ResNet 61.2, DenseNet 62.4의 Accuracy로 실제 향암쌈채에 대한 진단이 이루어졌다. 그리고, 원본 데이터에 증강 데이터를 추가한 성능개선 실험 결과, 원본 데이터만으로 학습된 모델 중 ResNet은 86.5에서 88.2로 DenseNet은 92.9에서 94.7로 향상되었다. 이 연구를 통해 부족한 수경재배 관련 데이터에 대한 증강 데이터의 활용 가능성과 그 영향력을 확인하였다.

Abstract

The need for big data analysis and artificial intelligence for smart agriculture is continuously requested. It is essential to collect and label large amounts of quality data for artificial intelligence research, but there is a relative lack of data in the hydroponic cultivation area. In this paper the performance of growth diagnostic ML model with original anti-cancer leaf lettuce data set was checked. Then ML model development and model performance improvement experiment was proceeded with augmented data. First, ML model learning and testing was conducted using DCGAN data only. As a result, the accuracy of ResNet was 61.2, DenseNet was 62.4. And model performance improvement experiment was proceeded by adding augmented data to the original data. As a result, the accuracy of ResNet increase from 86.5 to 88.2 and DenseNet from 92.9 to 94.7. In these experiments the possibility of using augmented data and its influence are studied for developing and improving the performance of diagnostic ML model.

Keywords

hydroponic conditions, anti-cancer leaf lettuce, machine learning, deep learning, data augmentation

* 국민대학교 데이터사이언스학과 박사 수료
- ORCID:<https://orcid.org/0009-0007-1570-8596>

• Received: Aug 29, 2023, Revised: Oct. 12, 2023, Accepted: Oct. 15, 2023
• Corresponding Author: Kyujin Lee
AI-DX Division, Kyowon, Eulji-ro 51, Jung-gu, Seoul, Korea
Tel.: +82-2-397-1616, Email: anfang@naver.com

1. 서 론

식물 질병은 복잡하고 가변적이며 식물 성장에 큰 영향을 미친다. 식물 질병 통제의 핵심은 적시에 정확하게 해당 질병을 감지하는 것으로, 질병의 손상 부위와 유형을 식별해야 한다. 식물 질병의 주요 원인 중 하나로 990종 이상의 식물 바이러스가 존재하며, 바이러스, 진균 또는 생리학적 병변에 감염된 식물은 변형, 퇴색, 꼬임, 썩음 등 다양한 현상을 발생시킨다. 하지만 증상이 뚜렷하게 구별되지 않고 복잡성 및 유사성, 기술, 문화 등의 원인으로 모든 지역에서 식물의 질병을 정확하게 진단하고 관리하는 것은 매우 어렵다[1].

이에 대한 해결책으로 ICT 기술을 활용한 스마트 농업에 대한 관심이 증가하고 있으며, 인공지능을 활용한 연구가 활발하게 진행되고 있다. 스마트 농업에 인공지능을 도입하기 위해서는 학습에 필요한 빅데이터가 요구되기에, 국내의 경우 관계부처 합동으로 빅데이터/인공지능 기반 스마트 농업 확산 종합대책(안)을 마련하고 관련 인프라 구축 및 스마트 농업 거점 육성 등의 노력을 병행하고 있다[2]. 하지만, 이렇게 연구용으로 공개된 데이터는 주로 스마트 온실 작물 및 노지 작물로 구성되는데, 이는 국내 스마트 농업 산업이 시설 원예, 과수원, 노지 분야에 집중되어 있고 환경, 시설, 제어를 위한 센서 기반의 생장 환경 관리 및 시설물 제어를 통해서 스마트농업의 목적을 달성하는 것이 주된 목표이기 때문이다[3]. 반면 스마트팜 기술을 적용한 수경재배의 경우 생육 환경 정보 및 작물 이미지 등의 데이터를 기반한 연구가 진행되어야 하지만, 초기 환경 구축 비용과 양액 문제로 인해 농가사업으로 확산되지 않았으며 공개된 데이터가 상대적으로 부족하다[4].

하지만, 2019년 말부터 시작된 COVID 19 장기화에 따라 실내에서 농작물을 재배하는 것이 새로운 취미 생활로 주목받기 시작했으며, 식물 재배기를 활용한 수경재배에 대한 관심도 자연스럽게 높아지고 있다. 국내에서는 2020년을 기점으로 더 많은 기업들이 식물재배기를 출시하고 있으며, 출시된 식물재배기용 수경 작물들은 스마트팜 팩토리에서 채소 모종 단계부터 생산하고 있다[5]. 이러한 수경재배 역시 생육 환경 정보 및 작물 이미지 등의 데이터

를 기반한 인공지능 연구가 필요한 부분이나, 공개된 데이터가 절대적으로 부족하여 수경재배 식물의 데이터 확보 및 인공지능을 활용한 연구에 한계가 있다[3][4]. 머신러닝 모델을 실제 산업 및 연구에 적용하기 위해서는 모종의 생육 특징을 학습시킬 수 있는 큰 규모의 데이터 수집이 필요하나, 데이터 수집 기간, 전문가 라벨링에 드는 시간 및 비용, 모종 불량률의 불균형 비율 등의 이유로 충분한 데이터 확보가 어려운 현실이다.

위와 같은 상황을 고려하여 본 연구에서는, 테스트 베드를 활용한 선행 연구[6]에서 확보된 향암쌈채 모종 원본 이미지 데이터가 머신러닝 기반의 불량 모종 질병 분류에 적용 가능한지 여부를 확인한 후 원본 데이터를 생성AI 알고리즘 중 DCGAN을 이용해 생성한 증강 데이터의 활용 가능성을 확인하는 실험을 진행하였다. 본 논문의 3.1 장에서는 선행연구를 통해 수집된 향암 쌈채 모종의 14일차 이미지 원본 데이터를 활용하여 이미지 분류 Backbone 모델을 적용한 불량 모종 진단 실험을 진행하여 원본 데이터의 머신러닝 모델 개발 가능성을 확인한다. 이후 원본 데이터를 생성AI를 통해 증강하여 머신러닝 기반의 학습을 진행하고 불량 모종 진단 모델에의 사용 가능성을 평가하는 실험을 진행하였다. 먼저 수집된 데이터를 DCGAN을 통해 학습, 이미지 데이터를 생성하고 원본 클래스와 유사한 특징을 가진 데이터를 전문가가 육안으로 선별하여 증강 데이터 세트를 확보하였다. 3.2 장에서는 선별된 DCGAN 생성 데이터를 바탕으로 향암쌈채 질병 분류 모델 개발 및 성능개선 시 증강 데이터의 활용 가능성을 확인하기 위한 2개의 실험을 진행하였다. 먼저 DCGAN 증강 데이터를 이미지 분류 Backbone 모델에 학습 데이터로 활용하여 모종 불량 진단 정확도를 측정하여 질병 분류가 가능한 것을 확인하였다. 이후 DCGAN 생성 데이터와 원본 데이터를 함께 학습한 모델을 확보한 후, 원본 테스트 데이터를 활용한 향암쌈채모종의 불량 진단 정확도를 측정하는 실험을 진행하였다. 해당 연구결과, 상대적으로 데이터가 부족한 수경재배 영역에서 증강 데이터 기반의 머신러닝 분류 모델의 개발 가능성과 증강 데이터 적용 시 분류 모델의 성능 향상을 확인하였다.

II. 관련 연구

2.1 식물 성장 데이터 수집 관련

식물 성장 및 질병 통제의 핵심은 시기에 맞춰 정확하게 생육 현황을 확인하고 질병을 감지하는 것으로, 질병의 손상 부위와 질병의 유형을 식별해야 한다. 이를 위해서 식물이 제대로 성장할 것인지 예측하거나 질병의 발생을 빠르고 쉽고 정확하게 감지하는 방법이 필요하다. 인공지능을 활용한 식물 성장 및 질병 예측 연구가 활발히 진행됨으로 인공지능 학습에 필요한 데이터 세트의 확보는 무엇보다 중요하다. 데이터 세트는 머신러닝 모델을 구축하기 위한 기초 Input으로 고품질 데이터 세트는 모델 학습결과를 향상시키는 경향이 있다. 또한 신경망의 품질과 예측 정확도도 효과적으로 향상될 수 있다. 네트워크 모델의 일반화 능력, 더 높은 품질의 모델은 식물 질병의 종류와 피해 정도를 정확하게 구별하고, 과학적인 통제 조치를 결정할 수 있게 한다.

S. Mingyue et al.의 연구[1]에서는 식물 질병과 관련하여 공개적으로 사용 가능한 식물 질병 이미지 데이터 세트 및 사이트를 소개하였다. 공개 데이터 세트 프로젝트인 PlantVillage, 식물병리학 관련 CVPR 2020-FGVC7 등 12종 이상의 식물 질병 이미지 데이터 세트와 웹사이트들을 소개하여 딥러닝 모델 구축을 위한 기초자료로 제안하였다. 또한, 딥러닝을 포함한 머신러닝 모델을 활용한 식물 질병 판정과 관련된 연구들을 정리하여 향후 발전 방향에 대해 제시하였다. 연구보다는 조사에 가까워 한계가 있으나, 머신러닝 적용을 위한 식물 질병 데이터 세트 및 관련 연구들을 모아 놓은 것에 의미가 있다.

E. Kim의 연구[7]에서는 RGB-D 센서를 통해 획득한 RGB영상과 Depth영상 데이터를 확보하였고, 이를 이용하여 사과의 픽셀 좌표와 깊이 센서를 통해 얻은 3차원 깊이 정보를 병합한 데이터로 과실 비대 예측을 진행하였다. 영상 내 사과 객체의 개수, 검출 상자의 위치를 찾기 위해 EfficientDet D2 객체검출 모델을 적용하였다.

S. Hong의 연구[8]에서는 농정원 수집 스마트 팜 농가 데이터(토마토)에 기반, 다중 회귀 분석, 랜덤 포레스트, 딥러닝 ConvLSTM을 적용한 성장량 예측과 생산량 예측을 수행, MAE, RSME, R2 등 지표를 통해 가장 좋은 성능의 모델을 선정 제안하였다. 스마트팜 코리아의 빅데이터(토마토, 파프리카, 오이 등 7가지 작물 생육정보, 환경정보, 경영정보 구축)에서 생육정보와 환경정보가 동시에 존재하고, 경작 기간에 빈 데이터, 환경정보 누락분이 없는 전체 245개의 농가 중 4개의 농가 데이터를 활용하였다. 작기 별 외부 환경 요인이 매우 달라, 일관된 조건하에서 분석하기 위해 작기가 최대한 겹치는 기간의 데이터를 선택하였으며, 환경 데이터의 경우 시간 단위 측정되어, 주간의 평균, 최소, 최대값을 구하여 각 환경 정보에 대한 데이터를 생성하였다. 학습 데이터는 주간 성장량 및 생산량, 환경정보가 포함되어 있는 데이터로 최종 구성하여 모델학습 및 성능 평가를 수행하였다. ConvLSTM 모델의 성능이 총 5번중 3번 성능이 높았고, 나머지 2번은 MLR모델의 성능이 더 좋은 것으로 확인하였다. 머신러닝을 적용한 생산량 및 성장량 예측을 진행하여 그 가능성을 확인했으나, 농업데이터의 수집 과정에서 일반 농가에서 수집된 데이터를 사용함에 따라 데이터의 일관성이 부족했다는 한계가 있었다.

2.2 이미지 증강 관련

C. Shorten and T. M. Khoshgoftaar는 논문[9]에서 Deep learning을 위한 Image data augmentation (데이터 증강)에 대한 연구를 통해 다양한 기법을 소개하고 Data Augmentation을 위한 설계 고려사항을 제시하였다. 논문에서 정의한 내용에 따르면, 'Data Augmentation은 데이터의 양을 늘리기 위해 원본에 각종 변환을 적용하여 데이터의 개수를 증강시키는 기법으로, Image Augmentation은 크게 Image Manipulation 기법과 Model Based 기법으로 나눌 수 있다. Image Manipulation은 Horizontal flipping, Color space augmentations, Random cropping 등과 같은 단순한 방법과 Geometric transformations, Kernel filters, mixing images, Random erasing, Feature space

augmentation와 같이 복잡한 방법으로 적용할 수 있다. Model Based Data Augmentation은 딥러닝 모델을 활용한 증강 방법으로, Adversarial training, GAN based augmentation, Neural Style Transfer와 Meta-learning schemes 이 있다. Optimal Strategy는 최적의 Manipulation의 조합 정책을 찾아내는 방법으로 오버헤드에 대한 부담을 줄이고자 이미지 회전, 뒤집기와 같은 기본적인 증강 기법 중에서 최고의 성능 조합을 자동으로 찾아주는 기법이다.

Ranford et al.는 논문[10]에서 GAN의 Generator와 Discriminator 구조에 CNN을 적용한 DCGAN(Deep Convolutional Generative Adversarial Networks)을 제안하였다. GAN의 우수한 아이디어에도 학습 구조의 불안정과 Neural Network가 가지는 블랙박스의 한계가 있었다. 이의 개선을 위해 이미지의 특징을 추출하는데 특화된 CNN 모델 구조를 도입하였는데, Convolution network를 사용해 Discriminator를 구성하고 Deconvolution network로 Generator를 구성한 GAN이다. 이를 통해 DCGAN이 학습한 Filter를 Visualize하여 이미지 생성에 미치는 영향을 파악할 수 있으며 GAN 대비 높은 성능을 보였다.

J. Park et al.은 클래스 분포가 불균형한 데이터 확보 상황에서 토마토 질병분류기의 성능을 향상시키기 위해, DCGAN을 활용하여 데이터를 증강하고 이에 대한 영향을 평가하는 실험을 진행하였다[11]. DCGAN기반의 이미지 증강의 영향력을 확인하기 위해 다양한 클래스 불균형 상황 및 데이터 확장 환경에서 토마토 질병분류 모델의 성능을 검증하였다. 먼저Plat-disease식물질병 데이터 세트의 토마토 질병 데이터 4종류와 healthy 이미지 데이터를 클래스별로 4,000장을 확보한 Base 실험을 진행한 후, 램덤 오버샘플링 및 하나의 질병에 대해 1/10 비율로 데이터의 양을 줄인 Imbalance 상황을 가정하여 실험하였다. 마지막으로, DCGAN모델을 통한 데이터 확장 실험을 위해 Imbalance 데이터 400장을 학습하여 3,600장의 fake 이미지를 생성하여 클래스 데이터의 균형을 맞춘 실험을 진행하였다. 연구에서 데이터 세트의 클래스 불균형 시에 CNN기반 분류기 성능이 저하되며, 오버샘플링 및 DCGAN을 활용한 데이터 증강을 통한 성능 향상을 확인하였다.

특히 해당 연구에서는 DCGAN을 이용한 이미지 데이터 확장을 통해 최대 30%의 정확도가 높아짐을 확인하였다.

2.3 인공지능을 활용한 식물 생장 연구

G. Geetharamani and A. Pandian의 연구[12]에서는 심층 컨볼루션 신경망(Deep CNN)을 기반으로 한 새로운 식물 잎 질병 식별 모델을 제안하였다. 39가지 종류의 식물 잎과 배경 이미지를 가진 개방형 데이터 세트를 사용하여 훈련을 진행하였고, 이미지 플립, 감마 보정, 노이즈 주입, 주성분 분석, 색상 확대, 회전 및 스케일링의 6가지 유형의 데이터 증강(Data augmentation) 방법을 사용하였고, data augmentation을 사용하면 모델의 성능을 높일 수 있다고 제안하였다. 시뮬레이션을 통해 제안된 모델은 96.46%의 분류정확도를 달성하였고 전통적인 기계 학습의 접근법보다 높은 정확도를 나타낸 것으로 확인하였다.

J. K. Kim et al.는 영상처리 기법과 딥러닝 기술을 사용한 채소 등급 자동 분류를 위한 연구를 진행하였다[13]. 농가의 오이를 동일 배경하에서 촬영하여 이미지 데이터 세트를 확보하고 데이터 증강 기법을 통해 학습 데이터 세트를 확보하였다. SVM과 CNN, VGGNet 등을 사용하여 오이를 3개 등급으로 분류하기 위해 머신러닝 알고리즘의 파라미터, 하이퍼파라미터를 변경시켜 더 좋은 성능을 확보하는 연구를 진행하였다.

J. Chen et al.는 식물 잎 질병의 식별을 위한 딥러닝 모델 적용의 효과성에 대한 연구를 진행하였다[14]. 이를 위해 ImageNet 및 Inception 모듈에서 사전 훈련(Pre-trained)된 VGGNet을 적용하였다. 가중치를 무작위로 초기화하여 처음부터 학습을 진행하는 대신, 레이블이 지정된 대규모 ImageNet 데이터 세트에서 사전 훈련된 네트워크의 가중치를 적용한 전이 학습(Transfer learning)을 진행하였다. 연구에서 제안된 방식으로 공개 데이터 세트에 적용하여 91.83%의 검증 정확도를 확인하였고, 복잡한 배경을 가진 벼 이미지의 분류 예측의 평균 정확도도 92.00%로 높게 측정되었다.

III. 불량 모종 진단을 위한 증강 데이터 활용 연구

농작물의 안정적인 생산을 위해, 스마트 농업 분야에도 빅데이터 분석 기술 및 인공지능을 활용한 연구의 필요성이 부각되고 있다. 인공지능 연구를 위해서는 대용량의 고품질 데이터 확보 및 라벨링이 필수적이나, 수경재배 영역에서의 데이터가 상대적으로 부족한 현황이다.

3.1 증강 데이터 확보

본 논문에서는, 수경재배 식물의 일종인 향암쌈채 모종 원본 데이터로 학습한 인공지능 분류 모델의 성능을 먼저 확인하고, 증강 데이터를 활용한 모델 개발 및 성능개선에 대한 실험을 통해 부족한 수경재배 데이터의 해결 가능성을 확인하려 하였다.

먼저 본 연구 진행을 위해 원본 향암쌈채 생육 모종 데이터를 확보하였다. 데이터는 선행연구[6]의 실험을 통해 확보된 것으로 수경 재배 시설의 환경과 동일한 조건으로 구축된 테스트 베드에서 수집된 생육 Cycle 14일차 이미지 데이터로, 현장전문가가 모종 생육상태를 판단하여 라벨링 하였으며 302*290 사이즈, 총 833 장으로 최종 Leaf lettuce healthy(정상) 200개, Leaf lettuce late budding(발아불량) 179개, Leaf lettuce not settled(들뜬 묘) 193개, Leaf lettuce poor growth(생육불량) 166개, Leaf lettuce shrink(오그라들) 95개의 5개 클래스로 구분된다. 확보된 데이터는 학습용 8.0과 테스트용 2.0 정도의 비율로 분리하여 준비하였다.

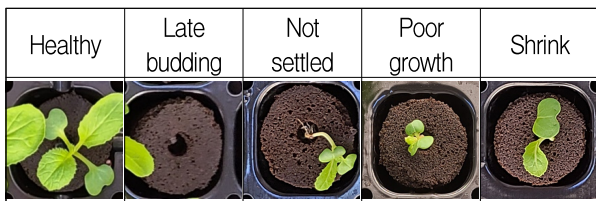


그림 1. 수집된 모종 잎 이미지
Fig. 1. Collected leaf lettuce images

먼저 수집된 향암쌈채 이미지 데이터의 머신러닝 적용 가능성을 확인하기 위해, 대표적인 Backbone

분류 모델인 ResNet, DenseNet에 이미지 데이터를 학습시켜 5개 질병 클래스에 대한 불량 모종 진단 실험 (300epoch, learning rate 0.001 기준)을 진행한 결과, ResNet 86.5, DenseNet 92.9의 높은 값으로 모종 분류 정확도가 측정되었다. 학습 데이터의 양이 적고 클래스간 불균형이 존재함에도 높은 성능을 나타낸 것으로 판단되며 이는 확보된 향암쌈채 모종 이미지 데이터가 머신러닝 학습에 적합한 수준의 품질을 가지고 있다는 것을 의미한다.

이어서 해당 원본 데이터를 활용한 실험을 진행하였다. 머신러닝 분류 모델의 성능을 높이기 위해서는 학습 데이터의 품질 및 클래스별 충분한 데이터 확보가 매우 중요하다. 하지만 동일한 환경하에서의 데이터 수집에는 전문가의 참여, 비용 및 시간 소요 등의 한계가 있어 충분한 양의 데이터 확보가 쉽지 않다. 향암쌈채 모종 불량률은 10% 수준으로 클래스 간 불균형이 높아 현재의 원본 데이터 수량을 확보하는 데에도 어려움이 있었다. 이러한 부분의 해결 가능성을 확인하기 위해 확보된 향암쌈채 원본 이미지 데이터를 증강하여 머신러닝 모델 개발 및 성능 개선을 평가하는 연구를 진행하였다.

향암쌈채 모종 이미지 생성에는 DCGAN을 사용하였다. 본 논문에서는 확보된 향암쌈채 원본 데이터 세트를 학습시켜 각 클래스 별로 증강을 진행하는 BinaryGAN 형태로 진행하였다. DCGAN으로 생성한 이미지는 각 클래스 당 3,600장으로, 원본 이미지와 유사한 특징을 가진 데이터를 1인의 전문가가 육안으로 관찰하여 선택하는 방식으로 모종불량 클래스 별로 200장의 데이터를 선별하였다.

3.2 증강 데이터 기반 분류 모델

선별된 DCGAN 증강 데이터를 활용한 향암쌈채 질병 분류 모델 개발 및 모델 성능 개선 가능성을 확인하기 위해 2개의 실험을 진행하였다. 첫 번째 실험은 Backbone 분류 모델을 증강 데이터만으로 학습한 후 원본 테스트 데이터에 대한 불량 모종 진단 정확도를 측정하는 실험을 진행하였다. 두 번째 실험은 DCGAN 생성 데이터와 원본 데이터를 합하여 Backbone 모델에 학습하고 원본 테스트 데

이터로 검증하는 실험을 하였다. DCGAN 생성데이터로 학습한 분류 모델의 Pre-trained 모델 활용 가능성과 원본 데이터가 부족한 경우 생성 데이터를 이용한 성능 개선이 가능한지를 확인하려 하였다.

첫 번째, 항암쌈채 모종 불량 클래스 별 DCGAN 증강 이미지 200장, 총 1,000장의 데이터를 ResNet, DenseNet의 Backbone 모델에 적용하여 300 epoch, 0.1, 0.001, 0.001의 Learning rate에서 학습한 후 원본 테스트 세트로 검증하는 실험을 하였다.

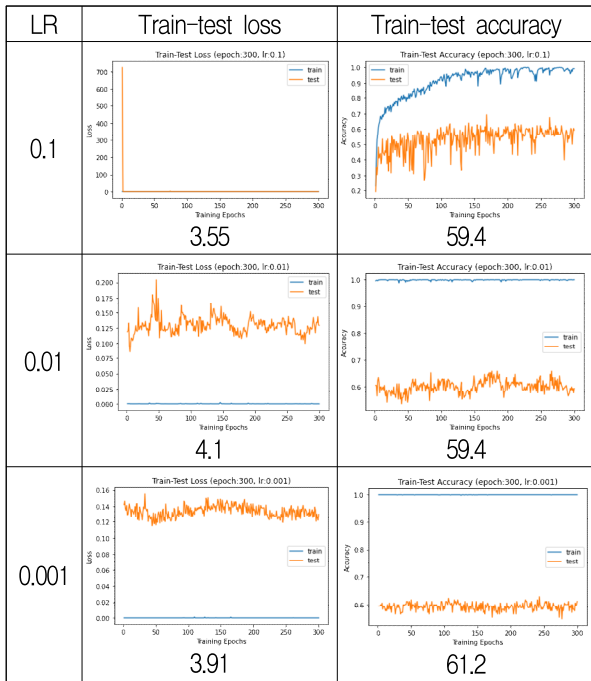


그림 2. ResNet의 손실과 정확도 그래프
Fig. 2. Loss and accuracy of ResNet

ResNet에 증강 데이터를 학습시킨 후, 원본 항암쌈채 데이터에 대한 불량 모종 진단 실험 결과 Learning rate 0.001에서 loss 3.91, accuracy 61.2로 가장 높게 측정되었다. DenseNet은 Learning rate 0.1에서 loss 1.34, accuracy 62.4로 가장 높게 측정되었다. 실험 결과, DCGAN 증강 이미지만으로 학습한 모델은 ResNet 61.2, DenseNet 62.4의 Accuracy로 실제 항암쌈채 불량 진단이 이루어졌다. 수경재배 항암쌈채류와 관련된 데이터가 부족한 상황에서 DCGAN 생성 데이터 세트와 Pre-trained 모델을 활용한 다양한 연구가 가능할 것으로 판단된다.

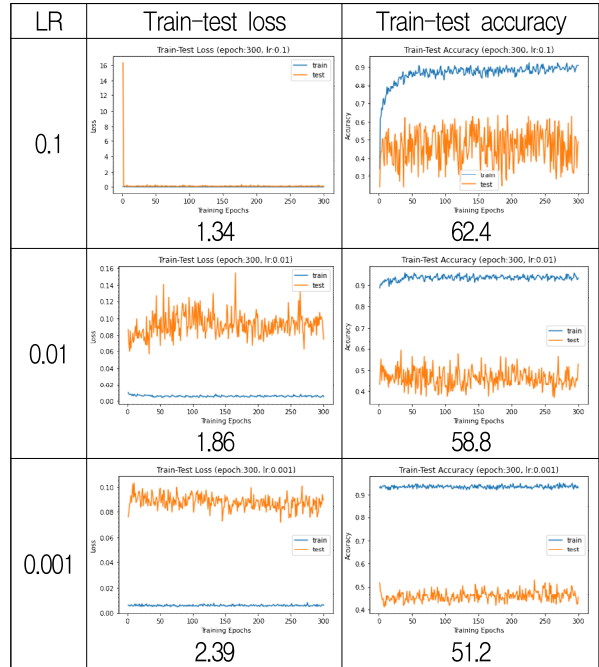


그림 3. DenseNet의 손실과 정확도 그래프
Fig. 3. Loss and accuracy of DenseNet

3.3 증강 데이터 활용 모델 성능 개선

DCGAN 생성 데이터의 활용 측면에서, 원본 데이터에 생성 데이터를 추가하였을 때의 모델 성능 개선에 대한 실험을 진행하였다. 증강 데이터의 영향을 파악하기 위해, 학습 데이터 구성 시 원본 대비 증강 데이터의 비중이 큰 경우와 적은 경우를 Test Case로 한 실험을 진행하고 이를 비교하였다.

표 1. 원본 vs 증강데이터 반영 건수
Table 1. Real vs augment data counts

	Case 1. 60% ratio			Case 2. 25% ratio		
	Orig.	Aug.	Sum	Orig.	Aug.	Sum
healthy	159	200	359	159	50	209
late budding	142	200	342	142	50	192
not settled	155	200	355	155	50	205
poor growth	133	200	333	133	50	183
shrink	74	200	274	74	25	99

ResNet, DenseNet의 Backbone 모델에 300 epoch, 0.1, 0.001, 0.001의 Learning rate 하에서 2가지 Case의 학습 데이터 세트에 대한 학습 및 원본 테스트 세트에 대한 불량 모종 진단 실험을 진행하였다. 먼저 ResNet에 대한 실험 결과, Case 1의 learning rate 0.001에서 Accuracy가 88.2로 가장 높게 측정되었다.

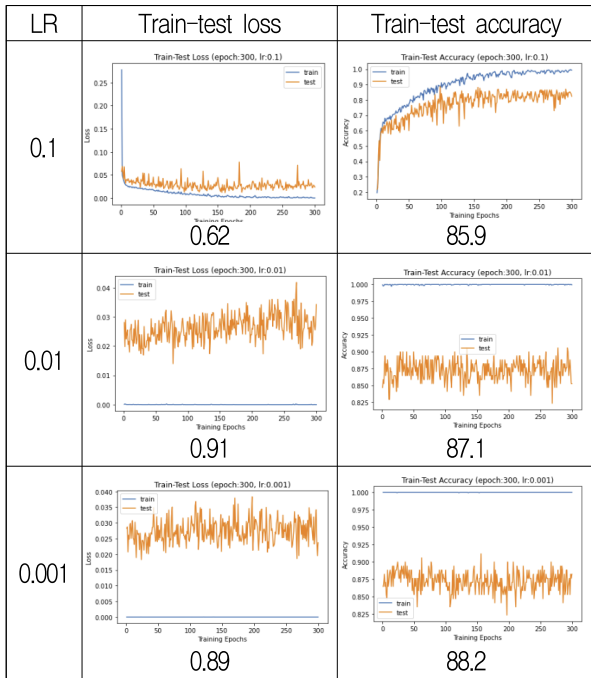


그림 4. Case 1에 대한 실험 결과 (ResNet)
Fig. 4. Test result of case 1 (ResNet)

DenseNet에 대한 실험 역시 동일한 설정으로 진행되었으며, Case 1의 learning rate 0.001에서 94.7로 가장 높은 Accuracy가 측정되었다.

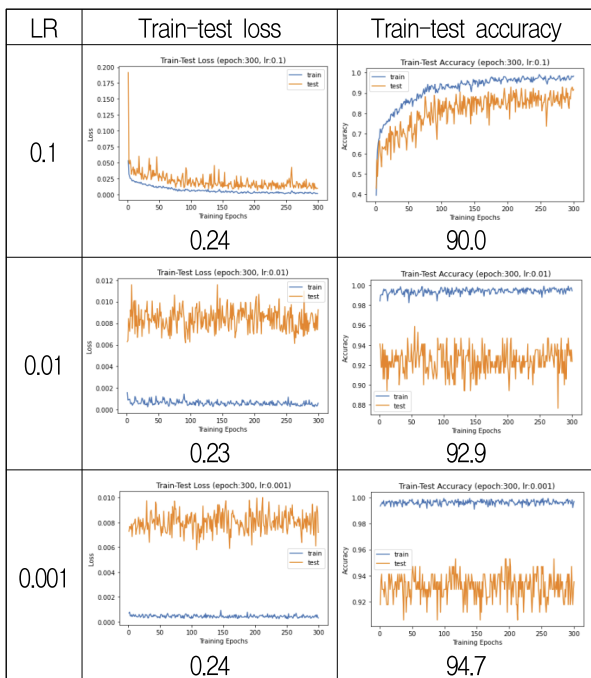


그림 5. Case 1에 대한 실험 결과 (DenseNet)
Fig. 5. Test result of case 1 (DenseNet)

DCGAN 생성 데이터를 학습 데이터로 추가하여 실험한 결과는 아래 표 2와 같다. Case 별로 Accuracy 측정값의 변화가 있었는데, DCGAN 증강 데이터의 품질이 떨어짐에도 적용 비중이 높은 경우에 더 높은 성능 향상이 발생하였다. 원본 데이터만으로 학습이 진행된 경우와 비교했을 때, ResNet은 86.5에서 88.2로 DenseNet 92.9에서 94.7로 향상되었다.

표 2. Case별 모델 성능 변화

Table 2. Changes of model performance by case

Case	Model	ResNet		DenseNet	
		LR	Loss	Acc.	Loss
Case 1	0.1	0.62	85.9	0.24	90.0
	0.01	0.91	87.1	0.23	92.9
	0.001	0.89	88.2	0.24	94.7
Case 2	0.1	0.97	78.8	0.32	88.2
	0.01	1.24	80.0	0.32	89.4
	0.001	1.43	78.8	0.34	88.2

IV. 결론 및 향후 과제

작물의 안정적인 생산을 위해 ICT 기술 기반의 스마트 농업에 대한 관심의 증가 뿐 아니라 빅데이터 분석 기술 및 인공지능을 활용한 연구의 필요성이 부각되고 있다[1]. 인공지능 연구를 위해서는 대용량의 고품질 데이터 확보 및 라벨링이 필수적이거나 수경재배 영역에서의 데이터가 상대적으로 부족한 현황이다.

본 실험에서는 수경재배 식물의 일종인 향암쌈채 모종 원본 데이터로 학습한 인공지능 분류 모델의 성능을 먼저 확인하고, 증강 데이터를 활용한 모델 개발 및 성능개선에 대한 실험을 통해 부족한 수경재배 데이터의 해결 가능성을 확인하려 하였다. 이를 위해 확보된 원본 향암쌈채 생육 모종 데이터를 Backbone 분류 모델인 ResNet, DenseNet에 학습시켜 5개 질병 클래스에 대한 불량 모종 진단 실험을 진행하여 머신러닝 학습 적합 여부를 먼저 확인하였다. 이후 원본 데이터를 DCGAN을 통해 증강하여 2개의 실험을 진행하였다. 먼저 DCGAN 증강 데이터만으로 모델 학습을 진행하였고, 실제 데이터에 대한 질병 분류가 가능한 것을 확인하였다.

그리고, 원본 데이터에 증강 데이터를 추가한 Case별 실험을 통해 증강 데이터를 활용한 모델 성능의 개선이 가능함을 확인하였다.

본 논문에서는, 상대적으로 인공지능 연구 데이터가 부족한 수경 재배 분야에서 특히 향암 싹채와 같이 낮은 불량률로 필요 데이터 확보가 어려운 경우, DCGAN과 같은 생성AI를 활용한 증강 데이터 기반의 머신러닝 분류 모델의 개발 가능성에 대한 실험과 부족한 원본 데이터에 증강 데이터를 적용한 분류 모델의 성능 향상이 가능함을 실험하였다는 부분에서 연구의 의의를 찾을 수 있다. 그러나, 비용과 시간, 전문가 참여 한계 등의 문제로 원본 데이터를 충분히 확보하기 어려워 생성AI를 통한 증강 데이터의 품질이 높지 않았던 점, 증강 데이터의 원본 유사도 측정의 한계, 더 다양한 Backbone 모델을 활용한 실험 일반화, 원본 데이터와 증강 데이터간의 비교 실험이 이루어지지 못한 점 등은 본 연구의 한계로 판단된다.

References

- [1] S. Mingyue, Z. Jianhua, F. Quan, C. Xiujuan, Z. Ning, and Z. Wenrong, "Research Progress of Deep Learning in Detection and Recognition of Plant Leaf Diseases", *Smart Agriculture*, Vol. 4, No. 1, pp. 29-46, Mar. 2022. <https://doi.org/10.12133/j.smartag.SA202202005>.
- [2] J. H. Kim, A. J. Kim, and J. K. Park, "Convergence of Smart Agricultural Technology and Weather Information", *Korea Meteorological Institute · Issue Paper*, Jan. 2022.
- [3] Y. C. Choi and I. H. Jang, "Smart Farm in the Era of the Fourth Industrial Revolution", *Information & Communications Magazine of KICS*, Vol. 36, No. 3, pp. 9-16, Mar. 2019.
- [4] H. Uchiyama, et al., "An easy-to-setup 3D phenotyping platform for KOMATSUNA dataset", *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, Italy, pp. 2038-2045, Oct. 2017. <https://doi.org/10.1109/ICCVW.2017.239>.
- [5] J. W. Shin, "Indoor smart farms", *Korea Institute of Science and Technology Information, Kisti · Asti Market Insight 2022-070*, 2022.
- [6] J. S. Kang, S. J. Oh, J. W. Shin, J. W. Shin, and K. J. Lee, "Plant data preparation system for growth prediction under hydroponic conditions using deep learning", *KICS Winter Conference 2022*, Gangwon, Korea, Vol. 77, No. 1, pp. 1600-1602, Feb. 2022.
- [7] E. Kim, "Development of fruit volume prediction and appropriate harvest season determination technology using RGBD sensor and deep learning", *Korea Society for Agricultural Machinery Fall Conference*, Vol. 26, No. 2, pp. 352-352, 2021.
- [8] S. Hong, "A Study on the Prediction Model for Tomato Production and Growth Using ConvLSTM", *The Journal of Korean Institute of Information Technology*, Vol. 18, No. 1, pp. 1-10, Jan. 2020. <https://doi.org/10.14801/jkiit.2020.18.1.1>.
- [9] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning", *Journal of Big Data*, pp. 1-48, Jul. 2019. <https://doi.org/10.1186/s40537-019-0197-0>.
- [10] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks", *arXiv preprint arXiv:1511.06434* (2015). <https://doi.org/10.48550/arXiv.1511.06434>
- [11] J. Park, H. Kim, and K. Kim, "Assessing Impact of DCGAN Image Data Augmentation for CNN based Tomato Disease Classification", *Digital Content Society*, Vol. 21, No. 5, pp. 959-967, May 2020. <https://doi.org/10.9728/dcs.2020.21.5.959>.
- [12] G. Geetharamani and A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network", *Computers & Electrical Engineering*, Vol. 76, pp. 323-338, Jun. 2019. <https://doi.org/10.1016/j.compeleceng.2019.04.011>.

- [13] J. K. Kim, W. H. Cho, M. H. Na, and M. H. Jeon, "Development of Automatic Classification System of Vegetables by Image Processing and Deep Learning", Journal of The Korean Data Analysis Society, Vol. 21, No. 1, pp. 63-73, Feb. 2019. <https://doi.org/10.37727/jkdas.2019.21.1.63>.
- [14] J. Chen, et al., "Using deep transfer learning for image-based plant disease identification", Computers and Electronics in Agriculture, Vol. 173, Jun. 2020. <https://doi.org/10.1016/j.compag.2020.105393>.

저자소개

이 규 진 (Kyujin Lee)



1998년 2월 : 한양대학교

화학공학과(학사)

2016년 2월 : 고려대학교

경영전문대학원(석사)

2020년 ~ 현재 : 국민대학교

데이터사이언스학과 박사 수료

관심분야 : 머신러닝, 딥러닝,

데이터 분석, 자연어처리, LLM, 응용시스템