

설명가능 인공지능을 활용한 식물 병해 심각도 정량적 접근법: 흰가루병 적용 사례

이종욱 · 이상연 · 장원준 · 정준각[†]

한양대학교 산업융합학부

A Quantitative Approach to Plant Disease Severity Using Explainable AI: A Case Study on Powdery Mildew

Jongwook Lee · Sangyeon Lee · Wonjun Jang · Junegak Joung

School of Interdisciplinary Industrial Studies, Hanyang University

This paper presents a novel approach for accurately quantifying plant disease severity. This is achieved by utilizing explainable artificial intelligence (eXplainable AI) and deep learning techniques, which play a crucial role in early plant disease management. This study focuses on generating severity scores for powdery mildew infection by not only predicting disease presence through a ResNet50-based classifier, but also interpreting the model's decision-making process using Grad-CAM. By highlighting the specific regions of the leaf that influence the diagnosis, this approach supports experts and farmers in making more informed and targeted intervention decisions. The significant contribution of this study lies in its ability to visualize lesion areas and convert them into quantifiable severity scores, allowing for objective and repeatable disease assessments. By spatially explaining the impact of input features through heatmap localization, Grad-CAM helps build trust in the model and supports more transparent applications of AI in precision agriculture.

Keywords: XAI, Grad-CAM, ResNet50, Powdery Mildew, Severity Assessment

1. 서론

최근 기후 변화와 시설 재배의 확대 등으로 농작물 병해충의 발생 빈도와 강도가 증가하고 있으며, 이에 따라 병해의 조기 진단과 효율적 방제에 대한 농업 현장의 수요가 높아지고 있다(Rani *et al.*, 2023). 특히 흰가루병은 가지, 딸기, 수박 등 다양한 작물에 광범위하게 발생하는 대표적인 곰팡이성 병해로, 고온다습한 환경에서 빠르게 전염되어 식물의 생육 및 수확량, 품질에 중대한 영향을 미친다(Glawe, 2008; Sharma, 2021).

실제로 흰가루병으로 인한 피해 사례는 다양한 작물에서 반복적으로 보고되고 있으며, 피해 규모도 상당하다. 딸기의 경우, 시설하우스 내 온난다습한 환경이 흰가루병 발생을 촉진

하여, 적절한 방제가 이루어지지 않으면 수확량이 최대 70%까지 감소할 수 있으며, 영국에서는 딸기 흰가루병으로 연간 약 400억 원 규모의 경제적 손실이 추정되기도 했다(Carew and Adams, 2011). 가지에서는 흰가루병의 초기 병반 면적률이 약 17.4%를 초과하면 수량 감소가 시작되며, 병반율이 33.5%에 도달할 경우 수확량이 급감하는 것으로 나타났다(Kim *et al.*, 2016). 적기에 방제를 실시하지 않을 경우 잎의 30% 이상이 병에 감염되어 심각한 피해가 유발되며, 반면 조기 방제를 통해 병반율을 3.5% 수준까지 억제할 수 있다는 보고도 있다(Kim *et al.*, 2016). 수박에서도 2008년 충북 진천의 사례처럼 흰가루병이 대규모로 발생할 경우 농가 소득이 30~50%까지 줄어드는 심각한 피해가 발생할 수 있다(Kang *et al.*, 2014). 이처럼 흰

이 논문은 한국연구재단의 지원을 받아 수행되었음 (NRF-2021S1A5A8065184).

[†] 연락저자 : 정준각 조교수, 04763 서울특별시 성동구 왕십리로 222, 제2공학관 503-2호 한양대학교 산업융합학부, Tel : 02-2220-2363,

Fax : 02-2220-2363, E-mail : june30@hanyang.ac.kr

2025년 7월 25일 접수; 2025년 11월 13일; 2025년 11월 20일 수정본 접수; 2025년 11월 21일 게재 확정.

가루병은 다양한 작물에 걸쳐 큰 피해를 유발하며, 발병 초기 단계에서의 조기 진단 및 신속한 방제가 피해 경감을 위한 핵심 전략임을 보여준다.

이러한 피해는 단순 생산량 손실을 넘어 방제 비용 및 노동력 증가로 이어지며, 특히 친환경·유기농 재배 환경에서는 약제 선택과 방제법 적용에 제약이 커 지속 가능한 병해 관리 방식에 대한 관심이 증가하고 있다. 이에 따라 농업 현장에서는 AI 기반 이미지 분석 기술을 활용한 병해 조기 진단이 점차 확대되고 있으며, ResNet, VGG, EfficientNet 등 딥러닝 기반 전이학습 모델을 활용한 병해 이미지 분류 연구도 활발히 진행되고 있다.

최근에는 병해의 유무뿐 아니라 병해의 심각도를 정량적으로 평가하려는 연구도 이뤄지고 있으며(Pukkela and Borra, 2017), 병반 면적 비율 계산이나, 등급 분류, 회귀 기반 예측 등 다양한 접근 방식이 제안되고 있다. 이러한 방법들은 정량적 평가의 기초를 제공하지만, 대부분 모델이 ‘블랙박스(black-box)’ 방식으로 작동하기 때문에, 실제로 어떤 병반 부위에 주목했는지에 대한 해석은 제공하지 않으며, 분류 결과에 대한 설명 가능성이나 판단 근거의 시각화 측면에서는 한계가 있다. 이에 따라 최근 농업 분야에서도 설명가능 인공지능(explainable AI; XAI) 기법 도입이 확대되고 있으며, Grad-CAM(Gradient-weighted Class Activation Mapping) 등을 활용하여 AI 모델이 실제 병반 부위에 집중하고 있는지를 시각적으로 설명할 수 있는 기반이 마련되고 있다(Alhammad *et al.*, 2025).

본 연구는 이러한 배경을 바탕으로, 식물 병해 진단의 정밀화 및 설명 가능성 확보를 위해 Grad-CAM 기반의 XAI 기법을 활용하는 새로운 방법을 제시한다. 본 연구는 실제 현장에서 표준화가 어렵고 확보에 많은 시간이 소요되는 ‘초기/중기/말기’와 같은 상세한 심각도 라벨을 모델 학습에 직접 사용하지 않는다. 구체적으로는, 상세 라벨 대신 확보가 용이한 ‘정상/질병’의 이진(Binary) 정보만을 이용해 ResNet50 분류 모델을 학습시킨다. 이후, Grad-CAM을 통해 모델이 질병으로 판단한 시각적 근거(활성 영역)를 추출하고, 해당 면적 정보를 기반으로 병의 진행 상태를 정량적으로 추론할 수 있는 심각도 점수를 도출한다. 이 상세 라벨 정보(초기/중기/말기)는 모델이 산출한 점수가 실제 병해 단계를 유의미하게 반영하는지 사후에 통계적으로 검증하는 데에만 활용하여, 제안된 방법론의 신뢰도를 확보하고자 한다.

이러한 접근은 단순히 병해 진단 신뢰도 향상을 넘어, AI 모델의 판단 근거를 사용자에게 설명하고, 실제 농업 의사결정에 활용 가능한 정량 정보를 제공함으로써, 농업 현장에서의 AI 기술 수용성과 활용 가능성을 크게 높일 수 있을 것으로 기대된다.

본 연구는 AI-Hub(<https://www.aihub.or.kr/>)에서 제공하는 「시설 작물 질병 진단 이미지」 데이터로 예측 모델을 구축하고 병해 심각도 점수를 도출하는 과정을 다룬다. 제2장에서는 기존 연구 대비 본 연구의 기여도를 구체적으로 설명하고자 한다.

제3장에서는 설명가능 인공지능 기법을 적용해 모델의 해석과 Grad-CAM을 활용한 병해 심각도 점수 도출 방법론을 다룬다. 제4장에서는 예측 모델 구축 결과 및 병해 심각도 점수를 도출한다. 이후 점수의 통계적 검증을 수행한다. 마지막으로, 제5장에서는 본 연구의 의의와 개선 방안을 제시한다.

2. 기존 문헌

2.1 식물 병해 진단을 위한 딥러닝 연구

식물 병해 진단을 위한 딥러닝 기반 연구는 2016년을 기점으로 급격히 확대되었다. 초기에는 이미지 분류(classification)에 특화된 CNN 모델을 활용하여 병해의 유무 또는 종류를 예측하는 방식이 주류를 이루었으며, AlexNet, VGG, ResNet 등 대표적인 아키텍처들이 주요 연구에 적용되었다.

Mohanty *et al.*(2016)은 PlantVillage 데이터셋을 활용하여 총 14종 작물과 26개 병해 클래스를 대상으로 AlexNet 및 GoogLeNet 모델을 비교 분석했으며, 약 99.35%의 높은 정확도를 달성했다. Sladojevic *et al.*(2016)은 자체 수집한 13종 병해 이미지에 CNN을 적용하여 평균 96.3%의 정확도를 기록하였고, 데이터 증강 및 사전학습의 효과를 입증했다.

이후 전이 학습 기법이 활발히 적용되면서, 실제 재배지에서 촬영된 이미지로도 성능을 확보하려는 시도가 이어졌다. Ramcharan *et al.*(2017)은 필드 환경에서 수집한 카사바 잎 이미지에 Inception V3 모델을 적용하여 5종 병해중 피해를 구별했고, 약 93%의 예측 정확도를 달성했다. Fuentes *et al.*(2017)은 Faster R-CNN과 R-FCN, SSD와 같은 객체 탐지 기법을 토마토 병해중 이미지에 적용하여, 병반의 위치를 식별함과 동시에 9개 병해·해충 클래스를 실시간 탐지했다.

최근에는 정량화와 정확도 향상을 동시에 추구하는 연구가 늘고 있다. Chen *et al.*(2022)은 채널 주의 메커니즘과 프루닝 전략을 결합한 CACPNET을 제안하여, 땅콩 병해와 PlantVillage 데이터셋 모두에서 97~99% 수준의 높은 정확도를 기록했다.

이러한 기존 연구들은 대부분 병해의 유무 또는 종류를 예측하는 데 초점을 맞추고 있으며, 모델이 내린 판단의 배경이나 병해의 심각도 수준까지는 고려하지 않는다는 공통된 한계를 지닌다. 이에 따라 병해의 진행 상태를 수치적으로 평가하거나, 예측 결과의 해석 가능성까지 아우르는 접근에 대한 필요성이 제기되고 있다.

2.2 식물 병해 심각도(Severity) 정량화 시도

식물 병해의 심각도를 정량적으로 평가하려는 연구는 최근 빠르게 발전하고 있다. 초기에는 병반 면적을 육안이나 간단한 이미지 분석으로 측정하는 접근이 주를 이루었으나, 딥러닝 기반 기법의 도입 이후 정량화의 정확도와 자동화 수준이

크게 향상되었다. 특히 영상 분할(Segmentation), 회귀 기반 정량화, 주의 메커니즘(Attention), 다양한 영상 센서의 활용 등 기술적 진보가 활발히 이루어지고 있다.

Lin *et al.*(2019)은 흰가루병이 발생한 오이 잎에서 U-Net 기반 딥러닝 모델을 이용해 병반 영역을 픽셀 단위로 분할하고, 해당 면적을 기반으로 병해율을 산출했다. 해당 연구는 영상 내 병반의 위치와 크기를 정확히 식별함으로써 정량적 병해 평가의 자동화 가능성을 입증했다.

Elliott *et al.*(2022)은 카사바 세균성 모무늬마름병을 대상으로, ImageJ 기반의 전통 영상 처리 기법과 머신러닝(SVM) 기반의 병반 분할 성능을 비교했다. 두 기법 모두 병반 면적 산출에는 성공했지만, 배경 처리 및 조건 설정에 따라 성능 편차가 발생했다. 이에 따라 딥러닝 기반의 세그멘테이션 모델이 관심을 끌게 되었고, Li *et al.*(2022)은 오이의 노균병을 대상으로 DeepLabV3+ 모델에 attention 최적화를 적용해 병반 영역을 분할하고 이를 기반으로 연속형 심각도 점수를 산출했다. 해당 연구는 R2=0.958의 높은 정확도를 달성하며 정량적 병해 평가에 딥러닝 세그멘테이션 기법이 유효함을 시사했다.

Singh *et al.*(2024)은 사과 흰가루병을 대상으로, Convolutional Autoencoder와 Few-Shot Learning을 결합한 PDSE-Lite 모델을 제안했다. 해당 모델은 소수의 라벨 데이터만으로 병반의 분포를 학습하고, 이를 기반으로 0~100% 범위의 연속형 심각도 점수를 도출했다. Zhao *et al.*(2025)은 자연광 조건에서 현장 촬영된 식물 병해 이미지를 처리할 수 있도록 경량화된 PDSNet 모델을 개발하여 병반 영역을 분할하고 병해 면적 비율을 정량화했다. 두 연구 모두 실제 현장 적용 가능성을 고려한 구조로 설계되어 실용성이 높다는 평가를 받았다.

그러나 U-Net, DeepLabV3+ 등 영상 분할(Segmentation) 모델을 기반으로 병반 면적을 직접 추정하는 기존 접근법들은 명확한 한계를 지닌다. 영상 분할 모델은 병반의 경계가 명확하고 이미지 내에서 일정 비중 이상을 차지할 때는 효과적이거나, 흰가루병과 같이 병반이 미세하게 흩뿌려져 있거나 발병 초기 단계인 경우 성능이 급격히 저하될 수 있다. 실제로 Zang *et al.*(2025)의 연구에서도 흰가루병 병반은 형태가 복잡하고 (complex morphology) 건강한 영역과의 경계가 모호하여 (blurred boundaries), U-Net, DeepLabV3+와 같은 기존 분할 모델의 정확도가 낮아지는 (low segmentation accuracy) 문제가 지적된 바 있다. 잎 전체 면적 대비 병반의 픽셀 수가 극히 적어, 모델이 이를 병반으로 탐지하지 못하거나 노이즈로 오인하기 쉽기 때문이다. 이처럼 신뢰도 낮은 분할 결과를 기반으로 산출된 병반 면적 비율은 병해 심각도를 정확히 반영하기 어렵다.

본 연구는 이러한 구조적 한계를 극복하기 위해, 병반의 면적을 직접 추정하는 접근에서 벗어나, 모델의 시각적 판단 근거를 활용한 정량화 방식을 제시했다. 기존의 세그멘테이션 기반 연구들이 병반의 명확한 경계나 면적 비율에 의존했다면, 본 연구는 AI 모델이 병징으로 인식한 시각적 집중도 (visual attention intensity)를 정량화함으로써, 병반이 작거나 경

계가 모호한 초기 감염 단계에서도 감염 신호를 민감하게 포착할 수 있음을 확인했다. 특히 본 연구의 정량화 지표는 병반의 넓이뿐 아니라 모델이 부여한 주의 강도(weighted activation)를 함께 반영함으로써, 면적이 좁더라도 감염 특징이 뚜렷한 사례를 구분할 수 있다. 이는 기존의 면적 기반 병해율 산정법이 간과하던 초기 감염·회박 감염 영역의 평가 한계를 보완하는 결과를 보였다. 또한, 이러한 방식은 고정밀 세그멘테이션 모델을 직접 학습하지 않고도 병해의 정량적 판단을 수행할 수 있어, 모델 경량화·학습 효율성 측면에서도 유의미한 대안이 될 수 있음을 입증했다.

2.3 식물 병해 진단에서의 설명가능 인공지능(XAI) 및 Grad-CAM 응용 사례

최근 식물 병해 진단 분야에서는 모델의 판단 근거를 시각적으로 설명할 수 있는 설명가능 인공지능(XAI) 기법이 활발히 도입되고 있다. 그중에서도 Grad-CAM은 이미지 분류 모델의 예측 근거가 된 시각적 주목 영역을 Heatmap 형태로 시각화할 수 있어, 식물 병해 진단 분야에서 널리 활용되고 있다.

예를 들어, Quach *et al.*(2024)은 토마토 건강 상태 모니터링 시스템에 Grad-CAM++을 적용하여, MobileNet 기반 병해 분류기의 설명력을 높였다. 이들은 시각화 기반 설명을 통해 사용자의 신뢰를 유도하고자 했다.

포도 잎 병해를 대상으로 한 Karim *et al.*(2024)의 연구에서도 Grad-CAM이 실시간 분류 결과의 타당성을 사용자에게 직관적으로 전달하는 데 활용되었으며, 알감자 병해를 대상으로 한 Alhammad *et al.*(2025)의 연구는 모델의 설명력을 향상시키기 위한 핵심 도구로서 Grad-CAM을 제시했다. 이들은 Grad-CAM이 단순한 시각화 수단을 넘어, 모델이 병반 중심을 기준으로 정확히 판단하고 있음을 정량적으로 평가하는 데 활용될 수 있음을 강조했다.

나아가 Rashid *et al.*(2025)은 오이 6종 병해 진단을 위한 앙상블 CNN 모델에서 Grad-CAM, Grad-CAM++, Eigen-CAM 등 다양한 시각화 기법을 비교 적용하여, 모델 해석성과 분류 정확도를 동시에 확보할 수 있음을 입증했다. Gopalan *et al.*(2025)은 옥수수 병해 분류에 ResNet152를 활용하고 Grad-CAM 시각화를 통해 모델이 실제 병반 부위에 집중하고 있음을 시각적으로 검증했다.

이와 같이 다수의 선행 연구들은 Grad-CAM을 통해 딥러닝 기반 병해 분류 모델의 시각적 판단 근거를 제시하고, 모델 해석력을 강화하고자 했다. 그러나 이들 연구는 대부분 정성적인 시각화에 머무르고, 이를 기반으로 한 정량적인 병해 정보 산출까지는 확장하지 않았다는 한계를 가진다. 즉, Grad-CAM이 제공하는 Heatmap을 통해 모델이 어디를 보았는지를 설명할 수는 있으나, 해당 정보를 수치화하여 병해 심각도를 정량적으로 도출하려는 시도는 거의 이루어지지 않았다.

이에 비해 본 연구는 Grad-CAM으로 시각화된 활성 영역의 넓이와 강도를 결합하여 병해 심각도 점수를 정량화하는 새로

은 방법론을 제안했다. 이 방식은 단순 시각화에 그치지 않고, AI 모델의 내부 판단 근거를 기반으로 병해 진행 수준까지 수치적으로 표현할 수 있는 확장적 XAI 응용으로 평가될 수 있다.

3. 연구 방법

본 연구의 전체 흐름은 <Figure 1>에 제시된 바와 같이, 데이터 수집 및 전처리, 병해 분류 모델 구축, XAI 적용, 그리고 심각도 점수 도출 단계로 구성된다.

3.1 데이터 수집 및 전처리

본 연구에서는 AI-Hub(<https://www.aihub.or.kr/>)에서 제공하는 「시설 작물 질병 진단 이미지」 데이터를 분석 대상으로 한다. 해당 데이터는 딸기, 가지, 단호박 등 12종의 주요 시설 작물을 대상으로, 정상 이미지 10,000개와 병해당 1,000개의 병해 이미지로 구성되어 있다. 병해 부위는 잎, 열매, 줄기 등으로 구분되며, 해당 부위가 이미지의 주요 영역을 차지하도록 설계되었다. 데이터는 이미지와 메타데이터가 짝을 이루며, 작물 코드, 병해 코드, 촬영 부위 등의 정보가 포함되어 있다. 데이터 수집은 2020년 9월부터 12월까지 실 재배지와 비전 솔루션 전문 기업인 라온피플이 주관하여 진행되었으며, 정상 이미지는 전국 약 50개소의 농가에서, 병해 이미지는 자연 발생 또는 인위 접종을 통해 확보했다. 본 연구에서는 이러한 이미지 데이터를 기반으로 작물의 병해 여부를 분류하고, 병해의 심각도 점수를 도출한다.

원시 데이터셋은 12종의 작물에 대한 방대한 이미지를 포함하고 있으나, 본 연구는 병반이 잎사귀에서 먼저 발생하는 흰가루병의 특성을 반영하여 흰가루병 데이터가 포함된 5종의

작물(가지, 딸기, 단호박, 수박, 참외)의 잎사귀 이미지만을 선별하여 분석 범위를 명확히 했다. 또한, 모델이 병해 심각도 점수 산출에만 집중할 수 있도록, 원본 이미지의 품질과 구성이 연구 목적에 부합하는지 검토했다. 이 과정에서 annotation으로 제공된 Bounding Box 정보를 활용해 주요 객체(잎)를 중심으로 이미지를 Crop하여 배경 노이즈의 영향을 최소화했다. 마지막으로, 통계적 검증의 신뢰도를 확보하고 모델 학습 시 발생할 수 있는 편향을 방지하기 위해, 클래스 불균형 문제를 해소하는 데이터 정제 및 균형화(balancing) 과정을 수행했다. 본 연구에서는 모델이 심각도 구간별 경계를 명확히 학습할 수 있도록 정상:초기:중기:말기 클래스 = 3:1:1:1 비율로 조정하는 것을 목표로 설정했다. 데이터의 하한선은 전체 클래스 중 가장 적은 이미지 수를 가진 집단(참외 흰가루병 말기 86장)을 기준으로 결정했다. 이를 기준으로 각 작물별 병해 단계(초기·중기·말기)는 각각 86장씩, 정상 이미지는 병해 이미지의 총합(86x3)을 반영해 258씩 랜덤 샘플링했다. 그 결과, 정상(1,290개)과 병해(1,290개) 클래스의 비율을 1:1로 맞추고 병해 이미지는 다시 심각도(초기/중기/말기)별로 각 430개씩, 5종의 작물별로도 각 516개씩 균등하게 표본을 추출하여 최종적으로 2,580개의 이미지로 연구 데이터셋을 구축했다. 이러한 선별 및 균형화 과정은 방대한 원시 데이터로부터 본 연구의 목적에 가장 적합하고 통계적으로 신뢰할 수 있는 데이터셋을 구성하기 위해 필수적이었다.

또한, 본 연구에서 5종의 작물을 개별 모델로 구축하지 않고 통합 데이터셋으로 구성한 이유는, 흰가루병의 시각적 병반이 작물 종류와 무관하게 매우 유사한 공통점을 보이기 때문이다. 또한, 개별 작물 단위로 데이터를 한정할 경우, 신뢰할 수 있는 심각도 점수 산출 모델을 구축하기에 데이터 수가 부족할 수 있다는 점을 고려했다. 데이터의 클래스 구성 요약은 <Table 1>, 데이터의 실제 예시는 <Figure 2>와 같다.

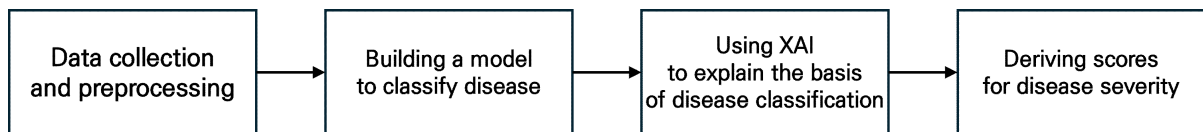


Figure 1. The Overall Process of the Proposed Approach

Table 1. Data Class Configuration Summary Table

Item	Class	Disease severity	Number of Images	Notes
Total	-	-	2,580	Five crop types, each with 516 samples
Class-wise Distribution	0(Normal)	-	1,290	
	1(Disease)	-	1,290	
Severity-wise Distribution	1(Disease)	1(Early)	430	
		2(Middle)	430	
		3(Late)	430	

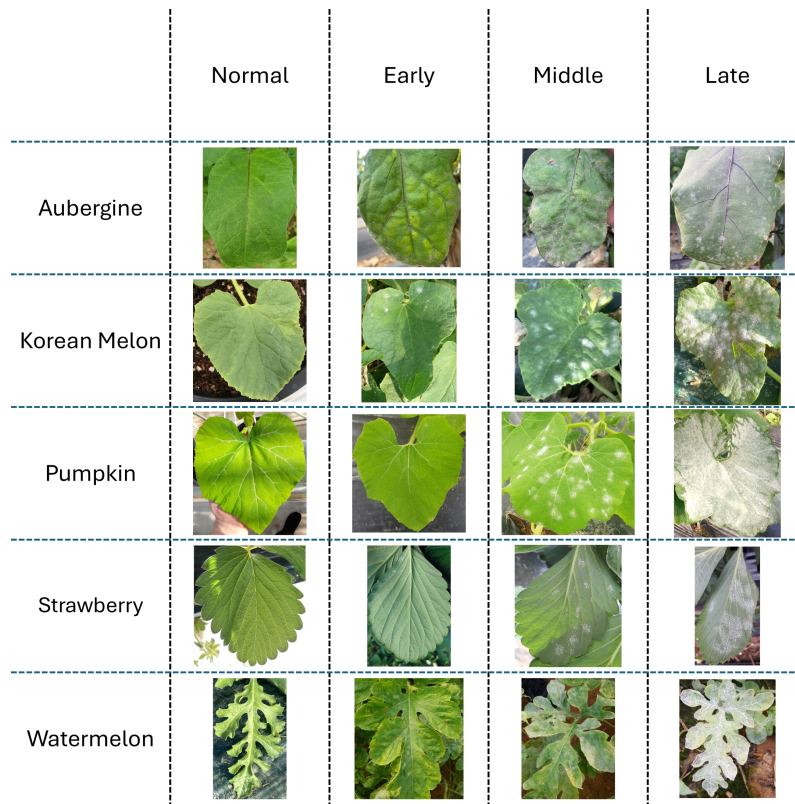


Figure 2. Real Data Example

3.2 병해 분류 모델 구축

식물 이미지에서 병해 유무를 분류하기 위한 분류 모델을 비교·선정하기 위해, VGG 16, ResNet-50, EfficientNet-B0, MobileNet-V2 총 4종의 CNN 기반 딥러닝 아키텍처를 활용하여 학습 및 성능 평가를 수행했다. VGG16은 단순한 계층 구조와 비교적 깊은 네트워크로 구성되어 있어, baseline 모델로서 성능 비교의 기준점으로 활용하기에 적합하다(Simonyan and Zisserman, 2015). ResNet-50은 잔차 연결(residual connection) 구조를 통해 gradient vanishing 문제를 완화하며, 심층 신경망 학습의 안정성을 확보할 수 있다(He *et al.*, 2016). EfficientNet-B0는 compound scaling 기법을 기반으로 모델의 깊이, 너비, 해상도를 균형 있게 확장함으로써, 높은 정확도와 계산 효율성을 동시에 달성하는 최신 구조이다(Tan and Le, 2019). MobileNet-V2는 depthwise separable convolution을 도입하여 경량화를 극대화하였으며, 제한된 연산 자원을 고려한 모바일 환경에서도 효율적인 추론이 가능하다(Sandler *et al.*, 2018).

각 모델의 학습을 위해 전체 데이터 셋에서 20%를 고정된 Test Set으로 분리하고, 나머지 80%는 5-fold 교차 검증 방식으로 나뉘어 훈련 셋과 검증 셋으로 구성했다. 그리고 훈련 셋과 검증 셋을 활용하여 각 분류 모델을 학습했으며, 학습된 모델을 사용하여 고정된 Test Set에 대한 분류를 수행, 결과를 바탕으로 성능을 평가했다. 각 모델은 동일한 데이터 분할 및 학습 설정 하에서 훈련되었으며, 모델의 분류 성능 평가는 정확도

(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score, 그리고 ROC-AUC 지표를 기준으로 진행했다. 다중 지표를 종합적으로 고려하되, 클래스 간 균형이 잘 유지된 데이터셋 특성을 반영하여 F1-score를 주요 기준 지표로 설정했다. F1-Score가 1에 가까울수록 정밀도와 재현율이 모두 높아 모델의 성능이 우수함을 의미한다. 각 지표는 단순한 분류 정확도를 넘어, 양성 클래스에 대한 탐지 성능(F1-score)을 반영하므로, 분류 문제에서 모델을 균형 있게 평가하는 데 유리하다.

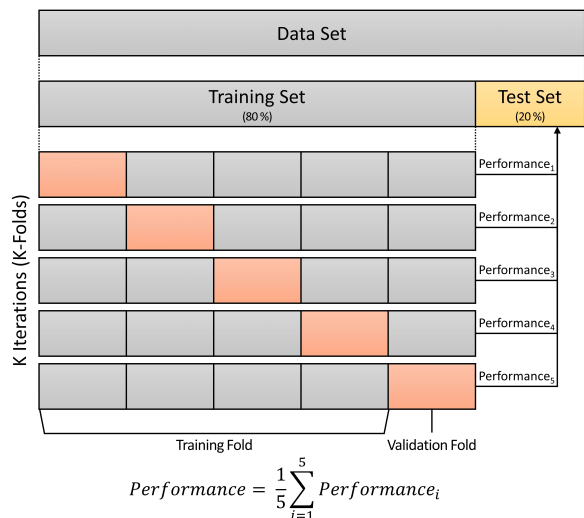


Figure 3. 5-fold Cross Validation

모델의 학습과 검증은 K-fold 교차 검증(K-fold cross validation, K=5)과 고정 테스트 셋을 결합한 방식으로 <Figure 3>과 같이 수행했다(Kohavi, 1995). 전체 데이터셋의 20%는 사전에 분리된 Test Set으로 확보하여 최종 성능 평가에만 활용했고, 나머지 80%는 K-fold 교차 검증을 통해 모델 학습 및 내부 검증에 사용했다. 또한, 본 연구는 클래스 간 불균형 문제를 고려하여 Stratified K-fold 방식을 적용했다. 이를 통해 각 fold 내 클래스 비율을 동일하게 유지함으로써 클래스 분포에 따른 편향을 최소화했다.

3.3 병해 분류의 근거 도출을 위한 XAI 적용

최종 선정된 블랙박스 모델의 의사 결정을 이해하기 위해 Grad-CAM을 활용하여 이미지 내 주요 활성 영역을 시각화했다. Grad-CAM(Gradient-weighted Class Activation Mapping)은 딥러닝 기반의 이미지 분류 모델이 특정 클래스에 대해 분류를 수행할 때, 입력 이미지의 어떤 영역에 주목했는지를 시각적으로 설명하는 방법이다. 이 기법은 분류 대상 클래스에 대한 Gradient와 마지막 합성곱 계층의 Feature Map을 결합하여 해당 클래스와 관련된 시각적 근거를 강조하는 히트맵을 생성한다. 이를 통해 모델이 어떤 시각적 정보를 바탕으로 결정을 내렸는지 직관적인 해석이 가능하다. Grad-CAM의 계산은 다음 수식으로 정의된다.

$$L_c^{Grad-CAM} = ReLU(\sum_{k=1}^K \alpha_k^c A^k) \quad (1)$$

$L_c^{Grad-CAM}$: 클래스 c에 대한 Grad-CAM 히트맵

A^k : 마지막 합성곱 층의 k번째 채널의 활성화 맵(activation map)

α_k^c : 클래스 c에 대한 채널 k의 가중치($= \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$)

Z: A^k 의 공간적 크기($H \times W$)

K: 마지막 합성곱 층의 채널 수

ReLU: 음수 값을 제거하여 중요 영역만 강조

Grad-CAM으로 생성된 히트맵은 입력 이미지의 공간적 특징 중 특정 클래스의 분류에 기여한 영역을 강조함으로써, 모델의 분류 과정에 대한 해석 가능성을 높여준다. 원시 이미지나 일반적인 분류 결과와는 달리, Grad-CAM 결과는 모델 내부의 중간 표현과 경사 정보를 활용해 중요한 영역을 시각화하기 때문에, 의사결정의 근거를 설명하거나 모델의 신뢰도를 판단하는 데 유용하게 활용될 수 있다.

3.4 병해 심각도 점수 도출

본 연구는 Grad-CAM HeatMap을 이용하여 병해 심각도 점

수를 정량적으로 도출하는 방법을 제시한다. 병해 심각도 점수를 도출하기 위해, 흰가루병 class에 대한 Grad-CAM Heat Map을 활용하여 각 이미지의 병해 심각도를 산출했다. Grad-CAM Heat Map의 픽셀값은 1에 가까울수록 흰가루병 클래스를 분류하기 위해 크게 영향을 미치는 것을 의미한다. 이를 통해 이미지 내에서 식물의 병해 유무 판단에 기여한 영역을 평가할 수 있다. 이러한 접근 방식은 병해가 심화됨에 따라 병반의 분포 범위와 크기가 점차 확대되는 흰가루병의 특성에 따라(Willocquet *et al.*, 2008), Heat Map의 강도를 병해 심각도 점수로 정량화하는 것을 가능하게 한다.

구체적인 방법론은 다음의 세 가지 단계의 마스킹 및 계산 절차를 수행했다. 먼저, Heat Map에서 유의미한 관심 영역만을 추출하기 위해, Heat Map의 픽셀 중 최댓값의 15% 이상인 값을 가진 픽셀만을 유효한 영역(Valid area mask)으로 정의했다. 이 임계값은 Grad-CAM의 원 저자가 실험적으로 제안한 값이며(Selvaraju *et al.*, 2017), 모델의 집중 영역을 보다 명확하게 분리하는 데 도움을 준다. 이는 다음의 수식으로 표현된다.

$$Valid\ area\ mask(i,j) = \begin{cases} 1, & \text{if } H(i,j) \geq 0.15 \times \max(H) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$H(i,j)$: Grad-CAM Heat Map의 각 픽셀값

$\max(H)$: Heat Map 내에서의 최댓값

다음으로, 입력 이미지에서 실제 잎사귀 객체가 존재하는 영역만을 분석 대상으로 삼기 위해, 불필요한 배경 정보를 제거하고 RGB 값이 [0,0,0]이 아닌 픽셀만을 주요 객체 영역(Non-black mask)으로 정의했다. 이 과정은 다음과 같다. 먼저, 입력된 RGB 이미지를 HSV 색공간으로 변환하여 채도(Saturation) 값을 1.5배로 증폭시키고, 각 RGB 채널에 CLAHE(Contrast Limited Adaptive Histogram Equalization)를 적용하여 전체적인 명암 대비를 향상시켰다(Zuiderveld, 1994). 이를 통해 배경과 객체 간 경계를 더욱 뚜렷하게 만들었다. 이후 SAM(Segment Anything Model)을 활용하여 객체 마스크를 자동으로 생성했으며(Kirillov *et al.*, 2023), 이 중 가장 넓은 면적을 차지하는 객체를 주요 피사체로 간주하고, 이를 기준으로 배경 영역을 제거했다. 그에 대한 예시는 <Figure 4>에 제시되어 있으며, 주요 객체 영역의 정의는 다음의 수식으로 표현된다.

$$Non-black\ mask(i,j) = \begin{cases} 1, & \text{if } P(i,j) \neq [0,0,0] \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$P(i,j)$: 이미지 각 픽셀의 [R,G,B] 값

최종적으로, 위에서 정의한 두 마스크의 교집합(AND 연산)을 통해, 모델이 실제 객체 영역 중에서 주목한 영역(Target mask)을 도출했다. 또한 해당 영역의 픽셀값을 곱하여 가중치

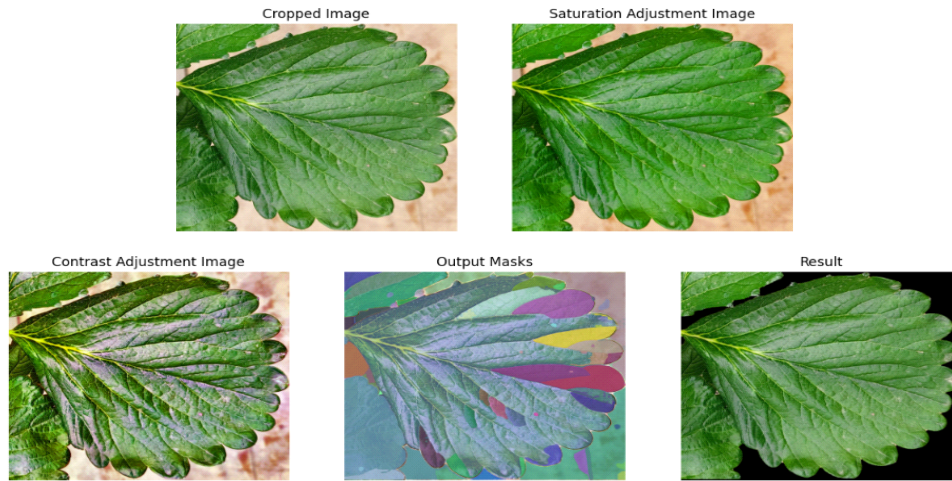


Figure 4. Example of Image Data Preprocessing

를 부여하고 실제 객체 영역의 픽셀 수로 정규화하여 병해 심각도 점수를 산출했다. 이는 다음의 수식으로 표현된다.

$$\text{Disease Severity Score} = \sum_i \sum_j \frac{\text{Target mask}(i,j) * H(i,j)}{\text{Non-black mask}(i,j)} \quad (4)$$

$\text{Target mask}(i,j)$: Valid area mask(i,j)
 • Non-black mask(i,j)

$H(i,j)$: Grad-CAM Heat Map의 각 픽셀값

또한 본 연구에서는 모델의 일반화 성능을 향상시키기 위해 K-Fold 교차 검증 방식을 적용하여 총 5개의 모델을 학습했고, 각 모델로부터 산출된 병해 심각도 점수에 대해 F1-score를 기준으로 가중평균을 수행했다.

4. 사례 연구

4.1 분류 모델 구축 및 선정

분류 모델의 선정을 위해 4가지 CNN 기반 딥러닝 아키텍처

로 종속 변수를 분류했다. 각 아키텍처별 성능 비교 결과는 <Table 2>와 같다. 모델별 분류 성능 비교 결과 ResNet50 모델이 F1-score에서 모든 모델 중 가장 높은 값(0.9837)을 기록했으며, ROC-AUC에서도 EfficientNet-B0(0.9990)에 근소하게 뒤처지지만, 여전히 매우 높은 수준(0.9989)의 값을 유지했다. 반면, EfficientNet-B0는 ROC-AUC에서는 가장 높았으나 F1-score는 ResNet50에 비해 낮은 수치(0.9814)를 보였다. 두 주요 지표를 동시에 고려했을 때 ResNet50의 성능이 가장 우수했으며, 보조지표인 정확도(0.9837), 정밀도(0.9842), 재현율(0.9834)에서도 최고 성능을 나타냈다. 따라서 본 연구에서는 F1-score의 우수성과 전반적인 지표 균형을 기반으로 ResNet50을 최종 분류 모델로 채택했다.

모델 학습 및 실험은 Google Colab Pro 환경에서 수행되었다. 학습에는 NVIDIA L4 GPU(24GB VRAM)가 제공되는 고성능 클라우드 환경을 사용했으며, 런타임 사양은 Intel Xeon CPU @ 2.20GHz, RAM 55.5GB로 구성되었다. 소프트웨어 환경은 Ubuntu 22.04 LTS 기반의 Python 3.12.12, PyTorch 2.8.0+cu126, TorchVision 0.23.0+cu126, CUDA 12.5, cuDNN 8.9로 구축되었다. 모델 학습에는 AdamW optimizer를 사용했고, 초기 학습률(learning rate)은 1×10^{-4} , batch size 32, epoch 20으로 설정했다. 또한 k-Fold 교차 검증($k = 5$)을 적용하여 fold

Table 2. Comparison of Performance Between Prediction Models

Model	Accuracy	Loss	ROC-AUC	Precision	Recall	F1-score
VGG16	0.9682 ± 0.0028	0.0777 ± 0.0095	0.9969 ± 0.0005	0.9695 ± 0.0022	0.9675 ± 0.0030	0.9681 ± 0.0028
ResNet50	0.9837 ± 0.0013	0.0475 ± 0.0042	0.9989 ± 0.0002	0.9842 ± 0.0013	0.9834 ± 0.0013	0.9837 ± 0.0013
EfficientNet-B0	0.9814 ± 0.0027	0.0424 ± 0.0024	0.9990 ± 0.0001	0.9815 ± 0.0027	0.9813 ± 0.0027	0.9814 ± 0.0027
MobileNet-V2	0.9686 ± 0.0028	0.0761 ± 0.0059	0.9972 ± 0.0003	0.9694 ± 0.0027	0.9680 ± 0.0029	0.9685 ± 0.0028

별 학습 및 검증을 수행함으로써, 모델의 일반화 성능을 향상시켰다. 학습률은 cosine annealing scheduler를 통해 점진적으로 감소시켰으며, early stopping(patience=10)을 설정하여 검증 손실(val_loss)의 개선이 일정 기간 지속되지 않을 경우, 학습을 조기 종료했다. 데이터 증강은 수평 반전(horizontal flip), 회전($\pm 15^\circ$), 밝기(brightness) 및 대비(contrast) 조정($\pm 20\%$)을 포함했으며, 모든 입력 이미지는 224×224 픽셀로 리사이즈 및 정규화했다. 이러한 환경 설정은 모델 간 성능 비교의 공정성과, 후속 연구에서의 재현 가능성을 높이기 위함이다.

4.2 XAI 기반 병해 분류 근거 해석

분류 모델의 해석과 병해 심각도 점수 도출을 위해 Grad-CAM을 적용했다. <Figure 5>은 입력 이미지(왼쪽), 해당 이미지에 대해 Grad-CAM을 적용한 Heat Map(가운데), 그리고 원본 이미지와 Heat Map을 중첩한 결과(오른쪽)를 함께 제시한 그림이다. Grad-CAM Heat Map에서는 붉은색에 가까울수록 모델이 해당 부위를 분류의 주요 근거로 활용하였음을 의미한다. 예시 이미지의 경우, 잎의 중심에서 오른쪽 아래 방향으로 확장된 붉은색 영역이 강조되었으며, 이는 모델이 해당 부위를 기반으로 병해 유무를 판단했음을 나타낸다.

분석 결과, 이미지 상에서 식별 가능한 병반 부위가 모델의 분류에 주요한 영향을 미치는 것으로 확인되었다. Heat Map 상에서 가장 높은 중요도를 보인 영역은 잎의 중앙 하단부로, 원본 이미지에서도 해당 부위에 흰색 반점이 분명히 나타나 있다. 이는 모델이 흰가루병의 시각적 특징 중 하나인 흰색 곰팡이 흔적을 주요 단서로 활용하여 감염 여부를 판단하고 있음을 보여준다.

반면, 잎의 상단이나 가장자리 부위는 Grad-CAM에서 상대적으로 낮은 주목도를 보였으며, 이는 해당 영역이 감염 여부 판단에 미치는 영향이 크지 않았음을 의미한다. 이러한 결과는 모델이 실제로 병반이 있는 시각적 특징에 주목하고 있다고 해석할 수 있다.

4.3 병해 심각도 점수 도출

본 연구에서 제시한 병해 심각도 점수는 모델이 흰가루병을 분류할 때 주목한 시각적 영역의 범위를 기준으로 하여, 병반의 면적과 모델의 집중도를 반영한다. 도출 과정의 예시는 <Figure 6>와 같다. 병해 감염 유무에 따라 병해 심각도 점수를 산출한 결과, 정상인 경우 최소 0, 최대 44.80, 평균 5.35, 중앙값 2.57로 나타났으며, 감염의 경우 최소 16.27, 최대 98.37, 평

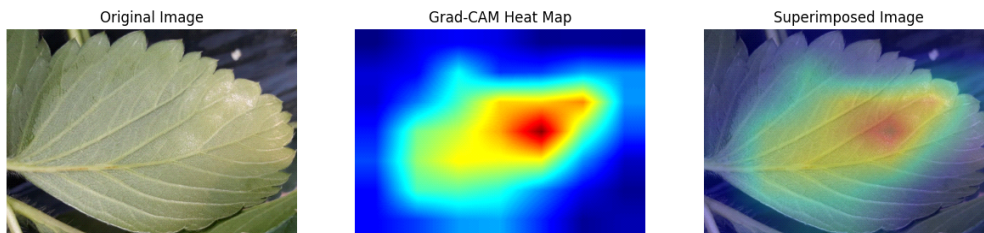


Figure 5. Example of Grad-CAM

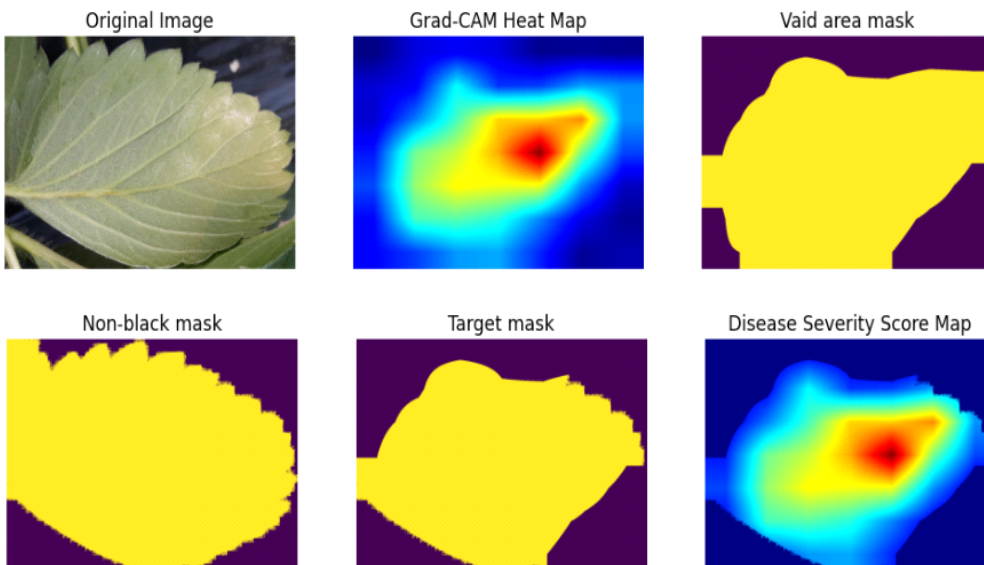


Figure 6. Example of Disease Severity Score Output Process

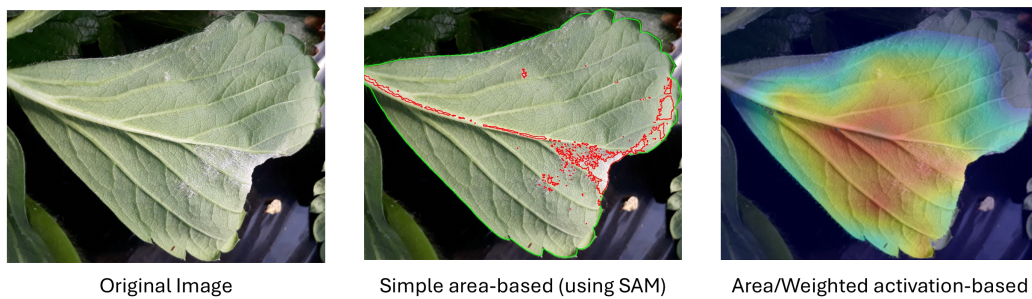


Figure 7. Example of Methodology Comparison

균 73.01, 중앙값 75.73으로 나타났다. 시각적으로 식별 가능한 병반이 넓게 퍼진 이미지일수록 Heat Map에서 강조되는 영역이 넓고, 이에 따라 높은 점수를 나타내는 경향이 관찰되었다. 반면, 병반이 미미하거나 존재하지 않는 이미지에서는 모델이 강하게 반응하는 영역이 제한되어 낮은 심각도 점수가 도출되었다.

한편, 육안상 병반의 면적이 넓게 분포하지 않더라도 모델은 높은 심각도 점수를 산출하는 사례도 관찰되었다. <Figure 7>은 흰가루병 말기 이미지임에도 불구하고, 병반이 잎 전체에 균일하게 퍼져 있지 않은 사례를 보여준다. <Figure 7>의 중앙은 단순 면적 기반 접근의 예시로, 병반의 물리적 분포만을 기준으로 면적을 추정한 것이다(본 연구의 방법론과 비교하기 위한 시각적 예시로 SAM을 활용). 이 경우 표면적으로 한정된 병반만을 인식하여 감염 범위를 과소평가하지만, <Figure 7>의 우측 결과에서는 병반 부위에 대한 활성화가 잎맥 주변으로 강하게 집중되어 모델이 시각적으로 미세한 감염 패턴까지 정확히 인식하고 있음을 확인할 수 있다. 이처럼 제안된 면적 및 집중도 기반 지표는 단순 병반 넓이만으로는 구분 어려운 감염이나 국소 감염 사례에서도 높은 집중도를 통해 심각한 감염을 식별할 수 있음을 의미한다.

4.4 검증

흰가루병 감염 여부를 분류하는 분류 모델을 기반으로, 모델이 주목한 영역을 활용하여 정량적인 병해 심각도 점수를 산출했다. 이 점수는 기존의 임상적 정의나 정립된 지표에 기반한 것이 아닌, 모델 내부 판단을 반영한 수치적 표현으로, 이를 통해 병해의 존재 여부나 진행 상태를 간접적으로 추정하는 데 목적이 있다.

하지만 이러한 방법론이 실제 감염 여부나 병해의 진행 상태를 정량적으로 반영하고 있는지 엄밀하게 증명하기 위해서는, 실제 임상에서 정의된 ‘병해 심각도 점수’가 존재해야 한다. 그러나 현재 흰가루병에 대해 객관적이고 표준화된 정량 지표가 존재하지 않기 때문에, 본 연구에서는 모델이 산출한 점수가 실제 감염 여부 및 진행 단계에 따라 유의미한 차이를 보이는지 통계적으로 분석함으로써 그 타당성을 검토하고자 했다.

(1) 감염 여부에 따른 통계적 분석

흰가루병 감염 여부에 따라 정상군(Normal)과 감염군(Infected)로 구분하여 두 집단 간의 분포 차이를 비교했다. 이를 통해 해당 점수가 실제 감염 여부의 반영 지표로 활용될 수 있는지 가능성을 검토하고자 했다.

<Figure 8>은 각 집단의 병해 심각도 점수 분포를 시각화한 것이다. 정상군 집단은 대부분 0~20점에 밀집되어 있으며, 급격한 감소 형태를 보이는 비대칭적인 분포를 나타낸다. 반면, 감염군 집단은 40~90점 범위에 넓게 퍼진 완만한 종형 분포를 나타냈다. 이로써 두 집단의 분포 양상이 명확히 다를 수 있었다.

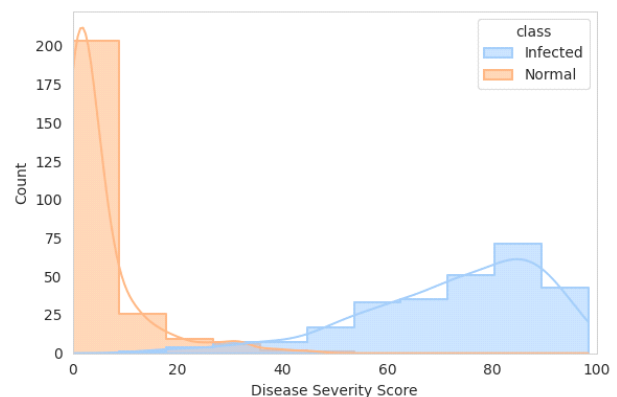


Figure 8. Histogram of Disease Severity Score

분포의 차이가 통계적으로 유의미한지를 평가하기에 앞서, 각 집단의 점수 분포가 정규성을 따르는지를 먼저 검정했다. 정규성을 만족할 경우에는 t-검정(t-test)과 같은 모수 검정을 사용할 수 있지만, 그렇지 않은 경우에는 비모수 검정(Mann-Whitney U test)을 적용해야 한다. 따라서 각 집단의 정규성을 판단하기 위해 정규성 검정(Shapiro-Wilk test)을 수행했다. 정규성 검정 결과, p-value가 유의수준 0.05보다 작으므로 귀무가설을 기각한다. 즉, 두 집단 모두 정규성을 만족하지 않는다는 결론을 도출할 수 있다.

정규성을 만족하지 않기 때문에, 두 집단 간의 중앙값 차이를 비교하기 위해 비모수 검정(Mann-Whitney U test)을 적용했다. 이 검정은 각 집단의 데이터를 하나의 순위 집합으로 변환

한 후, 집단 간 순위의 차이를 기반으로 분포의 차이를 검정하는 방법이다.

(1) 가설 설정:

- 귀무가설(H_0): 감염군 집단(M_2)의 병해 심각도 점수의 중앙값은 정상군 집단(M_1)의 병해 심각도 점수의 중앙값보다 작거나 같다($M_1 \geq M_2$).
- 대립가설(H_1): 감염군 집단(M_2)의 병해 심각도 점수의 중앙값은 정상군 집단(M_1)의 병해 심각도 점수의 중앙값보다 크다($M_1 < M_2$).

(2) 결과:

- p-value = 1.4e-85

비모수 검정 결과, p-value가 유의수준 0.05보다 작으므로 귀무가설을 기각하고 감염군 집단의 병해 심각도 점수가 정상군 집단보다 유의미하게 크다는 결론을 도출할 수 있었다. 이는 모델이 단순히 감염 여부에 따라 구분될 뿐만 아니라, 실제로 감염군에서 더 높은 값을 갖는 경향을 명확히 보여준다는 점에서 그 타당성을 강하게 뒷받침한다.

(2) 감염 진행 단계에 따른 통계적 분석

이어서, 흰가루병 감염 진행 단계에 따라 초기(Early), 중기(Middle), 말기(Late) 세 집단으로 세분화하여 각 단계별 병해 심각도 점수의 차이를 비교·분석했다. 이 분석은 분류 모델이 산출한 정량적인 점수가 단순히 감염 여부뿐 아니라 병해의 심각도 또한 정량적으로 반영하고 있는지를 검증하기 위한 목적을 가진다.

<Figure 9>은 단계별 병해 심각도 점수의 분포를 시각화한 Box plot이다. 전체적으로 진행 단계가 말기로 갈수록 중앙값이 점진적으로 상승하는 경향이 나타나며, 이는 모델이 산출한 점수가 실제 병해의 진행 상태와 관련성을 가질 수 있음을 나타낸다.

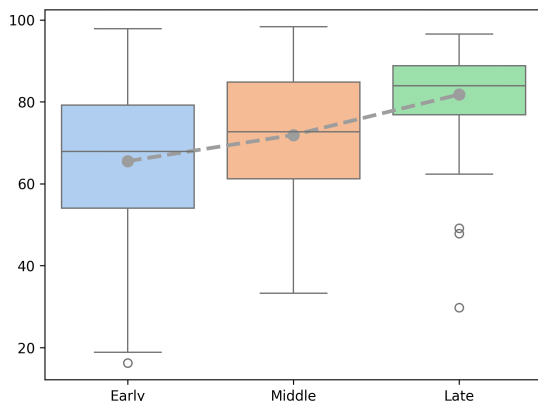


Figure 9. Box plot of Disease Severity Score

본격적인 통계 분석에 앞서, 각 집단의 점수가 정규분포를 따르는지 확인하기 위해 정규성 검정(Shapiro-Wilk) 검정을 수

행했다. 정규성 검정 결과, p-value가 유의수준 0.05보다 작으므로 귀무가설을 기각한다. 즉, 모두 정규분포를 따르지 않는 것으로 판단되었다. 이에 따라 세 그룹 간의 점수 차이를 비교하기 위해 비모수 일원분산 분석(Kruskal-Wallis H test)을 수행했다. 이는 정규성을 가정하지 않는 비모수 검정으로, 세 개 이상의 독립 집단 간 분포 차이를 검정할 수 있다. 비모수 일원분산 분석 결과, 유의수준 0.05 이하의 p-value가 확인되어 귀무가설은 기각되었고, 세 집단 간의 유의미한 차이가 존재함을 확인했다.

그 차이가 구체적으로 어느 집단 간에 존재하는지를 확인하기 위해 각 쌍별 비모수 검정(Mann-Whitney U test)을 수행했다.

(1) 가설 설정:

- 귀무가설(H_0): 비교 대상 집단 B의 병해 심각도 점수의 중앙값(M_2)은 집단 A의 중앙값(M_1)보다 작거나 같다($M_1 \geq M_2$).
- 대립가설(H_1): 비교 대상 집단 B의 병해 심각도 점수의 중앙값(M_2)은 집단 A의 중앙값(M_1)보다 크다($M_1 < M_2$).

(2) 결과:

- 초기(A), 중기(B) p-value = 2.3e-02
- 중기(A), 말기(B) p-value = 1.5e-06

비모수 검정 결과, 초기-중기, 중기-말기 조합 모두에서 유의미한 분포 차이가 나타났다. 특히 점수의 중앙값이 진행 단계에 따라 순차적으로 증가하는 경향이 뚜렷하게 관찰되었다. 이를 통계적으로 검증하기 위해 추가적으로 단조 증가 경향 검정(Jonckheere-Terpstra test)을 수행했다. 이 검정은 세 개 이상의 순서가 있는 집단 간에 단계적인 증가 또는 감소 경향이 존재하는지를 검정하는 비모수 검정 방법이다. 검정 결과, p-value가 유의수준 0.05보다 작으므로 귀무가설은 기각되며, 병해 진행 단계에 따라 점수가 단계적으로 증가하는 경향이 통계적으로 유의미함을 확인할 수 있었다.

이러한 분석 결과는 분류 모델이 단순히 감염 여부를 구분하는 데 그치지 않고, 감염의 심각도 수준까지 정량적으로 반영하고 있음을 나타낸다. 이는 모델이 산출한 점수가 흰가루병의 진행 상태를 평가하는 데 활용 가능한 정량적 지표로 기능할 수 있음을 보여준다.

(3) Grad-CAM 활성화 맵의 임계값 민감도 분석

Grad-CAM Heat Map에서 활성화 픽셀의 최댓값 대비 15% 이상을 유효 영역으로 정의하는 임계값(Threshold)은 모델이 산출하는 병해 심각도 점수에 직접적인 영향을 미칠 수 있다. 해당 값은 Grad-CAM의 원 저자(Selvaraju et al., 2017)가 제안한 실험적 기준을 기반으로 설정했으나, 본 연구에서는 이 임계값의 변화가 심각도 점수의 통계적 유의성에 미치는 영향을 분석하여 최적의 값을 검증했다.

이를 위해 임계값을 0%부터 30%까지 5% 단위로 변화시키

며 단계별로 산출된 심각도 점수에 대해 Jonckheere-Terpstra test를 수행했다. 이 검정은 병해 단계(초기→중기→말기)에 따른 단조 증가(monotonic trend)가 존재하는지를 평가하기 위한 비모수 검정이다. 분석 결과, p-value가 5%~20% 구간에서 일정하게 낮은 수준(10^{-10} 이하)을 유지했으나, 25% 이후 급격히 상승(2.23×10^{-9} → 2.20×10^{-8})하여 임계값이 과도하게 높을 경우 단계 간 점수 차이가 약화되는 경향을 보였다.

가장 낮은 p-value(3.84×10^{-10})를 기록한 20% 부근을 중심으로 15%~25% 구간을 1% 단위로 세분화하여 분석했다. 그 결과, p-value가 15%~20% 범위에서 전반적으로 낮게 유지되었고, 이 중 18%에서 최저값(3.41×10^{-10})을 기록했다. 이어서 18%를 중심으로 17%~19% 구간을 0.1% 단위로 다시 세분화하여 분석한 결과, 17.1%에서 p-value 3.32×10^{-10} 로 최저값을 나타냈으며, 17.1%~18.0% 구간에서도 유의성이 안정적으로 유지되는 것을 확인했다. 이는 <Figure 10>에서 확인할 수 있다.

이 결과는 Grad-CAM Heat Map의 유효 활성 영역이 전체 Heat Map의 약 17.1% 이상일 때, 병해 진행 단계 간 점수 차이가 가장 명확하게 구분됨을 의미한다. 그러나, Grad-CAM의 원 저자가 제안한 15% 임계값 또한 최적 임계값 탐색 없이 사용하여 최적값에 근사한 심각도를 도출할 수 있어 효율적인 임계값임을 확인했다.

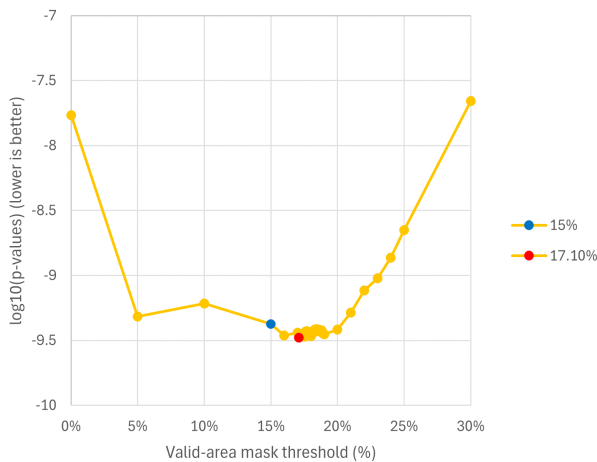


Figure 10. Graph of p-values by Threshold

5. 결론 및 시사점

본 연구에서는 설명가능 인공지능(XAI) 기법인 Grad-CAM을 활용하여 식물 병해의 감염 여부뿐 아니라 심각도까지 정량화할 수 있는 새로운 방법을 제시했다. 딸기, 가지, 수박 등 주요 작물에서 발생하는 흰가루병을 중심으로, ResNet50 기반의 분류 모델을 구축하고, Grad-CAM으로 시각화된 활성 영역을 통해 병반의 위치와 집중도를 반영한 병해 심각도 점수를 산출했다.

기존 연구들은 주로 병해 유무의 분류 또는 XAI 기법을 통한 시각적 설명에 초점을 맞췄다. 그러나 본 연구는 AI 모델의 분류 근거를 시각적으로 설명하는 동시에, 해당 정보를 기반으로 정량적인 심각도 지표를 도출함으로써 식물 병해 분류의 해석 가능성과 실용성을 동시에 제고했다. 특히, Grad-CAM Heat Map의 정량 정보와 병해 진행 단계 간 통계적 검증을 통해 모델이 산출한 점수가 실제 감염 수준과 유의미한 상관관계를 가진다는 점을 확인했다.

본 연구에서 제안한 정량적 심각도 점수는 실제 농업 현장의 의사결정을 지원하는 데 구체적으로 활용될 수 있다. 예를 들어, 점수 구간에 따른 임계값(Threshold) 기반의 의사결정 규칙을 설계하는 것이 가능하다. 가령, ‘심각도 점수 0.7’일 경우 즉각적인 약제 살포를 결정하고, ‘0.3~0.7’ 구간은 중점 모니터링 대상으로 분류하는 등, 방제 시기, 약제 농도, 방제 구역 범위를 정량적으로 결정하는 데 기여할 수 있다. 이러한 접근은 향후 농업 자동화 시스템의 의사결정 규칙 설계에 직접 적용될 수 있다. 나아가, 본 점수는 병해 조기 경보 시스템에 활용되거나, 시계열 데이터와 결합하여 병해 확산 예측 모델의 핵심 입력 변수(feature)로도 활용 가능할 것이다.

그러나 본 연구는 흰가루병이라는 특정 병해와 5종의 작물을 통합하여 모델을 구축했다는 점에서 몇 가지 한계를 가진다. 우선, 5종의 작물을 통합 분석하였기에 잎의 형태나 질감이 상이한 작물별 특성을 세밀하게 반영하지 못했다. 현재로서는 개별 작물 데이터 수가 적어, 작물별로 특화된 예측 모델을 구축할 경우 심각도 점수의 신뢰도를 확보하기 어렵다. 또한, 본 연구는 흰가루병을 대상으로 방법론의 유효성을 검증하였으므로, 제안된 점수 도출 방식이 다른 병해에도 일반화 가능할지에 대한 추가 검증이 필요하다. 표준화된 병해 심각도 기준이 부재한 상황에서, 향후 작물별 데이터 및 다양한 병해 데이터가 충분히 축적된다면, 개별 작물과 병해의 특성을 고려한 더 정교한 심각도 점수 도출 모델을 구축하는 확장 연구가 향후 중요한 과제가 될 것이다.

끝으로, 본 연구에서 제안한 Grad-CAM 기반 심각도 점수 도출 방법은 모델의 분류 결과를 보다 투명하게 제시하고, 농업 AI 시스템의 현장 수용성을 높이는 데 기여할 수 있는 실질적인 해법으로 평가된다. 향후 데이터의 축적 및 질적 향상을 통해 더욱 정밀하고 신뢰도 높은 병해 심각도 분류가 가능해질 것으로 전망된다.

참고문헌

- Alhammad, S. M., Khafaga, D. S., El-hady, W. M., Samy, F. M., and Hosny, K. M. (2025), Deep learning and explainable AI for classification of potato leaf diseases, *Frontiers in Artificial Intelligence*, 7, 1449329. <https://doi.org/10.3389/frai.2024.1449329>
- Carew, J. and Adams, L. (2011), Comparison of five non-pesticide mildew control products on strawberry, *Agriculture and Horticulture*

- Development Board, HDC Project SF 113, Edward Vinson Ltd., Faversham, Kent, UK.
- Chen, R., Qi, H., Liang, Y., and Yang, M. (2022), Identification of Plant Leaf Diseases by Deep Learning Based on Channel Attention and Channel Pruning, *Frontiers in Plant Science*, **13**, 1023515. <https://doi.org/10.3389/fpls.2022.1023515>
- Elliott, J., Meva, J., Fukai, S., and Leach, J. E. (2022), A comparison of ImageJ and machine learning based image analysis methods to measure cassava bacterial blight disease severity, *Plant Methods*, **18**, 86.
- Fuentes, A., Yoon, S., Kim, S. C., and Park, D. S. (2017), A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition, *Sensors*, **17**(9), 2022. <https://doi.org/10.3390/s17092022>
- Glawe, D. A. (2008), The Powdery Mildews: A Review of the World's Most Familiar (Yet Poorly Known) Plant Pathogens, *Annual Review of Phytopathology*, **46**, 27-51.
- Gopalan, K., Srinivasan, S., Pragya, Singh, M., Mathivanan, S. K., and Moorthy, U. (2025), Corn leaf disease diagnosis: Enhancing accuracy with ResNet152 and Grad-CAM for explainable AI, *BMC Plant Biology*, **25**, 440. <https://doi.org/10.1186/s12870-025-06386-0>
- He, K., Zhang, X., Ren, S., and Sun, J. (2016), Deep residual learning for image recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.
- Kang, H. J., Kim, Y. S., Han, B. T., Kim, T. I., Noh, J. W., Kim, Y. G., and Shin, H. D. (2014), Alternative fungicide spraying for the control of powdery mildew caused by *Sphaerotheca fusca* on greenhouse watermelon (*Citrullus lanatus*), *Research in Plant Disease*, **20**(1), 31-36.
- Karim, M. J., Goni, M. O. F., Nahiduzzaman, M., Ahsan, M., Haider, J., and Kowalski, M. (2024), Enhancing agriculture through real-time grape leaf disease classification via an edge device with a lightweight CNN architecture and Grad-CAM, *Scientific Reports*, **14**, 16022. <https://doi.org/10.1038/s41598-024-66989-9>
- Kim, J.-H., Lee, K.-K., Lim, J.-R., Kim, J., Choi, I.-Y., Jang, S.-J., Kim, J.-H., and Song, Y.-J. (2016), Yield Loss Assessment and Determination of Control Thresholds for Powdery Mildew of Eggplant (*Solanum melongena*), *Korean Journal of Pesticide Science*, **20**(2), 145-151.
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, A., Gustafson, L., Xiao, T., Whitehead, S., and Dollár, P. (2023), Segment Anything, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, **2023**, 1860-1872.
- Kohavi, R. (1995), A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, 1137-1145.
- Li, D., Lin, Z., Wang, Y., Huang, J., and Liu, Y. (2022), Attention-optimized DeepLabV3+ for automatic estimation of cucumber disease severity, *Plant Methods*, **18**, 109.
- Lin, C., Gong, L., Jiang, X., and Zhang, Y. (2019), Deep learning-based segmentation and quantification of cucumber powdery mildew using CNN, *Frontiers in Plant Science*, **10**, 155.
- Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016), Using Deep Learning for Image-Based Plant Disease Detection, *Frontiers in Plant Science*, **7**, 1419. <https://doi.org/10.3389/fpls.2016.01419>
- Pukkela, P. and Borra, S. (2017), Machine learning based plant leaf disease detection and severity assessment techniques: State-of-the-art, In: Classification in BioApps, *Lecture Notes in Computational Vision and Biomechanics*, **26**, 59-83, Springer, Cham.
- Quach, L.-D., Nguyen Quoc, K., Nguyen Quynh, A., Tran Ngoc, H., and Nguyen, T.-N. (2024), Tomato health monitoring system: Tomato classification, detection, and counting system based on YOLOv8 model with explainable MobileNet models using Grad-CAM++, *IEEE Access*, **12**, 9719-9734. <https://doi.org/10.1109/ACCESS.2024.3351805>
- Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., and Hughes, D. P. (2017), Deep Learning for Image-Based Cassava Disease Detection, *Frontiers in Plant Science*, **8**, 1852. <https://doi.org/10.3389/fpls.2017.01852>
- Rani, S., Singh, J., Kaur, A., and Kaur, P. (2023), Novel plant disease detection techniques - a brief review, *Molecular Biology Reports*, <https://doi.org/10.1007/s11033-023-08838-y>
- Rashid, M. R. A., Korim, M. A. E., Hasan, M., Ali, M. S., Islam, M. M., Jabid, T., Islam, R. U., and Islam, M. (2025), An ensemble learning framework with explainable AI for interpretable leaf disease detection, *Array*, **26**, 100386. <https://doi.org/10.1016/j.array.2025.100386>
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018), MobileNetV2: Inverted residuals and linear bottlenecks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4510-4520.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017), Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, 618-626. <https://doi.org/10.1109/ICCV.2017.74>
- Sharma, G. K. (2021), *Cucurbits Disease: Special Reference to Powdery Mildew*, Astral International, New Delhi, India.
- Simonyan, K. and Zisserman, A. (2015), Very deep convolutional networks for large-scale image recognition, *Proceedings of the International Conference on Learning Representations (ICLR)*, arXiv:1409.1556.
- Singh, R., Koundal, D., and Chouhan, D. S. (2024), PDSE-Lite: Lightweight framework for plant disease severity estimation based on convolutional autoencoder and few-shot learning, *Frontiers in Plant Science*, **15**, 1319894.
- Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., and Stefanovic, D. (2016), Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification, *Computational Intelligence and Neuroscience*, **2016**, 3289801. <https://doi.org/10.1155/2016/3289801>
- Tan, M. and Le, Q. V. (2019), EfficientNet: Rethinking model scaling for convolutional neural networks, *Proceedings of the International Conference on Machine Learning (ICML)*, 6105-6114.
- Willoquet, L., Sombardier, A., Blancard, D., Jolivet, J., and Savary, S. (2008), Spore dispersal and disease gradients in strawberry powdery mildew. In *Canadian Journal of Plant Pathology* (Vol. 30, No. 4, pp. 434-441). Taylor & Francis. <https://doi.org/10.1080/07060660809507541>
- Zang, H., Wang, C., Zhao, Q., Zhang, J., Wang, J., Zheng, G., and Li, G. (2025), Constructing segmentation method for wheat powdery mildew using deep learning, *Frontiers in Plant Science*, **16**, 1524283. <https://doi.org/10.3389/fpls.2025.1524283>
- Zhao, Y., Liu, J., Zhang, R., Sun, X., and Wang, Z. (2025), Plant disease segmentation networks for fast automatic severity estimation under natural field scenarios, *Agriculture*, **15**(6), 583.
- Zuiderveld, K. (1994), Contrast limited adaptive histogram equalization.

In P. S. Heckbert (Ed.), *Graphics Gems IV* (pp. 474-485), Academic Press.

저자소개

이종욱: 한양대학교 산업융합학부 정보공학전공 학사 과정에 재학 중이다. 관심 연구 분야는 이미지 모델에 특화된 설명가능 인공지능 기법을 활용한 사람 중심 설명 지원이다.

이상연: 한양대학교 산업융합학부 정보공학전공 학사 과정에 재학 중이다. 관심 연구 분야는 머신러닝을 활용한 데이터 분석과 예측 모델 개발이다.

장원준: 한양대학교 산업융합학부 정보공학전공 학사 과정에 재학 중이다. 관심 연구 분야는 설명가능 인공지능 기법을 활용한 모델의 해석과 이해이다.

정준각: 포항공과대학교 산업경영공학과에서 2013년 학사, 2019년 석박사 통합 학위를 취득하였다. 일리노이 대학교 어바나-샴페인과 울산과학기술원 산업공학과에서 박사후연구원으로 근무했으며, 현재 한양대학교 산업융합학부에서 조교수로 재직 중이다. 관심 연구 분야는 텍스트 마이닝, 품질 데이터 분석, 설명가능 인공지능 응용이다.