

딥러닝 기술을 활용한 복숭아 ‘미황’의 성숙도 자동 분류

이상준^{1†} · 신미희^{2†} · 수간디 히루시카³ · 우펙샤 달샤니³ · 이슬기⁴ · 조정건⁴ · 장시형⁴ · 조병관¹ · 김진국^{2,5*} 

¹충남대학교 농업생명과학대학 바이오시스템 기계공학과, ²경상국립대학교 농업생명과학연구원,

³경상국립대학교대학원 응용생명과학부, ⁴국립원예특작과학원 과수과,

⁵경상국립대학교 농업생명과학대학 원예과학부

Prediction of Fruit Maturity of ‘Mihwang’ Peach using Deep Learning

Sang Jun Lee^{1†}, Mi Hee Shin^{2†}, L. Sugandhi Hirushika Jayasooriya³,
W.M. Upeksha Darshani Wijethunga³, Seul Ki Lee⁴, Jung Gun Cho⁴, Si Hyeong Jang⁴,
Byoung-Kwan Cho¹, and Jin Gook Kim^{2,5*} 

¹Department of Biosystems Machinery Engineering, Chungnam National University, Daejeon 34134, Korea

²Institute of Agriculture and Life Sciences, Gyeongsang National University, Jinju 52828, Korea

³Division of Applied Life Science, Gyeongsang National University, Jinju 52828, Korea

⁴Fruit Research Division, National Institute of Horticultural and Herbal Science, Wanju 55365, Korea

⁵Division of Horticultural Science, College of Agriculture and Life Science, Gyeongsang National University, Jinju 52828, Korea

*Corresponding author: jgkim119@gnu.ac.kr

[†]These authors contributed equally to this work.

Received: March 30, 2023

Revised: May 10, 2023

Accepted: June 21, 2023

 OPEN ACCESS



HORTICULTURAL SCIENCE and TECHNOLOGY
42(1):80-93, 2024
URL: <http://www.hst-j.org>

pISSN : 1226-8763
eISSN : 2465-8588

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright©2024 Korean Society for Horticultural Science.

This work was carried out with the support of “Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ0156462023)” Rural Development Administration, Republic of Korea.

Abstract

Peach must be delivered to market when at their proper ripeness, as its fruit quality declines quickly after harvest. Therefore, it is necessary to consider suitable ripeness for consumption and distribution. However, research on ripeness judgments for peaches in the orchard is scarce. This study used deep learning technology to develop a ripeness classification model for ‘Mihwang’ peaches. A dataset was prepared using 2,800 images, each taken from a peach orchard (outside dataset) and a laboratory (inside dataset) with the same fruit. The dataset was constructed based on the harvest date of the peaches and the peach apex’s skin color (a^* value). It uses three classes, immature, ripe, and overripe, according to the classification criteria of the two datasets. The model was trained with a ratio of 7:2:1 of training data, validation data, and test data, and image data augmentation was carried out to improve the diversity of the data and to solve any imbalances. Among EfficientNet, YOLOv5, and Vision Transformer, the deep learning model recorded the best classification model performance on EfficientNet. Based on the classification model and performance evaluation index, the harvest-date-based classification model achieved the highest accuracy of 100%. The classification model based on the apex color a^* value of peaches showed high accuracy with a minimum rate of 94.7% and a maximum rate of 98.2%. The peach ripeness classification model developed in this study can be used for determining the proper time for the mechanical harvesting of fruit from an orchard.

Additional key words: EfficientNet, fruit firmness, robot harvester, Vision Transformer, YOLOv5

서 언

최근 들어 인공지능 기술의 확산, 빅데이터 생산과 처리를 위한 네트워크 구축, 스마트팜 연구 인프라 확장으로 인해 원예작물 생산현장에서의 노동력 부족 문제, 생산성 향상, 품질 개선을 위한 다양한 연구 결과가 보고 되고 있으며, 이와 같은 연구로 딥러닝을 활용한 사과 과원에서 수채와 과실의 생육 단계 판별(Bang et al., 2022), 복숭아 기계수확을 위한 영상 이미지 분석(Shin et al., 2022), 합성곱 신경망을 이용한 복숭아 외관 품질 평가(Lee et al., 2020), 식물공장에서 상추의 정식 후 일수 예측(Baek et al., 2023), 시설재배 토마토의 수량 예측 및 성숙도 예측(Seo et al., 2021), 양액시설재배 환경에서의 파프리카의 증산률 평가(Nam et al., 2019), 한라봉 가격 예측 모델 연구(Jung and Cho, 2022)를 통해 재배 및 수확작업의 자동화, 신속한 수량예측과 가격예측, 품질 손실을 최소화할 수 있는 기술들을 농업생산 전체에 확대할 수 있을 것으로 기대된다.

본 연구에서는 RGB 이미지를 활용하여 복숭아 로봇 수확을 위한 과실숙도 분류 모델을 개발하기 위해 딥러닝 이미지 분류 알고리즘을 활용하여 모델을 구축하였다. 이미지 분류 작업은 합성곱 신경망을 활용하여 객체의 특징을 추출하는 Convolutional Neural Network(CNN)(LeCun et al., 1989)에 기반을 두고 있다. CNN은 이미지를 인식하기 위한 패턴을 찾는 것에 유용하고 공간 정보를 유지한 채 학습을 할 수 있어 객체의 형태와 크기에 연관된 특징을 추출할 수 있다. CNN 알고리즘 개발 이후 하드웨어의 성능 향상과 GPU 병렬 처리 방식의 활용으로 딥러닝 기술은 컴퓨터 비전 분야에서 우수한 성능을 증명하였고, 이후 Alexnet(Krizhevsky et al., 2017), GoogleNet(Szegedy et al., 2015) 그리고 VGG(Simonyan and Zisserman, 2014)와 같은 다양한 딥러닝 모델들이 개발되었다.

하지만 컴퓨터 비전 모델들은 고정된 크기의 convolution filter size를 사용하기 때문에 receptive field를 벗어난 pixel과의 연관성을 배울 수 없다. 이러한 단점을 자연어 처리의 Self-Attention과 Transformer(Vaswani et al., 2017)를 활용하여 해결하였고 Vision Transformer(Dosovitskiy et al., 2020)와 Swin-transformer(Liu et al., 2021)등의 모델이 개발되었다.

본 연구에서는 최근 이미지 분류에서 우수한 성능을 보였던 EfficientNet(Tan and Le, 2019), 빠른 추론 속도와 높은 정확도로 실시간 탐지 분야에서 우수한 성능을 보인 YOLOv5(Bae et al., 2022), 컴퓨터 비전 영역을 자연어 처리 영역에 접목시켜 개발한 Vision Transformer를 활용하여 복숭아 기계수확 수확 시 적용가능한 복숭아 숙도 분류 모델을 개발하고자 실시하였다.

재료 및 방법

Data Sets

날짜 기준 복숭아 분류

본 연구의 실험 샘플은 경상남도 진주시 문산면 소재 과원에서 재배중인 무봉지 재배 복숭아 ‘미황’을 사용하였다. 복숭아 ‘미황’ 데이터는 실험실 내부(lab data) 및 야외(field data), 두 가지 조건에서 촬영하였다(Fig. 1). 실험실 내부에서 촬영한 데이터는 6,500K 주광색 60W LED 조명 아래에서 SONY 알파 A5000(SONY Inc., Japan) 카메라와 Samsung Galaxy S10 smart phone(Samsung Inc., Korea)을 활용하여 수집하였고, 야외에서 촬영된 데이터는 오전 9시부터 10시 사이에 태양광 아래에서 Samsung Galaxy S10 smart phone과 iPhone 11 Pro smart phone(Apple Inc., California, USA)을 활용하여 수집하였다.

복숭아 ‘미황’ 샘플을 수집하기 위해 4개의 나무를 선정하였고, 각 나무 당 5개의 복숭아를 선택하였다. 해당 과실을 대상으

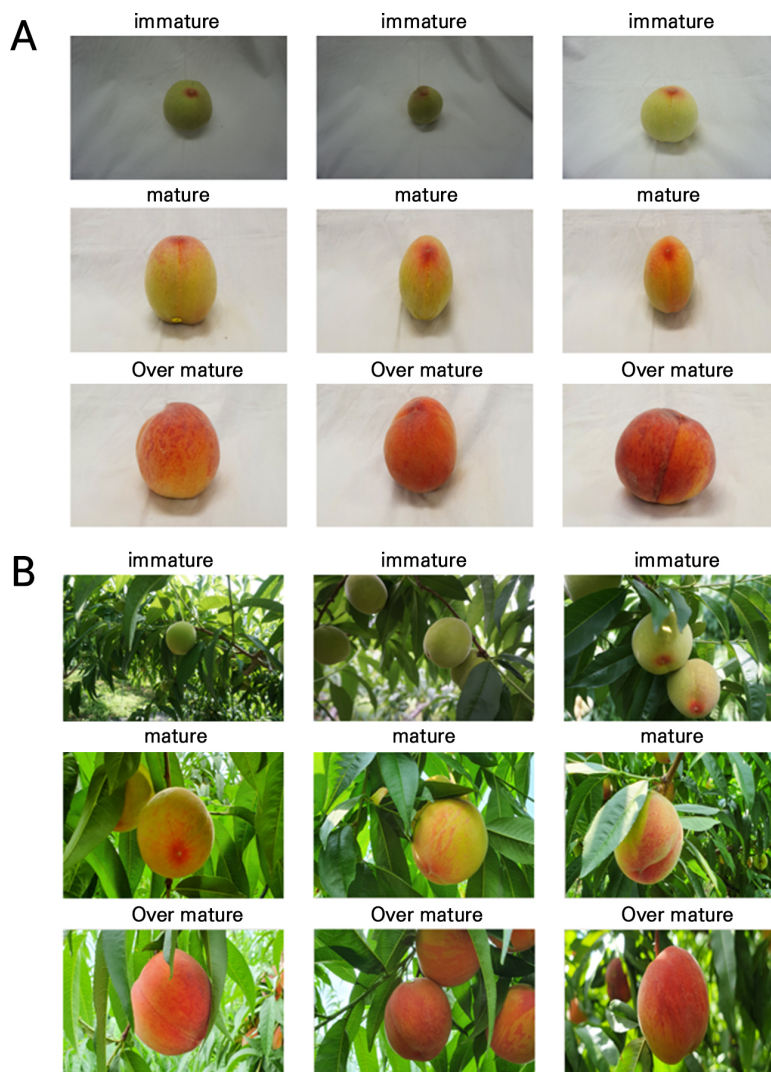


Fig. 1. Sample images of 'Mihwang' peaches using deep learning: lab data (A) and field data (B).

로 20가지의 서로 다른 방향에서 사진을 촬영하여 실험실 내부 및 야외에서 각각 400장의 컬러 영상을 취득하였다. 복숭아 수확은 2022년 6월 10, 13, 16, 20, 23, 27, 30일에 실시하여 내부 및 외부 조건에서 각각 2,800장의 이미지 샘플을 취득하였다. 복숭아 샘플의 class는 과실 성숙 정도를 기준으로 10, 13, 16일은 미숙, 20, 23일은 적숙 그리고 27, 30일은 과숙으로 나누었다.

딥러닝 모델 학습용 데이터는 실험실에서 촬영된 데이터 2,800장, 야외에서 촬영된 데이터 2,800장 그리고 그 둘을 포함한 종합 데이터 5,600장 총 3가지로 구성하였다. 각 데이터는 미숙, 적숙 그리고 과숙 총 3개의 class로 구성하였다(Table 1). Train, Validation 그리고 Test 데이터는 7:2:1의 비율로 나누었고, 데이터가 부족하여 underfitting이 일어나는 것을 방지하기 위해 pytorch(Paszke et al., 2019)의 albumentations(Buslaev et al., 2020) 라이브러리를 활용하여 image augmentation을 실시하였다. 태양광의 조도는 수확 시간에 따라 변한다. 조도와 성숙 정도에 따라 복숭아의 색상이 변하므로 색에 대한 변화를 중요한 특징으로 고려하여 색과 관련된 augmentation을 위주로 이미지를 추가 생성하였다. 기존의 이미지에 Random Brightness, Random Contrast, Hue Saturation Value, Random Gamma, CLAHE(Reza, 2004) 등을 적용하여 Train과 Validation 데이터는 5할 증가되었으며 딥러닝에 활용된 최종 데이터 셋은 Table 1과 같다.

복숭아 과정(apex) 색도 a^* 에 따른 복숭아 분류

복숭아는 성숙 정도에 따라 과실의 색이 초록색에서 붉은색으로 변하므로 과정부(apex)의 색도 a^* 를 기준으로 붉은색의 정도에 따라 숙도를 세 종류로 구분하였고 Train, Validation 그리고 Test 데이터를 Table 2와 같이 7:2:1의 비율로 분류하였다. 색도 a^* 값은 색도계(CR-400, Minolta Co., Tokyo, Japan)를 이용하여 CIE $L^*a^*b^*$ 모드에서 측정하였다. 색도 a^* 의 값이 15 미만이면 미숙, 15이상 21미만이면 적숙 그리고 21 이상은 과숙으로 분류하였다. 복숭아 이미지는 Fig. 1과 같은 이미지를 사용하였기 때문에, 이미지의 화질 및 총 개수는 같지만 분류 기준이 달라 각 class 별 이미지 개수에서 차이가 발생하였다. 총 미숙 1240장, 적숙 540장 그리고 과숙 1020장으로 적숙과 미숙 이미지는 약 두 배의 차이가 발생하였다. 데이터 수량의 차이가 크다면 unbalance 문제가 발생하는데, 이는 소규모의 데이터는 학습이 제대로 이루어지지 않고 딥러닝 학습 시 손실을 감소시키기 위해 대부분 분포도가 높은 class로 예측하여 해당 class에 대한 overfitting 문제를 야기할 수 있다. 그렇기 때문에 data augmentation을 활용하여 unbalance 문제를 해결하였다. Class는 색도 a^* 를 기준으로 분류를 하였기 때문에 albumentations 라이브러리에서, 숙도 분류 기준인 색에 영향을 주지 않는 Horizontal Flip, ShiftScaleRotate를 활용하여 가장 적은 적숙의 양을 두 배로 증가시켜 데이터의 균형을 맞춰주었고, 데이터의 underfitting 예방 및 데이터의 다양성 향상을 위해 Random Brightness, Random Contrast, Hue Saturation Value, RandomGamma 그리고 CLAHE 등을 적용하여 Train과 Validation 데이터의 5할을 증가시켰다. 딥러닝 학습에 활용된 최종 데이터 셋은 Table 2와 같다.

Deep Learning Model

모델 학습 환경 및 하이퍼파라미터 설정

본 연구는 Ubuntu 20.04.4 LTS, Linux version 5.15.0-48-generic, Intel® Xeon® Gold 6230R CPU @ 2.10GHZ, NVIDIA RTX A6000 GPU, 32GB DDR4 RAM으로 구성된 환경에서 전이 학습(transfer learning)(Shaha and Pawar, 2018)을 활용하여 복숭아 숙도 분류 모델을 학습하였다. 최적의 학습 모델을 구축하기 위해 batch size, learning rate, epoch 그리고 optimizer 등의 하이퍼 파라미터를 설정하였다. Epoch은 학습의 반복 횟수를 의미하고 Batch size는 전체 학습 데이터 셋을 작은 그룹으로 나누었을 때, 하나의 그룹에 속하는 데이터의 개수이다. Optimizer는 손실을 최소화하여 가중치를 업데이트하는 알고리즘이며 learning rate는 알고리즘 업데이트 시 보폭을 제어해주는 역할을 한다. Batch size는 32, Learning rate

Table 1. Dataset according to the harvest date of 'Mihwang' peaches

	Lab data			Field data			Total data		
	Immature	Mature	Over-mature	Immature	Mature	Over-mature	Immature	Mature	Over-mature
Train	1260	840	840	1260	840	840	2520	1680	1680
Val	360	240	240	360	240	240	720	480	480
Test	120	80	80	120	80	80	240	160	160
Total	1740	1160	1160	1740	1160	1160	3480	2320	2320

Table 2. Dataset according to the skin color on an apex area of 'Mihwang' peaches

	Lab data			Field data			Total data		
	Immature	Mature	Over-mature	Immature	Mature	Over-mature	Immature	Mature	Over-mature
Train	1302	1134	1071	1302	1134	1071	2604	2268	2142
Val	372	324	306	372	324	306	744	648	612
Test	124	54	102	124	54	102	248	108	204
Total	1798	1512	1479	1798	1512	1479	3596	3024	2958

는 $1e-6$, Epoch은 50 그리고 Optimizer는 AdamW(Loshchilov and Hutter, 2017)으로 최적의 하이퍼파라미터를 설정하였다. 복숭아 색도 a^* 을 기준으로 분류한 데이터 셋은 학습량 및 데이터의 다양성이 증가하였다고 판단해 epoch을 100으로 변경하였다.

EfficientNet-b7

모델의 성능 향상에는 다양한 방법이 존재하지만, 큰 범주에서 3가지로 분류할 수 있다. Network의 depth를 깊게 하는 것, channel width를 늘리는 것 그리고 input 이미지의 resolution을 높여 모델의 성능을 향상시킬 수 있는 방법이 있는데, EfficientNet은 이 3가지의 최적의 조합을 AutoML을 활용하여 도출하였고 조합을 효율적으로 만들 수 있도록 하는 compound scaling 방법을 제안하였다. Compound scaling은 세 가지 변수 depth, width 그리고 resolution에 곱할 factor를 동시에 고려하여 최적의 scaling을 찾는 것을 목표로 하며 모델 및 factor 값의 크기를 기준으로 EfficientNet을 b0부터 b7까지 나누었다(Bae et al., 2022). 해당 연구에서는 우수한 정확도를 달성하기 위하여 가장 큰 factor 값을 갖는 EfficientNet-b7을 활용하여 모델을 구축하였다. 해당 모델은 실험실 내부의 이미지 데이터 셋, 야외의 이미지 데이터 셋 그리고 두 개의 이미지를 합친 데이터 셋을 활용하여 학습을 진행하였고 모두 20 초반의 epoch에서부터 안정적인 양상으로 학습되는 것을 알 수 있다. 색도 a^* 값을 기준으로 분류된 데이터 셋은 60에서 80사이의 범위내에 최적의 가중치가 존재하였다.

YOLOv5

YOLOv5 모델은 빠른 속도가 장점인 1 stage detector YOLO에서 발전된 모델이다. 초기 YOLO는 속도가 빠른 장점 대신 정확도가 상대적으로 낮다는 단점이 있었는데, YOLOv5는 빠른 속도를 유지한 상태로 모델의 정확도를 높여 다양한 분야에서 활용이 되었고 특히 자율주행과 같은 실시간 탐지 모델 개발에 주로 사용되었다. YOLOv5 모델은 빠른 속도와 정확도를 향상시키기 위해 CSPNet backbone을 활용하였다(Kim et al., 2021). CSPNet은 정확도를 유지하면서 경량화가 이루어져 CNN의 학습능력을 강화하고, 각 계층의 연산량을 균등하게 분배해서 연산 bottleneck을 없앨 수 있으며 메모리 cost를 효과적으로 줄여 정확도 향상과 inference time을 감소시킨다(Wang et al., 2020). YOLOv5 모델은 depth multiple과 width multiple의 차이에 따라 s, m, l, x, s6, m6, l6, x6로 나뉘어 여러 버전이 존재하며 s 모델은 빠르지만 정확도가 낮고 x6까지 점진적으로 정확도는 향상되는 반면 속도는 감소한다(Kim et al., 2021). 본 연구에서는 높은 정확도를 달성하기 위해 YOLOv5x6 모델을 활용하여 복숭아 속도 분류 모델을 구축하였다. YOLOv5 또한 세 종류의 데이터 셋을 활용하여 학습 및 검증을 실시하였고 세 경우의 데이터 모두 후반의 epoch에서 최적의 가중치로 수렴하는 것을 알 수 있다. 과정보다 색도 a^* 값 기준 복숭아 분류 데이터에서 또한 후반의 epoch에서 최적의 가중치로 수렴하는 것을 알 수 있다.

Vision Transformer

컴퓨터 비전 영역에서 딥러닝 이미지 분류 알고리즘은 주로 CNN 기반의 모델이 우수한 성능을 기록하였다. 컴퓨터 비전과 자연어 처리의 알고리즘은 개별적인 분야에서 발전을 이룩하였으나 2017년 Self-attention 기법을 활용한 자연어 처리의 Transformer 모델이 다양한 분야에서 우수한 성능을 달성하며 자연어 처리 모델을 활용한 이미지 분류 알고리즘이 개발되기 시작했다. CNN은 고정된 크기의 convolution filter를 사용하기 때문에 receptive field 밖에 있는 pixel과의 연관성을 학습하지 못하지만 Transformer 모델은 특정 pixel과 나머지 모든 pixel과의 연관성을 학습하여 객체에 대해 정밀한 특징을 추출할 수 있다. 본 연구에서 사용한 Vision Transformer는 Transformer의 Encoder 부분인 Self-attention을 응용하여 개발된 모델로 이미지를 여러 개의 patch로 쪼개 특징을 추출한 뒤 flatten 작업 후 Transformer의 encoder에 넣고 classifier를 붙여 학습한다(Vaswani et al., 2017). Large-scale 컴퓨터 비전 데이터 셋에서 CNN을 활용하지 않고 우수한 성능을 달성한 모델이지만 적은 데이터 셋으로 사전 학습 시 우수한 성능을 기대할 수 없기 때문에 많은 양의 데이터 셋이 모델 학습에 중요한 요소로 요구된다.

다. 모델의 학습 양상을 보았을 때, 통합 데이터 학습에서 불안정한 양상으로 학습이 진행되는 것을 알 수 있고 후반의 epoch에서 최적의 가중치로 수렴하는 것을 알 수 있다. 과정부 색도 a^* 를 기준으로 분류한 데이터 셋 또한 후반의 epoch에서 최적의 가중치로 수렴하는 것을 알 수 있다.

모델 성능 평가 지표

Test set을 활용하여 복숭아 숙도 분류 모델의 성능을 검증하였는데, 데이터가 불균형 할 경우 사용된 성능 검증 지표에 따라 결과가 달라질 수 있다. 그러므로 Accuracy, Precision, Recall 그리고 F1-score 등 다양한 성능 검증 지표를 사용하여 불균형 데이터에 대한 예측 결과의 신뢰도를 보정하고 예측 결과에 대한 보완점을 파악할 수 있다.

Accuracy는 전체 샘플의 개수 중 올바르게 예측된 샘플 개수의 비율이다. 하지만 해당 지표를 신뢰하기 위해서는 데이터가 균형을 이루어야 한다. Precision은 정답으로 예측한 샘플 중 실제로 정답인 샘플의 비율이고 recall은 실제 정답 중 모델이 정답이라고 예측한 비율이다. F1-score은 precision과 recall의 조화 평균을 의미한다. TP는 True Positive, TN는 True Negative, FP는 False Positive, FN는 False Negative을 의미한다.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1-score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (4)$$

결과 및 고찰

복숭아 수확일 기준에 따른 데이터 셋 분류 결과

본 연구에서는 복숭아 과원에서 수확용 로봇에 적용될 숙도 분류 알고리즘 개발을 위해 무봉지 재배 복숭아를 대상으로 실험을 진행하였다. 복숭아 ‘미황’ 과실을 대상으로 만개 후 6월 10일부터 30일까지 과실 숙도의 변화를 조사하여 데이터 셋을 구축하였고, 그 중 일부를 Test dataset으로 설정하여 학습에서 제외 후 모델의 성능 평가를 실시하였다(Table 3). EfficientNet-b7 모델의 Test set에 대한 성능 평가 결과는 세 종류의 데이터 셋에서 모두 Accuracy 100%, Precision 100%, Recall 100%, F1-score 100%의 결과를 보였다. YOLOv5 또한 세 종류의 데이터 셋에서 모두 Accuracy 100%, Precision 100%, Recall 100%, F1-score 100%를 보였다. Vision Transformer는 실험실 내부에서 촬영한 데이터에 대하여 Accuracy 96.8%, Precision 96.5%, Recall 96.3% 그리고 F1-score 96.4%를 보였다. 야외 데이터에 대해서는 Accuracy 98.2%, Precision 97.9%, Recall 97.9% 그리고 F1-score는 97.9%이었고, 통합 데이터에 대해서는 Accuracy 94.6%, Precision 94.2%, Recall 93.8% 그리고 F1-score 94.0%를 보였다.

EfficientNet과 YOLOv5 모델의 Test set 결과(Table 3)가 100%를 달성하였기 때문에 Overfitting을 의심할 수 있다. Overfitting이 발생하였다면 Validation Loss는 Training Loss와 반대로 증가하는 추세를 보인다. 그러나 두 모델의 학습 그래

Table 3. Model performance outcomes of EfficientNet, YOLOv5 and Vision Transformer with the harvest date of 'Mihwang' peaches

Model	Maturity	Accuracy	Precision	Recall	F1-score
EfficientNet-b7	Inside	100.0%	100.0%	100.0%	100.0%
	Outside	100.0%	100.0%	100.0%	100.0%
	Combined	100.0%	100.0%	100.0%	100.0%
YOLOv5	Inside	100.0%	100.0%	100.0%	100.0%
	Outside	100.0%	100.0%	100.0%	100.0%
	Combined	100.0%	100.0%	100.0%	100.0%
Vision Transformer	Inside	96.8%	96.5%	96.3%	96.4%
	Outside	98.2%	97.9%	97.9%	97.9%
	Combined	94.6%	94.2%	93.8%	94.0%

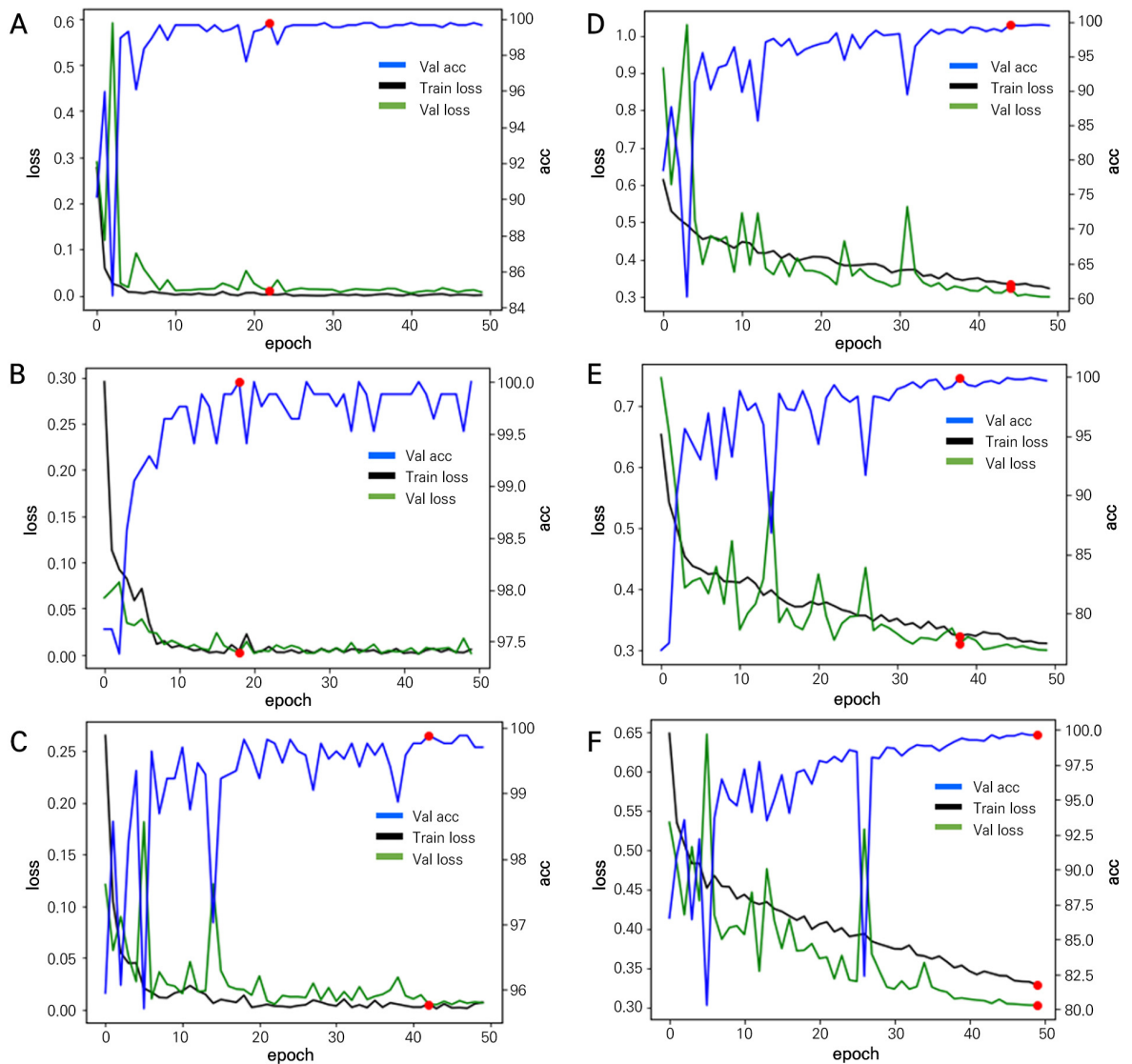


Fig. 2. Convergence rates for the number of iterations for the two networks of EfficientNet - inside (A), outside (B), and combined (C) dataset training aspect, and YOLOv5 - inside (D), outside (E), and combined (F) dataset training aspect.

프 양상은 Train Loss와 Validation Loss 모두 감소하는 형태이다(Fig. 2). 그러므로 EfficientNet과 YOLOv5 모델에서 Overfitting 현상은 일어나지 않은 것으로 판단된다. 하지만 YOLOv5는 불안정한 학습 그래프 양상을 나타내고 Loss의 편차가 심한 것으로 보았을 때 추후 연구에서 안정적인 학습을 위해 batch size 및 learning rate의 재조정이 필요하고 정확한 local minimum 혹은 global minimum으로의 수렴 지점을 파악하기 위해 epoch의 수를 늘리는 것이 분류 모델 정확도 향상에 유의미한 영향을 끼칠 것으로 판단된다.

Vision Transformer 모델의 Test set 결과(Table 3) EfficientNet 및 YOLOv5 모델보다 낮은 정확도를 기록하였고 F1-score에서 최대 6%의 정확도가 감소하였다. Vision Transformer 모델을 활용하여 학습을 진행한 양상을 분석해보면, 학습이 불안정하고 최적의 local minimum 혹은 global minimum이 모두 후반부의 epoch에서 최적값으로 수렴하였다(Fig. 3). 다른 모델들과 마찬가지로 향후 연구에서 epoch의 수를 증가시키고 batch size 및 learning rate를 재조정하여 분류 모델 정확

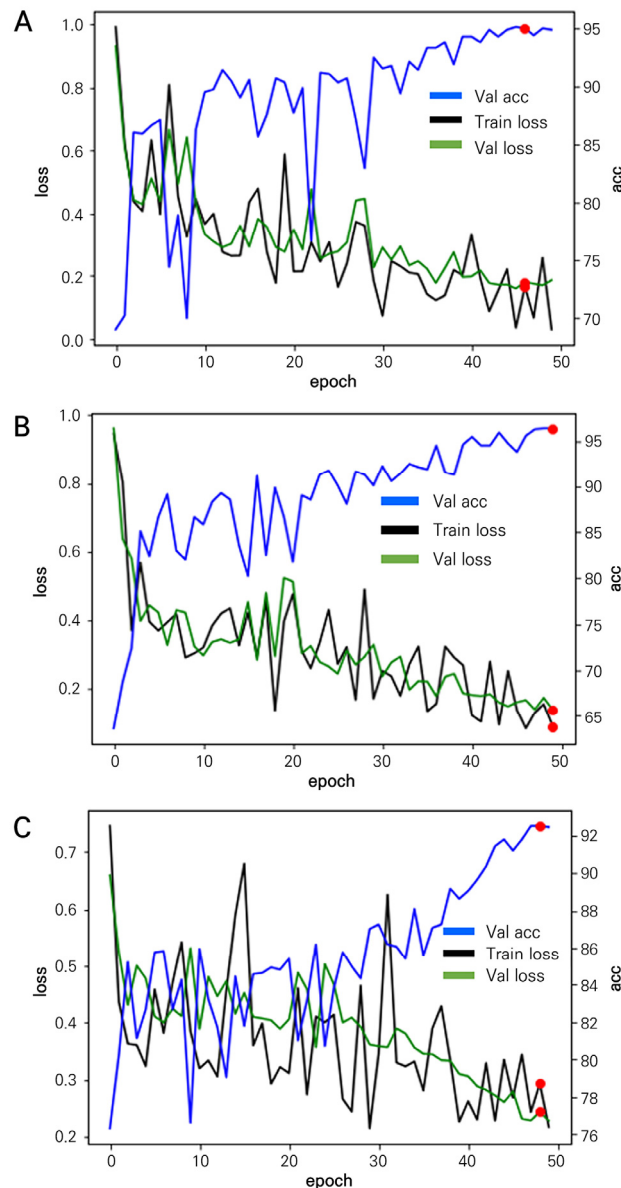


Fig. 3. Convergence rates for the number of iterations for the networks of Vision Transformer - inside (A), outside (B), and the combined (C) dataset training aspect.

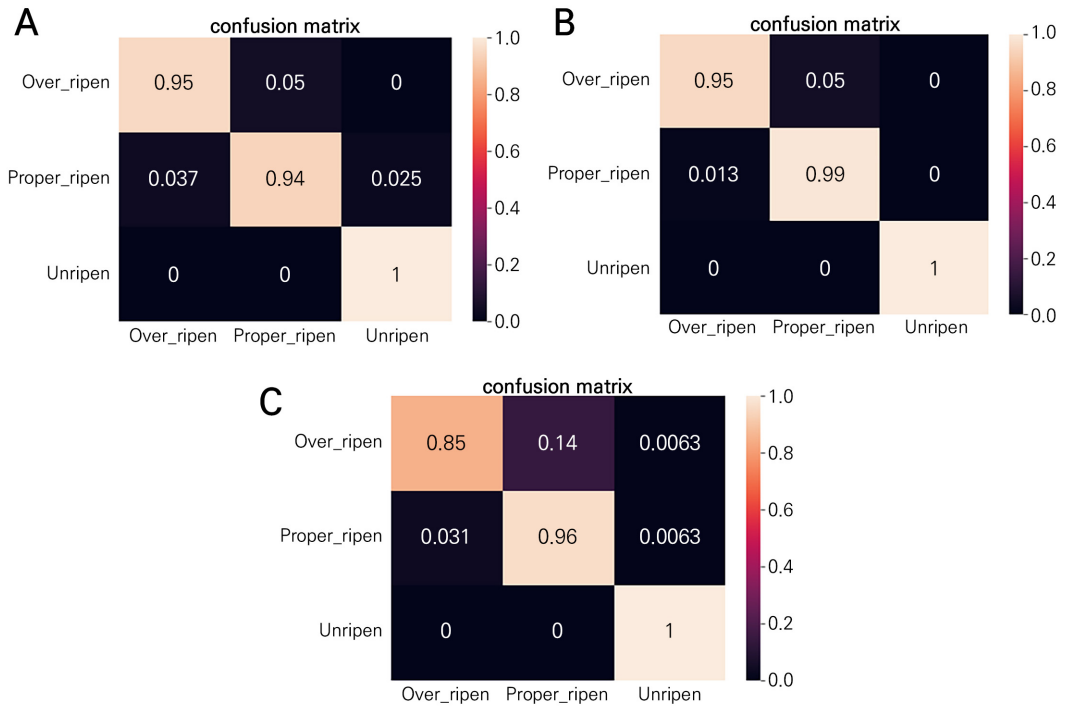


Fig. 4. Confusion matrix of the inside dataset (A), outside dataset (B), and combined dataset.

도 향상을 계획하였다. 실험실 내부 데이터는 confusion matrix(Fig. 4A)에서 보여지는 것과 같이 적숙과 과숙을 혼동하는 오류가 발생하였다. 야외 데이터(Fig. 4B)는 실험실 내부의 데이터와 비슷한 양상을 보여주었는데, 과숙 복숭아를 적숙으로 오인하는 분류 결과가 가장 많은 비율을 차지하였다. 통합 데이터(Fig. 4C)에서 특히 과숙 복숭아를 적숙으로 오인하는 분류 결과가 가장 많은 비율을 차지하였고 일부 적숙과 과숙을 미숙 복숭아로 분류하는 오류도 발생하였다. 대부분 미숙 복숭아의 껍질은 초록색이기 때문에 적숙과 과숙에 비교하여 뚜렷한 특징을 갖고 있어 적은 비율의 분류 오류를 차지한다. 하지만 적숙에서 과숙으로 숙성중인 시기에 촬영된 복숭아는 조명 혹은 태양빛에 의해 영향을 받아 유사한 과피 색으로 표현될 수 있어 가장 많은 오류가 발생한 것으로 추측된다(Fig. 4A). Vision Transformer 모델은 다른 두 모델과 비교하여 상대적으로 높은 오류 비율을 기록하였는데, 충분하지 않은 epoch 수와 많은 양의 데이터를 필요로 하는 모델의 특성상, 적은 학습 데이터 셋으로 인하여 발생한 것으로 판단된다. 학습 데이터 양을 증가시켜 데이터 셋의 모집단 및 다양성이 증가한다면, 향후 분류 모델 성능 개선에 큰 기여를 할 것으로 판단된다.

복숭아 과정부(apex) 색도 a^* 에 따른 데이터 셋 분류 결과

Table 2와 같은 Test set을 활용하여 색도 a^* 를 기준으로 분류된 데이터 셋의 딥러닝 학습 모델을 검증하였다. 복숭아 과정부의 색은 동일한 날에 촬영되었더라도 개체별로 상이할 수 있어 데이터의 다양성을 증가시켰다. 다양성 및 개수가 증가된 딥러닝 학습 양상은 local minimum 혹은 global minimum에 대한 수렴 속도가 늦어지고 날짜 기준으로 분류한 모델과 비교하여 더욱 높은 loss를 기록하였다. EfficientNet 학습 양상 그래프에서 실험실 내부 데이터 및 통합 데이터에 대한 loss가 일정 수치 이하로 감소하지 않는 것을 알 수 있고, YOLOv5와 Vision Transformer에서 또한 날짜 기준 데이터 셋의 모델과 비교하여 높은 loss를 기록하는 것을 확인하였다(Fig. 5). 색도 a^* 로 분류한 데이터 셋에서 일정 수치 이하로 loss가 감소하지 않는 것은, 촬영된 이미지 중 과 정부가 보이지 않거나 혹은 과 정부만 붉은색을 보이고 그 외의 부분에서 미숙과 같은 초록색을 띄는 과실이 존재하는 등의 사항들이 모델 학습에 영향을 끼친 것으로 판단된다. YOLOv5와 Vision Transformer의 그래프 양상을 보면

(Fig. 5), train loss와 val loss는 지속적으로 감소하는 것을 알 수 있는데, Epoch을 증가시켜 모델을 학습한다면 분류 정확도 향상 및 loss 값의 감소를 기대할 수 있을 것으로 판단된다. 위에서 언급한 다양한 이유들로 인해, 색도 a^* 기준 모델의 분류 정확도는 날짜 기준 분류보다 낮은 정확도를 기록하였다. 모델 성능 평가 기준은 Accuracy, Precision, Recall 그리고 F1-score로 설정하였고 3개의 모델 중 EfficientNet에서 가장 우수한 분류 정확도를 확인할 수 있었다(Table 4). EfficientNet은 실험실 내부 데이터에서 96.4%, 95.6%, 95.3%, 95.4%, 야외 데이터에서 98.2%, 97.9%, 97.8%, 97.8% 그리고 통합 데이터에서 95.9%, 95.2%, 94.3%, 94.7%의 분류 정확도를 달성하였다. 야외 데이터 보다 통합 데이터에서 정확도가 감소한 것으로 보았을 때, 실험실 내부 데이터 셋 중 위에서 언급한 것과 같이 명확한 분류 특징을 확인하기 다소 어려운 이미지들이 있어 정확도에 영향을 끼친 것으로 판단된다.

YOLOv5와 Vision Transformer 모델 또한 이와 같은 맥락으로 정확도가 감소한 것으로 사료된다. 세 가지 모델의 confusion matrix를 참고하였을 때(Figs. 6, 7 and 8), 적숙에서 가장 낮은 분류 정확도를 달성하였고 대다수의 오류는 적숙과 과숙을 혼동하여 발생하였다. 색도 a^* 값을 기준으로 데이터를 분류하였을 때, 적숙과 과숙 사이에 흡사한 이미지가 가장 많이 분포하였다. 색도 a^* 값은 측정하는 도구, 측정 위치, 측정 각도 및 측정하는 사람 등에 따라 달라질 가능성이 있는데, 적숙과 과숙의 색이 유사한 만큼 색도 a^* 값도 유사하여 다수의 흡사한 이미지가 존재하였다. 그렇기 때문에 딥러닝 모델 학습 시 몇몇 샘플에 대해 적숙과 과숙을 완전히 구별 가능한 특징 탐색에 어려움이 있어 분류 정확도의 감소를 야기한 것으로 판단된다. 본 연구에서 발생한 다양한 문제점들을 극복하기 위해 복숭아 데이터의 개수 및 다양성을 향상시키고 명확한 특징으로 구분할 수

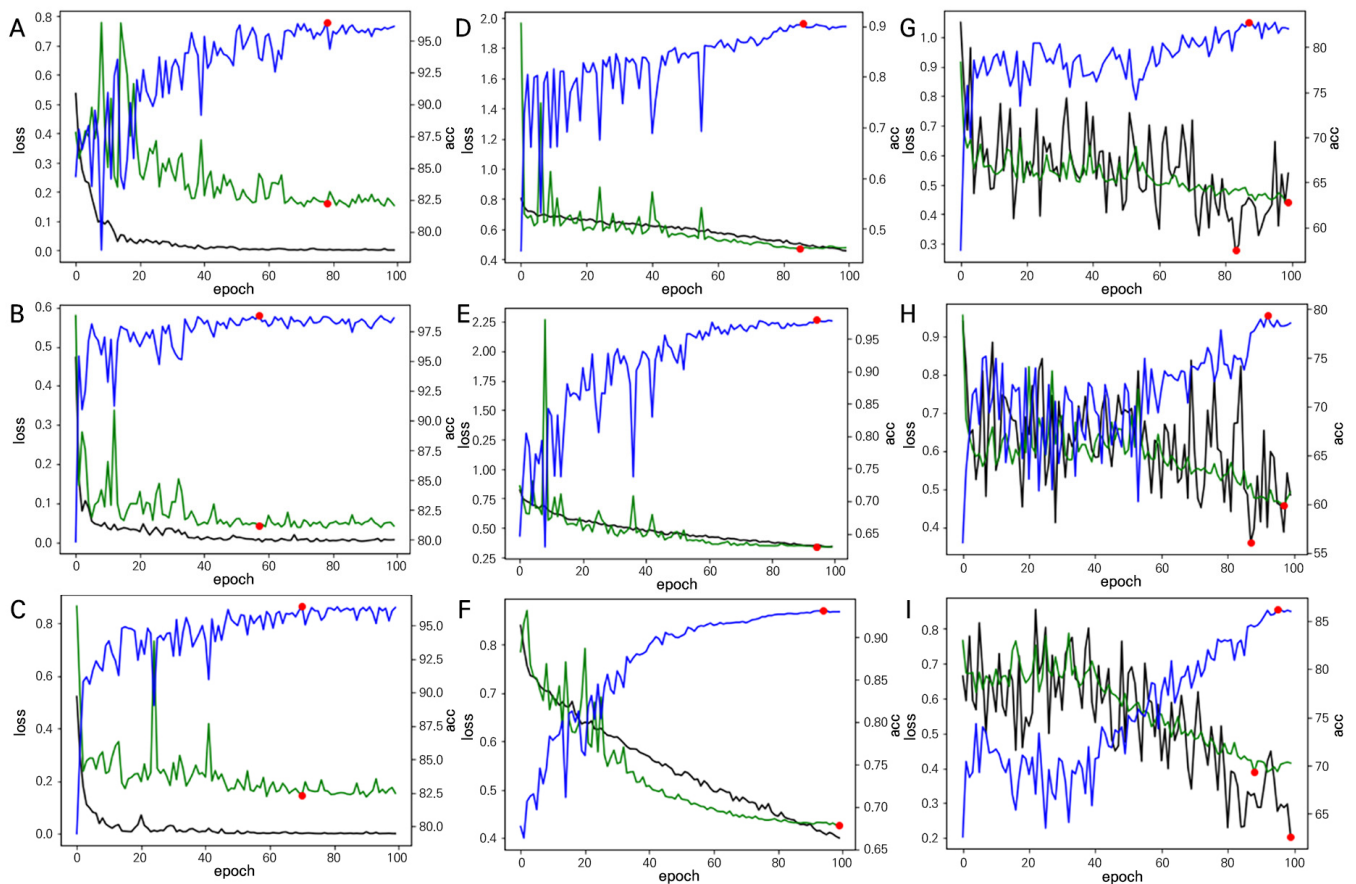


Fig. 5. Convergence rates for the number of iterations for the three networks of EfficientNet - inside (A), outside (B), and combined (C) dataset training aspect, YOLOv5 - inside (D), outside (E), and combined (F) dataset training aspect, and Vision Transformer - inside (G), outside (H), and combined (I) dataset training aspect.

있는 기준점을 찾으며 최신 딥러닝 모델 적용 및 하이퍼파라미터를 개선한다면 추후 복숭아 분류 모델 개발에서 분류 모델 성능 향상이 될 수 있을 것으로 사료된다.

본 연구에서 사용한 세 개의 딥러닝 모델 중 EfficientNet 모델이 가장 우수한 성능을 달성하였고, 수확 날짜 기준 분류 성능 평가 지표에서 100%의 정확도를, 색도 a^* 값 기준 분류 성능 평가에서는 최저 94.7%, 최고 98.2%의 정확도를 기록하였다. 색도 a^* 값 기준 분류 데이터 셋에서 적숙 class에 대한 분류 오류가 가장 높은 비율을 차지하였다. 이미지에서 복숭아의 과정부가 보이지 않거나, 색도 a^* 값으로 나누기에 모호한 이미지가 존재하여 적숙과 과숙의 명확한 분류 특징을 찾지 못하는 등의 이유로 분류 오류가 발생하였다고 판단된다. 이미지의 촬영 환경의 통일성, 구분 기준의 명확성, 기준의 정확한 촬영 및 데이터의

Table 4. Model performance outcomes of EfficientNet, YOLOv5 and Vision Transformer with the skin color of an apex area of 'Mihwang' peaches

Model	Maturity	Accuracy	Precision	Recall	F1-score
EfficientNet-b7	Inside	96.4%	95.6%	95.3%	95.4%
	Outside	98.2%	97.9%	97.8%	97.8%
	Combined	95.9%	95.2%	94.3%	94.7%
YOLOv5	Inside	92.8%	92.5%	90.4%	91.4%
	Outside	96.1%	95.3%	95.3%	95.3%
	Combined	93.9%	91.5%	92.5%	91.9%
Vision Transformer	Inside	84.3%	80.7%	80.3%	80.5%
	Outside	85.0%	82.4%	83.2%	82.8%
	Combined	88.2%	85.0%	86.1%	85.5%

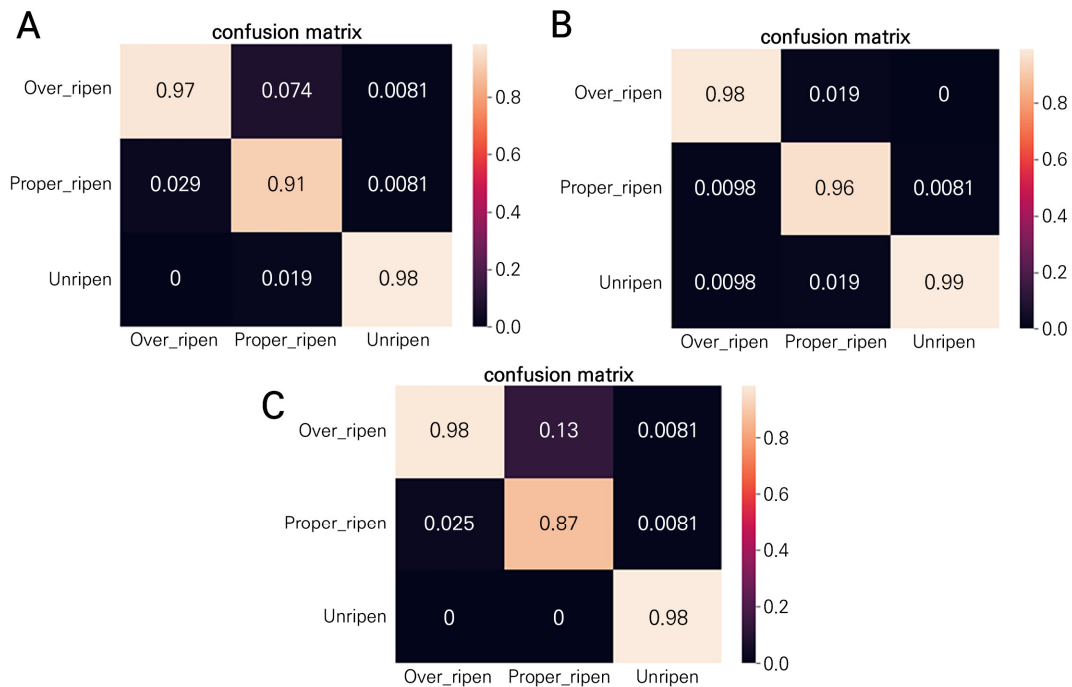


Fig. 6. EfficientNet confusion matrix: inside (A), outside (B), and combined (C).

증량과 다양성을 확보한다면 추후 연구에서 분류 모델의 정확도 향상에 기여하고 과실 상품성 및 저장성을 향상시키며 복숭아 기계수확 적기 판정 연구에 적합한 모델로 선정될 것으로 사료된다.

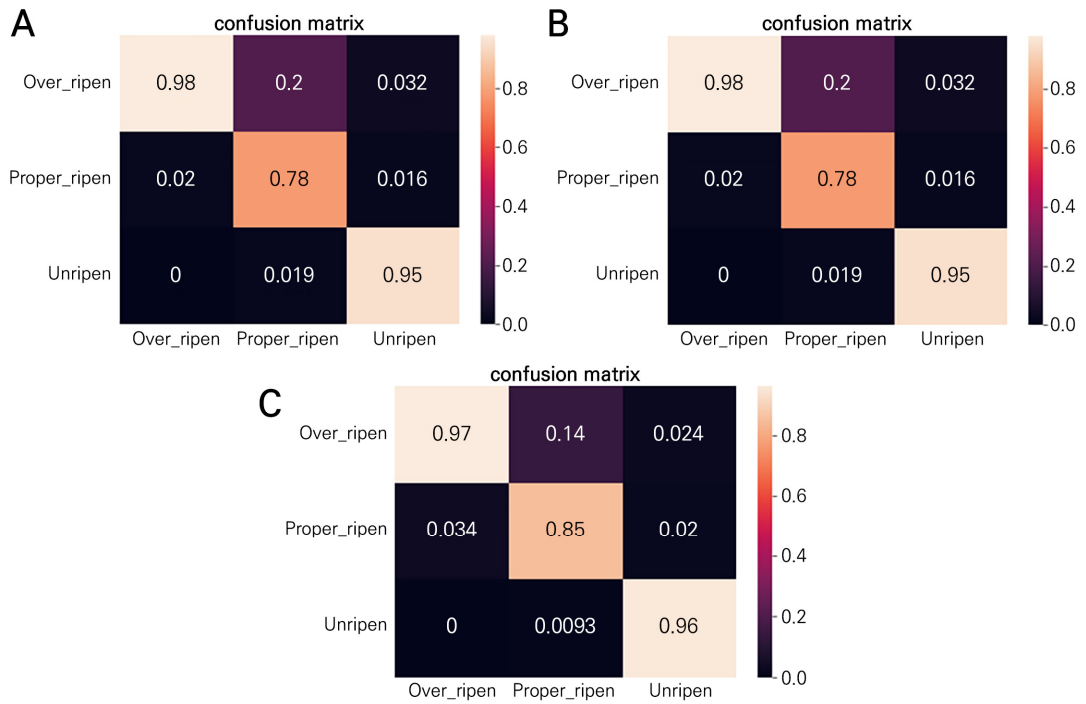


Fig. 7. YOLOv5 confusion matrix: inside (A), outside (B), and combined (C).

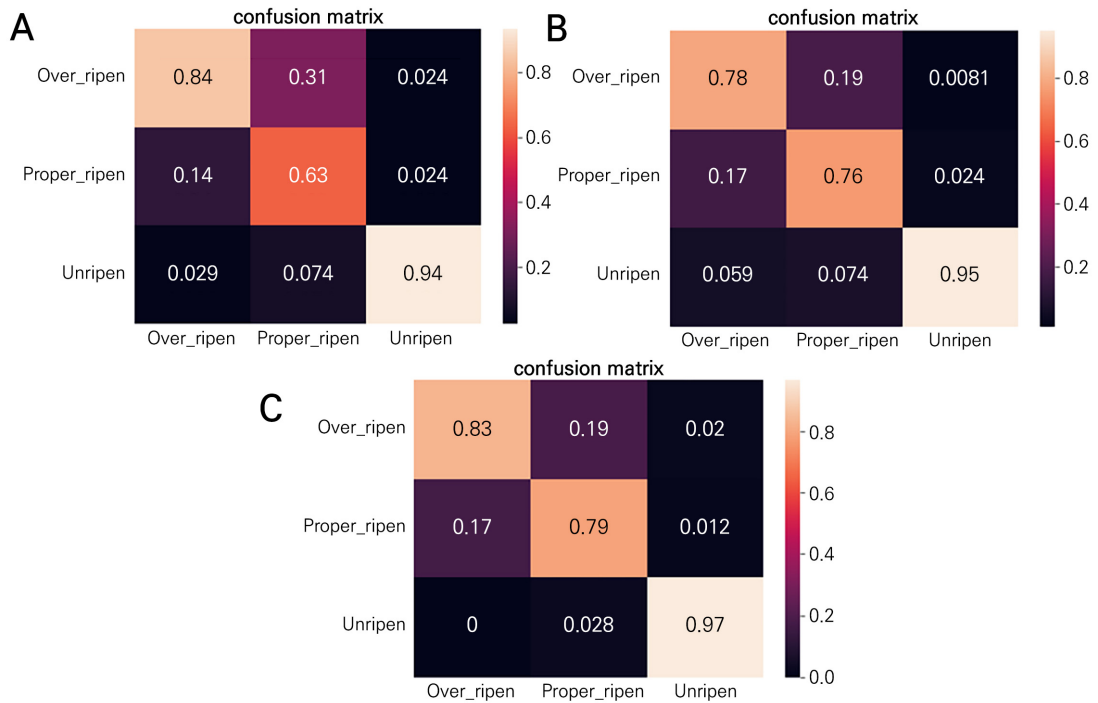


Fig. 8. Vision Transformer confusion matrix: inside (A), outside (B), and combined (C).

초 록

소비자에게 전달되는 복숭아는 속도에 따라서 품질이 달라지기 때문에 섭취하기에 적합한 속도를 고려하여 유통하는 과정이 필요하다. 또한, 속도는 복숭아의 상품성 및 저장성에 영향을 미칠 수 있어 적합한 수확 시기를 선정하는 작업이 요구되지만, 현재 노지 과수 작목의 속도 판별에 대한 국내 연구는 미미한 실정이다. 그렇기 때문에 본 연구에서는 딥러닝 객체 탐지 분류 모델을 활용하여 복숭아 '미황'에 대한 속도 분류 모델을 개발하였다. 실험실 내부 및 야외에서 촬영된 각 2,800장의 이미지를 활용하여 데이터 셋을 구축하였고, 수확 날짜 및 복숭아 과정부(apex)의 색도 a^* 값을 기준으로 하는 두 개의 데이터 셋으로 구성하여 각 셋의 구분 기준에 따라 미숙, 적숙 그리고 과숙 3개의 class로 분류하였다. Train : Validation : Test 데이터 셋은 7 : 2 : 1의 비율로 분류하였고 데이터의 다양성 향상 및 unbalance를 해결하기 위해 augmentation을 실시하였다. 딥러닝 모델은 EfficientNet, YOLOv5 그리고 Vision Transformer를 활용하였으며 EfficientNet에서 가장 우수한 분류 모델 성능을 기록하였다. 날짜 기준 분류 모델은 분류 모델 성능 평가 지표 기준 최저 및 최대 100%의 정확도를 달성하였고, 색도 a^* 값 기준 분류 모델은 최저 94.7%, 최대 98.2%의 높은 정확도를 보였다. 본 연구에서 개발된 객체 탐지 기반 복숭아 속도 분류 모델은 향후 노지 과수 작목의 속도 분류를 통한 기계수확 적기 판정 작업에 활용될 수 있을 것으로 판단된다.

추가 주요어 : EfficientNet, 과실경도, 로봇수확, Vision Transformer, YOLOv5

Literature Cited

- Bae SM, Park JS, Ju CY, Lee DH (2022) Selective face de-identification scheme using multiple face recognition and classification techniques. IV International Conference on Advances in Computer Technology 2022, Inf Sci Commun (CTISC): IEEE, Suzhou, China, pp 1-8. doi:10.1109/CTISC54888.2022.9849764
- Baek Y, Sul S, Cho YY (2023) Estimation of days after transplanting using an artificial intelligence CNN (Convolutional Neural Network) model in a closed-type plant factory. Hortic Sci Technol 41:81-90. doi:10.7235/HORT.20230008
- Bang JH, Park J, Park SW, Kim JY, Jung SH, Sim CB (2022) A system for determining the growth stage of fruit tree using a deep learning-based object detection model. Smart Mdeia J 11:9-18. doi:10.30693/SMJ.2022.11.4.9
- Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA (2020) Albumentations: fast and flexible image augmentations. Information 11:125. doi:10.3390/info11020125
- Dosovitskiy A, Beyer A, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, et al. (2020) An image is worth 16x16 words: Transformers for image recognition at scale. Cornell University, New York, USA. doi:10.48550/arXiv.2010.11929
- Jung, DH, Cho, YY (2022) A prediction model for Hallabong tangor product prices using LSTM (Long Short-term Memory) network. Hortic Sci Technol 40:571-577. doi:10.7235/HORT.20220051
- Kim I, Lee MG, Jeon Y (2021) Comparative analysis of defect detection using YOLO of deep learning. J KSMTE 30:514-519. doi:10.7735/ksmte.2021.30.6.513
- Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60:84-90. doi:10.1145/3065386
- LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Backpropagation applied to handwritten zip code recognition. Neural Comput 1:541-551. doi:10.1162/neco.1989.1.4.541
- Lee HJ, Lee WS, Choi IH, Lee CK (2020) Detection model of fruit epidermal defects using YOLOv3: A case of peach. Information System Review 22:113-124. doi:10.14329/isr.2020.22.1.113
- Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: Hierarchical vision transformer using shifted windows. IEEE/CVF International Conference on Computer Vision 2021, pp 10012-10022. doi:10.1109/ICCV48922.2021.00986
- Loshchilov I, Hutter F (2017) Decoupled weight decay regularization. arXiv preprint. doi:10.48550/arXiv.1711.05101
- Nam DS, Moon T, Lee JW, Son JE (2019) Estimating transpiration rates of hydroponically-grown paprika via an artificial neural network using aerial and root-zone environments and growth factors in greenhouses. Hortic Environ Biotechnol 60:913-923. doi:10.1007/s13580-019-00183-z
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, et al. (2019) Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process Syst, Vol 32. doi:10.48550/arXiv.1912.01703
- Reza AM (2004) Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. J VLSI Signal Process Syst Signal Image Video Technol 38:35-44. doi:10.1023/B:VLSI.0000028532.53893.82

- Seo D, Kim KC, Lee M, Kwon KD, Kim G** (2021) Research on tomato flowers and fruits object detection model in greenhouse environment using deep learning J Korean Inst Commu Infor Sci 46:2072-2077. doi:10.7840/kics.2021.46.11.2072
- Shaha M, Pawar M** (2018) Transfer learning for image classification. II International Conference on Electronics 2018, Commun. Aerosp. Technol (ICECA), 2018: IEEE, pp 656-660. doi:10.1109/ICECA.2018.8474802
- Shin MH, Jang KE, Lee SK, Cho JG, Song SJ, Kim JG** (2022) Grading of harvested 'Mihwang' peach maturity with convolutional neural network. J Bio-Envir Con 31:270-278. doi:10.12791/KSBEC.2022.31.4.270
- Simonyan K, Zisserman A** (2014) Very deep convolutional networks for large-scale image recognition. Cornell University, New York, USA. arXiv:1409.1556. doi:10.48550/arXiv.1409.1556
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A** (2015) Going deeper with convolutions. IEEE Conference on Computer Vision and Pattern Recognition, MA, USA, pp 1-9. doi:10.1109/CVPR.2015.7298594
- Tan M, Le Q** (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning. PMLR, California, USA, pp 6105-6114. doi:10.48550/arXiv.1905.11946
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I** (2017) Attention is all you need. Advances in Neural Information Processing Systems 30, California, USA, doi:10.48550/arXiv.1706.03762
- Wang CY, Liao HYM, Wu YH, Chen PY, Hsieh JW, Yeh IH** (2020) CSPNet: A new backbone that can enhance learning capability of CNN. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, WA, USA, pp 390-391. doi:10.1109/CVPRW50498.2020.00203