A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Accurate Prediction of the Fruit Development Stage of Sweet Pepper Using an Ensemble Model of Convolutional and Fully Connected Neural Networks

합성곱 신경망과 완전 연결 계층의 앙상블 모델을 이용한 정밀한 파프리카 과실 발달 단계 예측

BY

JUNYOUNG PARK

FEBRUARY, 2021

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF AGRICULTURE, FORESTRY AND BIORESOURCES
THE GRADUATE SCHOOL
OF SEOUL NATIONAL UNIVERSITY

Accurate Prediction of the Fruit Development Stage of Sweet Pepper Using an Ensemble Model of Convolutional and Fully Connected Neural Networks

UNDER THE DIRECTION OF DR. JUNG EEK SON
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

BY
JUNYOUNG PARK

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF AGRICULTURE, FORESTRY AND BIORESOURCES
THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

FEBRUARY, 2021

APPROVED AS A QUALIFIED THESIS OF JUNYOUNG PARK
FOR THE DEGREE OF MASTER OF SCIENCE
BY THE COMMITTEE MEMBERS

CHAIRMAN: Kim, Kisun
KI SUN KIM, PH.D.

VICE-CHAIRMAN: Son, Jung Eek
JUNG EEK SON, PH.D.

MEMBER: Hee Jae Lee
HEE JAE LEE, PH.D.

Accurate Prediction of the Fruit Development Stage of Sweet Pepper Using an Ensemble Model of Convolutional and Fully Connected Neural Networks

Junyoung Park

Department of Agriculture, Forestry, and Bioresources, Graduate School of Seoul National University

ABSTRACT

Accurate detection of individual fruits and prediction of their development stages enable growers to efficiently allocate labor and manage strategically. However, the prediction of the fruit development stage is challenging, especially in sweet peppers, because the fruit harvest is discrete and its immature stage is indistinguishable. An ensemble model of convolutional and fully connected neural networks was developed to detect sweet pepper (Capsicum annuum L.) fruits in images and predict their development stages. The plants were grown in four rows in a greenhouse, and images were collected in each row. Plant growth and environmental data were collected every minute and month, respectively. For predicting the fruit stage, an ensemble of convolutional neural network (CNN) and multilayer perceptron (MLP) models were used. The fruit development stage was classified into immature, breaking, and mature stages with a CNN using images. Moreover, the immature stage was internally divided into four stages with an MLP. The plant growth and environmental data and the information from the CNN output were used for the MLP input. That is, a total of six stages were classified using the CNN–MLP ensemble model. The ensemble model showed good agreement in predicting fruit development stages. The average accuracy of the six stages was F1 score = 0.77 and IoU = 0.86. The CNN-only model could classify the mature and breaking stages well, but the immature stages were not distinguished, while the MLP-only model could hardly classify the fruit stage except the immature stages. The most influential factors in classification were the data obtained from CNN and the plant growth and environment data, which contributed to the improvement of model accuracy. The ensemble models can help in appropriate labor allocation and strategic management by detecting individual fruits in images and predicting precise fruit development stages.

***Additional keywords*:** deep learning, harvest prediction, labor allocation, machine learning, precise management

Student Number: 2019-28124

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Strategic crop management considering crop growth can improve crop quality and yield (Panda Hoogenboom, Paz, 2010). Environmental conditions from before fruit set to harvest determine fruit quality and yield, and the most influential factors can vary depending on the fruit physiological stage (Marti & Mills, 1991; Pagamas & Nawata, 2008). Prediction of fruit physiological stages is crucial for strategic and precise fruit management because the harvesting time is closely related to the fruit stage. Therefore, labor can be efficiently allocated at the right time and place (Chen et al., 2019).

Fruit physiological stages have been studied along with detection and classification (Cárdenas-Pérez et al., 2017; Koirala, Walsh, Wang, McCarthy, 2019a). The fruits of strawberry and mango are easily detected, and their developmental stages can be classified well because the fruits are protruding and have characteristics different from the leaves (Sa et al., 2016). However, detecting individual fruits showing vague changes in fruit development stages can be challenging. Sweet pepper is one of the most prevalent fruiting vegetables grown in greenhouses (Bozokalfa & Kilic, 2010). However, high-wired plants usually grow in densely cultivated conditions, in which fruits can be covered with leaves having a similar color and texture (Nwachukwu, Mbagwu, Onyeji, 2007).

Physiological characteristics can also be an obstacle for precise prediction of the fruit development stage of sweet pepper. In general, the fruit requires 45–55 days from fruit set to harvest. The growth rate of fruit is highest near 20 days after anthesis, and the color change occurs just a week before harvest (Marcelis and Baan Hofman-Eijer, 1995; Tijskens et al., 2016). In particular, the sweet pepper belongs to the group of nonclimacteric agricultural products, and thus, the fruits are usually not harvested before their color change (Camelo, 2004). Except at the beginning and the end of the physiological time, it is difficult to classify the fruit development stage there is no visible change. In addition, yield patterns are usually irregular, displaying peak-and-valley harvest patterns with periods of high yields followed by periods of low yields in greenhouses (Lin & Hill, 2007). Fruit harvest is hardly predictable with these patterns. Therefore, a method that can accurately detect fruit and classify its developmental stage is required.

In previous studies, current and preceding yields, temperature, and solar radiation were used to detect sweet pepper fruits and predict harvest time (Sauviller, Baets, Verlinden, Nicolai, 2009; Verlinden, Nicolaï, Sauviller, Baets, 2005). Machine learning such as time series complex regression models and neurofuzzy networks has also been applied (Qaddoum, Hines, Illiescu, 2011; Verroens et al., 2006). Recently, convolutional neural networks (CNNs) have been used for object detection and have shown high performance in computer vision. However, these studies predicted whole fruits for a limited growth period before harvest rather than individual fruits for a whole fruit development stage.

Previous studies using CNN models had limitations in predicting a maximum of three stages based on distinct external features (Chen, Lu, Liu, Li, Qian, 2020; Tian et al., 2019). Since CNNs for computer vision learn from external features in images, specific crop growth and environmental changes cannot be considered. For strategic management and labor allocation planning, crop growth and environmental data are required to exhibit more subdivided stages in fruit development. Ensemble models use multiclassifiers that can provide different views or feature representations with strong complementarity. Thus, combining these factors can enhance the performance and accuracy of the model compared to using a single model (Lecun, Bengio, Hinton, 2015; Petrakova, Affenzeller, Merkurjeva, 2016; Ren, Zhang, Suganthan, 2016). Furthermore, the ensemble model can consider plant growth and environmental data. The objectives of this study were to detect sweet pepper fruits in images and predict the fruit development stage using an ensemble model of convolutional and fully connected neural network models.

LITERATURE REVIEW

Greenhouse environment

Greenhouse plant empowerment is the controlling the plant balances based on monitoring, rather than feelings and subjective visual assessment of the crop and the climate. Greenhouse led to new insights on managing radiation, temperature, humidity, fertilizing carbon, and nutrition control. To achieve more production with higher quality using less energy (Book 2018).

Sweet pepper fruits characteristics

Sweet pepper fruits expand early in the biological time, near 20 days after anthesis shows peak of fresh weight increase (Marcelis and Baan Hofman-Eijer 1995). And length growth terminate around 20-30 days after anthesis (Tijskens et al. 2016). Color in sweet pepper fruits changes abruptly, just a week before harvest (Yamada et al. 2019). Because sweet pepper belongs to non-climacteric fruits, color changing progress progress only obtained while fruit is attached to the plants(López Camelo and Food and Agriculture Organization of the United Nations. 2004).

Furthermore, sweet pepper fruit has peak-and-valley harvest pattern, with periods of high yields followed by periods of low yields (Marcelis et al. 2004). The nature of peaks-and-valleys in yield is not fully understood.

Predicting harvest using deep learning

To predict harvest of sweet pepper, simple approaches using regression analysis (Sauviller et al. 2009) to more complex approaches like using time series (Al-Halimi and Moussa 2015), neuro-fuzzy networks (Qaddoum et al. 2011) and neural networks. Verroens et al. uses current and preceding yields to predict future yields and Sauviller et al. uses temperature and radiation environments. Lin et al. tried to characterize to predict yield by using neural network model. According to Lin et al. important inputs which affects neural network model accuracy was internal biological factors including crop age, node position and coloration than external environmental factors like temperature and light (Verroens et al. 2006; Lin and Hill 2007; Sauviller et al. 2009). Verroens et al. founds current and preceding yields were influenced future yields more than known environmental factors. Preceding yields was the most important factors and light environment was followed, influencing on sweet pepper weekly yields. In contrast, average day temperature had a significant effect on the fruit developmental time. These models could estimate yields weekly prediction but these methods are only available on whole yields, not for individual fruits.

Convolutional neural network (CNN)

In recent year, convolutional neural network has been widely used and developing in image processing tasks. CNN can train translational invariant patterns, to detect objects in the image, wherever the objects are positioned. And also able to detects layers of increasingly complex patterns to extract complex visual concepts. In detecting frameworks, classification and localization progress are combined into a single system to detect object and draw bounding boxes around objects in images.

You-Only-Look-Once (YOLO), a new approach to object detection was developed to detect object by single shot in the image. In this model, single CNN is able to divides image into multiple region and predicts bounding boxes and probabilities for each region. Bounding boxes are weighted by the predicted probabilities (Redmon et al. 2016). YOLOv2 improves the network structure and uses a convolution layer to replace the fully connected layer in the output layer of YOLO. And YOLOv3 extract features by using three different scales for box predictions by the concept of feature pyramids (Lin et al. 2017). In order to capture more meaningful and fine-grained information, the feature maps from lower layers are merged with up-sampled feature map from higher layers and processed further (Koirala et al. 2019a). In this model, data augmentation improved robustness in train model by adding artificially generated images which were manipulated or combined existing training images. These techniques can be applied without affecting inference speed (Bochkovskiy et al. 2020).

In agriculture, Plant organs were identified and localized using CNN (Pound et al. 2016). And also applicated in fruit harvesting robot (Ringdahl et al. 2019) and plant diseases (Mohanty et al. 2016; Kussul et al. 2017).

Multi-layer perceptrons (MLP)

A multilayer perceptron (MLP) is a network mapping input data to the output data in a feedforward manner (Atkinson and Tatnall 1997). MLP model was constructed with interconnected nodes multiple layers including input layer, hidden layer and output layers. Each the layers are fully connected to the preceding layer and succeeding layer (Del Frate et al. 2007). To distinguish not linearly separable data, outputs of each node are followed by a nonlinear activation (Pacifici et al. 2009). Equation, the output activation $a^{(l+1)}$ at layer l + 1 is derived by the input activation $a^{(l)}$:

$$a^{(l+1)} = \sigma(w^{(l)} a^{(l)} + b^{(l)}) \qquad (1)$$

Where l corresponds to a specific layer, $w^{(l)}$ and $b^{(l)}$ denote the weight and bias, respectively, at layer l, and r represents the nonlinear activation operation (e.g., sigmoid, hyperbolic tangent, and rectified linear units) function. For an multilayer perceptron, the first input layer is $a^{(1)} = x$ while the last output layer is:

$$h_{w,b}(x) = a^{(m)} \qquad (2)$$

The weights w and bias b in Eq. (2) are learned by supervised training using a backpropagation algorithm to approximate an unknown input-output relation (Del Frate et al., 2007). The objective function is to minimize the difference between the predicted outputs and the desired outputs:

$$J(W,b;x,y) = \frac{1}{2}\left\|h_{w,b}(x) - y\right\|^2 \qquad (3)$$

In agriculture, MLP was used to interpolate greenhouse environmental factors (Moon et al. 2019a). It could manage a big data obtained from the greenhouse for analyzing future environment.

MATERIALS AND METHODS

Cultivation conditions

The data were collected from a Venlo-type greenhouse at Seoul National University, Suwon, Korea (37.3°N 127.0°E). Sweet pepper (*Capsicum annuum* L.) plants were transplanted on February 26, 2020, and the cultivation ended on July 7, 2020. In the cultivation area, four growing rows, with six rockwool slabs, each having four plants, were grown (Fig. 1), and the planting density was 3.3 plants m^{-2}. In the greenhouse, the temperature was maintained at 22.6–32.0°C. The electrical conductivity and pH of the nutrient solutions were maintained at 2.6–3.0 dS m^{-1} and 5.5–6.5 dS m^{-1}, respectively. Integrated solar radiation was applied for irrigation control. The plants were maintained on two main stems, which were vertically trellised into a "V" canopy system (Jovicich, Cantliffe, Stoffella, 2004). The fruits were harvested three times a week, when the surfaces of the fruits were mostly colored.

Image data collection

Images were acquired three times a week at 2:00 PM (KST) from April 6 (the first fruit set) to Jun 24, 2020. Image acquisition was performed before fruit harvest. A total of 7,682 images were collected from each side of the plants under natural light conditions using a smartphone (iPhone 11, Apple Inc., Cupertino, CA, USA) with 1,920×1,088 resolution using a stabilizer. To include the bottom to top plants in the image as much as possible, angles were up to 74° with 130–cm height and 100–cm distance from the row (Fig. 1). All images were resized to dimensions of 1,024×1,024 with 96 dots per inch (dpi) resolution.

Environmental data collection

Temperature, relative humidity, and solar radiation in the greenhouse were measured every minute, and the number of plant nodes was measured every month. Individual fruit harvest dates were observed three times a week after the image data were collected.

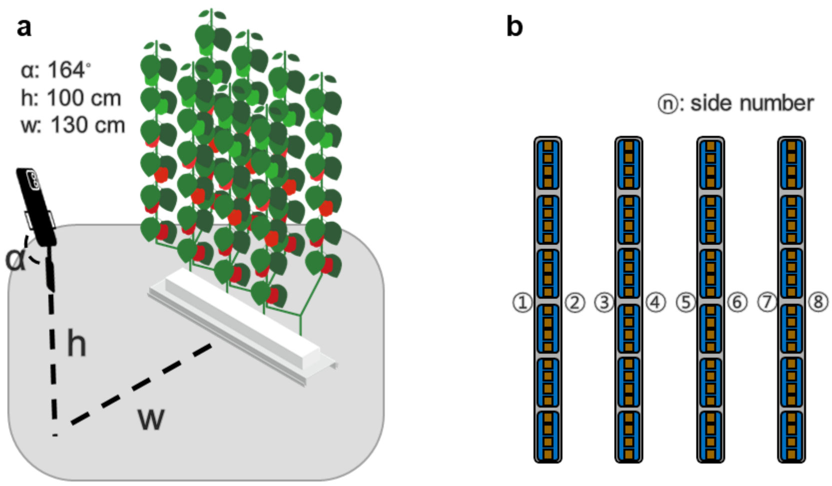

Fig. 1. Sketch of an image acquisition condition for sweet pepper fruits (a) and the numbered locations for image acquisition (b).

Classification of fruit development stage

The fruits were categorized into three major stages: immature, breaking, and mature (Fig. 2). Additionally, the immature stage was separated into four internal stages by the days before harvest (DBH). The stages were divided according to the physiological change and DBH in fruits. Immature stage 1 (IM_1) was approximately 35 DBH when fruit expansion started. Immature stage 2 (IM_2) was 25–35 DBH when the size expansion was almost finished. Immature stage 3 (IM_3) was 15–25 DBH when the fruit reached full size. Immature stage 4 (IM_4) was from 15 DBH to breaking, the last stage of immature. The breaking stage (B) was approximately 7 DBH when the fruit color changed from a purple state to a green and red mottled state due to carotenoid synthesis and chlorophyll reduction (Yamada, Nakayama, Shibata, Kishimoto, Ikeda, 2019). The mature stage (M) was the harvest period when the surface of fruits was 95% colored (Verlinden et al., 2005).

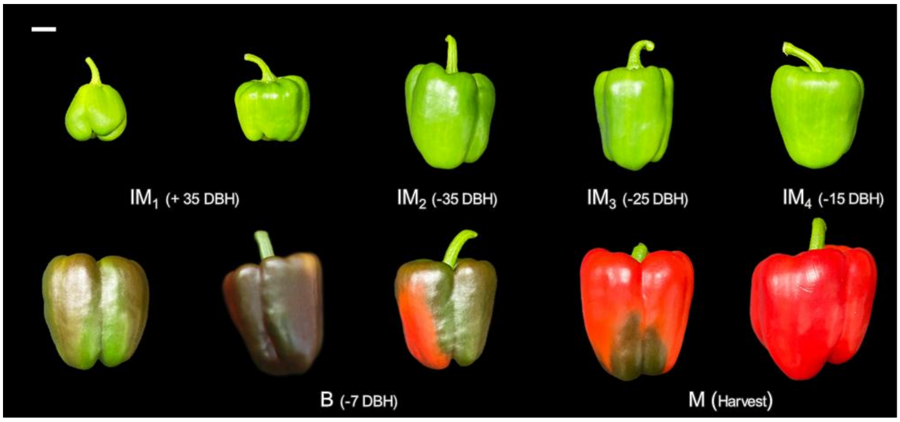

Fig. 2. Fruit development stages of sweet pepper with the days before harvest (DBH) (white line = 2 cm). $IM_{1\text{-}4}$, B, and M indicate the fruit development stages of immature 1-4, breaking, and mature stages, respectively.

CNN-only model

A pretrained CNN model for object detection called the YOLOv5 model was used, which was constructed with 89 million parameters and 284 layers (Yap et al., 2020). The You-Only Look-Once (YOLO) model shows real-time object detection with high accuracy by combining predicted bounding boxes and class probabilities directly from full images by using a single neural network architecture (Redmon et al., 2016). Inputs for CNN algorithms were sweet pepper images, and the six fruit stages and fruit locations were set as outputs.

MLP-only model

Plant growth and environmental data were used in the MLP model to predict the six fruit stages (Table 1). Environmental data, including temperature, relative humidity, and solar radiation were used. The number of nodes, crop age, current and preceding yield as plant growth data were used for predicting fruit physiological stages following previous studies (Sauviller et al., 2009; Lin & Hill, 2007; Al-Halimi & Moussa, 2015).

MLP is a neural network algorithm that learns the pattern of data from several layers connected with perceptrons. In this study, the structure of the MLP model was modified based on the previous MLP for greenhouse environments (Moon et al., 2019). The model was constructed with six fully connected layers, and the rectified linear unit (ReLU) function was used as an activation function. AdamOptimizer was used to train the model (Kingma & Ba, 2015). The coefficients were modified to construct the model solving the regression problem. All programs were based on Python language (v. 3.6.7, Python Software Foundation, Wilmington, NC, USA) and its library. The model was constructed using TensorFlow (v. 2.0.0), which was used for the computation (Abadi et al., 2016).

Table 1. Measured environmental and plant growth data used as inputs to the multilayer perceptron (MLP) model.

Input data (unit)	Range
Inside daily average temperature (°C)	22.6–32.0
Inside daily average relative humidity (%)	20.6–79.8
Inside daily average radiation (J cm^{-2})	5.4–76.2
Cumulative temperature for a week (°C)	169–211
Cumulative radiation for a week (J cm^{-2})	2637–6838
Cumulative temperature for 2 weeks (°C)	353–416
Cumulative radiation for 2 weeks (J cm^{-2})	7737–11774
Cumulative temperature for 3 weeks (°C)	543–618
Cumulative light for 3 weeks (J cm^{-2})	12429–17405
Days after transplanting	42–121
Number of nodes	14–31
Yield	0–90
Cumulative yield	5–827

Ensemble model of CNN and MLP

The structure of the CNN part of the ensemble model was the same as the structure of the YOLOv5 algorithms used in the CNN-only model. The CNN part was used to classify three major stages: immature, breaking, and mature. The CNN outputs were used as inputs to the MLP model (Fig. 3). The CNN outputs were box location, classified stage, and fruit number of each stage. The distances from immature fruits to the lowest breaking and mature points were calculated from the CNN outputs, which were also used as inputs (Table 2). In addition, the inputs to the MLP-only model were also added to the inputs of the ensemble model of CNN and MLP.

Table 2. Measured environmental and plant growth data used as inputs to the multilayer perceptron (MLP) model.

Input data	Range
CNN labelled stages	1–3
Labelled box coordinates	0–100
Box area	0.02–414
Number of immature stage in image	0–15
Number of breaking stage in image	0–12
Number of mature stage in image	0–11
Labelled box distance to lowest breaking coordinates	-1–1
Labelled box distance to lowest mature coordinates	-1–1

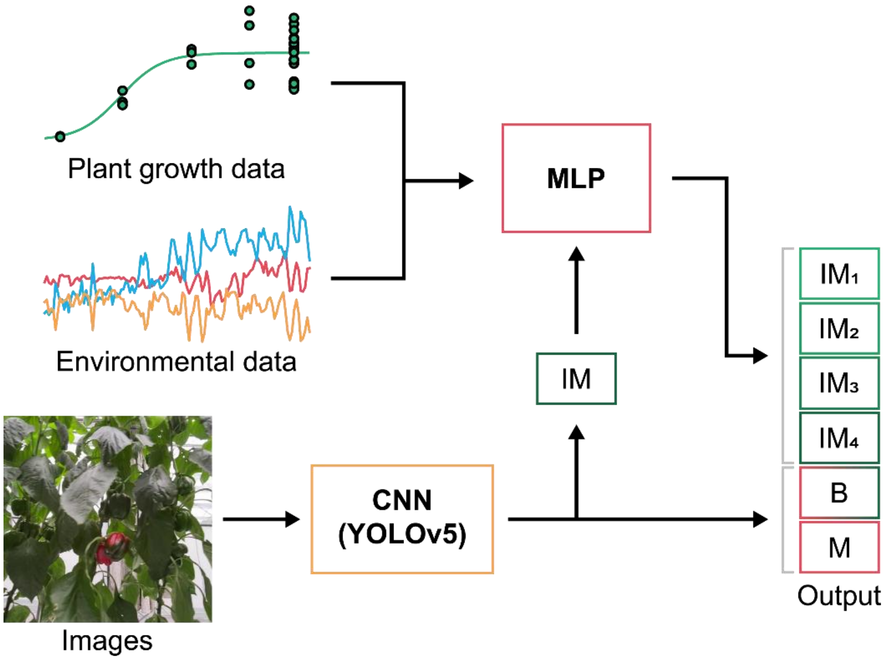

Fig. 3. Schematic diagram for predicting fruit development stages of sweet pepper using the convolutional neural network (CNN)-only (a), the multilayer perceptron (MLP)-only (b), or the ensemble (c) models. $IM_{1\text{-}4}$, B, and M indicate the fruit development stages of immature 1-4, breaking, and mature stages, respectively.

Model training

In the CNN-only model, 1,268 randomly selected images of the plants were used to train the detection model. The training and validation sets were divided at a 7:3 ratio, and 200 images were used for the test set. In the images, each fruit was manually labeled at the pixel level, and a total of 3,782 fruits were detected (Table 3). The corresponding annotated files were saved in json format.

Individual plants had fruits at different stages. A total of 16,600 fruits were used to train the MLP-only model with plant growth and environmental data. The training and validation sets were divided at a 7:3 ratio, and 3,321 classified stages were used for the test set. In the ensemble model of CNN and MLP, a total of 7,682 images were used for model training, and the data were divided at the same ratio as the CNN-only model. The number of labeled boxes was 35,958 with each stage. To see the importance of the factor, input data were excluded one by one to determine which factors affected the accuracy in classifying stages.

Table 3. The number of data points used for the convolutional neural network (CNN)-only, the multilayer perceptron (MLP)-only, and the ensemble models at different fruit development stages of sweet peppers. IM_{1-4}, B, and M indicate the fruit development stages of immature 1-4, breaking, and mature stages, respectively.

Model	Number of data						
	IM_1	IM_2	IM_3	IM_4	B	M	Total
CNN-only	644	910	450	506	614	658	3,782
MLP-only	437	2,005	2,981	1,450	3,083	2,862	16,600
Ensemble	23,287				7,557	5,114	35,958

Evaluation metrics

For the model evaluation, precision, recall, and F1 score were used in both CNN and MLP models to evaluate the classification (Equations 1–3) (Ganesh, Volle, Burks, Mehta, 2019).

$$Prescision = \frac{T_P}{T_P + F_P} \tag{1}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{2}$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

The classification result has four categories: true positive (T_P), the number of correctly detected objects; true negative (T_N), the result of correct rejection; false positive (F_P), the number of false data points detected; and false negatives (F_N), the number of missed objects. A precision score equal to 1.0 means that all detected objects were correctly classified. A recall score equal to 1.0 means that all the correctly classified objects were detected in the image. A precision recall score can be shown in the curve by using the average values of the precision across all recall values (Ni., Li, Jiang, Takeda, 2020).

To evaluate the detection performance, the intersection over union (IoU) score was used. IoU is the area ratio between the intersection and union of the detection and the ground truth. These scores can vary from 0.0 to 1.0, which means no overlap at 0.0 and full overlap at 1.0. Model performance is excellent above an IoU of 0.9, good enough at approximately 0.7, and poor near 0.4 (Koirala, Walsh, Wang, McCarthy, 2019b).

RESULTS

Accuracy of model

Using the CNN-only model, recall scores were remarkably low in all immature stages (Fig. 4). The MLP-only model also had low accuracies in both precision and recall. Particularly in IM_4, the model was totally unable to distinguish the stages with precision and showed recall scores of 0.26 and 0.13, respectively. However, in the ensemble model, precision and recall scores were within acceptable ranges in all stages. F1 scores at the mature and breaking stages were 0.90 and 0.86, respectively (Fig. 5). However, except for these stages, the accuracy became low in internal immature stages as the stage moved far from harvest. The lowest stage was IM_1 with an F1 score of 0.29. The CNN-only model was not able to distinguish the immature internal stages with a mean F1 score of 0.56. The mean F1 score for all stages using the CNN-only model was 0.56.

The MLP-only model showed the lowest accuracy among the three models, with an F1 score of 0.49. F1 scores for all stages were below 0.6 except IM_1, which showed 0.84. Comparing the model accuracies for prediction, the trained CNN-only model showed a high accuracy in breaking and mature stages, while the trained MLP-only model showed a high accuracy for the early stage in immature stages. The ensemble model overcame the shortcomings of a single model, adequately balancing these two advantages. In undistinguishable internal immature stages, the F1 score increased adequately compared to the MLP-only model. In IM_4, the F1 score significantly increased the most from 0.17 to 0.70. The mean F1 score for all stages using the ensemble model was 0.77.

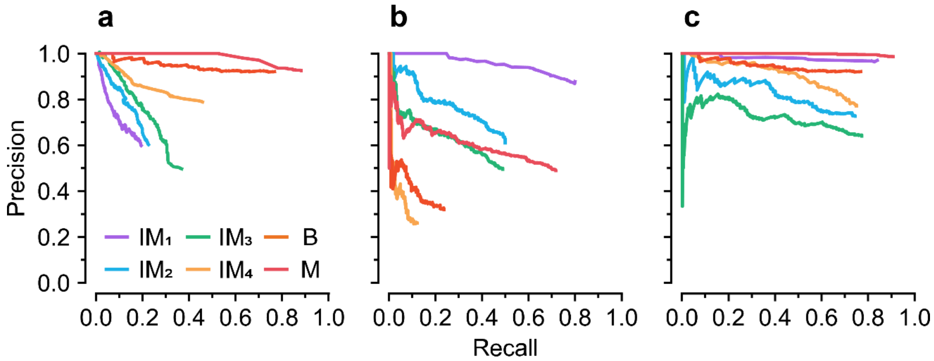

Fig. 4. Precision and recall scores of the convolutional neural network (CNN)-only (a), the multilayer perceptron (MLP)-only (b), and the ensemble (c) models. IM_{1-4}, B, and M indicate the fruit development stages of immature 1-4, breaking, and mature stages, respectively.

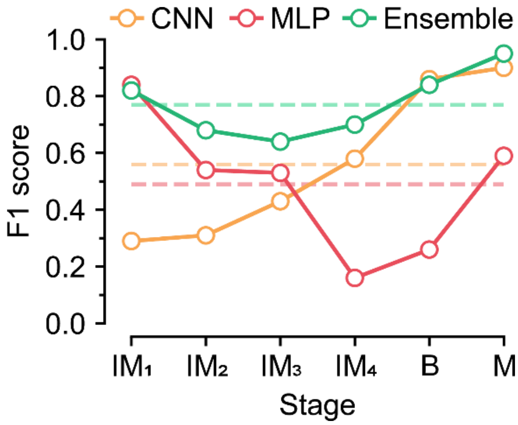

Fig. 5. Comparison of F1 scores among the convolutional neural network (CNN)-only, the multilayer perceptron (MLP)-only, and the ensemble models. IM_{1-4}, B, and M indicate the fruit development stages of immature 1-4, breaking, and mature stages, respectively. Dashed lines represent average F1 scores.

t-Distributed Stochastic Neighbor Embedding (t-SNE) result

The t-SNE results showed classes by each cluster of points (Fig. 6). IM_1 was clustered well but overlapped in IM_2. In the results, the points in IM_3 were most scattered to IM_2 and IM_4. Since the t-SNE result showed the final stages divided by the MLP model, both breaking and mature points, which were divided by the CNN model, were obviously separated without scattered points.

Relative correlations between classified stages and input factors are shown in Fig. 7. Temperature showed no differences over stage, and relative humidity showed few distinct features in immature initial stages. However, the cumulative radiation increased when the stage moved close to the harvest. Plant growth data had more distinct features with DAT and node number, but yield data did not show any differences depending on fruit stage. The data obtained by the CNN model showed distinct features from the values directly obtained (e.g., label box coordinates, number of each stage, and computed distance from immature fruits to the lowest breaking and mature points) for classifying stages. Particularly in IM_4, the data obtained by the CNN model showed a close relationship with divided stages.

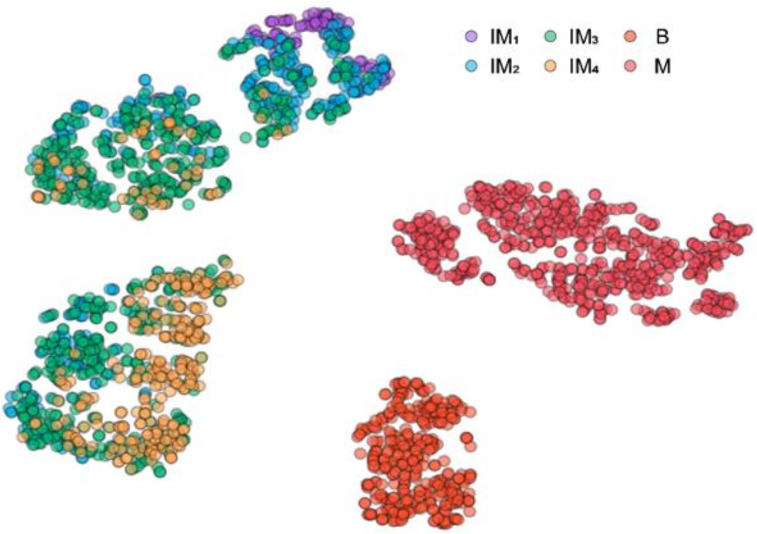

Fig. 6. t-Distributed stochastic neighbor embedding (t-SNE) results for dividing the fruit development stage of sweet pepper. IM_{1-4}, B, and M indicate the fruit development stages of immature 1-4, breaking, and mature stages, respectively.

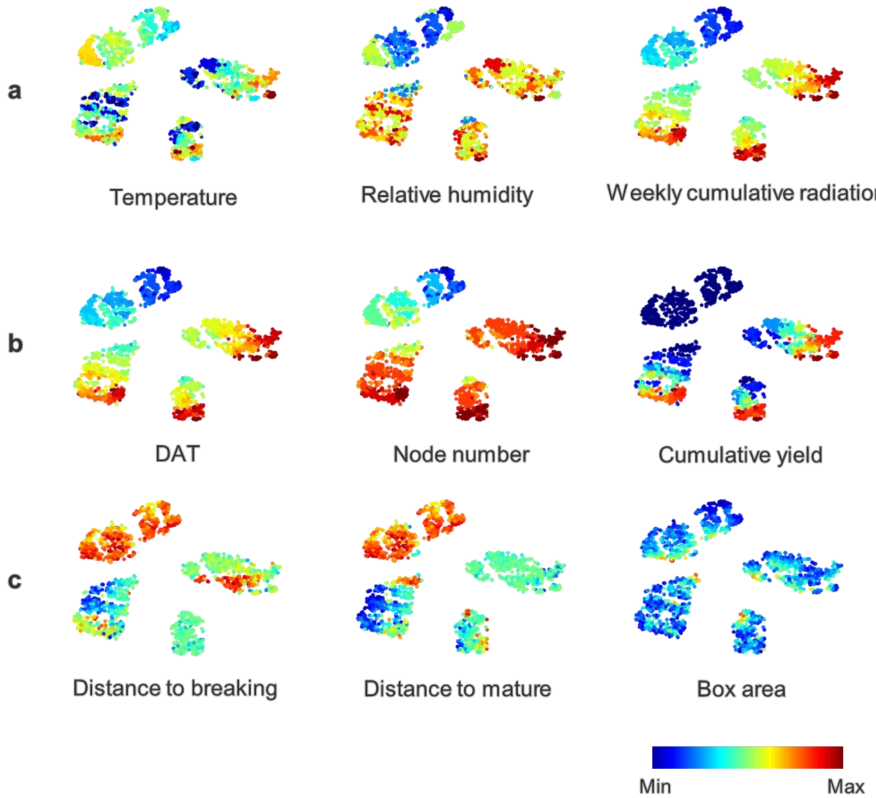

Fig. 7. t-Distributed stochastic neighbor embedding (t-SNE) results indicating the relative relevance of each factor with fruit development stages of sweet pepper related to the environment (a), plant growth (b), and the convolutional neural network (CNN)-obtained (c) data.

Model accuracy by data

Using environment, plant growth, and CNN output data together resulted in the highest F1 score of 0.77. Excluding each factor, the F1 score changed, as shown in Fig. 8. The reduction in accuracy was 0.04 when excluding environmental data and 0.29 when excluding CNN output data, indicating that the CNN output data most affected the classification of fruit development stages.

Application of the ensemble model to fruit detection

With or without breaking and mature stages, the model could classify the fruit development stage when the image contained only immature stages or completely mixed conditions (Fig. 9a). In any light conditions of the image, including backlight or poor light, the model also performed well under natural light conditions. However, the model could not accurately predict when the CNN model failed to detect fruits or incorrectly classified the stage, and thus prediction could not be reversed (Fig. 9b).

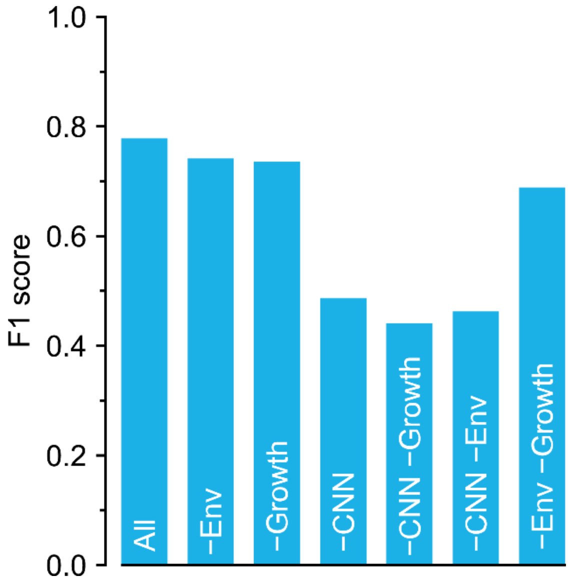

Fig. 8. F1 score change after subtracting data from all factors. Standard F1 score is 0.77. Env, growth, and CNN represent the environment, plant growth, and convolutional neural network-obtained data, respectively.

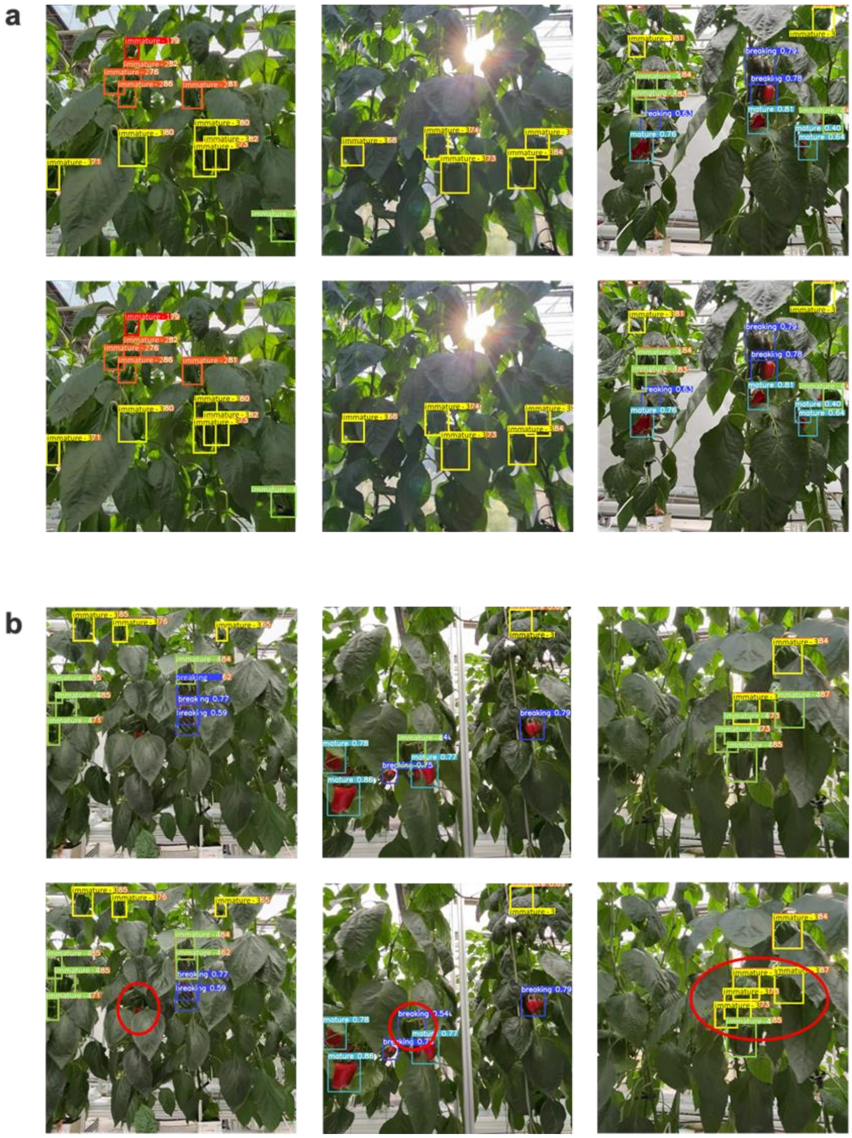

Fig. 9. Good (a) and incorrect (b) examples of detected fruit images and classified fruit development stages of sweet pepper. The upper and lower images in each example are the ground truth and the model detected and classified, respectively.

DISCUSSION

Evaluation of model accuracies

In this study, sweet pepper fruits were detected, and their stages could be predicted based on DBH using the ensemble model of CNN and MLP. In the CNN-only model, mature and breaking stages were adequately classified. These stages can be distinguished by color because the CNN model comprehends external features (O'Shea & Nash, 2015). Internal immature stages may have low accuracy because convolutional layers reduce image sizes, and multiple convolution layers make the image size smaller. The reduced size could cause low accuracy in detecting small sweet pepper fruits in earlier stages (Chen et al., 2020).

Comparing the CNN-only model and the CNN part of the ensemble model, the difference in F1 score was due mainly to the number of stages to predict (Fig. 5). In this study, fruit stages were more specifically classified than in a previous study (Dandavate & Patodkar, 2020). In the CNN-only model, all six stages were classified at once, while the CNN part of the ensemble model classified three stages. The performance of the CNN-only model may decrease because the model is more likely to classify other classes (Tian et al., 2019). The MLP-only model using only environmental and plant growth data was not suitable for classifying the fruit stage (Fig. 5), and additional data are required (Lin & Hill, 2007). Furthermore, model details such as the number of layers, number of perceptrons, and optimized coefficients may have the possibility of improving accuracy (Moon et al., 2019). Even though these details are adjusted, the MLP-only model seemed to have limits in increasing accuracy.

In the first step of the ensemble model, the prediction result of the CNN part was robust enough to predict the three stages of immature, breaking, and mature with high accuracy (F1 score = 0.91 and IoU = 0.86). The performance of the YOLOv5 model has been proven in various fields (Yap et al., 2020; Li et al., 2021), and its robust accuracy was achieved by a large image dataset. From the CNN results, the accuracy for the immature stages was adequate to continue the next MLP step. Comparing the MLP part of the ensemble model and the MLP-only model, there was no structural change between the two models. Since the model structure was the same, the changes in input that contained spatial information seemed to be the main reason for the increased model accuracy.

Factors affecting model classification

In classifying the fruit stages, the model did not depend on environmental data and could predict without environmental data (Fig. 8). Previous studies have shown that environmental data are not a major factor for predicting fruit stages (Tijskens et al., 2016; Verlinden et al., 2005). The accumulation of environmental factors could be contained in plant growth data. These correlated factors enabled confound in model training. Training duplicated data in a model can reduce the importance of environmental data compared to plant growth data (Lin & Hill, 2007). Therefore, additional steps are needed to remove less-related data and add newly acquired data considering data duplication.

In the maturation of sweet pepper fruits, specific proteins such as polygalacturonase, cellulase and β-galactosidase affect the harvest date regardless of the maturity stage (Biles et al., 2019). Environments regulate protein synthesis, which is a key factor in the development and ripening of pepper fruits (Batista-Silva et al., 2018). Fruit growth is optimal between 25°C and 30°C and slows down below 15°C (Mateos et al., 2013). Relative humidity is correlated with flower abortion (Marcelis, Heuvelink, Baan Hofman-Eijer, Den Bakker, Xue, 2004) and fruit growth. However, there was no significant correlation in fruit growth in the range of 55 to 80% (Baër & Smeets, 1978; Barker, 1989). In this study, temperature and relative humidity were controlled and made the gap between fruit stages smaller (Fig. 7); however, uncontrolled sunlight was an important factor for the model to classify the stages. Differences in solar radiation may affect the accumulation of carotenoids and flavonoids and accelerate the degradation of chlorophyll in immature fruits by affecting the synthesis of hormones (Pizarro & Stange, 2009; Zoratti, Karppinen, Luengo Escobar, Häggman, Jaakola, 2014; Cruz et al., 2018).

The CNN-obtained data, compared with the MLP-only model and the ensemble model, increased the accuracy most to predict IM_4 among the immature stages. In sweet peppers, anthesis and fruit set occur sequentially from bottom to top, and thus fruit maturation occurs in the same direction. In this study, immature stages close to breaking and mature stages were highly affected in classifying the fruit stages. Spatial information from the distance to mature and breaking data contributed to the classification of later immature stages. Among the input factors, the CNN-obtained data were the most important factor for classifying the stages. However, the results of subtracting plant growth and environmental data have proven to contribute to making robust models.

Fruit detection by the ensemble model

The fruits with dense and overlapping growing conditions showed low performance for classifying the fruit stage (Fig. 9). In the CNN-obtained data, the distances to breaking and mature stages were calculated using spatial information from the CNN model. Because the distance was calculated using the y-axis, the x-axis distance was not considered, which might affect the model having problems, especially in horizontally positioned fruit images. This study showed that the vertical information of the plant is one of the most important factors in the classification stage and can be trained from images. Horizontal information seems to be added easily from the CNN part of the ensemble model and can help classify fruits under dense and overlapping conditions. For robust models, adding detailed location information and images that were taken under various conditions will enable an increase in the model accuracy.

CONCLUSION

Individual fruit development stages of sweet pepper based on the days before harvest were classified with an ensemble model of a convolutional neural network and a multilayer perceptron. The fruits were detected, and a total of six fruit development stages were accurately predicted by using images, plant growth, and environmental data. Among the compared models, the ensemble model showed the highest accuracy with F1 score = 0.77 and IoU = 0.86. The CNN-only model performed well in classifying the mature and breaking stages when the features were obvious, but the model could not distinguish immature subclasses. The MLP-only model was better in classifying the immature classes, but accuracies were not adequate for classifying each stage. The ensemble model compensated for the weaknesses of the individual models, and the trained model showed the highest accuracy for all six fruit development stages. The model was available in common greenhouses with a common device under natural light conditions. Therefore, the trained model can help in appropriate labor allocation and strategic management.

LITERATURE CITED

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, et al (2016) TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv Preprint*, arXiv:1603.04467.

Al-Halimi R, Moussa M (2015) Long-Term Yield Prediction of Greenhouse Sweet Pepper Crops. *GSTF Journal on Agricultural Engineering*, *2*, 1-7.

Baër J, Smeets L (1978) Effect of relative humidity on fruit set and seed set in pepper (*Capsicum annuum* L.). *Netherlands Journal of Agricultural Science*, *26*, 59-63.

Barker JC (1989) The effects of air humidity on growth and fruit production of sweet pepper (*Capsicum annuum* L.). *Journal of Horticultural Science*, *64*, 41-46.

Batista-Silva W, Nascimento VL, Medeiros DB, Nunes-Nesi A, Ribeiro DM, Zsögön A, Araújo WL (2018) Modifications in organic acid profiles during fruit development and ripening: correlation or causation? *Frontiers in Plant Science*, *9*, 1689.

Bozokalfa MK, Kilic M (2010) Mathematical modeling in the estimation of pepper (*Capsicum annuum* L.) fruit volume. *Chilean Journal of Agricultural Research*, *70*, 626-632.

Camelo AFL (2004) Manual for the preparation and sale of fruits and vegetables : from field to market. Italy:Food & Agriculture Organization, (Vol. 151).

Cárdenas-Pérez S, Chanona-Pérez J, Méndez-Méndez JV, Calderón-Domínguez G, López-Santiago R, Perea-Flores MJ, Arzate-Vázquez I (2017) Evaluation of the ripening stages of apple (Golden Delicious) by means of computer vision system. *Biosystems Engineering*, *159*, 46-58.

Chen W, Lu S, Liu B, Li G, Qian T (2020) Detecting Citrus in Orchard Environment by Using Improved YOLOv4. *Scientific Programming*, 2020.

Chen Y, Lee WS, Gan H, Peres N, Fraisse C, Zhang Y, He Y (2019) Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. *Remote Sensing*, 11, 1584.

Cruz AB, Bianchetti RE, Alves FRR, Purgatto E, Peres LEP, Rossi M, Freschi L (2018) Light, ethylene and auxin signaling interaction regulates carotenoid biosynthesis during tomato fruit ripening. *Frontiers in Plant Science*, *9*, 1370.

Dandavate R, Patodkar V (2020) CNN and data augmentation based fruit classification model. In Fourth International Conference on I-SMAC (pp. 784-787).

Ganesh P, Volle K, Burks TF, Mehta SS (2019) Deep orange: Mask R-CNN based orange detection and segmentation. *IFAC-PapersOnLine*, 52, 70-75.

Jovicich E, Cantliffe DJ, Stoffella PJ (2004) Fruit yield and quality of greenhouse-grown bell pepper as influenced by density, container, and trellis system. *HortTechnology*, 14, 507-513.

Kingma DP, Ba JL (2015) Adam: A method for stochastic optimization. *arXiv Preprint*, arXiv:1412.6980.

Koirala A, Walsh KB, Wang Z, McCarthy C (2019a) Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO'. *Precision Agriculture*, *20,* 1107-1135.

Koirala A, Walsh KB, Wang Z, McCarthy C (2019b) Deep learning–Method overview and review of use for fruit detection and yield estimation. *Computers and Electronics in Agriculture*, *162,* 219-234.

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature*, *521,* 436-444.

Li S, Gu X, Xu X, Xu D, Zhang T, Liu Z, Dong Q (2021) Detection of concealed cracks from ground penetrating radar images based on deep learning algorithm. *Construction and Building Materials*, *273,* 121949.

Lin WC, Hill BD (2007) Neural network modelling of fruit colour and crop variables to predict harvest dates of greenhouse-grown sweet peppers. *Canadian Journal of Plant* Science, *87,* 137-143.

Marcelis LFM, Baan Hofman-Eijer LR (1995) Growth analysis of sweet pepper fruits (*Capsicum annuum* L.). *Acta Horticulturae*, (412), 470–478.

Marcelis LFM, Heuvelink E, Baan Hofman-Eijer LR, Den Bakker J, Xue LB (2004) Flower and fruit abortion in sweet pepper in relation to source and sink strength. *Journal of Experimental Botany*, *55*, 2261-2268.

Marti HR, Mills HA (1991) Nutrient uptake and yield of sweet pepper as affected by stage of development and N form. *Journal of Plant Nutrition*, *14*, 1165-1175.

Mateos, R. M., Jiménez, A., Román, P., Romojaro, F., Bacarizo, S., Leterrier, M., Gómez, M., Sevilla, F., Del Río, L.A., Francisco J., et al. (2013). Antioxidant systems from pepper (*Capsicum annuum* L.): involvement in the response to temperature changes in ripe fruits. *International Journal of Molecular Sciences*, *14*, 9556-9580.

Moon T, Hong S, Choi HY, Jung DH, Chang SH, Son JE (2019) Interpolation of greenhouse environment data using multilayer perceptron. *Computers and Electronics in Agriculture*, *166*, 105023.

Ni X, Li C, Jiang H, Takeda F (2020) Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield. *Horticulture Research*, *7*, 1-14.

Nwachukwu CU, Mbagwu FN, Onyeji AN (2007) Morphological and leaf epidermal features of Capsicum annuum and Capsicum frutescens Solanaceae. *Nature and Science*, *5*, 54-60.

O'Shea K, Nash R (2015) An introduction to convolutional neural networks. *arXiv Preprint*, arXiv:1511.08458.

Pagamas P, Nawata E (2008) Sensitive stages of fruit and seed development of chili pepper (*Capsicum annuum* L. var. Shishito) exposed to high-temperature stress. *Scientia Horticulturae*, *117*, 21-25.

Panda SS, Hoogenboom G, Paz JO (2010) Remote sensing and geospatial technological applications for site-specific management of fruit and nut crops: A review. *Remote Sensing*, *2*, 1973-1997.

Petrakova A, Affenzeller M, Merkurjeva G (2016) Heterogeneous versus homogeneous machine learning ensembles. *Information Technology and Management Science*, *18*, 135-140.

Pizarro L, Stange C (2009) Light-dependent regulation of carotenoid biosynthesis in plants. *Ciencia e Investigación Agraria*, *36*, 143-162.

Qaddoum K, Hines E, Illiescu D (2011) Adaptive neuro-fuzzy modeling for crop yield prediction. *Parameters*, *16*, 17.

Ren Y, Zhang L, Suganthan PN (2016) Ensemble classification and regression-recent developments, applications and future directions. *IEEE Computational Intelligence Magazine*, *11*, 41-53.

Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C (2016) Deepfruits: A fruit detection system using deep neural networks. *Sensors*, *16*, 1222.

Sauviller C, Baets W, Verlinden BE, Nicolai BM (2008) Predicting the weekly yield fluctuations of greenhouse bell pepper. In I International Symposium on Horticulture in Europe 817 (pp. 261-268).

Tian Y, Yang G, Wang Z, Wang H, Li E, Liang Z (2019) Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and Electronics in Agriculture*, *157*, 417-426.

Tijskens LMM, Unuk T, Okello RCO, Wubs AM, Šuštar V, Šumak D, Schouten RE (2016) From fruitlet to harvest: Modelling and predicting size and its distributions for tomato, apple and pepper fruit. *Scientia Horticulturae*, *204*, 54-64.

Verlinden BE, Nicolaï BM, Sauviller C, Baets W (2005) Bell pepper production prediction based on colour development distribution, solar radiation and glass house temperature data. In III International Symposium on Applications of Modelling as an Innovative Technology in the Agri-Food Chain, (pp. 375-380).

Verroens P, Verlinden BE, Sauviller C, Lammertyn J, De Ketelaere B, Nicolai BM (2006) Time series analysis of Capsicum annuum fruit production cultivated in greenhouse. In III International Symposium on Models for Plant Growth, Environmental Control and Farm Management in Protected Cultivation 718 (pp. 97-104).

Yamada Y, Nakayama M, Shibata H, Kishimoto S, Ikeda T (2019) Anthocyanin Production and Enzymatic Degradation during the

Development of Dark Purple and Lilac Paprika Fruit. *Journal of the American Society for Horticultural Science*, *144*, 329-338.

Yap MH, Hachiuma R, Alavi A, Brungel R, Goyal M, Zhu H, Cassidy B, Ruckert J, Olshansky M, Huang X, et al (2020) Deep learning in diabetic foot ulcers detection: A comprehensive evaluation. *arXiv Preprint*, arXiv:2010.03341.

Zoratti L, Karppinen K, Escobar AL, Häggman H, Jaakola L (2014) Light-controlled flavonoid biosynthesis in fruits. *Frontiers in Plant Science*, *5*, 534.

ABSTRACT IN KOREAN

온실에서는 고부가가치에 열매를 맺는 작물을 효율적으로 관리하는 것이 중요하다. 개별 과실을 감지하고 그것의 발달 단계를 예측함으로써 재배자가 노동력을 적재적소에 할당하고, 전략적인 관리를 할 수 있다. 그러나 파프리카의 과실 발달 단계를 예측하는 것은 과실 수확량이 불연속적이고, 미성숙 단계에서 과실 간 나타나는 외부적인 특징 차이를 구별하기 어렵기 때문에 쉽지 않다. 이 연구의 목적은 합성곱 신경망과 완전 연결 계층의 앙상블 모델을 이용하여 이미지에서 파프리카 과실을 찾아내고 과실 발달 단계를 예측하는 것이다. 실험용 온실에서 파프리카 (*Capsicum annuum* L.)를 4줄로 재배하였고, 각 줄에 양면에서 식물 이미지를 수집 하였다. 2020년 4월 6일부터 6월 24일까지 환경 데이터는 분 마다, 식물 생장 데이터는 월 마다 수집되었다. 과실 발달 단계는 이미지에서 합성곱 신경망을 이용하여 미성숙, 변화 중, 성숙 3 단계로 구분하였고, 미성숙 단계는 완전 연결 계층을 이용하여 다시 세부적으로 4 단계로 구분 하였다. 환경, 식물 생장 데이터 및 합성곱 신경망의 출력 정보가 완전 연결 계층에 입력으로 사용되었다. 즉, 총 6 개의 과실 발달 단계가 앙상블 모델을 이용하여 분류되었다. 앙상블 모델은 과실 발달 단계를 예측하는 데

좋은 성능을 보였다. 총 6 단계의 과실 발달 단계 분류에 평균 정확도는 F1 점수 = 0.77, IoU = 0.86이다. 합성곱 신경망만을 이용한 모델은 성숙 단계와 변화 중 단계를 잘 분류 할 수 있었지만 미성숙 단계를 구별하지 못하였다. 완전 연결 계층만을 이용한 모델은 미성숙 단계를 제외하고 과실 단계를 거의 분류 할 수 없었다. 과실 발달 단계의 분류에 가장 큰 영향을 미치는 요인은 합성곱 신경망의 출력 정보였고, 환경 및 식물 생장 데이터는 모델 정확도 향상에 기여했다. 본 연구 결과는 추후 다양한 환경에 이미지에서 개별 과실을 찾아내고, 정확한 과실 발달 단계를 예측함으로써 적절한 노동력 할당 및 전략적 관리에 도움이 될 수 있을 것으로 사료된다.

추가 주요어: 머신 러닝, 과실 성숙도, 수확 시기

학 번: 2019-28124