

Article

Detection of Tip-Burn Stress on Lettuce Grown in an Indoor Environment Using Deep Learning Algorithms

Munirah Hayati Hamidon ¹ and Tofael Ahamed ^{2,*}¹ Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8577, Japan² Faculty of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8577, Japan

* Correspondence: tofael.ahamed.gp@u.tsukuba.ac.jp

Abstract: Lettuce grown in indoor farms under fully artificial light is susceptible to a physiological disorder known as tip-burn. A vital factor that controls plant growth in indoor farms is the ability to adjust the growing environment to promote faster crop growth. However, this rapid growth process exacerbates the tip-burn problem, especially for lettuce. This paper presents an automated detection of tip-burn lettuce grown indoors using a deep-learning algorithm based on a one-stage object detector. The tip-burn lettuce images were captured under various light and indoor background conditions (under white, red, and blue LEDs). After augmentation, a total of 2333 images were generated and used for training using three different one-stage detectors, namely, CenterNet, YOLOv4, and YOLOv5. In the training dataset, all the models exhibited a mean average precision (mAP) greater than 80% except for YOLOv4. The most accurate model for detecting tip-burns was YOLOv5, which had the highest mAP of 82.8%. The performance of the trained models was also evaluated on the images taken under different indoor farm light settings, including white, red, and blue LEDs. Again, YOLOv5 was significantly better than CenterNet and YOLOv4. Therefore, detecting tip-burn on lettuce grown in indoor farms under different lighting conditions can be recognized by using deep-learning algorithms with a reliable overall accuracy. Early detection of tip-burn can help growers readjust the lighting and controlled environment parameters to increase the freshness of lettuce grown in plant factories.

Keywords: tip-burn; lettuce; indoor farming; deep learning; YOLO; CenterNet

Citation: Hamidon, M.H.; Ahamed, T. Detection of Tip-Burn Stress on Lettuce Grown in an Indoor Environment Using Deep Learning Algorithms. *Sensors* **2022**, *22*, 7251. <https://doi.org/10.3390/s22197251>

Academic Editors: Viacheslav Adamchuk and Abdul M. Mouazen

Received: 20 August 2022

Accepted: 20 September 2022

Published: 24 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Indoor vertical farms have been developed to grow fresh high-quality vegetables in buildings without being restricted by extreme climate or land availability limitations. Indoor farms serve as a significant alternative or supplement to conventional agriculture to meet the demands of major cities seeking fresh, safe, and locally grown veggies. To promote faster crop growth and to maximize total production within limited cultivation areas, the plants grown on indoor farms are highly dependent on artificial light sources [1]. Lettuce is one of the most widely planted vegetables grown in indoor farms, not only because of its nutritional content, but also due to its short growth cycle and high planting density. However, lettuce grown with this rapid growth process is prone to a physiological disorder known as tip-burn. Tip-burn is a major problem for most vegetable cultivation under a controlled environment [2], especially the completely closed environment of an indoor farm equipped with artificial light [3,4].

The primary cause of tip-burn stress which occurs on plants cultivated in completely closed environments, such as indoor farms, is calcium deficiency [5,6]. The deficiency is not because of a lack of calcium in the supply nutrients but rather is caused by the inability of calcium to enter the rapidly developing younger leaves. Commonly, the deficiency symptoms appear first on these younger leaves as calcium is one of the immobile nutrients that helps in leaf formation and growth. Due to the rapid growth changes, the leaf grows faster than the supply of calcium reaching the growth areas, which limits the plant's

ability to translocate an appropriate amount of calcium to a specific portion of the leaves. Additionally, environmental conditions, including a poorly formed root system, high light intensity, high electrical conductivity (EC), insufficient air movement, especially between the plants, and fluctuating temperatures and humidity also contribute to the incidence of tip-burn [1,2,5].

The typical symptom of tip-burn is when necrotic (brown) spots can be seen on the tips and margins of the rapidly developing young leaves of the lettuce. The affected leaves deform and cannot grow properly as they expand, which reduces the quality of the lettuce and significantly affects its commercial value [1–3]. Therefore, the detection of tip-burns in an early stage is crucial so that a proper and timely treatment process can be performed. Currently, a visual assessment is the primary method used by experts and growers to identify tip-burn problems [7]. This method may often result in poor decisions because it is highly prone to human error, which will negatively impact agricultural products. Furthermore, particularly under the complicated and condensed growth conditions of an indoor farming system, the availability of specialists in assessing such issues may be limited and difficult. Thus, using computer or machine vision systems to automatically identify tip-burn problems is the most impactful method for early and proper treatment processes to maintain leaf quality.

Several machine vision applications have been developed to identify plant diseases or stress, such as the imaging method based on visible and near-infrared reflectance [8]. These methods combine spectral information with machine vision information. However, a detection process based on reflectance is not suitable for plant growth in dense conditions inside indoor environments due to the difficulty in image acquisition, environment constraint, and accessibility [8]. Additionally, a high implementation cost, time-consuming process, and the requirement for laboratory equipment setup make it ineffective for automatic and real-time identification of plant diseases [9]. In the training or learning process of image classification or object detection, deep-learning-based convolutional neural network has the potential for high accuracy of prediction with minimum preprocessing [8,10].

Deep neural network models can be implemented in the plant factories for near real-time detection of tip-burn. The evaluation of these models may provide new insight related to tip-burn detection, particularly in choosing the best deep-learning model for the relevant task. Furthermore, having an automated tip-burn detection for indoor farming or plant factories can alert growers to manipulate the controlled growing environment or regulate the nutrient supply system by increasing the calcium concentration, especially in inner leaves. Different indoor farming systems may adopt different light conditions, as they impact plant growth, physiology, and quality [11]. There are different spectrums of lighting that is also required for plant growth in the indoor plant environment. Thus, it is highly necessary to detect tip-burn at the different lighting conditions for prompt treatment of leaves and to provide nutrients so that they reach the deficient areas.

Therefore, the purpose of this research is to detect tip-burn disease of lettuce from single images captured under different light conditions (colors) in an indoor plant growing system. The light condition and background of the images are required to adjust with indoor plant growth environment under white, red, and blue LED colors. To achieve the purpose of this study, images were collected from the different lighting conditions and the images were trained with three different established one-stage detection models: CenterNet, YOLOv4, and YOLOv5. In this regard, this paper is organized into several sections: Section 1 highlights the tip-burn disease problem of indoor systems. In Section 2, relevant works and the potential of deep learning are discussed. Section 3 presents the dataset used for training and testing to predict tip-burn in lettuce plants. Sections 4 and 5 discuss the results of applying deep-learning models for tip-burn detection. Finally, Section 6 summarizes this work by presenting concluding remarks on the results and future research.

2. Related Works

A convolutional neural network (CNN) is a subset of the deep-learning techniques that are frequently used on image data to perform a variety of tasks, such as segmentation, object detection, and image classification. Generally, deep-learning-based detection algorithms can be divided into two-stage object detection and one-stage object detection. The core principle of a two-stage object detection is based on the use of a selective search algorithm to create a region proposal in the image for the targeted object, which is subsequently classified using a CNN [12]. Some popular models of this method include SPP-Net [13], RCNN [14], Fast R-CNN [15], and Faster R-CNN [16]. High detection accuracy may be attained with these approaches, but the drawbacks are the complexity of the network that require a longer training time and result in a diminished detection speed. In contrast, a one-stage detection method predicts all the bounding boxes in only a single run through the neural network [12]. Examples of the one-stage detector are SSD [17], YOLO [18], and CenterNet [19]. These one-stage detectors are not only able to reach high accuracies, but also have a faster processing speed [12], making them notable in the field of agriculture, where plant images are collected and utilized to classify plant species [20–22], count plants or fruits [23,24], identify pests [25–27] and weeds [28,29], and detect diseases [26,27,30–34].

Despite the fact that there are many deep-learning studies on the detection of different plant diseases, most of them have focused on plant detection in outdoor environments. Several public datasets have greatly contributed to plant disease detection, such as PlantVillage [20], PlantDoc [30], and PlantLeaves [35]. However, all the publicly accessible datasets only have available images of plant diseases that were cultivated in an outdoor environment. No public datasets are found for diseases or for stress analyses in indoor environments, particularly in indoor farming or plant factories, where plants are cultivated in multilayer structures that are closely clustered and are under artificial light [36,37].

Early detection of tip-burns in lettuce in the highly dense growing conditions of indoor environments is of great importance in reducing the cost of manual identification and improving lettuce quality and yield. Based on our literature search, there are very few studies on automatic tip-burn detection specifically for indoor farms. Shimamura et al. (2019) introduced tip-burn identification in plant factories using GoogLeNet for two classifications of tip-burn types from single lettuce images [7]. The images were collected under white lighting and a white background. Instead of single plant images, Gozzovelli et al. (2021) conducted tip-burn detection based on images of very dense plant canopies in plant factories, and Wasserstein generative adversarial network (WGAN) was applied to solve the problem of dataset imbalance (between healthy and unhealthy lettuces). YOLOv2 backbone Darknet-19 was used to detect tip-burns in their study [36]. Most recently, Franchetti and Pirri (2022) developed a new method for tip-burn stress detection and localization that was also deployed on plant canopy images. This technique used classification and self-supervised segmentation to locate and closely segment the stressed regions using ImageNet-1000 backbone Resnet-50V2 [37]. All of the studies collected the data on the uniform setup background, especially lighting conditions [7,36,37].

3. Materials and Methods

3.1. Plant Material and Cultivation Condition

Due to the limited availability of a public dataset of tip-burn lettuce grown in an indoor environment, to acquire the dataset, the lettuce plants were grown under conditions that can manifest tip-burns. As mentioned before, several conditions contribute to tip-burn in plants cultivated on indoor farms, including high temperature, high light intensity, extended day length, poor air flow, and a high concentration of nutrient solution [1,3,4,7]. Therefore, in this experiment, the aforementioned considerations were taken into account while estimating and adjusting the parameters for growing tip-burn lettuce.

The experiments were conducted in a small-scale cultivation room located at the Bioproduction and Machinery Laboratory, Tsukuba-Plant Innovation Research Center (T-PIRC), University of Tsukuba, Japan, in the spring season during the period from March to

June 2022. Green leaf lettuce seeds (*Lactuca sativa*) (Sakata Seed Corporation, Yokohama, Japan) were sown in hydroponic sponges for 14 days before they were transplanted to the hydroponic setup based on the nutrient film technique (NFT). Two sets of hydroponic systems were constructed from food-grade polyvinyl chloride (PVC) pipes, with each set consisting of three layers. The hydroponic systems were occupied with 189 total lettuce plants. The growing cycle was repeated twice. The lettuce plants were cultivated under a fully artificial light source with a combination of red, blue, and white LEDs, which contained wavelengths that are suitable for the plant photosynthesis process.

The lettuce cultivated in a solution culture had its roots immersed in a hydroponic nutrient solution (Hyponica Liquid Fertilizer, Kyowa Co., Ltd., Takatsuki, Japan). Initially, all the lettuce plants were cultivated normally with the standardized nutrient solution for the first two weeks after they were transplanted into the system. Then, the deficient solutions were induced through the system through a supply of nutrient solutions that had higher concentrations of nitrogen but were deficient in calcium. To obtain faster symptoms of tip-burn lettuce, the nutrient pH and EC were also adjusted within the range between 6.2–7 and 2.0–2.5 mS/cm, respectively, every three days to make the nutrient parameters fluctuate. The photoperiod was set to 24 h/day during the vegetative stage and 20 h/day during the harvest stage. The temperature and humidity observed throughout the growing period were 18–23 °C and 48–56%, respectively. The experiment was carried out until the plants showed symptoms of tip-burn.

3.2. Data Collection

Images of lettuce plants with tip-burn spots were collected starting from the first day they were visible by eye observation until the harvesting period. As tip-burn manifests on the tip of the leaves, the images were captured from the top of each infected plant with different angles, LED colors, and distances. The images were taken using a smartphone camera (Samsung Galaxy A50, Samsung Electronics Co., Ltd., Suwon, Korea) with aspect ratios of 1:1 and 3:4. Each image is 1080 × 2340 pixels. The total initial collection dataset of tip-burn lettuces was 538 images. The discolored leaf tips, brownish or blackish, indicated tip-burn (Figure 1).

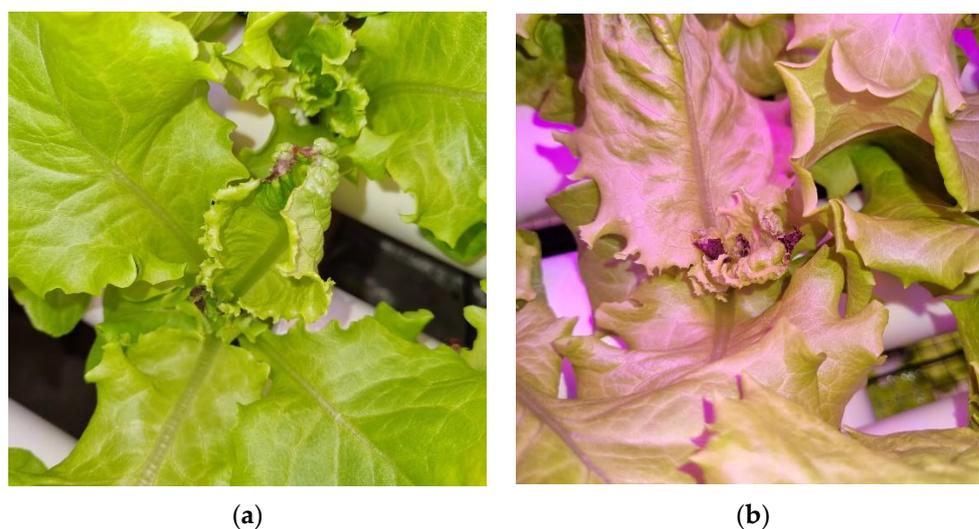


Figure 1. Example of captured images of tip-burn lettuce: (a) under white LEDs; (b) under red/blue LEDs.

3.3. Data Preparation

3.3.1. Data Labeling

Before labeling, we resized all the images to a uniform size and constant resolution of 640 × 640. The resized images were then uniformly numbered. After that, we performed

the labeling process using an open-source and free image annotation tool called LabelImg based on Python and Qt. Every visible and clear tip-burn spot was manually labeled by a rectangular bounding-box. Each output training image had a corresponding .txt file, containing the object class and coordinates of the bounding box of the upper left and lower right corners for each labeled tip-burn spot. In this experiment, tip-burn spots that were too small or highly indistinct were ignored and not labeled to prevent the possibility of these samples from degrading the neural network detection performance. The .txt file dataset was used for training in YOLOv4 and YOLOv5. Meanwhile, for CenterNet, we converted the .txt dataset into JSON format.

3.3.2. Data Augmentation and Splitting

A large dataset is required when training using a deep-learning algorithm as the model must extract and learn features from the images to identify and localize the targeted attributes. However, collecting a large dataset is very challenging and time consuming. Therefore, a data augmentation approach was performed to enhance our minimal dataset with the aim of being sufficiently able to be used to develop a reliable detection model [38]. In the experiment, several types of data augmentation were randomly used to increase the tip-burn images dataset. The data augmentations included brightness changes, image flips (mirrored original image horizontally or vertically), image rotations (rotated original image at 90° clockwise or counterclockwise), and shift (shifted original image horizontally and vertically and wrapped image around by the same image) (Figure 2).



Figure 2. The data augmentation performed in this experiment: (a) original image; (b) increase image brightness; (c) flipped image vertically; (d) rotated image 90° clockwise; (e) shifted image horizontally and vertically.

The final dataset was expanded to a total of 2333 images as a result of the augmentation process. Table 1 presents the number of images that were split into training, validation, and test datasets.

Table 1. Data splitting for training.

Model	Train	Validation	Test
CenterNet	1750	433	150
YOLOv4-5	1750	433	150

3.4. Training Process

3.4.1. CenterNet

CenterNet is an anchor-free target detection network which is an improvised model of CornerNet. The CenterNet-improved working principle makes them faster and more accurate than CornerNet. CenterNet recognizes each target object as a triplet of key points instead of a pair, as in CornerNet, which produces better precision and recall. In CenterNet, cascade corner pooling and center pooling were developed to enhance the data gathered from both the top-left and bottom-right corners and of an object to identify data from the targeted areas more clearly. The architecture of CenterNet begins with the input image entering the convolution neural network and employing cascade corner pooling to generate corner heatmaps, and center pooling to generate center heatmaps to the center point

(heatmap), offset and boxes with three branches for prediction, to obtain the results. Then, a pair of detected corners and corresponding embeddings were utilized, as in CornerNet, to predict a potential bounding box. Finally, the final bounding boxes were identified by using the detected center key points [19] (Figure 3).

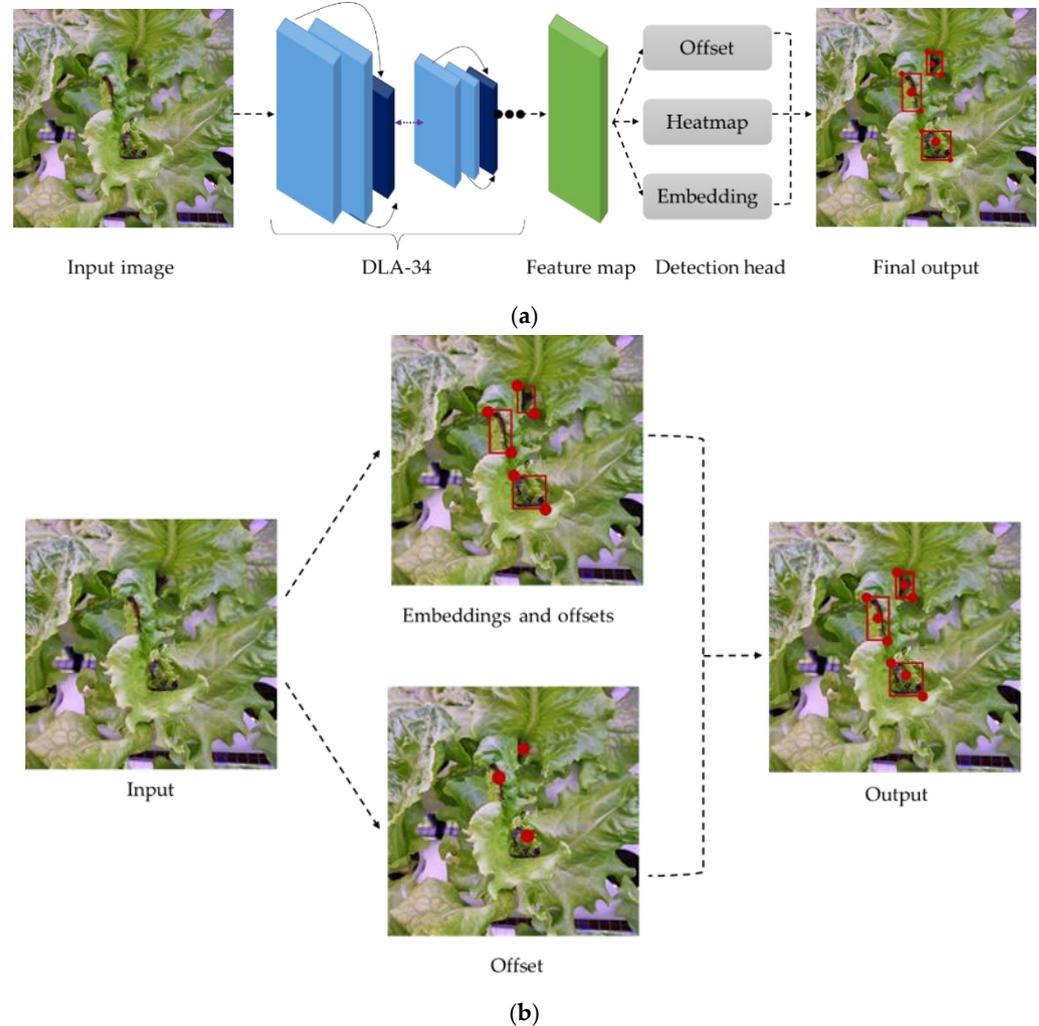


Figure 3. Working principle of CenterNet model (a) detection process of CenterNet based on backbone DLA-34; (b) tip-burn detection based on triplet key points.

3.4.2. YOLOv4 and YOLOv5

The YOLO model is basically an object detector based on bounding boxes. During the detection process, the input image is segmented uniformly into equal grids. If the target is within the grid, the model generates a predicted bounding box and an associated confidence score. Then, the target for a particular object class is recognized when the center of the target-class ground truth lies within a specific grid (Figure 4). YOLOv4 is an improved model based on the backbone of YOLOv3, namely Darknet-53, to improve the accuracy of detecting small objects. A residual block skip connection and upsampling were included in Darknet-53, which greatly enhanced the algorithm's accuracy. YOLOv4 further updates the network structure from YOLOv3 by changing to Cross Stage Partial Darknet-53 (CSPDarknet-53), where it utilizes convolution in the output layer. YOLOv4 additionally comes with batch normalization, a high-resolution classifier, and other tuning parameters to improve the detection result. The YOLOv4 model also employs multiscale prediction for detecting the final target, resulting in a better result in detecting small targets with high accuracy and speed. Additionally, YOLOv4 also offers a bag of freebies and a bag of

specials to increase the performance of the algorithm. The bag of freebies includes complete intersection over union (CIoU) and is mostly related to the different data augmentations, including mosaic and self-adversarial training (SAT). The main purpose of SAT is to identify the region in the image that the network depends on the most during training and then modify the image to disguise this dependency. This teaches the network to generalize to other new features when finding the target class. Bags of specials, on the other hand, consist of distance IoU- non-maximum suppression (DIoU-NMS) and the additional activation function, mish activation. The working architecture of YOLOv4 starts with inputting images into the CSPDarknet-53 for feature extraction; then, it is sent to path aggregation network (PANnet) to extract information in layers near the input by transmitting features to the detector from multiple backbone levels [39].

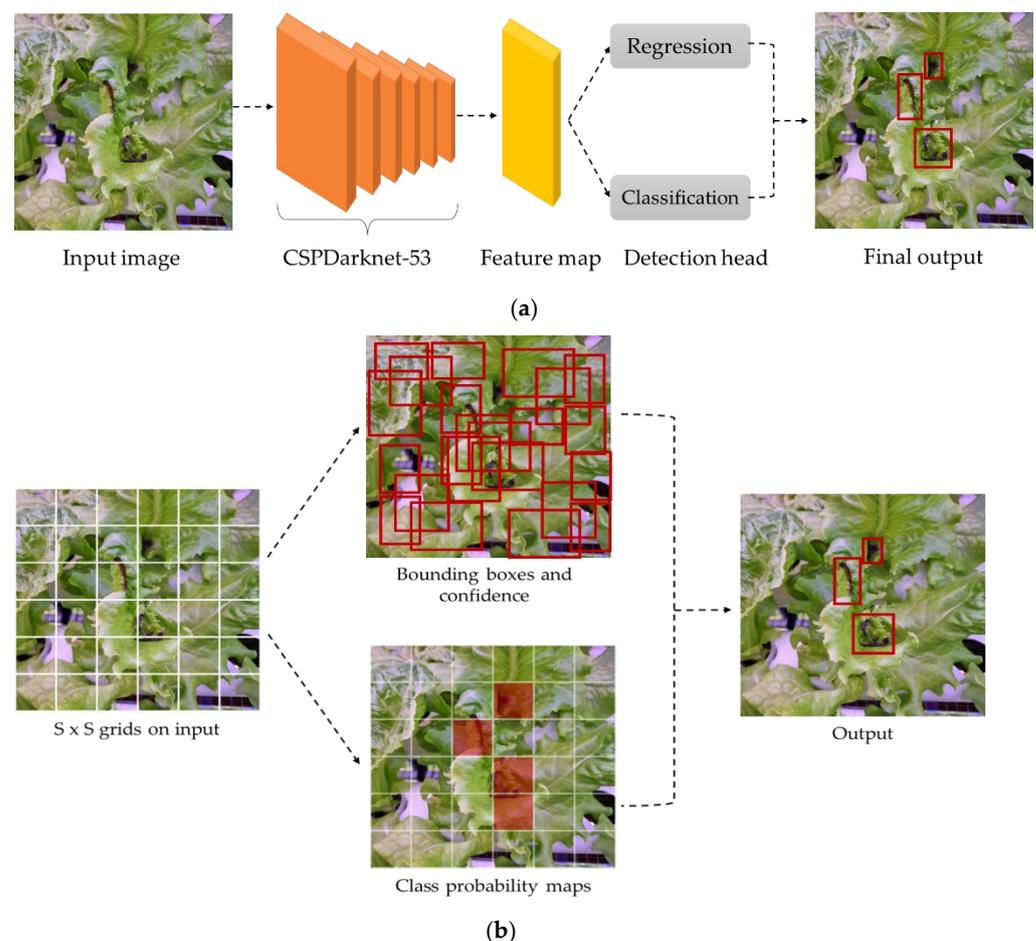


Figure 4. Working principle of the YOLO model: (a) detection process of YOLO model based on backbone CSPDarknet-53; (b) tip-burn detection based on bounding boxes.

YOLOv5 is a lighter version of the previous YOLO models that employs the PyTorch framework rather than the Darknet framework. It also uses CSPDarknet-53 as the backbone, similar to YOLOv4. The main difference in YOLOv5 is that it replaces the first three layers in the backbone of the YOLOv3 algorithm with a focus layer. The aim of the focus layer is to minimize the model size by eliminating certain parameter layers and parameters, reducing floating point operations per second (FLOPS) and CUDA memory, and increase the forward and backward speeds while minimizing the impact on the mean average precision (mAP) [40]. This process speeds up inference speed, improves accuracy, and reduces the computational load. In addition, YOLOv5 also utilizes PANet as its head part. The most significant improvements from YOLOv5 are auto-learning bounding box anchoring.

Training of the YOLOv4 and CenterNet models was conducted on a desktop PC with an NVIDIA[®] GTX 1650[™] with 4 GB GPU and an Intel[®] Xeon[™] E5-1607 CPU with 32 GB of RAM memory. The YOLOv5 model was trained on the cloud platform known as Google Colaboratory, a web-integrated development environment (IDE) with the GPU of Tesla P100-PCIE 16GB.

3.5. Performance Metrics

In our dataset, the target for the deep-learning models was to detect the tip-burn spot, hence we only had a single class, labelled as tip-burn. Outside of the target area or bounding box was predicted as background. Several performance metrics were used to evaluate the performance of the model, including intersection over union (IoU), precision, recall, and mean average precision (mAP). These metrics are based on PASCAL VOC which are well-known for use in object detection [41]. IoU is a measure of the distance between the predicted box and the ground truth box which ranges from zero (no overlap) to one (complete overlap). It highlights the preciseness of the algorithm in detecting the tip-burn. With the IoU, the total number of true-positives (TP), false-positives (FP), and false-negatives (FN) were determined. In this experiment, a TP is a tip-burn spot that was detected as a tip-burn. A FP is when another object, i.e., background, was detected as a tip-burn, while FN is when the tip-burn spot is not detected. Precision and recall are the two metrics that are commonly used to evaluate objection models. Precision determines the accuracy and preciseness of TP detection (Equation (1)), whereas recall indicates the effectiveness of the trained model in identifying all the TPs (Equation (2)). The mAP is the area under the precision and recall curve (Equation (3)).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{mAP} = \frac{1}{C} \sum_{k=1}^T P(k) \Delta R(k) \quad (3)$$

where C is total class numbers, T is IoU threshold numbers, k is the IoU threshold, P(k) is the precision, and R(k) is the recall.

Loss function in object detection is often used to indicate the degree of discrepancy between the predicted value and the true value of the model. There are three main losses in YOLOv5 denoted as: bounding box loss (box_loss), objectness loss (obj_loss), and classification loss (cls_loss). Box_loss represents bounding box regression loss to evaluate the preciseness of predicted bounding box on the target object (Equation (4)). Obj_loss is an objectness loss used to measure a confidence that the object falls in the proposed region of interest (Equation (5)). Cls_loss is a classification loss (Equation (6)).

$$\text{Box_loss} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{i,j}^{\text{obj}} \text{bj} (2 - w_i \times h_i) \left[(x_i \times \hat{x}_i^j)^2 + (y_i \times \hat{y}_i^j)^2 + (w_i \times \hat{w}_i^j)^2 + (h_i \times \hat{h}_i^j)^2 \right] \quad (4)$$

$$\text{Obj_loss} = \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{i,j}^{\text{noobj}} (c_i - \hat{c}_j)^2 + \lambda_{\text{obj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{i,j}^{\text{obj}} (c_i - \hat{c}_j)^2 \quad (5)$$

$$\text{Cls_loss} = \lambda_{\text{class}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{i,j}^{\text{obj}} \sum_{C \in \text{classes}} p_i(c) \log(\hat{p}_1(c)) \quad (6)$$

where S^2 is the grids number, B is bounding boxes in each grid, λ_{coord} is the coefficient of position loss, \hat{x} and \hat{y} denote the target true central position, \hat{w} is the target width, and \hat{h} is the target height. λ_{noobj} is the coefficient of no object existing in the bounding boxes. c_i and

\hat{c}_j represent the true confidence of bounding box and predicted confidence of bounding box, respectively. λ_{class} is the coefficient of classes loss. $p_i(c)$ is defined as the class probability of the target, and $\hat{p}_1(c)$ is the true value of the class.

Therefore, the total loss function is accumulated by using Equations (4) to (6) and expressed as:

$$\text{Loss} = \text{Box_loss} + \text{Obj_loss} + \text{Cls_loss} \quad (7)$$

4. Results

4.1. Training

The total number of datasets in this study was 2333, from which 1750 images were used for training, 433 images for validation, and the remaining 150 images for testing. The training settings for the three models were applied differently according to the model and dataset suitability (Table 2).

Table 2. Configuration for training for selected deep learning algorithms.

	Model	Input Size	Learning Rate	Batch Size	Epoch
CenterNet	ctdet_coco_dla_2x	416 × 416	0.00025	24	30
YOLOv4	-	416 × 416	0.0001	16	4000
YOLOv5	YOLOv5s	416 × 416	0.01	32	200

Table 3 shows the comparison of the validation results at 50% IoU for the three models. CenterNet shows a recall value of 58% with mAP of 81.2%. The YOLOv5 model yielded the highest training accuracy, with mAP of 84.1% and a recall score of 79.4%. The mAP value at the same IoU level for YOLOv4 was 76.2%, which was the lowest among all the models. Therefore, the YOLOv5 model had a relatively better performance in training than both the CenterNet and YOLOv4 models.

Table 3. Comparison of recall and mAP between CenterNet, YOLOv4, and YOLOv5 deep-learning algorithms.

Model	Recall (%)	mAP (%)
CenterNet	58.0	81.2
YOLOv4	74.0	76.2
YOLOv5	79.4	84.1

Figure 5 shows the results of losses and metrics obtained from the training and validation process using YOLOv5. Initially, the losses showed a rapid decline when reached at around 89 epochs from a total of 200 epochs. In this study, the classification loss was constant at 0 as we only trained for a single class, tip-burn. At the same time, as the losses decreased, the precision, recall, and mAP continued to increase before they reached the plateau phase. The training was stopped early at 189 epochs as no further improvement was observed (Figure 5).

4.2. Testing for Detection Accuracy

It is important to precisely detect tip-burn spots on lettuce to develop an effective automated tip-burn detection method for lettuce grown indoors. To evaluate the detection model, we tested 150 images by using the best trained weight obtained from each model. From the testing results, YOLOv5 shows the highest accuracy at 82.8% followed by CenterNet at 78.1%. YOLOv4 gives the lowest detection accuracy of only 67.6% (Table 4). The results indicate that the YOLOv5 model demonstrates better detection accuracy compared with CenterNet and YOLOv4.

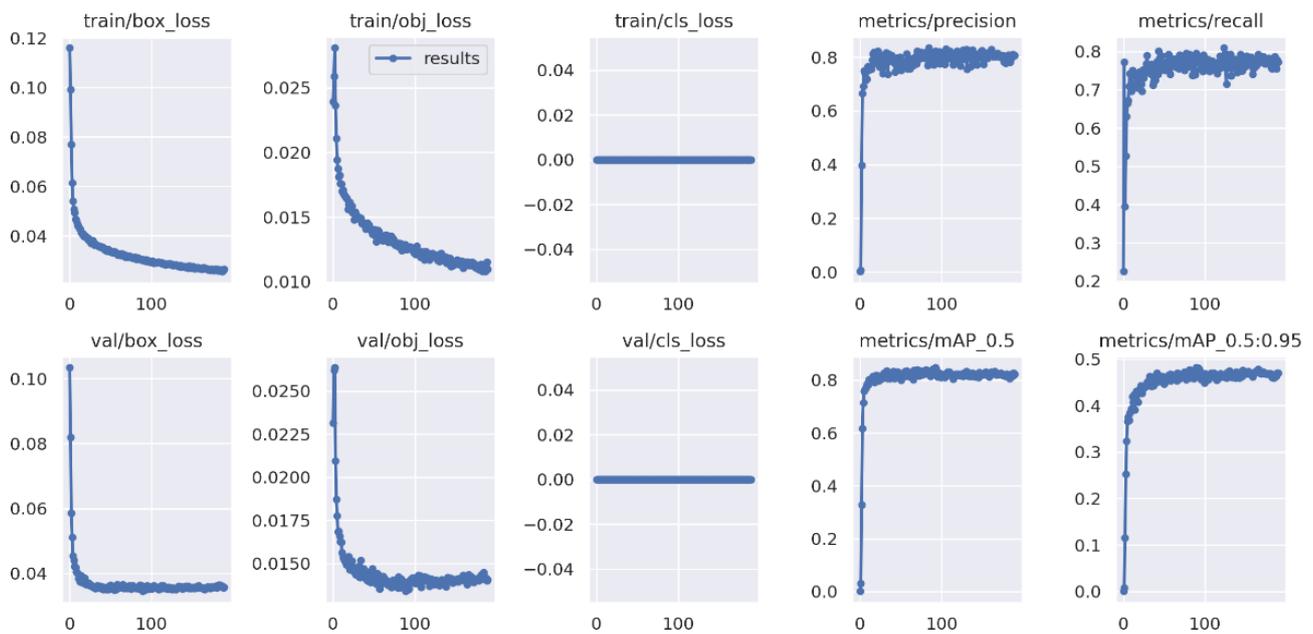


Figure 5. Training and validation results obtained from YOLOv5.

Table 4. Comparison detection of mAP between CenterNet, YOLOv4, and YOLOv5 deep-learning algorithms.

Model	mAP (%)
CenterNet	78.1
YOLOv4	67.6
YOLOv5	82.8

Figure 6 shows examples of tip-burn detection based on an image acquired under white LEDs. All models were able to detect all the tip-burn spots on the lettuce under white light conditions (Figure 6). However, YOLOv5 (Figure 6d) and YOLOv4 (Figure 6h) had misdetection of two tip-burn spots, whereas CenterNet only missed one spot (Figure 6f). This misdetection was plausible given that the very small tip-burn spots made it challenging for the models to detect.

Detection under red/blue light conditions is shown in Figure 7. The YOLOv5 had false positives that falsely detected tip-burns that were not in the image (Figure 7c). CenterNet had one misdetection (Figure 7e). On the other hand, both YOLOv5 (Figure 7d) and CenterNet (Figure 7f) missed the detection of one tip-burn. Meanwhile, YOLOv4 struggled to detect all the tip-burn spots accurately in Figure 7h, where none were detected. Some of these misdetections occurred due to an overlapping between the leaves in the background of the exact tip-burn spot. Moreover, this is understandable under this light condition; it is very difficult for even a human to identify where the exact tip-burns are in the image.

Figure 8 shows tip-burn detection under combination of white, red, and blue LEDs. The total of TP from the two proposed test images were eight (Figure 8a,b). Both YOLOv5 (Figure 8c) and YOLOv4 (Figure 8g) were able to detect all the tip-burn spots accurately, while CenterNet missed one (Figure 8e). YOLOv5 had a FP and one misdetection (Figure 8d). CenterNet had only one misdetection (Figure 8f), whereas YOLOv4 had two missed spots (Figure 8h).

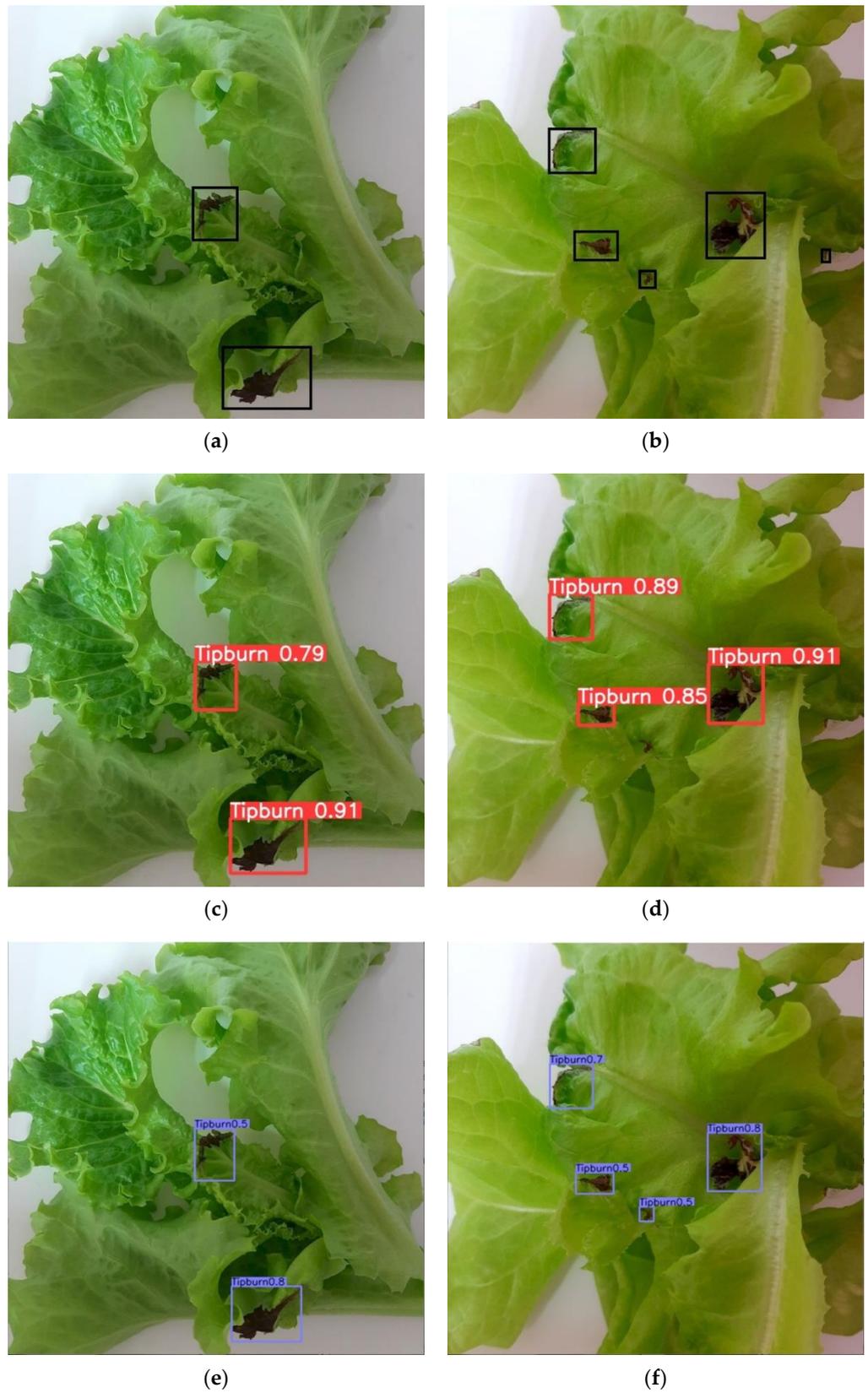


Figure 6. Cont.



Figure 6. Detection of tip-burn under white light conditions from (a,b) manually labelled; (c,d) YOLOv5; (e,f) CenterNet; (g,h) YOLOv4.

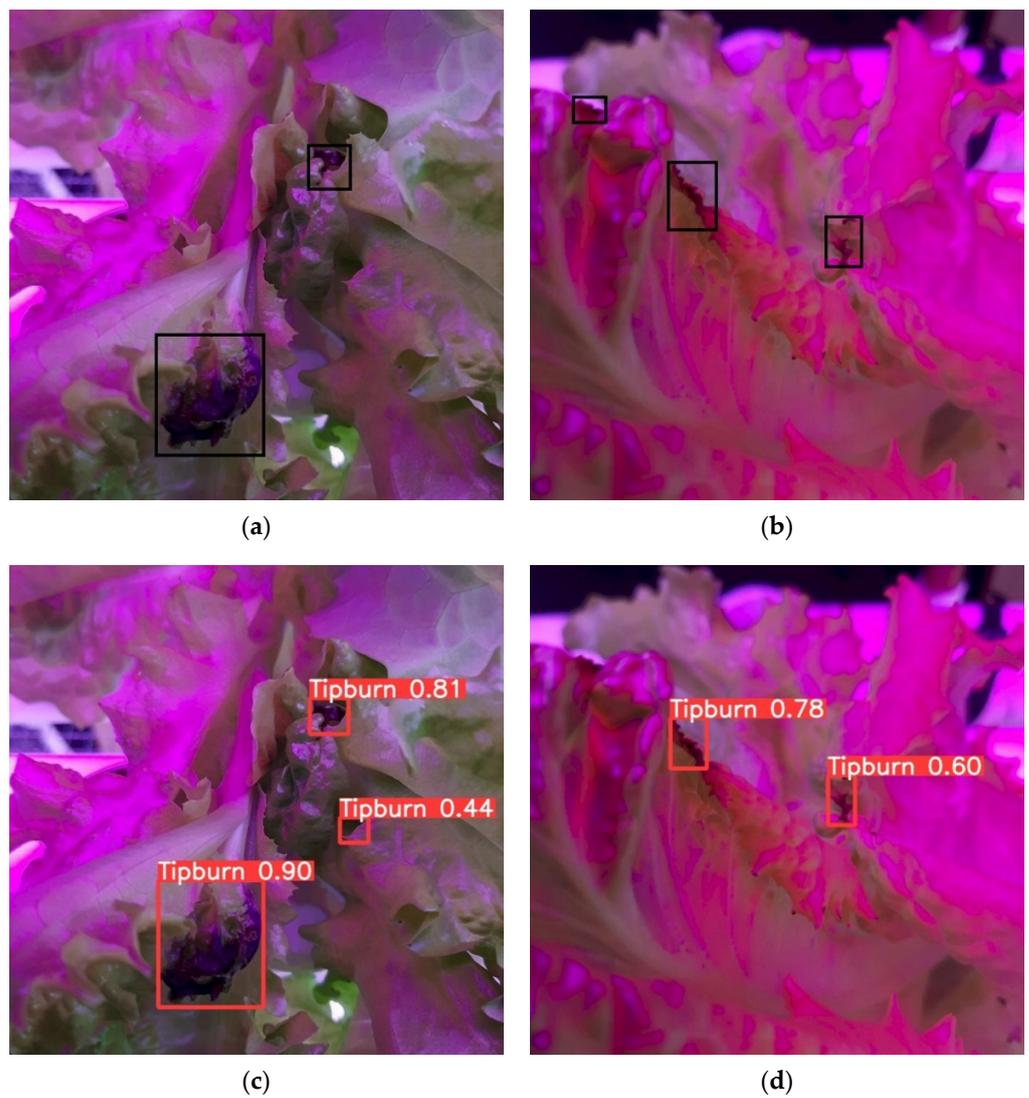
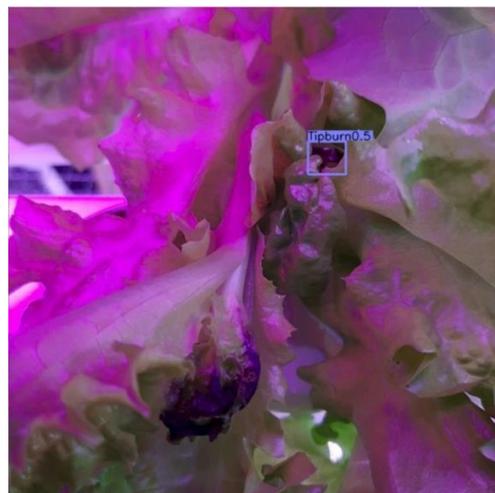
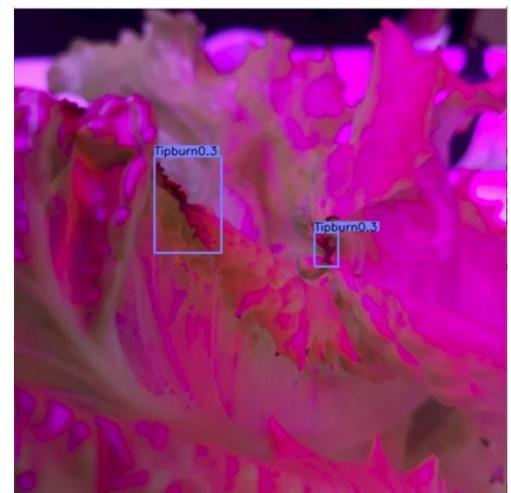


Figure 7. Cont.



(e)



(f)



(g)



(h)

Figure 7. Detection of tip-burn under red/blue light conditions from (a,b) manually labelled; (c,d) YOLOv5; (e,f) CenterNet; (g,h) YOLOv4.



(a)



(b)

Figure 8. Cont.

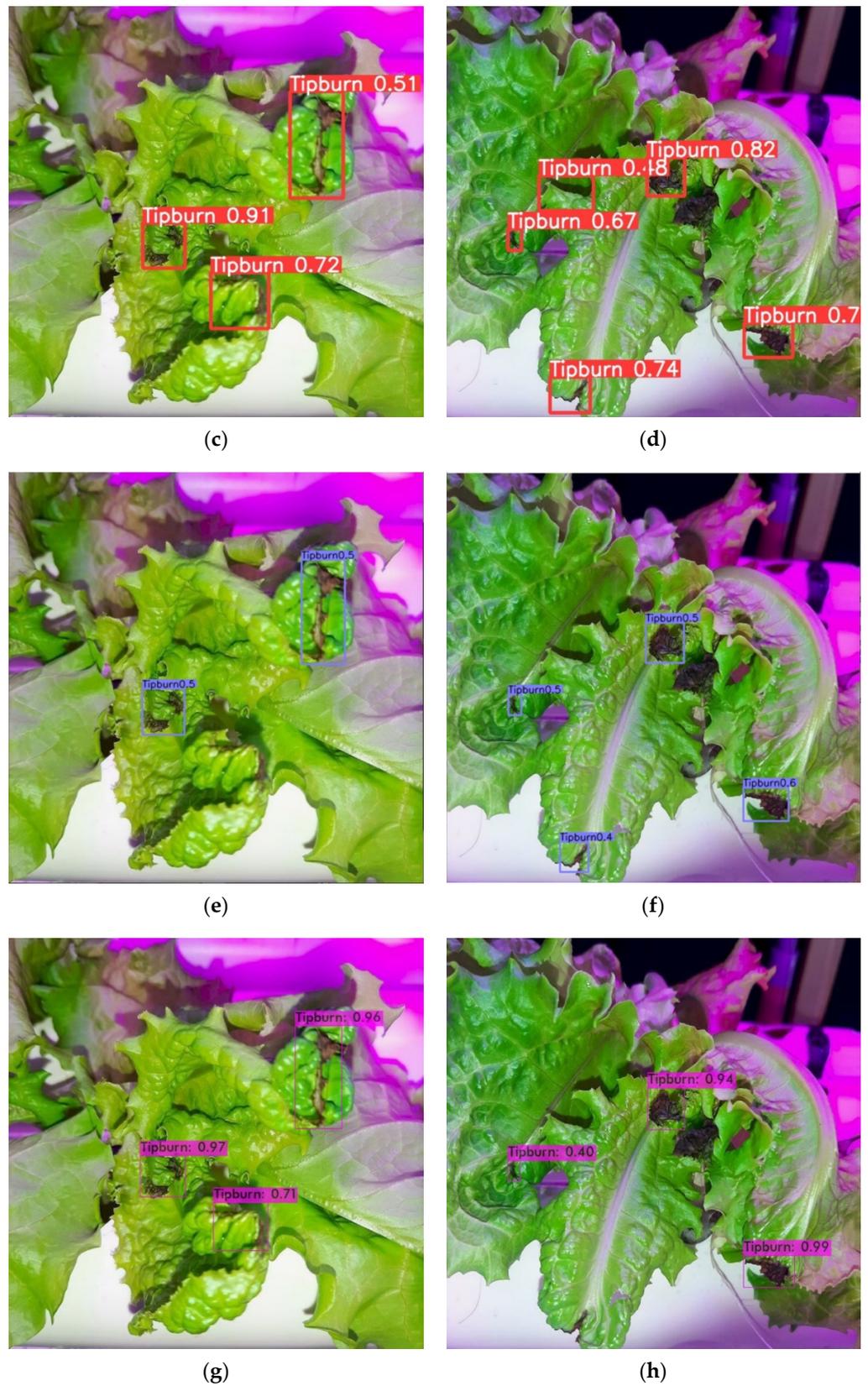


Figure 8. Detection of tip-burn under white/red/blue light conditions from (a,b) manually labelled; (c,d) YOLOv5; (e,f) CenterNet; (g,h) YOLOv4.

5. Discussion

Plant stress and disease detection is important not only for plants grown in the outdoor environment, but also for plants grown indoors. Although the main advantage of growing plants indoors is the potential to promote rapid growth for fast returns, it also has drawbacks, particularly the occurrence of tip-burn leaves [1]. In this paper, the detection of tip-burn lettuce grown on an indoor farm using three different deep-learning models based on a one-stage detector was performed. The models used in the study were CenterNet, YOLOv4, and YOLOv5. Overall, YOLOv5 outperformed the other two tested models with the highest mAP of 82.8%.

However, for tip-burn detection between different light conditions, most of the models could not perform well under red/blue LEDs. One of the reasons is due to the complex scenes, especially under red and blue lights, where the color of the tip-burn is difficult to differentiate from the background leaves, eventually confusing the models and producing a misdetection. Additionally, as mentioned before, it is very difficult for even a human to see and detect the tip-burn under these light conditions. In this case, a thermal camera may be useful for collecting a dataset under different light conditions, as it is not affected by visible light [42]. Apart from that, there may also be some errors and inconsistencies during the labeling process with certain tip-burn locations not being labeled properly.

Datasets are the most important part of every deep-learning algorithm. The quantity and quality of the input data are very crucial to generate the best model and most efficient system. In this paper, the dataset used for the training is relatively small compared with other deep-learning datasets. In this experiment, we prepared all the tip-burn datasets from several batches of plant growth, starting from seeds' germination, seedling, and growing in an indoor environment. It is very difficult and time consuming to generate the tip-burn dataset from batches after batches in the indoor plant growing systems. Moreover, the generation of the tip-burn is also randomly occurring where the deficiency is observed. Furthermore, the training dataset contained more images of tip-burn lettuce under white light than under the other light conditions. The imbalanced dataset may have a bias toward white light conditions during training, causing the models to lack enough data to learn and resulting in the low detection of tip-burns under red/blue LED conditions. We believe that the detection accuracy may be improved by collecting and appropriately labeling more images, particularly by balancing the image numbers under red/blue light conditions. Since growing batch data collection is difficult, we recommend exploring advanced data augmentation strategies, such as generative adversarial network (GAN), to produce additional high-quality artificial datasets. The GAN method can produce artificial images realistically as brand-new data compared with traditional augmentation methods such as flipping or rotating [36]. On the other hand, instead of using well-known performance metrics, it is also recommended to choose performance metrics based on dataset conditions either balanced or imbalanced dataset, which can provide more comprehensive perspective on the performance of the deep-learning models [43,44].

From this study, it is also noteworthy that a complex model with large parameter numbers, such as YOLOv4, had the lowest overall accuracy. YOLOv5 and CenterNet were both smaller and lighter than YOLOv4, but they could produce better accuracy. As stated in Table 3, we chose the YOLOv5s model and CenterNet ctdet_coco_dla_2x model, which are both smaller version of their original models. This indicates that when the target class and dataset are minimal, employing a model with large parameters may not be effective and suitable. Therefore, utilizing a comparatively small model or network makes it still possible to achieve accurate results with less computing facilities to develop a commercial system for detection of tip-burn lettuces. The capability of this small model of YOLOv5 can be further utilized for real-time monitoring of tip-burn lettuce detection in indoor farms, allowing for prompt and accurate interventions for early detection of tip-burn, such as modifying the growing environment with suitable humidity, temperature, light settings, and air movement, and additionally, providing calcium nutrients to the plants in the indoor farming systems.

6. Conclusions

Lettuce plants cultivated in indoor environments are fully reliant on artificial lighting sources. The ability to modify growing environments and lighting conditions can help to accelerate plant growth. However, this rapid growth process prevents the developing leaves from receiving an adequate amount of calcium, hence intensifying the incidence of tip-burn on plants grown indoors. Therefore, in this study, a method for the detection of tip-burn lettuce cultivated in an indoor environment under different light conditions was developed. Images of tip-burn lettuces were collected under white, red, and blue LEDs and were used as training, validation, and testing datasets for a deep-learning detection method. The detection method used was based on one-stage detectors, namely, CenterNet, YOLOv4, and YOLOv5. Among the three tested models, YOLOv5 achieved the best accuracy with 84.1% mAP. Nevertheless, further improvements can be made by using a larger dataset with balanced conditions to increase the detection accuracy. We believe this study provides an additional foundation for the automation of plant disease or stress detection in indoor farming systems, particularly under the different growing light conditions. This work can be extended in the future by employing this model for real-time application.

Author Contributions: Conceptualization, M.H.H. and T.A.; methodology, M.H.H.; formal analysis, M.H.H.; investigation, M.H.H.; resources, T.A.; data curation, M.H.H.; writing—original draft preparation, M.H.H.; writing—review and editing, T.A.; visualization, M.H.H.; supervision, T.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset that was generated and analyzed during this study is available from the corresponding author upon reasonable request, but restrictions apply to the data reproducibility and commercially confident details.

Acknowledgments: The authors would like to thank the Tsukuba Plant Innovation Research Center (T-PIRC), University of Tsukuba, for providing facilities for conducting this research. Furthermore, authors also express gratitude to the MEXT, Japan for providing scholarships to pursue this research at the University of Tsukuba.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, J.G.; Choi, C.S.; Jang, Y.A.; Jang, S.W.; Lee, S.G.; Um, Y.C. Effects of air temperature and air flow rate control on the tipburn occurrence of leaf lettuce in a closed-type plant factory system. *Hortic. Environ. Biotechnol.* **2013**, *54*, 303–310. [\[CrossRef\]](#)
2. Cox, E.F.; McKee, J.M.T. A Comparison of Tipburn Susceptibility in Lettuce Under Field and Glasshouse Conditions. *J. Hortic. Sci.* **1976**, *51*, 117–122. [\[CrossRef\]](#)
3. Goto, E.; Takakura, T. Reduction of Lettuce Tipburn by Shortening Day/night Cycle. *J. Agric. Meteorol.* **2003**, *59*, 219–225. [\[CrossRef\]](#)
4. Son, J.E.; Takakura, T. Effect of EC of nutrient solution and light condition on transpiration and tipburn injury of lettuce in a plant factory. *J. Agric. Meteorol.* **1989**, *44*, 253–258. [\[CrossRef\]](#)
5. Saure, M. Causes of the tipburn disorder in leaves of vegetables. *Sci. Hortic.* **1998**, *76*, 131–147. [\[CrossRef\]](#)
6. Tibbitts, T.W.; Rao, R.R. Light intensity and duration in the development of lettuce tipburn. *Proc. Amer. Soc. Hort. Sci.* **1968**, *93*, 454–461.
7. Shimamura, S.; Uehara, K.; Koakutsu, S. Automatic Identification of Plant Physiological Disorders in Plant Factory Crops. *IEEJ Trans. Electron. Inf. Syst.* **2019**, *139*, 818–819. [\[CrossRef\]](#)
8. Boulent, J.; Foucher, S.; Theau, J.; St-Charles, P.-L. Convolutional Neural Networks for the Automatic Identification of Plant Diseases. *Front. Plant Sci.* **2019**, *10*, 941. [\[CrossRef\]](#)
9. Fang, Y.; Ramasamy, R.P. Current and Prospective Methods for Plant Disease Detection. *Biosensors* **2015**, *5*, 537–561. [\[CrossRef\]](#)
10. Abade, A.; Ferreira, P.A.; Vidal, F.D.B. Plant diseases recognition on images using convolutional neural networks: A systematic review. *Comput. Electron. Agric.* **2021**, *185*, 106125. [\[CrossRef\]](#)
11. Wong, C.E.; Teo, Z.W.N.; Shen, L.; Yu, H. Seeing the lights for leafy greens in indoor vertical farming. *Trends Food Sci. Technol.* **2020**, *106*, 48–63. [\[CrossRef\]](#)

12. Jiao, L.C.; Zhang, F.; Liu, F.; Yang, S.Y.; Li, L.L.; Feng, Z.X.; Qu, R. A Survey of Deep Learning-Based Object Detection. *IEEE Access* **2019**, *7*, 128837–128868. [[CrossRef](#)]
13. Luvizon, D.; Tabia, H.; Picard, D. SSP-Net: Scalable Sequential Pyramid Networks for Real-Time 3D Human Pose Regression. *arXiv* **2020**, arXiv:2009.01998. [[CrossRef](#)]
14. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [[CrossRef](#)]
15. Girshick, R. Fast r-cnn. *arXiv* **2015**, arXiv:1504.08083. [[CrossRef](#)]
16. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
17. Liu, W.; Dragomir Anguelov, D.E.; Szegedy, C.; Reed, S.E.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. *arXiv* **2015**, arXiv:1512.02325.
18. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]
19. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision 2019, Seoul, Korea, 27 October–2 November 2019; pp. 6569–6578.
20. Hughes, D.; Salathé, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv* **2015**, arXiv:1511.08060.
21. Li, G.; Kong, M.; Wang, S. Research on Plant Recognition Algorithm Based on YOLOV3 in Complex Scenes. In Proceedings of the 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI), Kunming, China, 17–19 September 2021; pp. 343–347. [[CrossRef](#)]
22. Hassan, A.; Islam, S.; Hasan, M.; Shorif, S.B.; Habib, T.; Uddin, M.S. Medicinal Plant Recognition from Leaf Images Using Deep Learning. In *Computer Vision and Machine Learning in Agriculture*; Springer: Singapore, 2022; Volume 2, pp. 137–154. [[CrossRef](#)]
23. Parico, A.; Ahamed, T. Real Time Pear Fruit Detection and Counting Using YOLOv4 Models and Deep SORT. *Sensors* **2021**, *21*, 4803. [[CrossRef](#)]
24. Lu, S.; Song, Z.; Chen, W.; Qian, T.; Zhang, Y.; Chen, M.; Li, G. Counting Dense Leaves under Natural Environments via an Improved Deep-Learning-Based Object Detection Algorithm. *Agriculture* **2021**, *11*, 1003. [[CrossRef](#)]
25. Chen, J.-W.; Lin, W.-J.; Cheng, H.-J.; Hung, C.-L.; Lin, C.-Y.; Chen, S.-P. A Smartphone-Based Application for Scale Pest Detection Using Multiple-Object Detection Methods. *Electronics* **2021**, *10*, 372. [[CrossRef](#)]
26. Li, D.; Ahmed, F.; Wu, N.; Sethi, A.I. YOLO-JD: A Deep Learning Network for Jute Diseases and Pests Detection from Images. *Plants* **2022**, *11*, 937. [[CrossRef](#)]
27. Liu, J.; Wang, X. Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network. *Front. Plant Sci.* **2020**, *11*, 898. [[CrossRef](#)]
28. Jin, X.; Che, J.; Chen, Y. Weed Identification Using Deep Learning and Image Processing in Vegetable Plantation. *IEEE Access* **2021**, *9*, 10940–10950. [[CrossRef](#)]
29. Jin, X.; Sun, Y.; Che, J.; Bagavathiannan, M.; Yu, J.; Chen, Y. A novel deep learning-based method for detection of weeds in vegetables. *Pest Manag. Sci.* **2022**, *78*, 1861–1869. [[CrossRef](#)]
30. Singh, D.; Jain, N.; Jain, P.; Kayal, P.; Kumawat, S.; Batra, N. PlantDoc: A dataset for visual plant disease detection. In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, Hyderabad, India, 5–7 January 2020; pp. 249–253. [[CrossRef](#)]
31. Sun, J.; Yang, Y.; He, X.; Wu, X. Northern Maize Leaf Blight Detection Under Complex Field Environment Based on Deep Learning. *IEEE Access* **2020**, *8*, 33679–33688. [[CrossRef](#)]
32. Wang, X.; Liu, J. Tomato anomalies detection in greenhouse scenarios based on YOLO-Dense. *Front. Plant Science.* **2021**, *12*, 634103. [[CrossRef](#)] [[PubMed](#)]
33. Rashid, J.; Khan, I.; Ali, G.; Almotiri, S.H.; AlGhamdi, M.A.; Masood, K. Multi-Level Deep Learning Model for Potato Leaf Disease Recognition. *Electronics* **2021**, *10*, 2064. [[CrossRef](#)]
34. Li, Y.; Sun, S.; Zhang, C.; Yang, G.; Ye, Q. One-Stage Disease Detection Method for Maize Leaf Based on Multi-Scale Feature Fusion. *Appl. Sci.* **2022**, *12*, 7960. [[CrossRef](#)]
35. Chouhan, S.S.; Singh, U.P.; Kaul, A.; Jain, S. A data repository of leaf images: Practice towards plant conservation with plant pathology. In Proceedings of the 2019 4th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 21 November 2019; pp. 700–707. [[CrossRef](#)]
36. Gozzovelli, R.; Franchetti, B.; Bekmurat, M.; Pirri, F. Tip-burn stress detection of lettuce canopy grown in Plant Factories. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1259–1268.
37. Franchetti, B.; Pirri, F. Detection and Localization of Tip-Burn on Large Lettuce Canopies. *Front. Plant Sci.* **2022**, *13*, 874035. [[CrossRef](#)]
38. Cui, X.; Goel, V.; Kingsbury, B. Data augmentation for deep convolutional neural network acoustic modeling. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, Australia, 19–24 April 2015; pp. 4545–4549. [[CrossRef](#)]

39. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
40. Glenn, R.J. YOLOv5 Focus Layer #3181. In *Ultralytics: Github*. 2021. Available online: <https://github.com/ultralytics/yolov5/discussions/3181m1> (accessed on 20 July 2022).
41. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2009**, *88*, 303–338. [[CrossRef](#)]
42. Jiang, A.; Noguchi, R.; Ahamed, T. Tree Trunk Recognition in Orchard Autonomous Operations under Different Light Conditions Using a Thermal Camera and Faster R-CNN. *Sensors* **2022**, *22*, 2065. [[CrossRef](#)] [[PubMed](#)]
43. Luque, A.; Carrasco, A.; Martín, A.; de las Heras, A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognit.* **2019**, *91*, 216–231. [[CrossRef](#)]
44. Luque, A.; Mazzoleni, M.; Carrasco, A.; Ferramosca, A. Visualizing Classification Results: Confusion Star and Confusion Gear. *IEEE Access* **2021**, *10*, 1659–1677. [[CrossRef](#)]