



# A two-stage deep-learning based segmentation model for crop disease quantification based on corn field imagery

L.G. Divyant<sup>a</sup>, Aanis Ahmad<sup>b</sup>, Dharmendra Saraswat<sup>c,\*</sup>

<sup>a</sup> Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur, West Bengal, India

<sup>b</sup> Elmore Family School of Electrical and Computer Engineering, Purdue University West Lafayette, Indiana, United States

<sup>c</sup> Agricultural and Biological Engineering, Purdue University West Lafayette, Indiana, United States



## ARTICLE INFO

### Keywords:

Plant diseases  
Deep learning  
Semantic segmentation  
SegNet  
UNet  
DeepLabV3+

## ABSTRACT

It is important to develop accurate disease management systems to identify and segment corn disease lesions and estimate their severity under complex field conditions. Although deep learning techniques are becoming increasingly popular to identify singular diseases, access to robust models for identifying multiple diseases and segmenting lesion areas for severity estimation under field conditions remain unsolved. In this study, a custom dataset consisting of handheld images of corn leaves infected with Gray Leaf Spot (GLS), Northern Leaf Blight (NLB), and Northern Leaf Spot (NLS) diseases, acquired under field conditions, was used to develop a novel two-stage semantic segmentation approach for identifying corn diseases and estimate their severity. Three semantic segmentation models were trained for each stage using SegNet, UNet, and DeepLabV3+ network architectures. Stage one used semantic segmentation to extract leaves from complex field backgrounds. In stage two, semantic segmentation was used to locate, identify, and calculate area coverage for disease lesions. After the models were trained, the best performance for stage one was observed from the UNet model, which achieved up to 0.9422 mean weighted intersection over union (mwIoU) and 0.8063 mean boundary F1-score (mBFScore). The best performance for stage two was observed from the DeepLabV3+ model, which could identify the disease lesions with a mwIoU of 0.7379 and mBFScore of 0.5351. Finally, severity was estimated by calculating the percentage of leaf area covered by disease lesions. In the test set, an R<sup>2</sup> value (coefficient of determination) of 0.96 was achieved, which denotes that the integrated (UNet-DeepLabV3+) model predicted the severity of three diseases very close to the actual observations. This study developed a novel two-stage deep learning-based approach to accurately identify three targeted corn diseases and estimate their severity to pave the way for developing a field-worthy disease management system.

## 1. Introduction

Corn (*Zea mays*) is an important global food crop and is the third leading crop based on planting area [1]. It is cultivated on about 200 million hectares globally and accounts for nearly 40% of global grain production [2]. In the US, corn production stood at 13 billion bushels from about 32 million hectares of land, at a national average of nearly 300 bushels per acre [2]. However, diseases are a major yield-limiting factor in corn production systems. Gray Leaf Spot (GLS), Northern Leaf Blight (NLB), and Northern Leaf Spot (NLS) are common foliar diseases of corn found in the Midwest US. Greyish tan, near rectangular lesions on the leaf are symptomatic of GLS. The symptoms of NLB include long, elliptical to cigar-shaped tan-brown lesions that form

parallel to leaf margins. NLS symptoms include greyish tan narrow or concentric circular lesions surrounded by a dark border.

Farmers rely on pesticides to treat crops and check the further spread of diseases to counter yield losses due to diseases. Proper management of diseases is important for both economic and ecological purposes. The Integrated Pest Management (IPM) programs have effectively reduced pesticide use without causing a reduction in yields or profits [3]. Accurately identifying diseases and estimating severity is crucial for assisting farmers in site-specific disease management (SSDM). With the recent advancements in computational resources and the evolution of Convolutional Neural Networks (CNNs), researchers rely on deep learning to develop disease management systems. Before the use of deep learning, traditional image processing techniques [4,5], spectral indices

\* Corresponding author.

E-mail address: [saraswat@purdue.edu](mailto:saraswat@purdue.edu) (D. Saraswat).

[6], region growth algorithms [7], and machine learning [8] was proposed for identifying diseases. However, deep learning-based solutions have become increasingly popular due to the advancement in computational resources, availability of large datasets, and rapid development of CNN architectures. Unlike traditional computer vision solutions that rely on identifying important features, deep learning uses CNNs to automatically learn important features within training data, resulting in a less biased model [9,10]. CNNs have overcome the limitations of conventional methods based on pattern recognition for feature extraction. Analyzing images using deep learning has provided better and faster results than traditional image processing techniques [11]. Many studies have reported promising results for disease identification and severity estimation using deep learning-based image classification across a range of crops, including rice [12,13], wheat [14,15], corn [16], tomato [17], apple [18,19], grapes [20], and tea [21]. Although promising results were reported for deep learning-based image classification for disease identification, image classification could not localize leaf lesions. In contrast to deep learning-based image classification, identifying the location of the disease lesions at the pixel level through segmentation is an important consideration [22]. Segmentation models create a pixel-wise mask for the desired objects in the image, providing a more granular understanding of the objects. To the best of our knowledge, limited studies have been conducted to segment disease lesions using deep learning-based semantic segmentation.

Semantic segmentation is a deep learning technique in which every image pixel is associated with a class label [23]. A semantic segmentation model proposed to segment powdery mildew regions on cucumber leaf images outperformed K-means and random forest algorithms [24]. Unfortunately, since the dataset comprised leaf images over a uniform background, the developed model failed to work under field conditions. A new architecture based on SegNet was proposed with different class weight ratios for semantic segmentation of late blight lesions and severity estimation in potato leaf [25]. The researchers also statistically compared the results of FCN (Fully Convolutional Network), PSPNet, and DeepLab models with the proposed architecture. The model achieved an intersection over union (IoU) score of 0.996 and 0.386 for background and lesion, respectively. Two recent studies [26,27] proposed a deep learning model capable of performing semantic segmentation of disease lesions in coffee and deployed the model on a smartphone application for Android. However, the effect of varying backgrounds was not taken into consideration. Thus, a key challenge for estimating plant disease severity is accurately segmenting the disease lesions under complex field conditions. A leaf instance segmentation model was added to address the challenge of removing background in field images [27].

Studies have been conducted to estimate the severity of diseases using deep learning-based image classification [28,29]. Some studies that used segmentation models for extracting plants from images acquired under field conditions reported a performance reduction due to complex backgrounds [30,31]. Therefore, background removal can increase the performance of deep learning-based disease identification models. Henceforth, it is important to highlight the necessity of fast and automatic background removal in leaf images [30]. Two CNN models were set up to compare the effect of background removal during pre-processing for identifying 56 diseases affecting 12 plant species [32]. The comparison displayed an increased accuracy for most plants when the background was removed. Although high accuracies were reported, such models could not localize disease lesions during identification. Moreover, most of the diseases affecting corn exhibit similar visual characteristics, making the traditional analysis methods difficult and tiresome.

This study attempts to overcome the limitations in existing research by proposing a novel two-stage deep learning-based semantic segmentation approach for disease identification and severity estimation to improve the reliability of plant disease management systems. Two main objectives identified for this study were: (1) to evaluate the performance

of three state-of-the-art semantic segmentation models using the SegNet, UNet, and DeepLabv3+ network architectures for corn leaf segmentation and multi-class lesion (GLS, NLB, and NLS) segmentation; and (2) build a two-stage model for severity estimation of the targeted diseases.

The remaining paper is organized as follows: Section 2 introduces the dataset and ground truth preparation, architectures of deep learning networks employed in this study, and the proposed two-stage model. Section 3 presents the results and discussions. Finally, the paper concludes with Section 4.

## 2. Materials and methods

### 2.1. Dataset

#### 2.1.1. Dataset and annotation masks

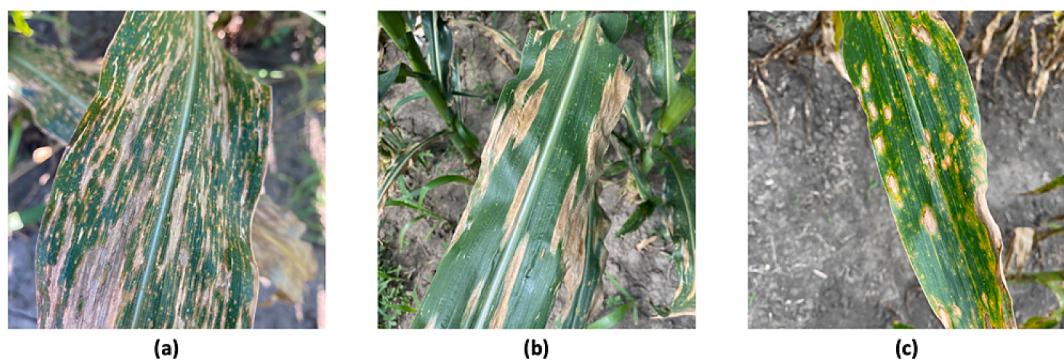
A custom corn disease dataset acquired under field conditions was used for developing the two-stage semantic segmentation model. The dataset consisted of field images of corn leaves infected by GLS, NLB, and NLS diseases [33]. A total of 1050 images of leaves affected by one or more diseases were collected at Purdue University's Agronomy Center for Research and Education (ACRE). Images were captured under varying illumination conditions in the field. Samples of leaf images in the dataset affected by GLS, NLB, and NLS are shown in Fig. 1. The size of the images was  $512 \times 512$  pixels, with 96 dpi resolution. The images featured various background elements: weeds, soil, leaves, stems, shadows, and human hands. From this set, 100 images were reserved for testing the final two-stage model that will be discussed in the later sections of the article. The remaining 950 images were utilized for training and testing the SegNet, UNet, and DeepLabv3+ networks for corn leaf segmentation and lesion segmentation.

#### 2.1.2. Corn leaf segmentation dataset

Among the 950 images assembled, 350 images were separated from the dataset for training and testing the corn leaf segmentation task models. All images were annotated using the MATLAB® Image Labeler app, which provided an interface with a wide range of annotation tools such as assisted freehand, smart polygon, and flood fill, thus facilitating the creation of ground truths manually and quickly. Examples of images in this dataset and their respective masks are shown in Fig. 2. The images featured diverse conditions associated with the background, shadows, blurring and sharpness effects, and target leaf orientation. For this task, the 350 images and their corresponding masks were divided into 250 image-mask pairs for training purposes and 100 for testing. Geometrical image data augmentations such as rotations ( $-90^\circ$  and  $+90^\circ$ ), horizontal and vertical mirroring, and translation along x- and y- axes were performed on the train set to increase its volume by eight times.

#### 2.1.3. Lesion segmentation dataset

The remaining 600 images of size  $512 \times 512$  pixels were used for lesion segmentation. Since this second task targeted segmenting the lesions from the extracted leaf portion (i.e., after background elimination in the first step), the background in this image dataset had to be removed before preparing the ground truth. A small code script was written in MATLAB® software to perform manual segmentation by calling individual images from the folder to the interface of the MATLAB® Image Segmenter app. After removing the background manually, the leaf portions consisting of GLS, NLB, and NLS lesions were annotated to their respective disease classes. Hence, there were four classes in this dataset, namely: "gls", "nlb", "nls", and "background". Examples of images in this dataset and their respective ground truths are shown in Fig. 3. A set of 450 images were used for training and the remaining 150 were for testing purposes. The augmentation methods mentioned in Section 2.1.1 were also performed on this modified dataset.



**Fig. 1.** Sample images of corn leaves affected by (a) Gray Leaf Spot (b) Northern Leaf Blight (c) Northern Leaf Spot.



**Fig. 2.** Sample images present in the corn leaf segmentation dataset.

## 2.2. Convolutional neural networks and deep learning architectures

Convolutional Neural Network (CNN) is a feed-forward neural network usually employed for image and visual analysis. The hidden layers in a typical CNN consist of a combination of one or more of the following: convolutional layers to extract features from the input or the previous layer; pooling layers that generalize the presence of features; batch normalization layers; activation function layers; and fully connected layers. The convolutional layers [34] are defined by the convolutional kernels or filters that help in transforming and highlighting the patterns in the input data. Pooling layers reduce the data dimensions by linking neuron clusters of the previous layer to a single neuron. Further, the image classification occurs in the fully connected layers, where the activations are processed as flattened matrices. Deep learning has reduced the necessity of time-consuming, complex feature engineering tasks [31,35]. Various network architectures have been proposed for efficient feature extraction. Some of the most successful architectures, among others, are the AlexNet [36], VGG [37], GoogLeNet [34], and ResNet [38].

Apart from image data collection and ground truth annotation, the remaining of this study comprised of three main steps: training and

evaluation of deep learning-based semantic segmentation networks, namely SegNet, UNet, and DeepLabv3+, for corn leaf semantic segmentation to eliminate background and disease lesion segmentation from corn leaf; and finally, disease severity estimation. Before disease severity estimation, a two-stage model was designed by combining the best models from the initial investigations for segmenting corn leaves and disease lesions. The novel deep learning-based two-stage semantic segmentation approach is illustrated in Fig. 4.

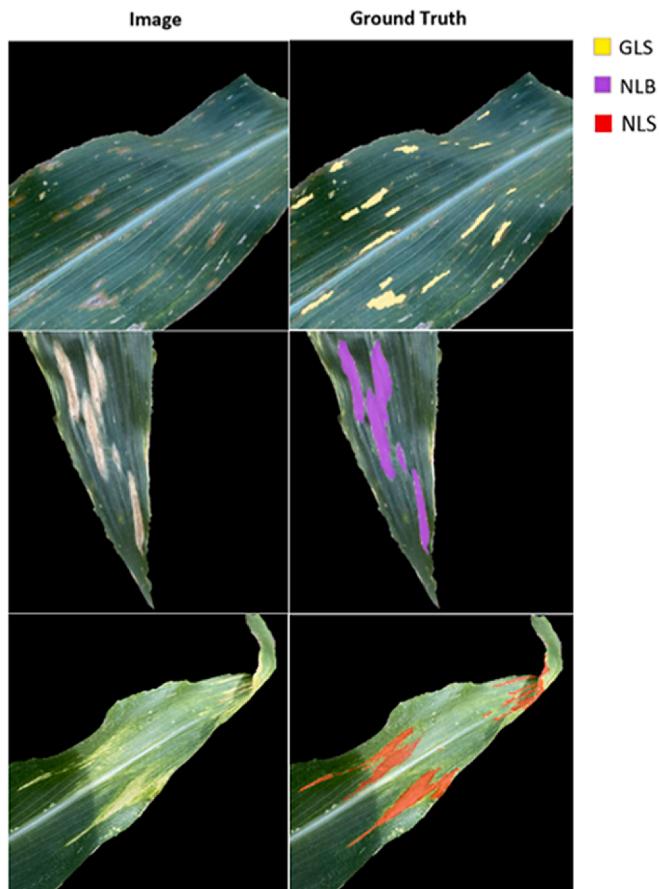
Three CNN-based architectures, namely, SegNet [39], UNet [40], and DeepLabv3+ [41], were evaluated for the semantic segmentation of (i) corn leaf from field images with complex backgrounds and (ii) lesion from the background eliminated leaf image. The study's network architectures are state-of-the-art, best-performing for pixel-level segmentation and localization in 2-dimensional images. Apart from their top-level performances, even with small datasets, their architectures and the networks' depth are easily adjustable [42]. The models are based on stacks of contracting paths called the encoder and expanding paths referred to as the decoder.

### 2.2.1. SegNet

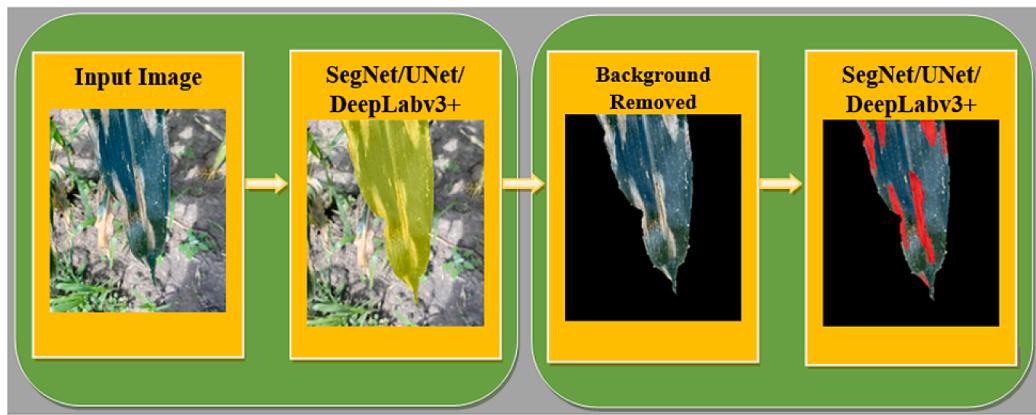
SegNet [39] comprises a group of encoders followed by their corresponding decoders, which feed into a softmax pixel classification layer. The proposed network operates on images of  $512 \times 512$  pixels and outputs a segmentation mask of the same size. Each encoder comprises two sets of convolutional, batch-normalization (BN) and ReLU layers, and a max-pooling layer at the end (Fig. 5). All convolutional layers had  $3 \times 3$  filters with  $1 \times 1$  stride; pooling layers performed  $2 \times 2$  max-pooling with  $2 \times 2$  strides and zero paddings. The convolutional layers in all the four encoders consisted of 64 filters. Thus, the input image of dimension  $512 \times 512 \times 3$  resulted in a set of feature maps of dimension  $32 \times 32 \times 64$  at the output of the encoder stacks. Each set begins with a max-unpooling layer in the decoder stack, followed by convolution, BN, and ReLU layers (Fig. 5). These unpooling layers draw feature information from the previous decoder and the pooling layer of the paired encoder. Similar to encoders, 64 filters were present in all the decoder's convolutional layers except the last one. The number of filters in this last layer corresponds to the number of classes. So, activations of size  $512 \times 512 \times 2$  for leaf segmentation and  $512 \times 512 \times 4$  for lesion segmentation networks (depending on where the model is employed) were derived at the end of the decoder stacks. It is then passed into the softmax and classification layers for pixel-wise classification.

### 2.2.2. UNet

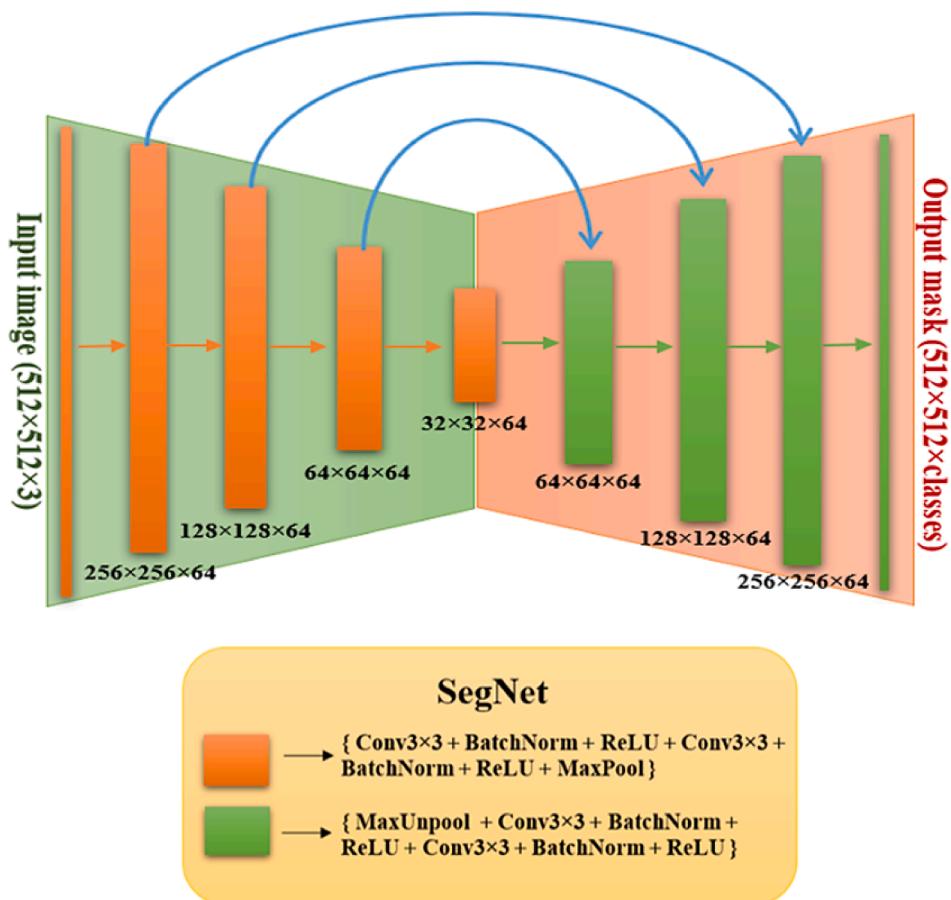
The architecture of UNet [40] closely resembles SegNet, comprising encoders and decoders. They differ in how the encoders' activations are transferred to the decoders. The UNet concatenates a copy of features at the end of each encoder with the features at the end of the corresponding decoder stage. Unlike UNet, SegNet transfers only the pooling indices to the decoder. Furthermore, decoders in UNet consist of transposed



**Fig. 3.** Sample images in the lesion segmentation dataset and their respective ground truths.



**Fig. 4.** Schematic illustration of deep-learning-related activities in the study: Step 1: segmentation of leaf from field images through a SegNet, UNet, or DeepLabv3+ model; Step 2: semantic segmentation of lesion spots from the image without the background, performed via SegNet, UNet or DeepLabv3+.



**Fig. 5.** Schematic diagram of the SegNet model. The decoder's unpooling layers (MaxUnpool) extract features from the previous layer and the corresponding encoder's pooling layer.

convolutional layers (deconvolutional layers) for an up-sampling operation whose weights must be learned during the training process. This is not the case with SegNet since the up-sampling of pooling indices received from the corresponding encoder is performed by the un-pooling layers.

The architecture of UNet is illustrated in Fig. 6. Each encoder contains two pairs of convolutional (stride  $1 \times 1$  and same padding) and ReLU layers, followed by a max-pooling layer. A decoder consists of a transposed convolutional layer, depth concatenation (with the activations of the corresponding encoder), and two convolutional layers. The encoders and decoders are merged in a bridge, a pair of convolutional and ReLU layers. A dropout layer with a 50% dropout rate was placed at the end of the bridge.

#### 2.2.3. DeepLabV3+

The FCNs [43] was commonly used for deep learning-based semantic segmentation tasks. However, the input activations map becomes smaller and smaller while passing through the convolution and pooling layers, leading to the loss of information in the images. The creators of the DeepLabv3+ model [41] addressed this issue with the introduction of parallel ‘atrous convolutions’ (or called dilated convolutions) and Atrous Spatial Pyramid Pooling (ASPP) layers [44]. The atrous convolutions give an effective field-of-view to the convolution function using a hyperparameter termed ‘atrous rate’, without impacting the computational resource requirement. It differs from the standard/usual convolution in that the filter is up-sampled by inserting zeros between successive filter values along every spatial dimension. This type of layer is useful in apprehending long-range information. The DeepLabv3+ model has atrous convolutions of atrous rates 18, 12, and 6.

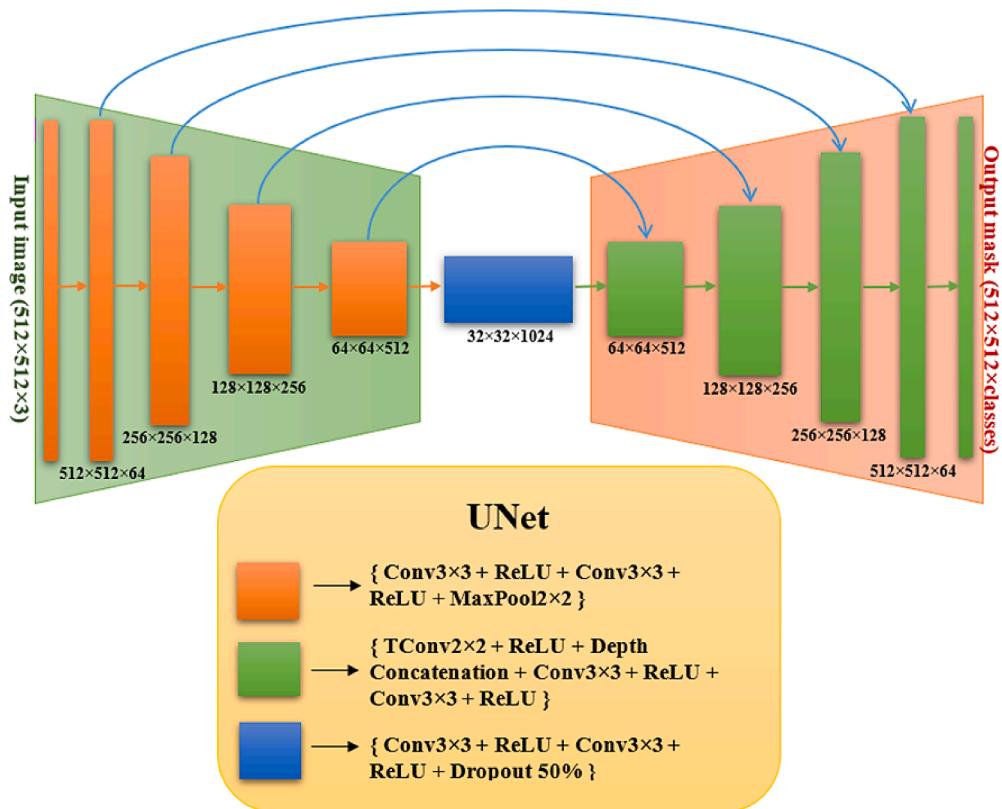
Thus, the architecture of DeepLabv3+ (Fig. 7) consists of four parts: (i) the backbone network for extracting features; (ii) Atrous convolutions in the last layers of this backbone network to control the size of the features; (iii) ASPP network for classification of each pixel; and finally, (iv) The prediction outcomes from the ASPP network is passed into  $1 \times 1$  convolution to get the final mask of the same size as that of image. In this study, the Xception [45] network was implemented as the backbone feature extraction network and used depth separable convolutions.

#### 2.3. Integrated model

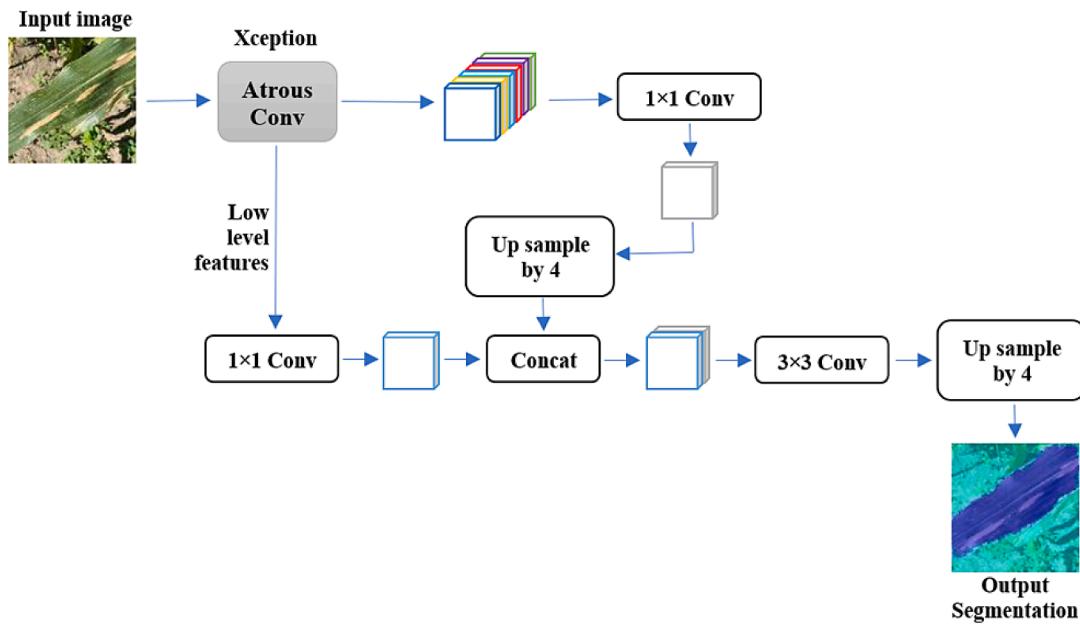
In this study, the validity of SegNet, UNet, and DeepLabv3+ networks was investigated for two separate goals: semantic segmentation of leaf from the complex background; semantic segmentation of lesions belonging to GLS, NLB, and NLS diseases from corn leaf images without field background. Semantic segmentation networks were integrated to segment corn leaves and disease spots in two stages. After analyzing the results, two networks – (i) the best of SegNet, UNet, and DeepLabv3+ for corn leaf segmentation task and (ii) the best among them for multiple disease lesions segmentation, were integrated to design a two-stage model for estimating the severity of the diseases (Fig. 4).

#### 2.4. Evaluation metrics

The evaluation metrics used in this study are introduced in this section. The performance of the leaf and lesion segmentation models was individually evaluated using the following four evaluation metrics: mean pixel accuracy ( $A_m$ ), mean weighted Intersection over Union (mwIoU), mean Boundary F1 Score (mBFScore), and Dice similarity



**Fig. 6.** Schematic diagram of the UNet model. The decoder comprises transposed convolution layers for up-sampling. The depth concatenation layer concatenates the features of the previous layer and the features at the end of the corresponding encoder.



**Fig. 7.** Schematic diagram of the DeepLabV3+ model. The Xception was adopted as the feature extraction network with Atrous convolutions.

coefficient (DSC).

#### 2.4.1. Mean pixel accuracy

In an image, the pixel accuracy for a particular class  $c$  is defined by Eq. (1), where TP, TN, FP, and FN are the number of true positive, true negative, false positive, and false negative pixels, respectively.

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

The mean accuracy ( $A_m$ ) is the average of all class accuracies over all the images. This quantity measures the overall performance for all the pixels in the image, describing the relative number of correctly classified

pixels. However, this poses a problem for evaluating lesion segmentation models since the number of pixels in the background class dominates the three disease classes. For  $N$  images and  $C$  classes, mean accuracy is given by Eq. (2).

$$A_m = \frac{1}{NC} \times \sum_{c=1}^C \sum_{n=1}^N A_{nc} \quad (2)$$

#### 2.4.2. Weighted mean intersection over union

Intersection over Union (IoU), or the Jaccard similarity coefficient or Jaccard index, measures the overlap between the ground-truth and predicted masks. The IoU can be determined via the ratio of the area of overlap and the area of the union of two masks (Fig. 8). Predicted masks that have heavily overlapped with the ground truth mask have more IoU value than ones with less overlap. In terms of TP, FP, and FN, IoU can be expressed as given in Eq. (3).

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

Intuitively, the mean IoU (mIoU) for an image is the average IoU value of the predicted mask with the ground truth mask for each class in the image. In this work, to reduce the biasing effect of IoU for dominating classes, the mIoU was weighted with the number of pixels in each class ( $p_c$ ) for all images in the dataset, called the mean weighted IoU (mwIoU), as in Eq. (4).

$$mwIoU = \frac{1}{NC} \times \sum_{c=1}^C p_c^{-1} \sum_{n=1}^N mIoU \quad (4)$$

#### 2.4.3. Mean boundary F1-score

The boundary F1 score (BFScore) measures the closeness of the predicted and ground truth segmentation boundaries. The BFScore is the harmonic mean of precision and recall values, defined with a distance error tolerance  $\epsilon$ , to determine if a pixel on the predicted mask's boundary matches the boundary of the ground truth mask. It is a contour-based evaluation metric, unlike IoU, which is a region-based metric. It is mathematically calculated as Eq. (5).

$$BFScore = \frac{(2 \times precision \times recall)}{(precision + recall)} \quad (5)$$

Here, precision is the ratio of the number of pixels in the boundary of the predicted mask that lie within a distance of  $\epsilon$  to the ground truth boundary and the total number of pixels in the predicted mask boundary. The recall is the ratio of the number of pixels in the boundary of the predicted mask that lies within the error tolerance distance to the ground truth boundary and the total number of pixels in the boundary of the ground truth mask. The mean BFScore (mBFScore) is the average BFScore of all the classes in all the images present in the dataset, expressed as Eq. (6).

$$mBEScore = \frac{1}{NC} \times \sum_{c=1}^C \sum_{n=1}^N BFScore \quad (6)$$

#### 2.4.4. Dice similarity coefficient

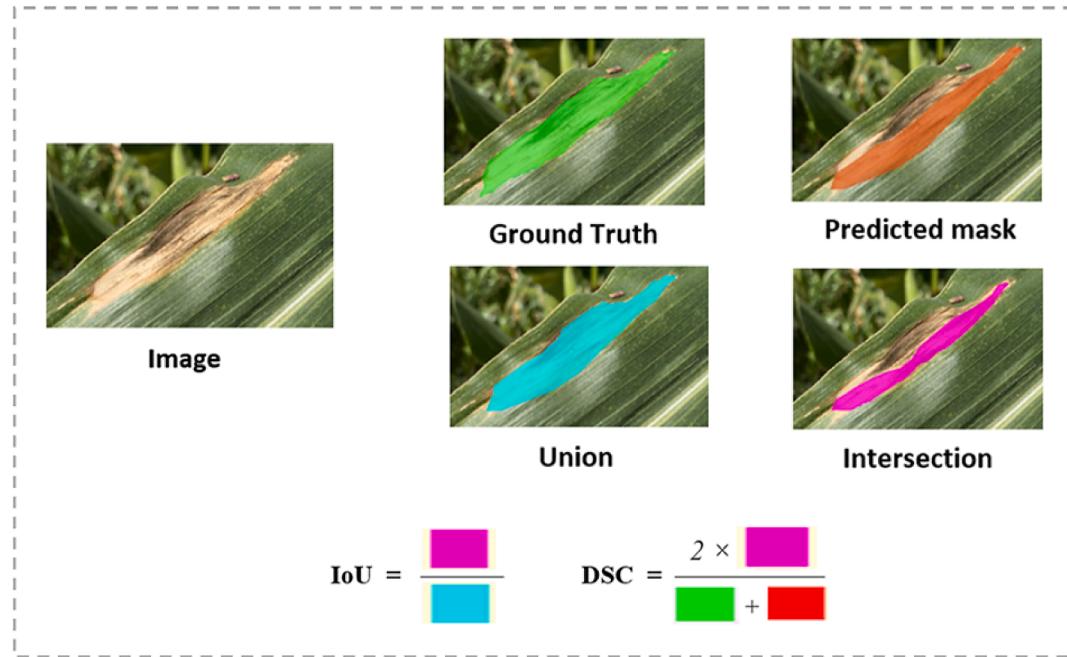
The Dice Similarity Coefficient (DSC), also called the Sørensen–Dice index measures similarity between two masks. The Dice index is defined in Eq. (7) in TP, FP, and FN. Unlike the Jaccard index, the TP values are counted twice while computing DSC. The DSC is related to the Jaccard Index (IoU) as defined by Eq. (8). Alternatively, it is the ratio of the intersection area to the sum of the ground truth area and predicted mask area (Fig. 8).

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (7)$$

$$DSC = \frac{2 \times IoU}{1 + IoU} \quad (8)$$

#### 2.5. Disease severity

The final result expected from the integrated model is the severity of the diseases present in the corn leaf of the field image. Since we performed segmentation in two stages, the severity of the lesion can be easily approximated. In addition to the evaluation metrics that were defined to analyze the model statistically, a metric to measure the



**Fig. 8.** Comparison between IoU and DSC as evaluation metrics. Union and intersection represent the union area and the area of intersection of ground truth and the

infection severity was also defined, as in Eq. (9).

$$\text{Severity} = \left( \frac{A_d}{A_l} \right) \times 100 \quad (9)$$

Here,  $A_d$  is the area of lesion belonging to a particular disease in the corn leaf of an image; and  $A_l$  is the total visible corn leaf area. It is to be noted that  $A_l$  and  $A_d$  are derived from the leaf segmentation and lesion segmentation networks, respectively (Fig. 9). To test the feasibility of the integrated approach, the estimated disease severity in images from the test dataset were statistically compared with the ground truth severity values using  $R^2$  values.

## 2.6. Hyperparameter options and training

The hyperparameters set during the training processes are provided in Table 1. Previous studies stressed the importance of assigning class weight ratio during semantic segmentation of plant disease lesions from the leaf [25]. Since the number of pixels belonging to the three disease classes was much less than the background, using a standard loss function would make the network perform well only on the dominant background class. This potential problem, called the ‘accuracy paradox’, would give a good overall performance but poor results over the disease classes. Thus, an optimal class weight ratio based on the median frequency of the classes was set as an additional hyperparameter for the lesion segmentation task.

## 3. Results and discussion

### 3.1. Performance of deep learning models for corn leaf segmentation

The first stage of the novel disease severity estimation approach was to accurately segment the corn leaf using deep learning. The results of the three models on the 100 test images separated for evaluating the leaf segmentation task are presented in Table 2 in terms of  $A_m$ , mwIoU, mBFScore, and DSC metrics. Among the evaluated models for this task, UNet performed the best compared with all the performance metrics, resulting in an  $A_m$  of 97.31% and a DSC of 0.7758. The performance of UNet was marginally better than SegNet and DeepLabv3+. The worst performance was observed for DeepLabV3+ with  $A_m$  of 89.33% and DSC of 0.6884. It can be inferred from the high mBFScore value of UNet ( $\sim 0.806$ ) that the boundaries of its prediction masks and the ground truths are very close to each other [30]. High values of mBFScore and DSC reflect that network’s potential to learn better representations. It was observed that segmentation of leaves affected with NLB from the backgrounds was challenging for all the models. The disease’s symptoms shared quite similar visual characteristics with the field conditions.

**Table 1**

Hyperparameters set during the training for deep learning models.

Hyperparameter	Value
Optimizer	sgdm
Initial Learn Rate	0.01
Maximum epochs	200
Early Stopping	yes
Mini batch size	32
Learn rate drop factor	0.05
Learn rate drop period	10
Momentum	0.9

**Table 2**

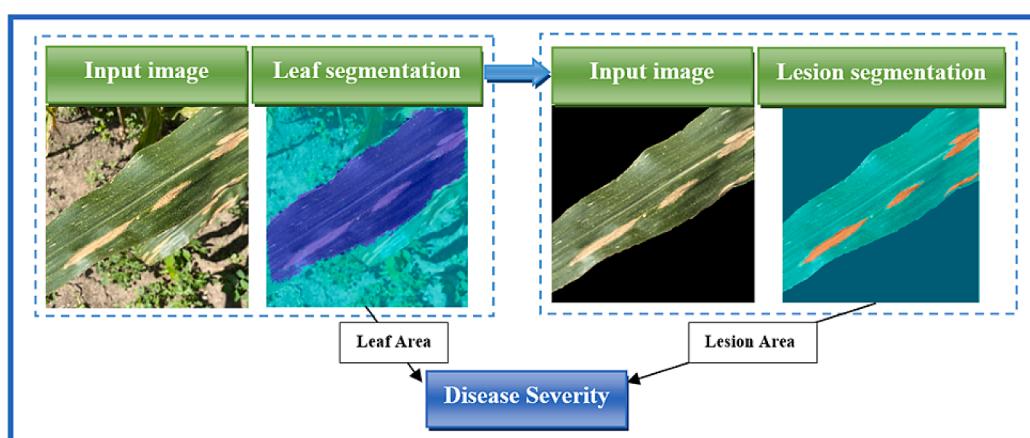
Results obtained for semantic segmentation of corn leaf with the implementation of SegNet, UNet, and DeepLabv3+ networks.

Network	$A_m$	mwIoU	mBFScore	DSC
SegNet	0.9683	0.9218	0.7942	0.7459
UNet	0.9731	<b>0.9422</b>	0.8063	<b>0.7758</b>
DeepLabv3+	0.8933	0.8731	0.7151	0.6884

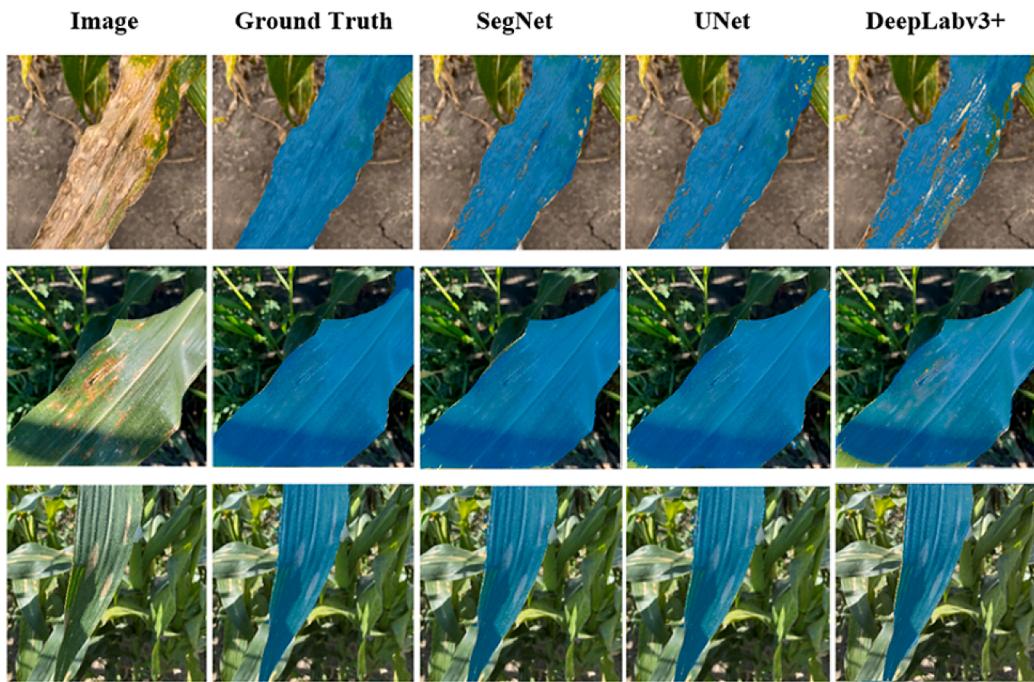
Since GLS and NLS’s disease spots were small compared to leaf tissues, the networks could separate the complete leaf portion without comprehending the spots. This desired trait was more prevailing in the predictions of the UNet model. The orientation of the maize images was not a concern for all three models as they produced consistent results.

The segmentation predictions of SegNet, UNet, and DeepLabv3+ architectures on a few images in the test set are presented in Fig. 10. The DeepLabv3+ network was not found to be generalizing over a range of leaf pixels, creating misclassified patches in the leaf segmentation masks. This was primarily at leaf portions that contained shadows and soil-like symptoms. Moreover, the segmentation masks from the three networks were not sharp at the leaf edges and image corners, which could be the reason for the reduction in mBFScore. Notably, in the predictions of SegNet, some pixel misclassifications were observed in images with maize leaf sharing its boundary/edge with other plant species in the background. This was more evident when the veins of these leaves in the image were running parallel, which is a usual occurrence in cornfields. The model suffered to recognize and distinguish because of similarity in multiple leaves’ colors and textural appearances. However, despite these meager drawbacks, the ultimate motive of this first stage segmentation to remove background objects of uninterest was well executed by all the models.

The implementations of UNet and SegNet presented good results compared to results obtained in the literature with the same network; for



**Fig. 9.** Illustration of disease severity estimation model using the two-stage semantic-segmentation approach.



**Fig. 10.** Examples of images, ground truth masks, and the results obtained from the implementations of SegNet, UNet, and DeepLabv3+ models for semantic segmentation of corn leaf.

example, an average IoU of 72.11% and pixel accuracy of 96.08% was achieved with the UNet network for powdery mildew lesion segmentation in cucumber leaves [24]. Few researchers were able to attain a better performance for a similar task. For instance, a modified UNet-based network – KijaniNet, was proposed for tomato leaf segmentation from complex backgrounds [30]. It resulted in a mean IoU of 0.9766 and mBFScore of 0.9439. However, the study was not improved further to identify and locate disease lesions. The dataset compiled for the present study contained diverse objects, including soils of different colors and textures, stones, stems, leaves of other corn crops and plant species, and varied environmental conditions. The networks implemented in the present study proved that corn leaves (with or without biotic stress) could be effectively segmented from field-acquired images with complex backgrounds. Besides, the results show that the UNet model has the edge over SegNet and DeepLabv3+ for leaf semantic segmentation and background elimination.

### 3.2. Performance of deep learning models for lesion segmentation

The same three networks, SegNet, UNet, and DeepLabV3+, were tested for lesion segmentation from the background eliminated leaf images. In addition to the previous analysis, the models investigated in this section function on the loss function with the introduction of class weighted ratio. The overall segmentation results of the networks on the test images separated from the dataset discussed in Section 2.1.2 are presented in Table 3. It is evident from the table that the performance of DeepLabv3+ is significantly better than SegNet and UNet for the task of lesion segmentation. The DeepLabv3+ network architecture achieved a

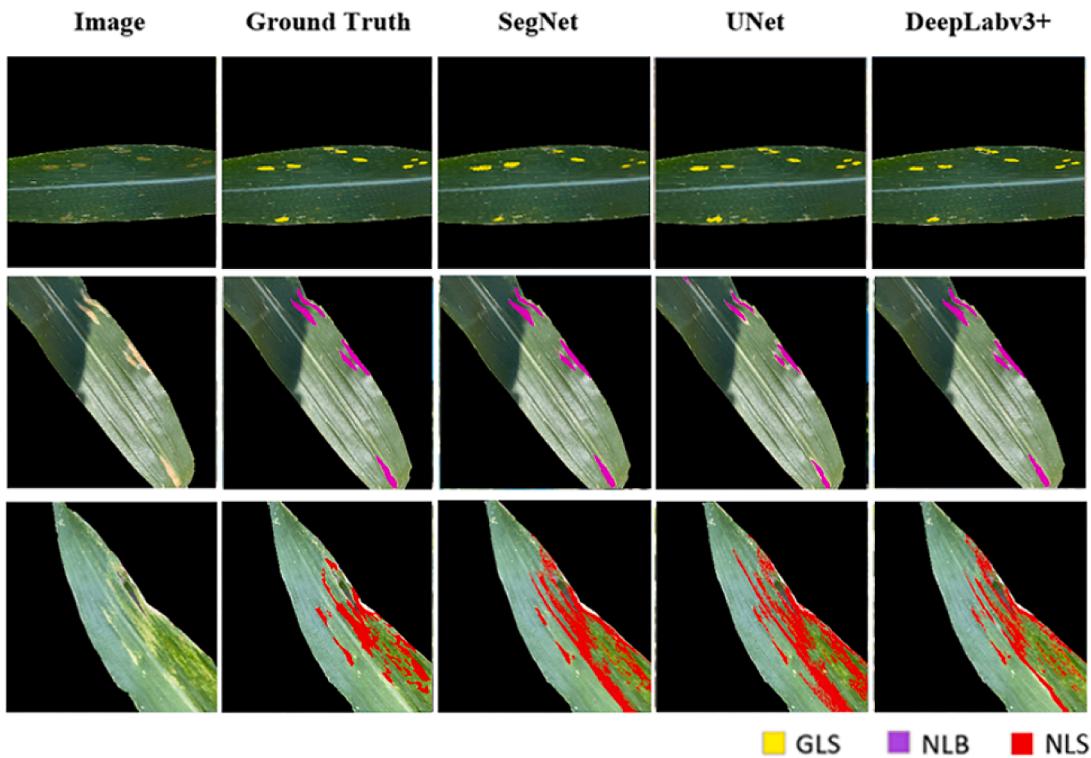
mwIoU value of 0.7379, whereas the mwIoU by SegNet and UNet were only 0.6919 and 0.6710, respectively. The higher values of mwIoU and mBFScore indicate that the predictions of DeepLabv3+ overlap with the ground truth very well and share close mask boundaries. The result of semantic segmentation of images affected by GLS, NLB, and NLS is presented in Fig. 11. Since the background dominated the lesion classes, the mean accuracy ( $A_m$ ) cannot provide a legitimate comparison of the performances on lesion segmentation [24]. Class-wise analysis was performed to check the potential of the networks on different diseases and symptoms. The class-wise pixel classification accuracy ( $A_{m, \text{class}}$ ), IoU, and mean boundary F1 score (mBFScore<sub>class</sub>) are shown in Table 4.

The analysis shows that among the three networks, DeepLabv3+ resulted in a good performance for segmenting lesions affected by NLS, with a class accuracy and IoU of 62.75% and 0.6357, respectively. The SegNet and UNet model performed very poorly in the NLS class. NLS symptoms include small, scattered, yellowish-brown lesions with irregular shapes running along the leaf veins. As reported in a few studies in the literature [46–48], the present study also highlights that the DeepLabv3+ network is more effective for segmenting small objects in the image. In a few images, the primary vein in corn leaves was incorrectly predicted as NLS disease spots due to resemblance with its symptoms. NLB experienced the most effective segmentation among the disease classes, recording class accuracy greater than 92% by all the models. The symptoms of NLB exhibit unique and more orderly shape, color, and textural features. Hence, the logic behind higher class mBFScore represents the sharpness of the estimated boundary and the proximity of actual and predicted boundaries. Furthermore, the GLS-infected lesions were also substantially classified, with class accuracies greater than 89% and IoU of more than 0.50. Some spots of GLS were incorrectly predicted as NLB when the lesion lacked the usual yellow coloring. Thus, good results were obtained for multi-class semantic segmentation of disease lesions, notably by adopting the DeepLabv3+ model.

**Table 3**

Performance analysis for semantic segmentation of lesion with the implementation of SegNet, UNet, and DeepLabv3+ networks.

Network	$A_m$	mwIoU	mBFScore	DSC
SegNet	0.9157	0.6919	0.4778	0.5183
UNet	0.9189	0.6710	0.4582	0.5058
DeepLabv3+	0.9237	<b>0.7379</b>	<b>0.5351</b>	<b>0.5731</b>



**Fig. 11.** Examples of images, ground truth masks, and the results obtained from the implementations of SegNet, UNet, and DeepLabv3+ models for semantic segmentation of GLS, NLB, and NLS lesions.

**Table 4**

Class-wise performance analysis of SegNet, UNet, and DeepLabv3+ for the task of lesion segmentation.

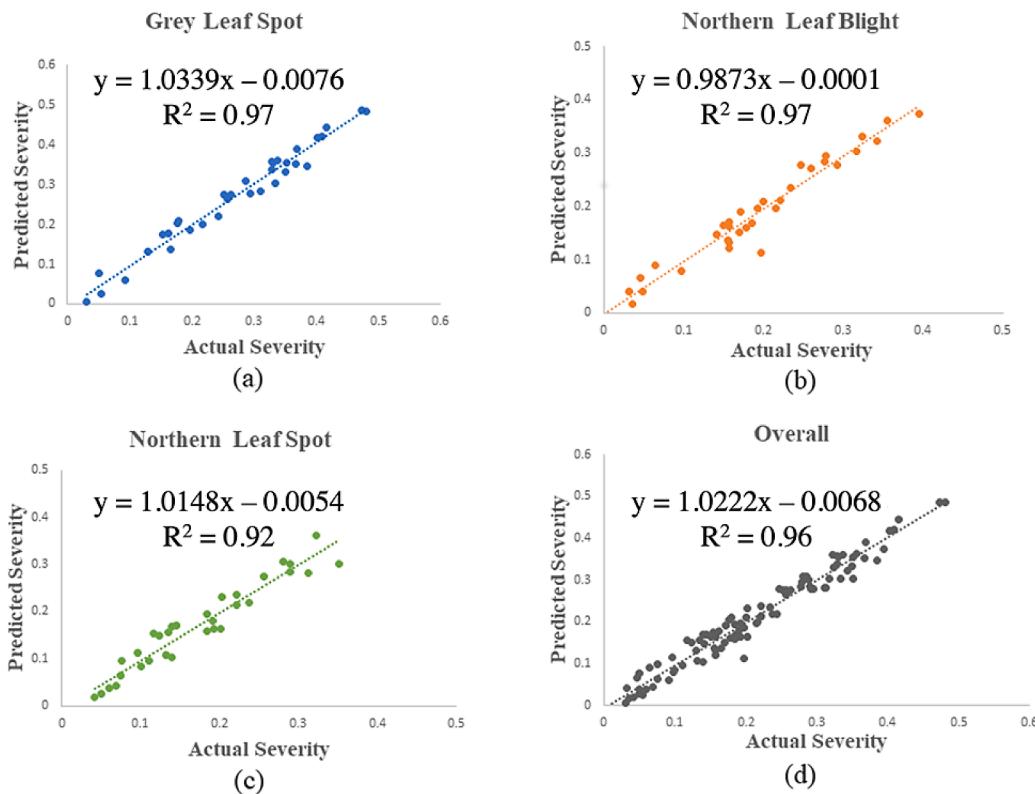
Network	Class	$A_{m, \text{class}}$	IoU	mBFScore <sub>class</sub>
SegNet	gls	0.9034	0.5191	0.4377
	nlb	0.9262	0.6716	0.5634
	nls	0.3994	0.4485	0.4083
	background	0.9679	0.8403	0.5505
UNet	gls	0.8937	0.5028	0.4272
	nlb	0.9429	0.6819	0.5576
	nls	0.4018	0.4347	0.3926
	background	0.9652	0.8394	0.5492
DeepLabv3+	gls	0.9287	0.5379	0.4498
	nlb	0.9468	0.7058	0.5804
	nls	<b>0.6275</b>	<b>0.6357</b>	<b>0.4946</b>
	background	0.9708	0.8490	0.5561

### 3.3. Performance of integrated model for severity estimation

The above experiments and results found that UNet and DeepLabv3+ were the best performing leaf segmentation and lesion segmentation models. Therefore, these two models were integrated to build a two-stage semantic segmentation model. The UNet aims to extract the corn leaf, thus eliminating the background, and the DeepLabV3+ performs semantic segmentation of disease lesions. To validate its feasibility, the model was tested with 100 test images. The severity of diseases in each image was computed using the ground truth mask of leaf annotation and lesion annotations. To estimate the severity, each disease's leaf area and lesion area were obtained from the UNet and DeepLabV3+ networks, respectively. The regression plots between the predicted disease severity and the actual severity, along with the determination coefficients ( $R^2$  value), are presented in Fig. 12. The plots in Fig. 12(a–c) and (d) represent the fitted relationship for each disease independently and altogether. High  $R^2$  values ( $> 0.90$ ) were observed for the three classes,

with an  $R^2$  value of 0.96 on the complete test set. The integrated model demonstrated  $R^2$  values around 0.97 for predicting the severity of GLS and NLB, and 0.92 for NLS.

The results of the integrated model proposed in this study exhibit strong performance compared to some highlighted studies in the literature (Table 5). A CNN-based semantic segmentation used within a previous study [24] for quantifying cucumber powdery mildew achieved 96%-pixel accuracy and IoU of 0.72. Another study obtained an IoU score of 0.386 for lesion class and an  $R^2$  value of 0.655 for quantifying late blight lesions in potato genotypes [25]. It is important to note that the deep-learning models proposed in the reported studies were applied to recognize only one disease. Although some studies estimated the severity of multiple diseases with an overall accuracy of 91% [49] and 97% [26], they were designed to operate on a homogeneous background. As an up-gradation, researchers accounted for the potential problem due to the complex background by extracting the leaf instances from the image [27]. However, the network had three stages for a severity classification approach. Thus, despite being computationally heavier for deployment, it was not very accurate for severity estimation compared to the semantic segmentation approach. Recently, another study proposed a severity estimation model based on the fusion of DeepLabv3+ for cucumber leaf segmentation and UNet for disease lesion segmentation [50]. This model demonstrated an accuracy of 93.27% for leaf segmentation and a DSC value of 0.69 for lesion segmentation. Despite a superior performance, no previous studies have evaluated the potential of such contemporary models for plant disease identification through a multi-stage segmentation technique. Overall, the DeepLabv3+ resulted in superior segmentation performance for small disease lesions present within images because of the presence of atrous convolution layers. The atrous convolutions can carry the features of small objects in the image over deeper layers, thus contributing to a superior performance.



**Fig. 12.** Linear relationship between true severity and the severity predicted by the UNet-DeepLabv3+ integrated model: (a) GLS; (b) NLB; (c) NLS; and (d) overall.

**Table 5**

Comparison of results of deep-learning-based segmentation models for plant disease severity estimation.

Targeted crop	Disease(s)	Best performing deep-learning architecture(s)	Results	Refs.
Cucumber	Powdery mildew	UNet	mIoU of 0.721 and pixel accuracy of 96.08%	[24]
Potato	Late blight	SegNet	IoU of 0.996 for background and 0.386 for lesion class.	[25]
Coffee	Leaf miner, rust, brown leaf spot, cercospora leaf spot	UNet	mIoU of 0.949 and an accuracy of 99.53%	[26]
Coffee	Leaf miner, leaf rust, brown leaf spot, cercospora leaf spot	UNet	mIoU of 0.943	[27]
Cucumber	Downy mildew, powdery mildew, cucumber virus disease	Integration of DeepLabV3+ and UNet	Pixel accuracy of 93.27% and disease classification accuracy of 92.85%.	[50]
Corn	GLS, NLB, NLS	Integration of UNet and DeepLabV3+	R <sup>2</sup> -value of 0.96 for disease severity estimation	This study

#### 4. Conclusion

In this study, the state-of-the-art 2D image segmentation networks, SegNet, UNet, and DeepLabV3+, were evaluated for the semantic segmentation of corn leaf from field images and GLS, NLB, and NLS disease lesions from the corn leaf. Based on the performance criteria used in the study, UNet and DeepLabV3+ were identified as the best network architectures for segmenting corn leaf and disease lesions, respectively. Excellent results were obtained: mwIoU of 0.9422 by UNet for segmenting leaf and mwIoU of 0.7379 by DeepLabV3+ for lesion segmentation. These two best-performing network architectures were integrated to build a novel two-stage semantic segmentation model. The resulting model predicted the severity of three diseases, namely GLS, NLB, and NLS, with an overall R<sup>2</sup> value of 0.96, with R<sup>2</sup> values ranging between 0.92 to 0.97 on individual disease classes. The proposed approach showed improved results for disease severity estimation when compared with previous studies. Overall, this study's proposed two-stage deep learning-based approach showed promising results for corn disease severity estimation for developing an effective field-ready disease management system.

#### CRediT authorship contribution statement

**L.G. Divyanth:** Conceptualization, Methodology, Data curation, Software, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Aanis Ahmad:** Conceptualization, Methodology, Data curation, Software, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Dharmendra**

**Saraswat:** Methodology, Writing – review & editing, Supervision.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This research was funded by Wabash Heartland Innovation Network (WHIN) grant number 18024589 and USDA-NIFA Hatch Fund number 1012501.

## References

- [1] P. Ranum, J.P. Peña-Rosas, M.N. Garcia-Casal, Global maize production, utilization, and consumption, *Ann. N. Y. Acad. Sci.* 1312 (2014) 105–112.
- [2] Food Outlook – Biannual Report on Global Food Markets, Food Outlook – Biannual Report on Global Food Markets (2022), <https://doi.org/10.4060/CB9427EN>.
- [3] L.E. Ehler, Integrated pest management (IPM): definition, historical development and implementation, and the other IPM, *Pest Manag. Sci.* 62 (2006) 787–789.
- [4] L.C. Ngugi, M. Abelwahab, M. Abo-Zahhad, Recent advances in image processing techniques for automated leaf pest and disease recognition—a review, *Inf. Process. Agric.* 8 (2021) 27–51.
- [5] Y. Nanehkaran, D. Zhang, J. Chen, Y. Tian, N. Al-Nabhan, Recognition of plant leaf diseases based on computer vision, *J. Ambient. Intell. Humaniz. Comput.* (2020) 1–18.
- [6] M.M. Ali, N.A. Bachik, N. Muhamdi, T.N.T. Yusof, C. Gomes, Non-destructive techniques of detecting plant diseases: a review, *Physiol. Mol. Plant Pathol.* 108 (2019), 101426.
- [7] V. Singh, A.K. Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques, *Inf. Process. Agric.* 4 (2017) 41–49.
- [8] S. Ramesh, R. Hebbar, M. Niveditha, R. Pooja, N. Shashank, P. Vinod, et al., Plant disease detection using machine learning, in: Proceedings of the international conference on design innovations for 3Cs compute communicate control (ICDI3C), 2018, pp. 41–45. Bangalore, India.
- [9] A. Abade, P.A. Ferreira, F. de Barros Vidal, Plant diseases recognition on images using convolutional neural networks: a systematic review, *Comput. Electron. Agric.* 185 (2021), 106125.
- [10] A. Ahmad, D. Saraswat, A. El Gamal, A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools, *Smart Agric. Technol.* (2022), 100083.
- [11] L. Li, S. Zhang, B. Wang, Plant disease detection and classification by deep learning – a review, *IEEE Access* 9 (2021) 56683–56698.
- [12] J. Chen, J. Chen, D. Zhang, Y. Sun, Y.A. Nanehkaran, Using deep transfer learning for image-based plant disease identification, *Comput. Electron. Agric.* 173 (2020), 105393.
- [13] J. Chen, D. Zhang, Y.A. Nanehkaran, D. Li, Detection of rice plant diseases based on deep transfer learning, *J. Sci. Food Agric.* 100 (2020) 3246–3256.
- [14] M. Azadbakht, D. Ashourloo, H. Aghighi, S. Radiom, A. Alimohammadi, Wheat leaf rust detection at canopy scale under different LAI levels using machine learning techniques, *Comput. Electron. Agric.* 156 (2019) 119–128.
- [15] A. Johannes, A. Picon, A. Alvarez-Gila, J. Echazarra, S. Rodriguez-Vaamonde, A. D. Navajas, A. Ortiz-Barredo, Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case, *Comput. Electron. Agric.* 138 (2017) 200–209.
- [16] X. Zhang, Y. Qiao, F. Meng, C. Fan, M. Zhang, Identification of maize leaf diseases using improved deep convolutional neural networks, *IEEE Access* 6 (2018) 30370–30377.
- [17] A. Fuentes, S. Yoon, S.C. Kim, D.S. Park, A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition, *Sensors* 17 (2017) 2022.
- [18] P. Jiang, Y. Chen, B. Liu, D. He, C. Liang, Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks, *IEEE Access* 7 (2019) 59069–59080.
- [19] Y. Zhong, M. Zhao, Research on deep learning in apple leaf disease recognition, *Comput. Electron. Agric.* 168 (2020), 105146.
- [20] M. Ji, L. Zhang, Q. Wu, Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks, *Inf. Process. Agric.* 7 (2020) 418–426.
- [21] B.C. Karmokar, M.S. Ullah, M.K. Siddiquee, K.M.R. Alam, Tea leaf diseases recognition using neural network ensemble, *Int. J. Comput. Appl.* 114 (2015).
- [22] J. Boulet, S. Fouquer, J. Théau, P.L. St-Charles, Convolutional neural networks for the automatic identification of plant diseases, *Front. Plant Sci.* 10 (2019) 941.
- [23] Y. Guo, Y. Liu, T. Georgiou, M.S. Lew, A review of semantic segmentation using deep neural networks, *Int. J. Multimed. Inf. Retr.* 7 (2018) 87–93.
- [24] K. Lin, L. Gong, Y. Huang, C. Liu, J. Pan, Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network, *Front. Plant Sci.* 10 (2019) 155.
- [25] J. Gao, J.C. Westergaard, E.H.R. Sundmark, M. Bagge, E. Liljeroth, E. Alexandersson, Automatic late blight lesion recognition and severity quantification based on field imagery of diverse potato genotypes by deep learning, *Knowl. Based Syst.* 214 (2021), 106723.
- [26] J.G. Esgario, P.B. de Castro, L.M. Tassis, R.A. Krohling, An app to assist farmers in the identification of diseases and pests of coffee leaves using deep learning, *Inf. Process. Agric.* 9 (2022) 38–47.
- [27] L.M. Tassis, J.E.T. de Souza, R.A. Krohling, A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images, *Comput. Electron. Agric.* 186 (2021), 106191.
- [28] Q. Zeng, X. Ma, B. Cheng, E. Zhou, W. Pang, GANs-based data augmentation for citrus disease severity detection using deep learning, *IEEE Access* 8 (2020) 172882–172891.
- [29] G. Wang, Y. Sun, J. Wang, Automatic image-based plant disease severity estimation using deep learning, *Comput. Intell. Neurosci.* 2017 (2017), 2917536.
- [30] L.C. Ngugi, M. Abdelwahab, M. Abo-Zahhad, Tomato leaf segmentation algorithms for mobile phone applications using deep learning, *Comput. Electron. Agric.* 178 (2020), 105788.
- [31] A. Abdalla, H. Cen, L. Wan, R. Rashid, H. Weng, W. Zhou, Y. He, Fine-tuning convolutional neural network with transfer learning for semantic segmentation of ground-level oilseed rape images in a field with high weed pressure, *Comput. Electron. Agric.* 167 (2019), 105091.
- [32] J.G.A. Barbedo, Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification, *Comput. Electron. Agric.* 153 (2018) 46–53.
- [33] A. Ahmad, D. Saraswat, A.E. Gamal, G. Johal, C.D.&S. dataset: handheld imagery dataset acquired under field conditions for corn disease identification and severity estimation, *arXiv preprint arXiv:2110.12084* (2021).
- [34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, MA, USA, 2015, pp. 1–9.
- [35] L.G. Divyanth, V. Chelladurai, M. Loganathan, D.S. Jayas, P. Soni, Identification of green gram (*Vigna radiata*) grains infested by *Callosobruchus maculatus* through X-ray imaging and GAN-based image augmentation, *J. Biosyst. Eng.* (2022) 1–16.
- [36] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural. Inf. Process. syst.* (2012) 25.
- [37] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).
- [38] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, NV, USA, 2016, pp. 770–778.
- [39] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (2017) 2481–2495.
- [40] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
- [41] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018, pp. 801–818.
- [42] I. Ulku, E. Akagündüz, A survey on deep learning-based architectures for semantic segmentation on 2D images, *Appl. Artif. Intell.* (2022) 1–45.
- [43] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer vision and Pattern Recognition, MA, USA, 2015, pp. 3431–3440.
- [44] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS, *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (2017) 834–848.
- [45] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.
- [46] B. Baheti, S. Innani, S. Gajre, S. Talbar, Semantic scene segmentation in unstructured environment with modified DeeplabV3+, *Pattern Recognit. Lett.* 138 (2020) 223–229.
- [47] D. Zhang, Y. Ding, P. Chen, X. Zhang, Z. Pan, D. Liang, Automatic extraction of wheat lodging area based on transfer learning method and DeeplabV3+ network, *Comput. Electron. Agric.* 179 (2020), 105845.
- [48] J. Kang, L. Liu, F. Zhang, C. Shen, N. Wang, L. Shao, Semantic segmentation model of cotton roots in-situ image based on attention mechanism, *Comput. Electron. Agric.* 189 (2021), 106370.
- [49] Q. Liang, S. Xiang, Y. Hu, G. Coppola, D. Zhang, W. Sun, PD<sup>2</sup>SE-net: computer-assisted plant disease diagnosis and severity estimation network, *Comput. Electron. Agric.* 157 (2019) 518–529.
- [50] C. Wang, P. Du, H. Wu, J. Li, C. Zhao, H. Zhu, A cucumber leaf disease severity classification method based on the fusion of DeeplabV3+ and U-Net, *Comput. Electron. Agric.* 189 (2021), 106373.