Automated Southern Leaf Blight Severity Grading of Corn Leaves in RGB Field Imagery

Chanae Ottley[1,2], Michael Kudenov[1,2], Peter Balint-Kurti[3], Ralph Dean[3], Cranos Williams[1,2,4]

[1]NC Plant Sciences Initiative, North Carolina State University, Raleigh, North Carolina

[2]Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, North Carolina

[3]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, North Carolina

[4]Department of Plant and Microbial Biology[4], North Carolina State University, Raleigh, North Carolina

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.

Plant stress phenotyping research has progressively addressed approaches for stress quantification. Deep learning techniques provide a means to develop objective and automated methods for identifying abiotic and biotic stress experienced in an uncontrolled environment by plants comparable to the traditional visual assessment conducted by an expert rater. This work demonstrates a computational pipeline capable of estimating the disease severity caused by southern corn leaf blight in images of field-grown maize plants compared to traditional practices involving visual assessment. We collected visible light (RGB) field images of maize leaves from lines susceptible and resistant to southern corn leaf blight and field severity scores representing an expert's assessment of the overall infection of the maize lines. We implemented a multistage approach for maize leaf and disease detection among field images and the quantification of leaf tissue infected by southern corn leaf blight. The multistage approach addresses the quantification task by (i) training a neural network on manually annotated maize leaves present in the RGB images, (ii) training a neural network on manually annotated southern corn leaf blight lesions present in the RGB images, and (iii) utilizing the identified regions of the disease to determine the proportion of infection on the leaves. The first two stages of the approach involved training two semantic segmentation models, U-Net and DeepLabv3. The DeepLabv3 model was observed to have the best leaf and lesion segmentation performance. The models achieved an average intersection over union value of 0.879 and 0.849 for the leaf and lesion segmentation tasks, respectively. The final stage of the pipeline estimated SCLB severity by

computing the proportion of SCLB lesion area to maize leaf area using the predictions of the previous stages. The DeepLabv3-DeepLabv3 integrated system achieved a correlation coefficient of 0.88 and 0.72 between the field severity scores collected over two field trials and estimated severity scores representing the infection rating of an RGB image with an individual leaf sampled from the plots that were graded by the rater. These results indicate that images of a single leaf sample provide consistent grading as would be done in the field for a plot of plants. This research could provide a decision support tool that utilizes high-throughput data to understand the plants' stress responses and improve the efficiency of disease management.

Keywords.
Computer vision, deep learning, DeepLabv3, U-Net, disease quantification, Southern Corn Leaf Blight, RGB Imaging

Introduction

Efficient identification of indicators of abiotic and biotic stresses experienced by plants is a prominent issue in agriculture. Crop losses are estimated to be $60 billion a year due to plant diseases, demonstrating the impact of biotic stresses on food security (Reddy, Sudarshana, Fuchs, Rao, & Thottappilly, 2009). These losses negatively affect the ability to meet the demand for increased agricultural productivity. Typically, infected plants are identified and scored based on visual inspection, which is subjective and time-consuming (Zhou, et al., 2020). Moreover, waiting for visible symptoms may result in the crops being too infected for recovery or may allow the spread of the disease to neighboring plants. Information that could aid in identifying and quantifying symptoms of infection among plants could be helpful for crop breeding to identify superior varieties that are more resistant, understand crop loss, and evaluate targeted treatment strategies (Bock, Barbedo, Del Ponte, Bohnenkamp, & Mahlein, 2020; Bock, Chiang, & Del Ponte, 2022). Therefore, there is a need for an objective and consistent platform for quantifying disease severity in plants.

State-of-the-art methods for plant stress phenotyping have progressed to imaging and remote-sensing technologies combined with machine learning to capture observable characteristics resulting from its genotype interacting with the environment (Boggess, et al., 2013; Bock, et al., 2021). Biotic stresses, such as southern corn leaf blight (SCLB), that have distinct visual characteristics are suitable subjects for image-based phenotyping studies. Southern corn leaf blight is a fungal disease caused by the pathogen *Cochliobolus heterostrophus* (anamorph, *Bipolaris maydis*) that produces necrotic lesions along the veins of the leaves (Bruns, 2017; Sermons & Balint-Kurti, 2018). While the disease favors a humid environment, its global presence contributes to substantial yield loss in corn-growing regions (Bruns, 2017; Kumar, et al., 2022).

Identification and detection research in stress phenotyping involves data analysis to identify a particular stress when present and then to classify images based on stress patterns. Identification and detection research can be extended to quantification research by leveraging methods to utilize quantitative or qualitative traits (Singh, Ganapathysubramanian, Singh, & & Sarkar, 2016). Stress quantification encompasses the estimation of disease severity by detecting and quantifying one or more diseases in plants based on the analysis of the plant's health. Disease severity can be defined as the measurement of plant disease expression based on disease symptoms experienced by a plant or plot (Bock, et al., 2021). Customarily, visual estimates of disease severity diagnosed by plant pathologists are represented on a nominal or ordinal scale (Bock, et al., 2021; Stevens, 1946; Bock, Chiang, & Del Ponte, 2022). However, attaining inter and intra-rater reliability can be challenging due to lack of experience or training, scale type, symptoms, preference for specific severities, color blindness, the environment, and the rater's speed (Bock, et al., 2021). Recent progress in machine learning for computer vision problems demonstrates the potential for greater accuracy, efficiency, and repeatability in image-based phenotyping.

Prior work typically used thresholding to analyze visible light (RGB) images taken by a standard

or smartphone camera and estimated severity on a qualitative or quantitative ordinal scale. In simplified cases, studies have worked with data where a single leaf of interest is against a simple, contrasting background (Pethybridge & Nelson 2015, Manso et al. 2019, Sibiya & Sumbwanyambe 2019). For example, the work of Pethybridge and Nelson (2015) and Manso et al. (2019) resulted in a smartphone application that estimated the diseased area using thresholding as a form of image segmentation to separate objects/areas of interest based on pixel intensity. Their proposed approaches enabled user interaction, or a function-derived value based on the data, to detect diseased regions prior to quantification (Pethybridge & Nelson, 2015; Manso, Knidel, Krohling, & Ventura, 2019). However, the methods may find the severity estimation task difficult due to factors such as inconsistent lighting and varying noise levels on images captured in uncontrolled settings.

Currently, studies have implemented supervised, deep learning algorithms built on convolutional neural networks (CNN) that learn from annotations to extract features for the classification of objects within the image or the entire image (Garg, Bhugra, & Lall, 2021; Wang, et al., 2021; Divyanth, Ahmad, & Saraswat, 2023). Garg, Bhugra, & Lall (2021), Wang et al. (2021), and Divyanth et al. (2023) implemented a two-stage, deep learning-based model to reliably distinguish leaves and lesions within field images. Garg, Bhugra, & Lall (2021) developed their two-stage model using the Cascaded MRCNN architecture to segment all corn leaf instances and northern leaf blight lesions in field images captured with an uncrewed aerial vehicle. The fused model consisting of the DeepLabv3+ and U-Net networks proposed by Wang et al. (2021) estimated the severity of cucumber downy mildew in images taken with a smartphone. Both models presented by Garg, Bhugra, & Lall (2021) and Wang et al. (2021) quantified the severity of a single disease. In contrast, Divyanth et al. (2023) applied their U-Net-DeepLabv3+ integrated model for multiple corn disease quantification. Garg, Bhugra, & Lall (2021) and Divyanth et al. (2023) reported a correlation coefficient of 0.73 and 0.96, respectively, between the estimated and ground truth ratios. Wang et al.'s method of disease severity estimation was assessed after binning the severity proportions into six classes and resulting in an accuracy greater than 90%. The results of these studies support the notion that a multistage approach built on semantic segmentation models to remove nuisance factors and identify objects of interest can aid in quantifying diseases. It is worth investigating how the information in images to estimate disease severity compares to the information available to a breeder during field trials.

This study presents an image-based approach intended to (i) assess the performance of two state-of-the-art semantic segmentation models, U-Net and DeepLabv3, (ii) develop a deep learning-based integrated model for maize leaf segmentation and SCLB lesion segmentation, (iii) estimate the intensity of southern corn leaf blight symptoms in images of field-grown maize plants, and (iv) compare the estimated severity scores to traditional practices involving visual assessment. The main contributions of this study are:

- An assessment of deep learning segmentation models for maize leaves and southern corn leaf blight lesions and the development of a multistage segmentation pipeline that achieved consistent segmentation results.
- An objective method of disease severity grading for southern corn leaf blight by calculating the proportion of infected area to leaf area and mapping the proportion to a monotonically decreasing function. The mapped proportions enabled the comparison of the estimated values to field data based on visual assessment on the same scale.

Materials and Methodology

Data Acquisition

The dataset used in this study was collected from field trials at the Central Crops Research Station in North Carolina during June to August in the 2019 and 2023 growing seasons. The inoculation and planting procedure of near-isogenic lines (NILs) of the maize inbred line B73 inoculated with *Cochliobolus heterostrophus* (anamorph, *Bipolaris maydis*) can be found in a protocol by Sermons

and Balint-Kurti (2018). After anthesis, a plant pathologist visually assessed the plants and recorded the degree of infection on a one to nine scale. A plant with a score of nine displayed no visual symptoms of the disease, whereas a score of one implies a dead plant (Sermons & Balint-Kurti, 2018). Each field severity score represented the average degree of infection among all the plants within the plot.

RGB images of leaf samples were taken with a NIKON D5600 digital camera or Samsung Galaxy S8 camera on or around the same day the plants were visually assessed. The original image sizes were 4000 by 6000 pixels with a 300-dpi resolution and 4032 by 3024 pixels with a 72-dpi resolution for the digital and smartphone cameras, respectively. Infield images captured diverse conditions associated with the background, shadows, blurring and sharpness effects, target leaf orientation, presence of multiple plant stresses, and other artifacts, such as clothing, paper, zip ties, and human hands. Fig 1 displays sample images from the dataset of maize leaves afflicted by SCLB. The photographer held a single maize leaf, which is attached to the uppermost ear of the plant, that would be the subject of analysis. They photographed about three randomly selected leaves for each plot of maize line per collection day. Although nearly 3,000 images were collected between the two growing seasons, only 396 images were manually annotated and used to train and test the deep learning models, U-Net and DeepLabv3, for the leaf and lesion segmentation tasks.

Images for both segmentation tasks were manually annotated using the VGG Image Annotator (VIA) (Dutta & Zisserman, 2019). VIA provides multiple tools for annotating images, enabling users to outline objects of interest. Among the 396 annotated images, 350 were used in the models' training process for leaf segmentation, and the remaining 46 were used to train the models for lesion segmentation. Each segmentation subset was randomly divided such that approximately 70% of the images were used for training (i.e., feature learning), 20% for validation (i.e., evaluation of the models' performance while training), and 10% for testing (i.e., evaluation of the models' performance after training).

Field severity scores and RGB images were associated with a near-isogenic line and the date of data collection. Incorporation of the remaining images in the last evaluation step depended on the availability of a severity score. After filtration, 1,904 images were available to evaluate the multistage pipeline by comparing predicted severity scores to the field severity scores recorded by the plant pathologist during the field trials.

Fig 1. Maize plants exhibit stress symptoms on leaves caused by the *Bipolaris maydis* pathogen. Images collected during the field trials of 2019 and 2023 display a variety of lighting and background conditions, as well as degrees of infection.

Maize Leaf Annotations

For the leaf segmentation task, all annotations were generated by placing points around the border of the target leaf with VIA's polyline tool. The annotated images exhibited diverse conditions with respect to the target leaf's position and orientation, lighting, object focus, and

background. Ground truth masks were created as binary images, where white pixels represented the leaf of interest and black pixels represented the background containing the remaining objects captured within the image. Sample images and the corresponding ground truth masks are displayed in Fig 2. The leaf dataset was divided into 70/20/10 percent splits, resulting in 245 image-mask pairs for training, 70 image-mask pairs for validation, and 35 image-mask pairs for testing. Images in the train and validation sets were resized to 512 by 512 pixels before training the deep learning models. Also, geometric data augmentation methods of horizontal flips and random rotations were applied only to the train set.

(A) Field Images

(B) Ground Truth Masks

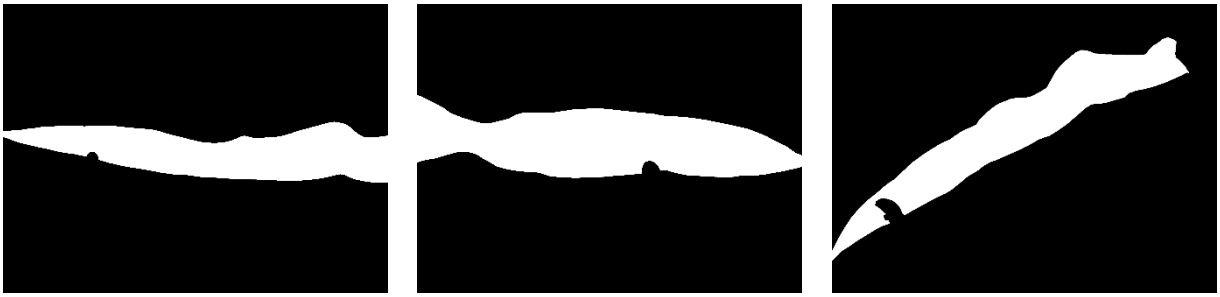
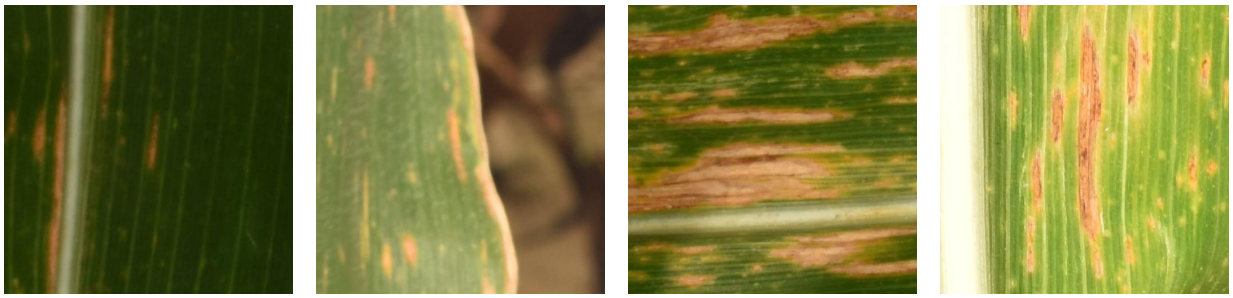


Fig 2. Example image and mask pairs utilized for training the stage 1 models to perform leaf segmentation. (A) Sample images from the dataset that were manually annotated to locate the object of interest, the maize leaf, within the image. (B) The generated masks of the maize leaf used as the ground truth to train the segmentation models: U-Net and DeepLabv3.

Southern Corn Leaf Blight Lesion Annotations

The plant pathologist manually annotated 46 images for lesion segmentation using VIA's ellipse tool. An ellipse was placed around the SCLB lesions present on the target leaf, resulting in 2,054 SCLB lesions. Unlike the previous task, the original images were not resized, but rather, segments around each lesion were cropped to avoid reducing the resolution and possibly losing information. Lesions smaller than 512 pixels in width and height were at the center of the generated cropped patches. Lesions that were longer or wider than 512 pixels were cropped into multiple patches every 256 pixels. This process generated 2,514 image-mask pairs for the lesion dataset (Fig 3). Ground truth masks displayed two classes of interest: 1) SCLB lesions (white pixels) and 2) the remaining leaf tissue not afflicted by SCLB and background (black pixels). It is worth noting that these masks were not comprehensive because not all SCLB lesions were annotated. The dataset was split into 74/19/07 percent splits, resulting in 1,863 image-mask pairs for training, 465 image-mask pairs for validation, and 186 image-mask pairs for testing. The same data augmentation methods performed on the leaf dataset were also implemented on this dataset.

(A) Field Images



(B) Ground Truth Masks

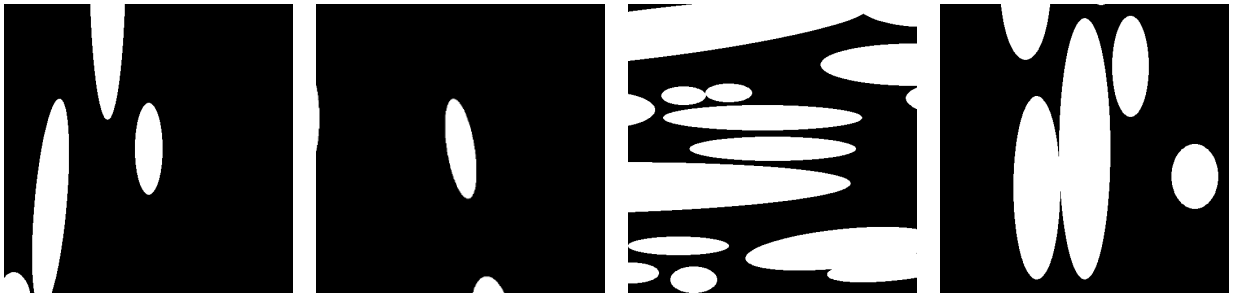


Fig 3. Example of images and their corresponding masks employed in the stage 2 training process to perform SCLB lesion segmentation. SCLB lesions present on the target leaf were manually annotated and the annotations were used to create the training dataset. (A) Cropped image patches containing at least one SCLB lesion. (B) The generated masks of SCLB lesions used as the ground truth to train the segmentation models: U-Net and DeepLabv3.

Proposed Multistage Pipeline

We designed a multistage pipeline for this study to estimate the SCLB severity on maize leaves in RGB images (Fig 3). The pipeline was formed by fusing the best-performing models based on their effectiveness to semantically segment maize leaves and SCLB lesions. First, field images were resized to 512 by 512 pixels before making predictions to separate the leaf from the background using the deep learning model selected for stage one. The output of stage one was scaled back to its original size, then partitioned into patches of 512 by 512 pixels by moving across the image 512 pixels at a time. For stage two, the deep learning model chosen to segment SCLB lesions was used to make predictions on the patches. After processing all the patches, they were restitched. The pipeline's ability to estimate SCLB severity using RGB images was evaluated based on the results of the 1,904 images that were not considered when making decisions about the networks' architecture for stages one and two. Stage three involves estimating the degree of affliction by computing the leaf and lesion areas after analyzing the results of the stages one and two models, respectively, and comparing the estimated scores to the field severity scores.

The deep learning models were trained using transfer learning to leverage the layers of the neural networks to learn features and knowledge gained from a previous task to improve generalization (Weiss, Khoshgoftaar, & Wang, 2016). The configuration for each neural network includes a stochastic gradient descent (sgd) optimizer, cosine decay rate with an initial rate of 0.1, batch size of 16 images, and trained for a total of 50 epochs. All training and validation processes were performed using Tensorflow’s deep learning framework on a CPU with an Intel Xeon Gold 6258R processor, taking approximately 1.5 to 2 hours to train the U-Net and DeepLabv3 models for leaf segmentation and 5 to 10 hours for the lesion segmentation. Prediction mask generation for an image took less than a second in stage one and 15 seconds to process the patches and restitch them to the original dimensions in stage two. The area calculations for all leaves and lesions were completed within 30 minutes.

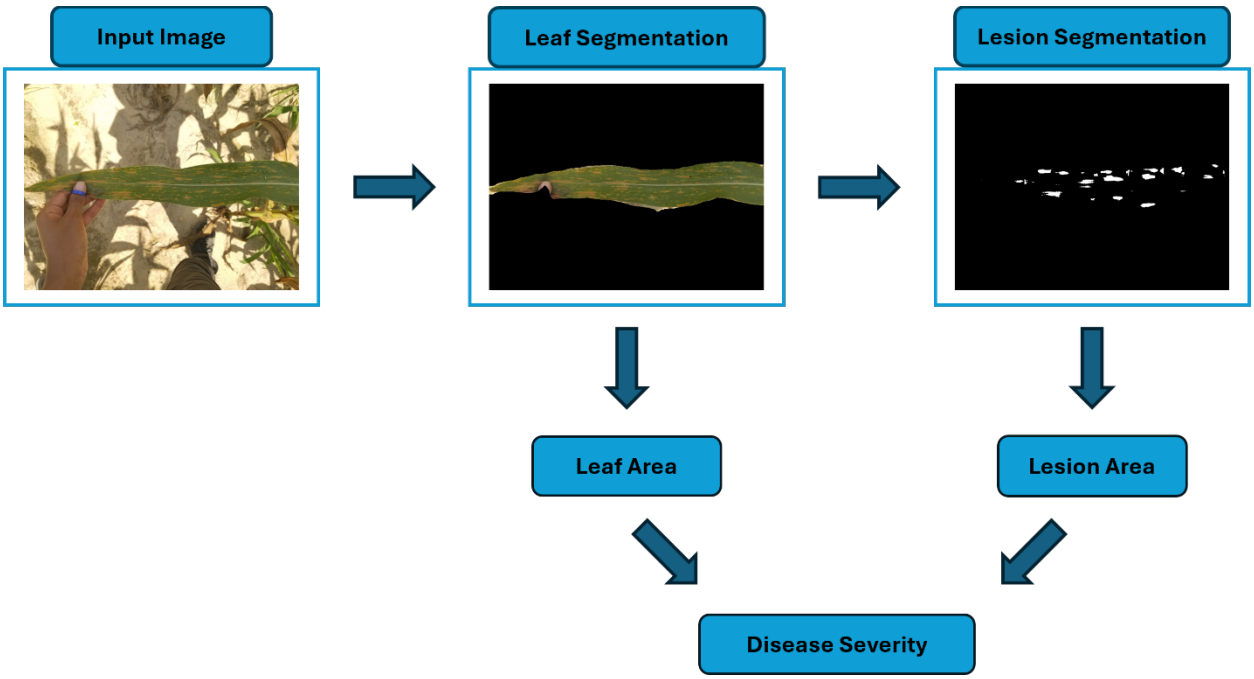


Fig 3. Graphical representation of the multi-stage segmentation approach to estimate SCLB disease severity.

Semantic Segmentation Models

In this study, two state-of-the-art networks, DeepLabv3 and U-Net, were trained on the RGB images collected during a maize field trial. We evaluated the networks' ability to perform object localization and semantic segmentation of a maize leaf in images captured under uncontrolled conditions and disease symptoms from the leaf-segmented image. The architecture of each network employed in this study was adjusted to use the MobileNet v2 model as the backbone for feature extraction. MobileNet v2 was designed to locate objects using an inverted residual layer that reduces the number of operations while maintaining accuracy (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018).

DeepLabv3

Chen et al. (2017) created the DeepLabv3 model that consists of two main structures: (1) Atrous Spatial Pyramid Pooling (ASPP) module and (2) Pyramid Scene Parsing Network (PSPNet) (Zhao, Shi, Qi, Wang, & Jia, 2017). The ASPP structure applies atrous (dilated) convolutions (Dumoulin & Visin, 2016) that adjusts the receptive field while extracting features at different strides, and the PSPNet performs multiple pooling operations (Chen, Papandreou, Schroff, & Adam, 2017). Atrous convolutions help prevent the diminishment of detailed, spatial information of feature maps after passing through the convolution and pooling layers. The authors claim incorporating the ASPP module created denser feature maps containing global context that improved performance (Chen, Papandreou, Schroff, & Adam, 2017). A performance comparison against state-of-the-art models in the PASCAL visual object classes (VOC) challenge establishes DeepLabv3 as a state-of-the-art model for semantic image segmentation.

U-Net

U-Net is a semantic segmentation model initially created by Ronneberger, Fischer and Brox (2015) to segment neuronal structures in biomedical images that eventually became a state-of-the-art model among other applications. It consists of a contracting path or encoder and an expanding path or decoder. The encoder contains blocks of convolution, activation, and pooling layers that transmit contextual information by increasing the outputs' resolution using upsampling operators (Ronneberger, Fischer, & Brox, 2015). The outputs of the encoder are fed through the decoder, which is symmetrical to the encoder and creates a U-shaped architecture to obtain

object localization through transposed convolutional (Dumoulin & Visin, 2016) layers. The authors state that a major contribution of U-Net is the accuracy of segmentation with fewer training images (Ronneberger, Fischer, & Brox, 2015). Unlike the DeepLabv3 model, U-Net recovers spatial resolution with transposed convolutions.

Table 1. Hyperparameters defined for training the deep learning models for both leaf and SCLB lesion segmentation tasks.

Hyperparameter	Value
Batch Size	16
Epochs	50
Initial Learning Rate	0.1
Momentum	0.9
Optimizer	sgd

Evaluation Metrics

A segmentation problem can be expressed in terms of pixel classification, where the results are evaluated based on true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) (Taha & Hanbury, 2015; Rahman & Wang, 2016). Subsequently, the Intersection over Union (IoU), also known as the Jaccard score, provides information on the models' ability to make pixel-wise classifications. IoU is a commonly used spatial metric in semantic segmentation that measures the similarity between samples by quantifying the ratio of overlap between the ground truth (Jaccard, 1912). Higher IoU values indicate a greater overlap between the ground truth and predicted masks. The results were compared and quantitatively assessed using the IoU score for each class (Equation 1) and the mean IoU (mIoU) score (Equation 2), which is calculated by computing the average IoU over all classes. If C denotes the number of classes, then the IoU and mIoU can be expressed as:

$$IoU_C = \frac{TP_C}{TP_C + FP_C + FN_C} \tag{1}$$

$$mIoU = \frac{1}{C}\sum_{C} IoU_C \tag{2}$$

Stages one and two outputs provided segmentation masks of the maize leaf and SCLB lesions, respectively. The disease severity was estimated by computing the area of leaf tissue and SCLB lesion tissue using the leaf and lesion segmentation results. The leaf's area was defined as the number of pixels in the predicted mask of the leaf segmentation, whereas the lesions' area was the number of pixels in the predicted mask of the lesion segmentation. Equation 3 computes the estimated disease severity for each image.

$$Disease\ Severity = \frac{Area_{SCLB}}{Area_{Leaf}} \tag{3}$$

Then, values of the proportion defined above were mapped using the generalized logistic function (Equation 4) and exponential function (Equation 5):

$$f(x) = \frac{L - A}{C + Qe^{-k(x-x_0)^{1/v}}} + A \tag{4}$$

$$g(x) = (L - A)e^{kx} + A \tag{5}$$

where L is the upper asymptote or the greatest desired y-value, k is the growth rate, A is the lower asymptote, and x_0 is the x-value of the function's midpoint (Birch 1999). The logistic and exponential functions were selected to map disease severity to a monotonically decreasing function (Fig 4). The predicted disease severity scores were mapped between the range of [1,9],

the same range of the scale used to rate disease severity among the field plots, using the following adjusted generalized logistic function (Equation 6) where L=9, A=1, x_0=0.5, k= -10, and the remaining coefficients set to 1 and the adjusted exponential function (Equation 7) where L=9, A=1, and k= -5.

$$f(x) = \frac{8}{1 + e^{10(x-0.5)}} + 1 \tag{6}$$

$$g(x) = 8e^{-5x} + 1 \tag{7}$$

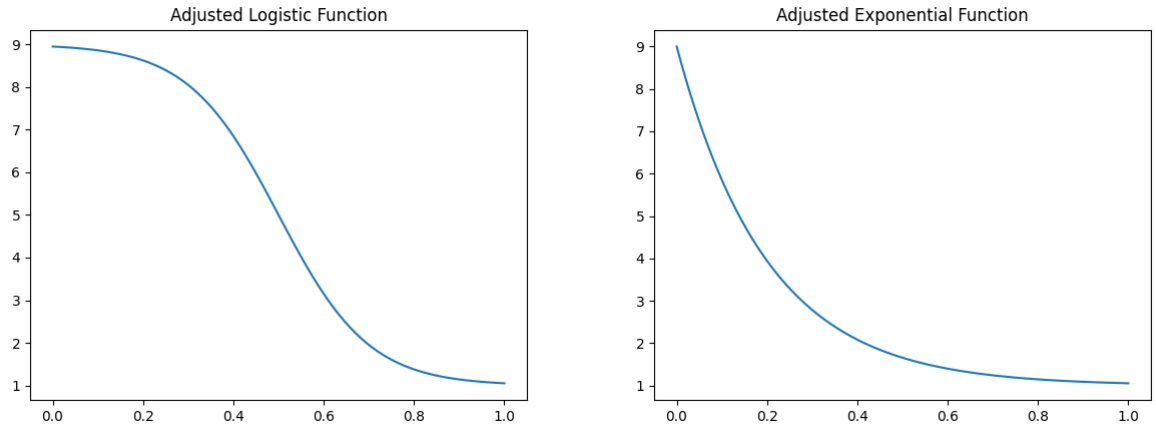


Fig 4. Adjusted generalized logistic and exponential functions used to map the estimated disease severity scores. The lower horizontal asymptote of the adjusted logistic and exponential functions occurs at 1, and the upper horizontal asymptote or maximum value within range occurs at 9.

Finally, the estimated disease severity was statistically compared to the field severity scores provided by a plant pathologist using a correlation analysis. The analysis was used to determine if images of a single leaf contained relevant information that could be used to provide accurate severity grading similar to infield visual assessment.

Results and Discussion

Leaf Segmentation

The objective of the first stage was to accurately segment the target maize leaf using deep learning models. The performance of each model was evaluated using a test set of 35 images, which was separated from the leaf dataset before training the models for segmentation. Table 2 presents the quantitative results for the U-Net and DeepLabv3 networks trained to distinguish the target maize leaf from the image's background. Based on the performance metrics (Table 2), the DeepLabv3 network was deemed the better model because it achieved a mean IoU of 0.879, a background class IoU of 0.966, and a leaf class IoU of 0.792. Meanwhile, the U-Net network obtained a mean IoU, background class IoU, and leaf class IoU of 0.707, 0.920, and 0.494, respectively. Since the background typically represents the majority of pixels within the images, biasing of the IoU towards the background class is expected. This biasing effect is more prominent in the U-Net network than in the DeepLabv3 network upon inspection of the individual class IoU values.

Table 2. Quantitative results of the U-Net and DeepLabv3 models for the maize leaf segmentation task to evaluate the performance based on the IoU evaluation metric.

Evaluation Metrics	U-Net	DeepLabv3
Background class IoU	0.920	**0.966**
Leaf class IoU	0.494	**0.792**
mIoU	0.707	**0.879**

Fig 5 provides example images from the test set and their corresponding predictions that allow the results of each network to be qualitatively analyzed. The networks' results tend to demonstrate orientation invariance. Misclassifications occurred among images that captured leaves sharing a boundary, large quantities of direct sunlight on the leaf, inconsistent lighting across the leaf, and holes in the leaves. Although the U-Net network could recognize maize leaves, it struggled with defining the target leaf's boundary, tip, and blade nearest to the sheath. Based on DeepLabv3's performance, we inferred the model learned leaf features that allowed for the repeatability of leaf segmentation despite the nuisances in the image.

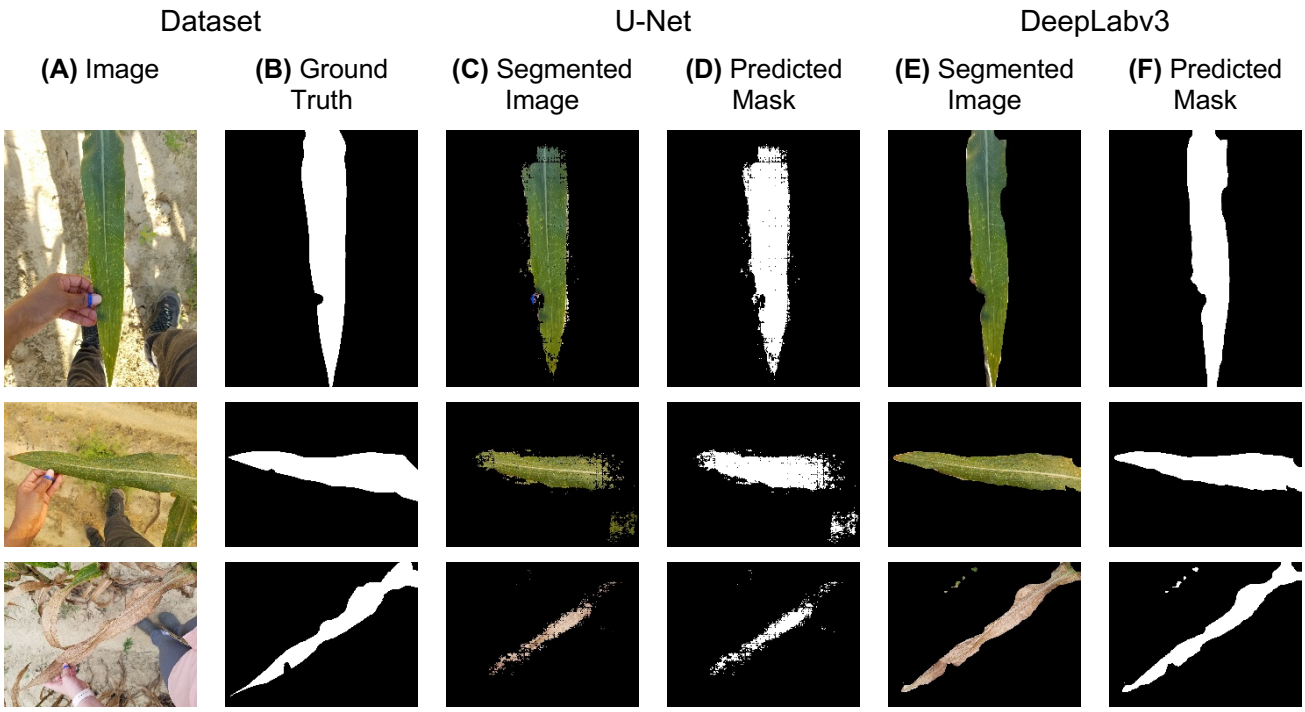


Fig 5. Qualitative results of the U-Net and DeepLabv3 models for the maize leaf segmentation task. (A) and (B) Sample images and ground truth mask pairs from the test set, respectively. (C) Segmented images created based on the (D) predicted masks for the U-Net model. (E) Segmented images created based on the (F) predicted masks for the DeepLabv3 model.

Lesion Segmentation

For the lesion segmentation task, the U-Net and DeepLabv3 networks were evaluated for their ability to identify SCLB from the leaf-segmented images produced in stage one. The evaluation metrics presented in Table 3 are the results of the test images that were separated from the dataset prior to training. Once again, the DeepLabv3 network outperformed the U-Net network for the lesion segmentation task. The DeepLabv3 network obtained a mean IoU, background class IoU, and SCLB lesion class IoU of 0.849, 0.919, and 0.779, respectively. The U-Net network achieved a mean IoU of 0.572, a background class IoU of 0.761, and a SCLB class IoU of 0.383.

Table 3. Quantitative results of the U-Net and DeepLabv3 models for the SCLB lesion segmentation task to evaluate the performance based on the IoU evaluation metric.

Evaluation Metrics	U-Net	DeepLabv3
Background class IoU	0.761	**0.919**
SCLB lesion class IoU	0.383	**0.779**
mIoU	0.572	**0.849**

Example images from the test set and their corresponding predictions of each network were qualitatively examined (Fig 6). Results indicate that both networks could identify lesions on the leaf regardless of orientation. However, the DeepLabv3 model learned to distinguish and segment SCLB lesions more precisely than the U-Net network. For example, the U-Net network often

segmented other foliar stress and the midrib, as seen in rows one and three of column C in Fig 6. We infer that the network learned to segment stress in a color thresholding manner by distinguishing plant stress with yellow or white coloring from leaf tissue and not patterns specific to SCLB. Since the annotations of SCLB lesions were not comprehensive, decisions about the network's architecture could not solely rely on a high mIoU value, and the model that performed better at SCLB lesion generalization was considered. Therefore, the DeepLabv3 network was selected as the best-performing network for stage two because it achieved greater IoU scores and seemed to be more specific at SCLB detection.

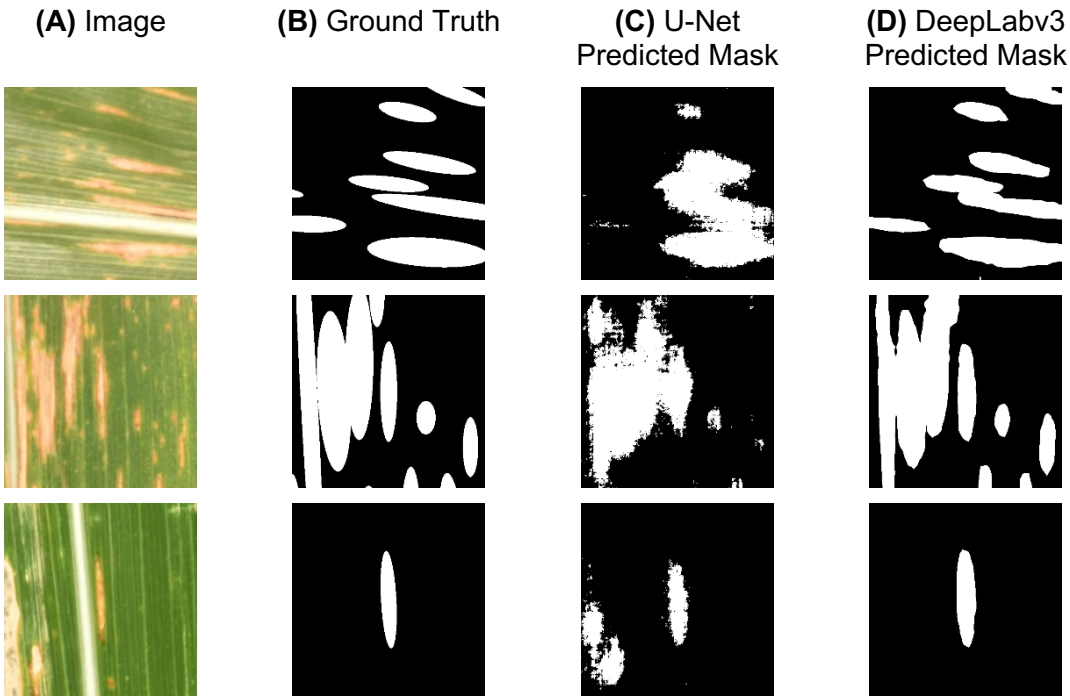


Fig 6. Qualitative results of the U-Net and DeepLabv3 models for the SCLB lesion segmentation task. (A) and (B) Sample images and ground truth mask pairs from the test set, respectively. Predicted masks of the cropped patches in column A for the (C) U-Net and (D) DeepLabv3 models.

Quantification

The analysis of the previous experiments determined that the DeepLabv3 model was better than the U-Net model at performing leaf and lesion segmentation. Thus, the first two stages of the pipeline integrated two DeepLabv3 models to extract the maize leaf by eliminating the background to remove nuisance elements within the image and semantic segmentation of the SCLB lesions. The fused model's feasibility and validity were tested on 1,904 images, 1,352 of which were collected during the 2019 field trial and 552 of which were collected during the 2023 field trial. Intensity estimation of SCLB symptoms was computed using the proportion of lesion area to leaf area calculated from the results of the lesion and leaf segmentation tasks, respectively. For every data collection day, images of about three randomly selected leaf samples were captured for each maize line. Before performing the correlation analysis, the average estimated severity proportion was computed for each maize line and collection day combination, which resulted in 468 and 184 estimated severity proportions for the 2019 and 2023 growing seasons, respectively. Then, the adjusted logistic and exponential functions mapped each proportion within the range of [1,9].

A correlation analysis for each field trial was performed between the field severity scores and the estimated disease severity scores to identify the absence or presence of a relationship; results are shown in Fig 7. Row A of Fig 7 shows the computed disease severity ratios of a single leaf and the field severity scores had a negative correlation ($r = -0.85$, $p \leq 0.001$ for 2019 and $r = -$

0.67, p ≤ 0.001 for 2023). A negative relationship was expected since maize plants rated with lower field severity scores would present more disease symptoms, resulting in greater disease severity ratios. The estimated severity scores mapped with the logistic function and the field severity scores had a positive correlation among the data for both growing seasons (r= 0.67, p ≤ 0.001 for 2019 and r= 0.52, p ≤ 0.001 for 2023). Any computed ratio less than or equal to 0.30 (row 7A) was mapped between 8 and 9 (row 7B). The correlation coefficient between the estimated severity scores mapped with the exponential and the field severity scores were r= 0.88 (p ≤ 0.001) and r= 0.72 (p ≤ 0.001) for the 2019 and 2023 growing seasons, respectively. The values mapped using the exponential function exhibited a stronger positive relationship than those mapped using the logistic function. However, there was substantial deviation among some of the field severity scores. For example, the field severity score of 7 during the 2019 field trial had estimated severity scores between 2 and 9. This result suggested that the pipeline learned some patterns to identify SCLB lesions, but the model may have struggled to locate most of the lesions within the images.

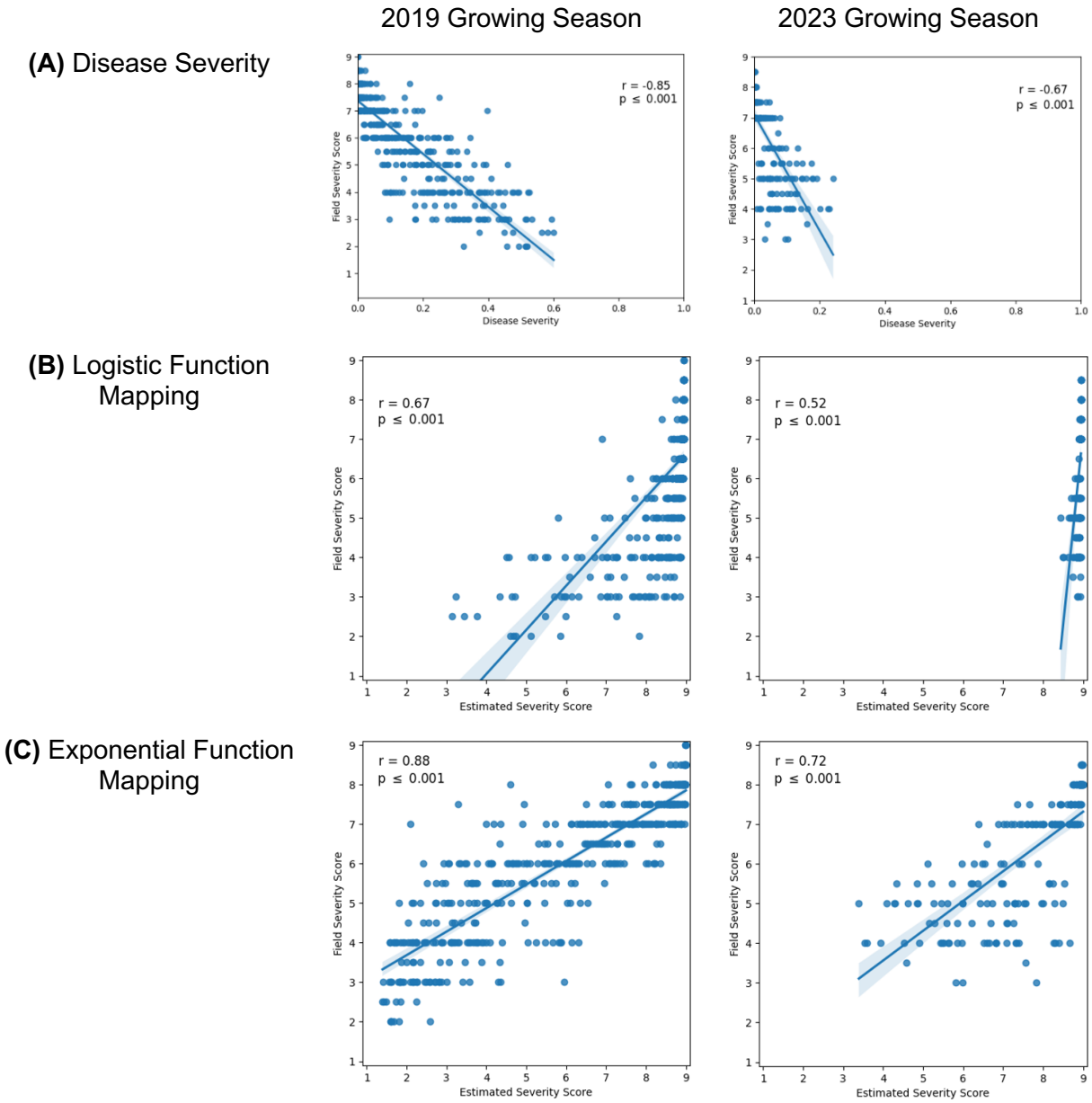


Fig 7. Correlation analysis between the estimated severity (x-axis) and field severity scores (y-axis) for the 2019 and 2023 growing seasons using the (A) disease severity proportions, (B) predicted scores mapped using the adjusted generalized logistic function, and (C) predicted scores mapped using the adjusted exponential function.

Conclusion

Our approach utilized state-of-the-art deep learning models U-Net and DeepLabv3, often implemented in semantic segmentation. The DeepLabv3 model was identified as the better network for segmenting maize leaves and SCLB lesions among field imagery. The model achieved a mean IoU of 0.879 for leaf segmentation and a mean IoU of 0.849 for lesion segmentation. Two DeepLabv3 models were fused into a multistage pipeline that could be used to estimate disease severity. The DeepLabv3-DeepLabv3 pipeline benefited from full-resolution images to predict the disease severity of SCLB with a correlation coefficient of 0.88 and 0.72 for the 2019 and 2023 growing seasons, respectively. This study proposed a non-destructive, deep learning-based approach that extends disease detection and localization to severity estimation and maps the results within the same range as the rating scale for visual assessment of plants. This demonstrates promising results for its use when developing an efficient disease management system. Future work involves improving the pipeline's generalization ability for SCLB lesion detection, monitoring lesion growth over time, and investigating whether the growth rates vary significantly among different near-isogenic lines.

Acknowledgments

The authors would like to thank the NCSU/NIH Biotech Training Program, S.A.S. Institute, Microsoft Corporation, and N.C. Plant Science Initiatives for the resources. Financial support was provided by the NCSU/NIH Biotech Training Program, grant number NIH 5T32GM133366. The authors would like to thank the Central Crops Research Station in North Carolina for providing support during the field trials. Also, we would like to thank William Hsieh, Sydney Gyurek, Sebastiano Busato, and Grace Vincent for helping to collect the field images and creating manual annotations for this paper.

References

Bock, C. H., Barbedo, J. G., Del Ponte, E. M., Bohnenkamp, D., & Mahlein, A. K. (2020). From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy. *Phytopathology Research*, 1-30.

Bock, C. H., Chiang, K. S., & Del Ponte, E. M. (2022). Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. *Tropical Plant Pathology, 47(1)*, 25-42.

Bock, C. H., Pethybridge, S. J., Barbedo, J. G., Esker, P. D., Mahlein, A. K., & Del Ponte, E. M. (2021). A phytopathometry glossary for the twenty-first century: towards consistency and precision in intra-and inter-disciplinary dialogues. *Tropical Plant Pathology*, 1-11.

Boggess, M. V., Lippolis, J. D., Hurkman, W. J., Fagerquist, C. K., Briggs, S. P., Gomes, A. V., & Bala, K. (2013). The need for agriculture phenotyping:"Moving from genotype to phenotype". *Journal of proteomics*, 20-39.

Bruns, H. A. (2017). Southern corn leaf blight: a story worth retelling. *Agronomy Journal, 109(4)*, 1218-1224.

Chen, L. C., Papandreou, G., S. F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Divyanth, L. G., Ahmad, A., & Saraswat, D. (2023). A two-stage deep-learning based segmentation model for crop disease quantification based on corn field imagery. *Smart Agricultural Technology, 3*, 100108.

Dutta, A., & Zisserman, A. (2019). The VIA annotation software for images, audio and video. *In Proceedings of the 27th ACM international conference on multimedia*, 2276-2279.

Garg, K., Bhugra, S., & Lall, B. (2021). Automatic quantification of plant disease from field image data using deep learning. *In Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 1965-1972.

Jaccard, P. (1912). The distribution of the flora in the alpine zone. 1. *New phytologist, 11(2)*, 37-50.

Kumar, B. C., Kumar, S., Bagaria, P., Sharma, M., Lahkar, C., Singh, B., . . . Jha, A. (2022). Maydis leaf blight of maize: Update on status, sustainable management and genetic architecture of its resistance. *Physiological and Molecular Plant Pathology, 121*, 101889.

Manso, G. L., Knidel, H., Krohling, R. A., & Ventura, J. A. (2019). A smartphone application to detection and classification of coffee leaf miner and coffee leaf rust. *arXiv preprint arXiv: 1904.00742*.

Pethybridge, S. J., & Nelson, S. C. (2015). Leaf Doctor: A new portable application for quantifying plant disease severity. *Plant disease*, 1310-1316.

Rahman, M. A., & Wang, Y. (2016). Optimizing intersection-over-union in deep neural networks for image segmentation. In International symposium on visual computing . *In International symposium on visual computing*, 234-244.

Reddy, D. V., Sudarshana, M. R., Fuchs, M., Rao, N. C., & Thottappilly, G. (2009). Genetically engineered virus-resistant plants in developing countries: current status and future prospects. *Advances in virus research*, 185-220.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 234-241.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510-4520.

Sermons, S. M., & Balint-Kurti, P. J. (2018). Large scale field inoculation and scoring of maize southern leaf blight and other maize foliar fungal diseases. *Bio-protocol, 8(5)*, e2745-e2745.

Singh, A., Ganapathysubramanian, B., Singh, A. K., & & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in plant science, 21(2)*, 110-124.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103(2684)*, 677-680.

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC medical imaging, 15*, 1-28.

Wang, C., Du, P., Wu, H., Li, J., Zhao, C., & Zhu, H. (2021). A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net. *Computers and Electronics in Agriculture, 189*, 106373.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big data 3*, 1-40.

Zhou, J., Zhou, J., Ye, H., Ali, M. L., Nguyen, H. T., & Chen, P. (2020). Classification of soybean leaf wilting due to drought stress using UAV-based imagery. *Computers and Electronics in Agriculture, 175*, 105576.