Check for updates

The Plant Phenome Journal

Original Research

Autonomous Detection of Plant Disease Symptoms Directly from Aerial Imagery

Harvey Wu, Tyr Wiesner-Hanks,* Ethan L. Stewart, Chad DeChant, Nicholas Kaczmar, Michael A. Gore, Rebecca J. Nelson,* and Hod Lipson

Core Ideas

- A deep learning model identified plant disease in UAV images with 95% accuracy.
- Transfer learning allowed for faster model optimization.
- This method detected plant disease symptoms at a very fine spatial scale.

The detection, diagnosis and quantification of plant diseases using digital technologies is an important research frontier. New and accurate methods would be an asset to growers, for whom early disease detection can mean the difference between successful intervention and massive losses, and plant breeders, who often must rely on time-consuming phenotyping by eye. We have developed such a method for detecting an important maize (*Zea mays* L.) disease. Northern leaf blight [NLB; causal agent *Setosphaeria turcica* (Luttrell) Leonard & Suggs] is a foliar disease of maize that causes significant yield losses. Accurately measuring NLB infection is necessary both for breeding more resistant maize lines and for guiding crop management decisions. Visual disease scoring in a large area is time-consuming and human evaluations are subjective and prone to error. In this work, we demonstrate an automated, high-throughput system for the detection of NLB in field images of maize plants. Through the use of an unmanned aerial vehicle (UAV) to acquire high resolution images, we trained a convolutional neural network (CNN) model on lower resolution sub-images, achieving 95.1% accuracy on a separate test set of sub-images. The CNN model was used to create interpretable heat maps of the original images, indicating the locations of putative lesions. Detecting lesions at a fine spatial scale allows for the potential of unprecedented high-resolution disease detection for plant breeding and crop management strategies.

A trained plant pathologist can often diagnose disease with fairly high confidence by looking at an image of symptomatic tissue. Such visual diagnosis does not scale well, however; one person cannot efficiently evaluate every plant in a large area, even if each plant could be clearly photographed. Disease surveillance and related applications can benefit from robust computer vision models that detect or classify objects in images. Trained using machine learning, wherein models are iteratively trained and tested on large amounts of human-generated data, such models have already demonstrated their usefulness for detection and classification of numerous plant diseases (Singh et al., 2016).

Convolutional neural networks (CNNs) are a class of machine learning models that are currently state of the art in many computer vision tasks, including object classification and detection. Part of this success lies in the ability of a CNN to perform automated feature extraction, as opposed to classical methods that may require handcrafted features (LeCun et al., 2015; Schmidhuber 2015). In a sense, training a CNN to expert-level performance crystallizes some of a pathologist's or crop scout's diagnostic capabilities to be shared with any number of growers at any time or location.

A number of recent studies have successfully used machine learning approaches, particularly CNNs, to detect plant diseases in images that can be acquired without a specialized camera (standard red–green–blue or RGB photos). Many studies have relied on the expert-curated set of over 50,000 plant disease images released through PlantVillage (described by Hughes and Salathe, 2016; utilized by Mohanty et al., 2016; Islam et al.,

Abbreviations: CNN, convolutional neural network; NLB, northern leaf blight; UAV, unmanned aerial vehicle.

H. Wu and C. DeChant, Dep. of Computer Science, Columbia Univ., New York, NY 10027; T. Wiesner-Hanks, E.L. Stewart, N. Kaczmar, M.A. Gore, and R.J. Nelson, Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell Univ., Ithaca, NY 14853; R.J. Nelson, Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell Univ., Ithaca, NY 14853; H. Lipson, Dep. of Mechanical Engineering and Institute of Data Science, Columbia Univ., New York, NY 10027. Harvey Wu and Tyr Wiesner-Hanks contributed equally to this work.

© 2019 The Author(s). This is an open access article distributed under the terms of the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Plant Phenome J. 2:190006 (2019)
doi:10.2135/tppj2019.03.0006

Received 19 Mar. 2019.
Accepted 8 Sept. 2019.
*Corresponding author (tyr.wiesnerhanks@pepsico.com; rjn7@cornell.edu).

2017; Wang et al., 2017; Barbedo 2018; Ferentinos, 2018; Too et al., 2019; Zhang et al., 2018). Due to its size and scope, this dataset has become a benchmark for new approaches in plant disease detection, analogous to the large datasets such as COCO (Lin et al., 2014) or ImageNet (Deng et al., 2009) used for many other computer vision tasks.

While the PlantVillage dataset contains mostly images of detached plant tissues on uniform backgrounds, several groups have successfully used machine learning to detect disease from images taken in the field (Mwebaze and Owomugisha, 2016; DeChant et al., 2017; Lu et al., 2017). Such images present additional challenges, such as variations in the field of view, overlap between plants, and extraneous sources of dead tissue. On the other hand, field images can be taken quickly, without the need to detach or isolate the disease tissue in question. Even without detaching plant tissue, however, taking images by hand is still prohibitively time consuming for many downstream uses. To rapidly detect disease across a large area, a faster imaging platform is needed.

The low cost and widespread availability of small unmanned aerial vehicles (UAVs) have made them an attractive option for imaging plants in the field. To our knowledge, there are only four published systems that can identify plant disease symptoms from UAV images via CNNs (Ha et al., 2017; Sugiura et al., 2018; Kerkech et al., 2018) or other machine learning methods (Tetila et al., 2017). These systems were applied to various crops across spatial scales, ranging from the entire plant (Ha et al., 2017; Sugiura et al., 2018) to regions of a plant (Kerkech et al., 2018) to individual leaves (Tetila et al., 2017).

Northern leaf blight (NLB) is a logical disease in which to attempt the challenge of fine-scale aerial identification, due to its conspicuous symptoms and economic importance. Northern leaf blight, also called northern corn leaf blight or NCLB, is a fungal foliar disease of maize that causes large, gray-brown necrotic lesions (Fig. 1). Identifying these lesions, often >1 cm wide and 5 cm long, in aerial images of maize plants is more tractable than it would be for diseases with smaller or more subtle symptoms. The economic impact of NLB also makes it an ideal target disease. Between 2012 and 2015, estimated annual yield loss due to NLB in the United States and Ontario rose sevenfold to 14 million tonnes (Mueller et al., 2016), an economic loss of roughly US$1.9 billion, accounting for one fourth of all estimated yield losses from disease that year.

Northern leaf blight losses can be mitigated via two main strategies: breeding of maize lines with improved NLB resistance and chemical control via fungicides, both of which rely on accurate detection and quantification of symptoms. To breed NLB-resistant maize, plant breeders must quantify resistance in field trials. This is currently achieved by human experts visually estimating disease severity by eye late in the growing season, a method subject to high inter- and intra-rater variation (Poland and Nelson, 2011). Incubation period, the time between inoculation and the first visible lesion, can be measured before flowering, allowing breeders to select both male and female parents and ideally achieve twice as much genetic gain per unit time (Smith and Kinsey, 1993; Carson, 1995). Measuring the incubation period is, however, very time consuming and not feasible for large trials.



Fig. 1. Representative image of northern leaf blight (NLB) symptoms.

Chemical control of NLB also relies on accurate detection and quantification. To decide whether or not to apply foliar fungicide, growers must estimate whether economic losses from NLB (and/or other diseases) would exceed the costs of fungicide application. Extension guides suggest scouting for NLB symptoms when plants are flowering, the time at which diseases and other environmental factors are most likely to decrease yield (Robertston and Mueller, 2007; Dewerff et al., 2019). Extensively scouting a large field of fully-grown maize plants is naturally quite difficult. Both breeding and management would be facilitated by methods that can rapidly detect and quantify NLB across a large area.

The major objective of this study was to demonstrate a proof-of-concept for an automated phenotyping system that combines the deep learning approach of DeChant et al. (2017) with UAV-based imagery to accurately identify maize disease symptoms by producing meaningful heat maps that predicted the presence of NLB lesions with high accuracy. To our knowledge, this is the first method capable of detecting diseased portions of plant leaves from UAV imagery.

Materials and Methods

Plant Preparation

All images were taken in a 2017 field experiment planted at the Cornell University Musgrave Research Farm in Aurora, NY. The experiment contained two replicates of 250 maize hybrids, consisting of diverse genotypes from the Genomes to Fields Initiative (www.genomes2fields.org; AlKhalifah et al., 2018)

25782703, 2019, 1, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.2135/cppj2019.03.0006, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and local checks chosen as representatives of hybrids commonly grown in central New York. Genotypes were arranged in two-row plots 5.64 m in length, with interrow spacing of 0.76 m and interplant spacing of 15.7 cm. The rainfed field trial was treated with a nitrogen and broadleaf herbicide application regimen standard for Central New York. Maize plants were inoculated at the V5 to V6 stage with both 0.5 mL of liquid spore suspension and ~2 mL of sorghum [*Sorghum bicolor* (L.) Moench.] grains colonized by *S. turcica* for 3 to 4 wk prior (DeChant et al., 2017).

Dataset Description

The dataset contains images of both infected and uninfected leaves taken during 10 flights between 22 and 84 d post-inoculation (Wiesner-Hanks et al., 2018). Images were collected using a Sony Alpha 6000 camera fitted with a Sony SEL55210 lens set to 210-mm focal length. The camera was mounted to a DJI Matrice 600 UAV flown at a speed of 1.5 m/s and 6 m above ground level. The UAV was programed to fly between waypoints set out in a serpentine fashion across the field, and the cameras built-in intervalometer was used to capture an image approximately once per second. There was no overlap between images. Flights were conducted under lighting conditions from overcast to partial cloud cover to full sun, which often changed during the course of a single flight (Fig. 2). The images thus contain a mixture of direct and diffuse lighting.

For each image, we annotated the semi-major axis of each NLB lesion using a custom ImageJ macro (Wiesner-Hanks et al., 2018). Images were first filtered automatically by Canny edge detection to remove blurry images and then filtered out manually during annotation if they contained no maize leaves, were out of focus or blurry, or were otherwise unsatisfactory. A total of 6267 images are included in the dataset: 3741 with lesions and 2526 without lesions. Most images contained at least two plants in part or in full, and images were not tied to genotypes or plots because they could not be geolocated with high enough precision. All images and annotations are available (Wiesner-Hanks and Brahimi, 2019).

Each set of images (infected or uninfected) was randomly divided into training, validation, and test sets at a ratio of 70:15:15. The test set was isolated from all aspects of model design, training, and hyperparameter tuning.

Model Architecture

The image analysis approach was based on the first two stages of the three-stage CNN pipeline developed by DeChant et al. (2017). Code for the model is available at https://github.com/Columbia-Creative-Machines-Lab/crops. In the first stage, a CNN was trained to predict whether small subregions of an image contained lesions (Fig. 3). The second stage used this CNN as a sliding window across the whole image to generate a heat map. Unlike the prior study (DeChant et al., 2017) that used a manually designed architecture, we used a Resnet-34 model (He et al., 2016) pretrained on ImageNet (Russakovsky et al., 2015) to "transfer learn" rather than training from scratch. ImageNet contains no images of NLB lesions, but due to the immense variety of visual information contained in the dataset, models are able to learn basic visual structure that is useful for detection tasks (Goodfellow et al., 2016). This technique brought significant time savings and improved accuracy. To perform transfer learning, a linear layer of output dimension two was appended to the pretrained Resnet-34 model, and all parameters of the model besides those of the new linear layer were fixed. The linear layer was trained for one epoch, and all remaining parameters were subsequently unfixed and trained jointly in later epochs.

A key distinction between hand-held images and UAV images lies in the altitude from which the pictures are taken. Whereas the individual lesions in hand-held images may occupy a large percentage of the pixels in an image, at heights of 6 m each lesion will be proportionally much smaller. When taking sub-images of 224 by 224 pixels, we found it was impossible for human experts to tell



Fig. 2. Four sample images from our dataset that were collected by an unmanned aerial vehicle (UAV). The two images on the left were captured in August 2017; the right images are from September of the same year. Although the images in each column are visually similar, the top row contains no lesions while each image in the bottom row contains seven.

25782703, 2019, 1, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.2135/cppj2019.03.0006, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

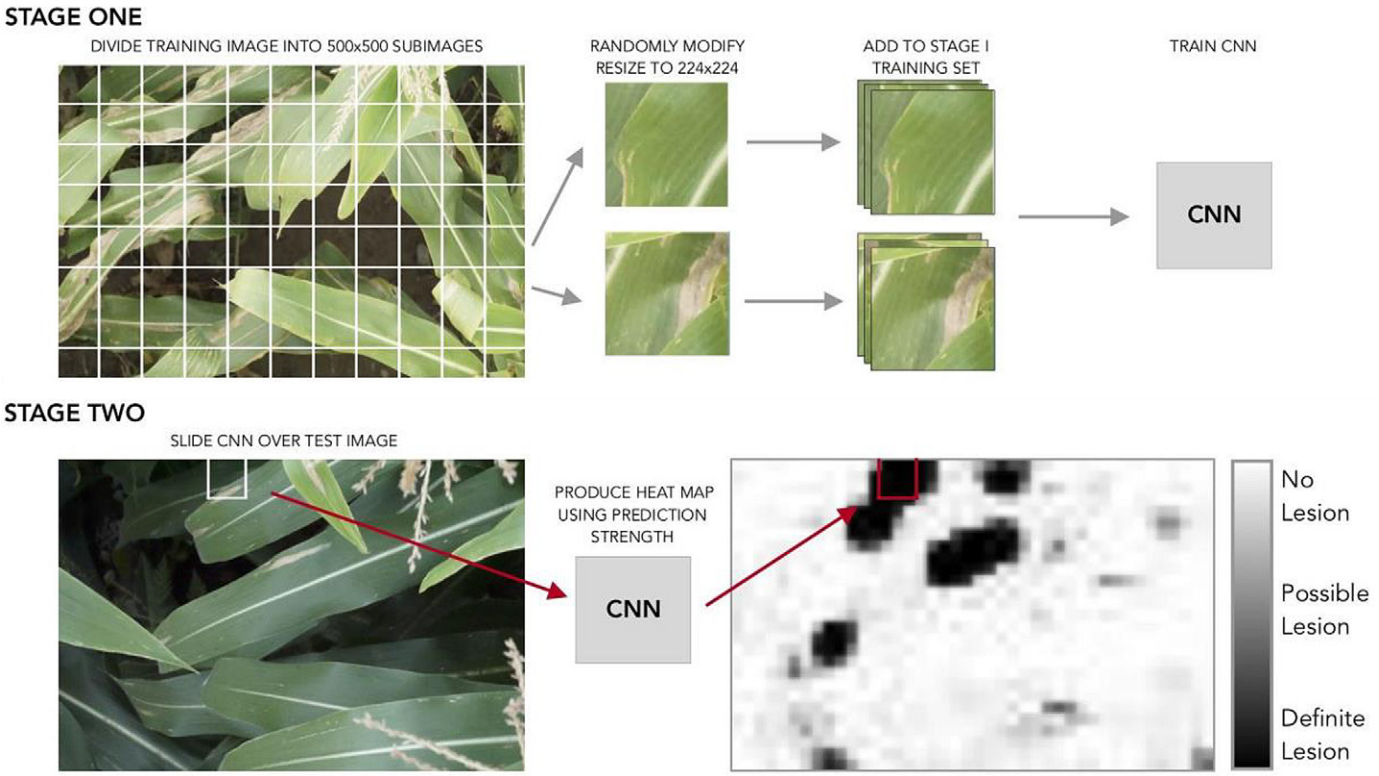


Fig. 3. Outline of the two-stage pipeline for detection of northern leaf blight lesions using convolutional neural networks (CNNs). The first stage (top) involves training a CNN backbone using data augmentation. First, 500- by 500-pixel sub-images are randomly cropped from all the original training images, which generally show several plants in part or in full. Random transformations such as rotations and flips are performed on each sub-image. The transformed sub-images are divided into those containing lesions and those without lesions and used to train a ResNet-34 with a stochastic gradient descent. The second stage (bottom) is heat map generation. Using a sliding window approach, crops of the original image (e.g., as indicated with the small white square) are fed into the CNN. The output of the CNN is used as the pixel value for that position in the heat map. In this image, the final heat map represents lesions as black and non-lesioned areas as white. The three main black areas in the heat map overlap the lesions in the image.

whether the sub-image contained a lesion or not, because there was not enough "context" in the image. We thus modified the procedure of DeChant et al. (2017) by taking 500 by 500 selections from each original image of 4000 by 6000, and labeling them according to whether the centermost 224 by 224 portion contained a lesion. The CNN was trained to classify the center portion but could use the surrounding image area as contextual information.

We added additional stochasticity with random flips, rotations, and zooms, post-processing steps that serve as a form of data augmentation. Zoom augmentation introduces a random variable X sampled from a discrete uniform distribution $[-50, 50]$. Instead of 500 by 500, sub-images were taken with dimension $(500 + X)$ by $(500 + X)$. After applying other post-processing steps, the sub-image was scaled back to 224 by 224 and added to the sub-image training set.

In Stage 2, we took a sliding window of 500 by 500 over the image, scaled it down to 224 by 224, and fed it into the trained Stage 1 CNN. The output of the Stage 1 CNN determined the strength of the region of the heat map. The step size for the sliding window was fixed at 40 pixels.

Experiments were performed on a Google Compute Engine instance with 16 GB of RAM and an NVidia P100 graphics processing unit, as well as a local machine with 16 GB of RAM and an NVidia 1080 GTX. To implement and execute experiments in Python, we used PyTorch (Paszke et al., 2017), a Python library that performs automatic differentiation over dynamic computation graphs; practically, this allows convenient training of arbitrary neural network architectures using backpropagation with gradient descent. Moreover, the PyTorch model zoo contains various models pretrained on the ImageNet (Russakovsky et al., 2015) dataset. We used a pretrained ResNet-34 architecture (He et al., 2016) as the backbone of the Stage 1 CNN. Initially, we treated the pretrained ResNet model as a feature extractor whose outputs were used as inputs to a simple linear classifier. To this end, we trained the model for one epoch, freezing all the weights but those of the final linear layer. All future epochs trained all parameters of the entire model, using the Adam variant of the stochastic gradient descent optimization algorithm (Kingma and Ba 2014) with a batch size of 80.

After training the CNN, we used it to generate heat maps on the original 4000 by 6000 images. We ran the CNN as a 224 by 224 sliding window with a given step size over each image and applied the softmax function to the outputs, with the effect of normalizing them so that they represented a probability distribution.

25782703, 2019, 1, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.2135/cppj2019.03.0006, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Collecting the component of the output corresponding to "with lesion," we generated a heat map based on the matrix of these selected values. Each pixel of the heat map represents an associated portion of the original image—the intensity, or the strength of the softmax output, represents the model's confidence in lesion presence.

Results

We analyzed 6267 images of maize leaves, comprising 3741 images containing NLB-infected leaves and 2526 images without lesions. On average, each image of infected leaves had 6.28 labeled lesions, totaling 25,508 lesions. Not all necrotic leaf tissue in the dataset was attributable to NLB; other causes included physical damage, natural lower leaf senescence, nitrogen deficiency, corn flea beetle feeding, and other foliar diseases such as northern corn leaf spot. Lesions present due to inoculation were comparable to those caused by natural infection in the non-inoculated batch, with similar color and shape.

Our next attempt, using our two stage CNN pipeline, was based on a sub-image generation process that output approximately 4.5 sub-images per lesion. This training set was roughly balanced between positive and negative samples. We were able to achieve an accuracy of 97.76% (number of correct predictions divided by the number of total samples) on a hold-out test set of sub-images, as well as 98.42% precision (number of correct positive predictions divided by the total number of positive predictions), and 97.85% recall (number of correct positive predictions over the sum of true positives and false negatives).

The heat maps generated by this initial model still produced many false positives and negatives (Table 1). This indicated that the training set did not accurately represent the empirical distribution and perhaps that the network was overfitting on the large numbers of sub-images for each lesion. We then reshuffled the data split, and changed the generation process such that only one image per lesion was emitted, together with many more negative samples (Table 2). This model had a lower accuracy of 95.1% (representative correct and incorrect classifications are shown in Fig. 4 and 5). However, it created much more interpretable heat maps (Fig. 6).

The first epoch was trained with all weights frozen except for the last, appended linear layer. The second epoch unfroze all the weights of the network; the fine-tuning that occurred during training allowed for a substantial increase in accuracy (Fig. 7).

Table 1. Confusion matrix for the final convolutional neural network (CNN) model on a hold-out test set of sub-images. There were 675 misclassifications out of 13,732 sub-images. The number of false negatives was large, but in the heat map generation setting this was acceptable; the CNN had many opportunities to make the correct prediction when used as a sliding window.

	Predicted positive	Predicted negative
True positive	2798	585
True negative	90	10,259

While examining the heat maps produced on the test set, we realized that some of the mistakes made by the final model, especially false positives, were actually mis-annotated items in the dataset (Fig. 8a). Another category of misclassifications belonged to images different than those seen in training, such as images of senesced leaves, or different irregular viewpoints (Fig. 8b).

Discussion

The disease detection system described here was able to detect disease symptoms from aerial images at a fine spatial scale with high accuracy. Existing models have detected and/or classified disease(s) in plant tissue at larger spatial scales, from individual leaves to entire plants. Direct comparisons between accuracies are confounded by many factors: the size, shape, and color of symptoms, the size and growth habit of the plant, which tissue is visibly infected, etc. As stated above, we selected NLB as a target disease because of its large, distinct symptoms, making it a more tenable target for accurate detection in images. With this caveat in mind, our system had accuracy comparable to or higher than previously described approaches. On the scale of whole plants, other systems were able to classify entire plants as virus infected or not with 93% accuracy in beet (*Beta vulgaris* L.; Ha et al., 2017) and 84% accuracy in potato (*Solanum tuberosum* L.; Sugiura et al., 2018). On the sub-plant scale, Kerkech et al. (2018) classified square patches of grape vines (*Vitis vinifera* L.) as diseased or not with 94 to 96% accuracy. Tetila et al. (2017) successfully implemented machine learning recognition at the finest spatial scale to date, diagnosing soybean [*Glycine max* (L.) Merr.] diseases on a leaf-by-leaf basis with peak accuracy of 98% when flying at an altitude of 1 to 2 m, comparable to that our initial model and higher than that of our final model. At similar altitudes, the two were comparable: our final model reached 95% accuracy from images taken at 6 m, while that of Tetila et al. (2017) reached 95% accuracy on images taken at 4 m and 85% on those taken at 8 m.

An advantage of the multistage pipeline was that it made full use of the high-resolution images. There was probably too much information in a full image for a single CNN to accurately predict the presence of lesions, especially considering the various naturally occurring objects with similar color and shape. Indeed, our heat map method showed significantly improved performance compared with a baseline network trained on scaled-down whole images. When we scaled down such an image, it was difficult even for human experts to identify lesions. By training a CNN on full-resolution fragments of these images, we could leverage both the

Table 2. Statistics for a dataset of sub-images randomly generated from the original images. This dataset was used to train the final model.

	Lesions	Non-lesions
Training	17,324	56,528
Validation	3730	10,404
Test	3384	10,350

25782703, 2019, 1, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.2135/cppj2019.03.0006, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



Fig. 4. Examples of test set images classified correctly by our final trained model: true positives (left); true negatives (right).

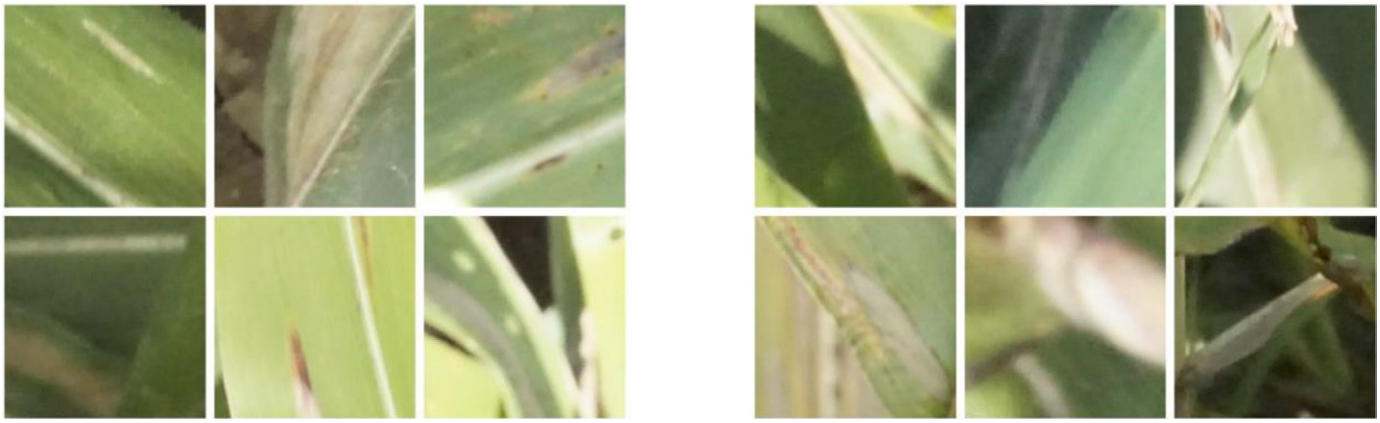


Fig. 5. Examples of test set images misclassified by our final trained model: false negatives (left); false positives (right).

power of CNNs and the large amount of information contained in the data. Breaking up images into pieces, as opposed to classifying entire images, was required partly due to hardware constraints; while the size of each image was around 8 MB, the parameters of the neural network (such as the weights of each convolutional filter) occupied the bulk of graphics processing unit memory during training.

Despite the differences between the NLB dataset and ImageNet (two classes instead of thousands, 6267 images rather than approximately 14 million), the low level features of the pretrained Resnet-34 transferred to the task successfully. Compared with the Stage 1 ensemble network (DeChant et al., 2017), which made use of individual CNNs with accuracies as low as 81%, the transfer learned ResNet model achieved accuracies >90% in only two epochs, shaving significant amounts of training time compared with training on the dataset from scratch.

The system described allowed detection of the presence of a disease lesion in 500 × 500 image fragments, allowing heat maps to be generated for base images of arbitrary size. These heat maps can be used for many tasks, such as early detection or incubation period

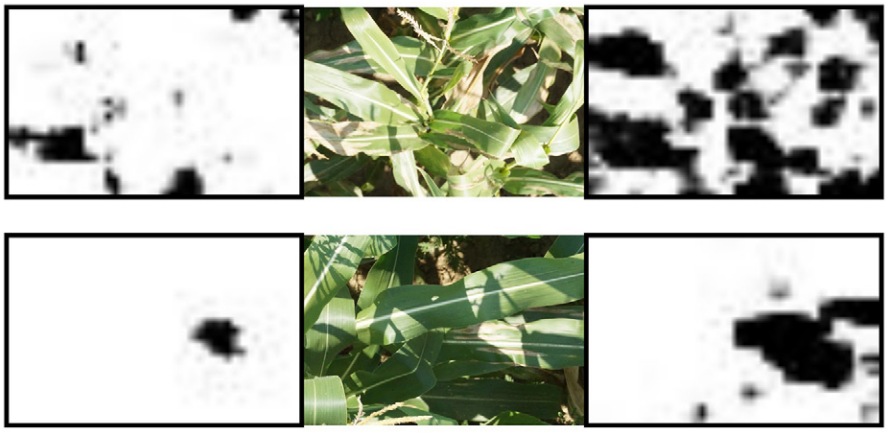


Fig. 6. A comparison between the first convolutional neural network (CNN) model (left, trained with 1:1 lesion/non-lesion ratio) and the final CNN model (right, trained with a ratio of about 1:4). The original image is shown in the middle. In both cases, the initial model failed to detect many smaller lesions, while the final model was much more sensitive.

25782703, 2019, 1, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.2135/cppj2019.03.0006, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

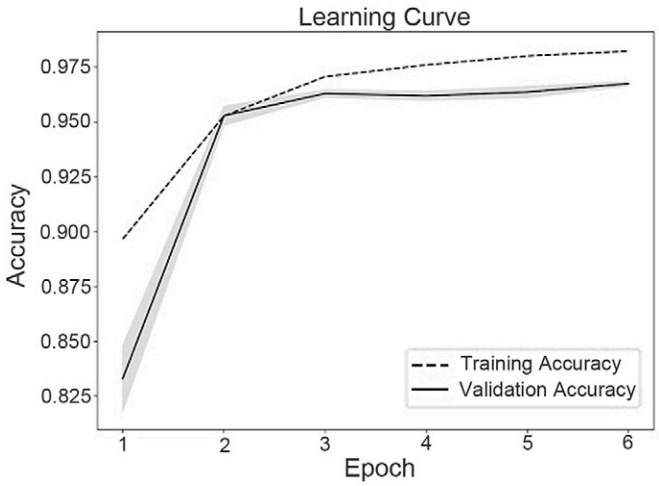


Fig. 7. Learning curve for the first convolutional neural network (CNN) model. We took three different random seeds and plotted the average. The standard deviation is shown in gray. Note that the first epoch trained only the last linear layer of the pretrained CNN, with the remaining weights fixed. The subsequent epochs allowed fine-tuning of all the weights.

evaluation; moreover, heat maps also create visual representations of disease severity, roughly quantifying the amount of disease damage. Using a UAV, data acquisition no longer has a large time cost; the limiting factor is having human experts label thousands of photographs. These annotations are, in a sense, a reusable resource, unlike time spent in the field evaluating disease by eye. Different groups can apply different machine learning methods or models to the same labeled training data to find the best approach for the task at hand. Annotated data from multiple years and locations can be combined to train a model for robust diagnosis or divided to train models that are specifically optimized for detecting disease symptoms under specific conditions: at a specific growth phase, in the presence of a particular abiotic stress, in cultivars with a certain plant architecture, etc. Much as we used a network pretrained on diverse images from ImageNet as a foundation for transfer learning, this model could be used as the basis for transfer learning for a new diagnostic task.

We used accuracy on the sub-image dataset as a proxy for performance but found that different sub-image generation processes could drastically change the quality of the final heat maps, even when accuracy itself was fairly high. For example, a network with >97% accuracy generated heat maps that were worse in terms of correspondence to lesion shape, size, and area than a network with 95% accurate predictions. While it was straightforward for humans to judge heat maps qualitatively, quantifying their performance was difficult, as the annotations used did not delineate lesion boundaries. Annotating images with polygons bounding entire lesions, rather than just lines, would be more time consuming but would ultimately allow the accuracy of the produced heat maps to be quantitatively measured and models selected that can identify symptoms in a "true to size" fashion.

Whether this model can be successfully transferred to other contexts remains to be seen. Although the experimental trial in question contained phenotypically diverse lines, all images came from a single year in a single field. The model may perform differently when applied to maize lines with very different appearance or morphology, different planting densities, abiotic stresses like nutrient deficiency or drought stress, etc. Furthermore, all images were taken in an artificially-inoculated field trial, in which NLB predominated and other diseases were fairly rare. This was useful for limiting misclassifications, but it is unclear how the model would respond to diseases with similar symptoms, such as Stewart's wilt or anthracnose leaf blight. Any potential user would presumably prefer a system that can distinguish between similar-looking diseases.

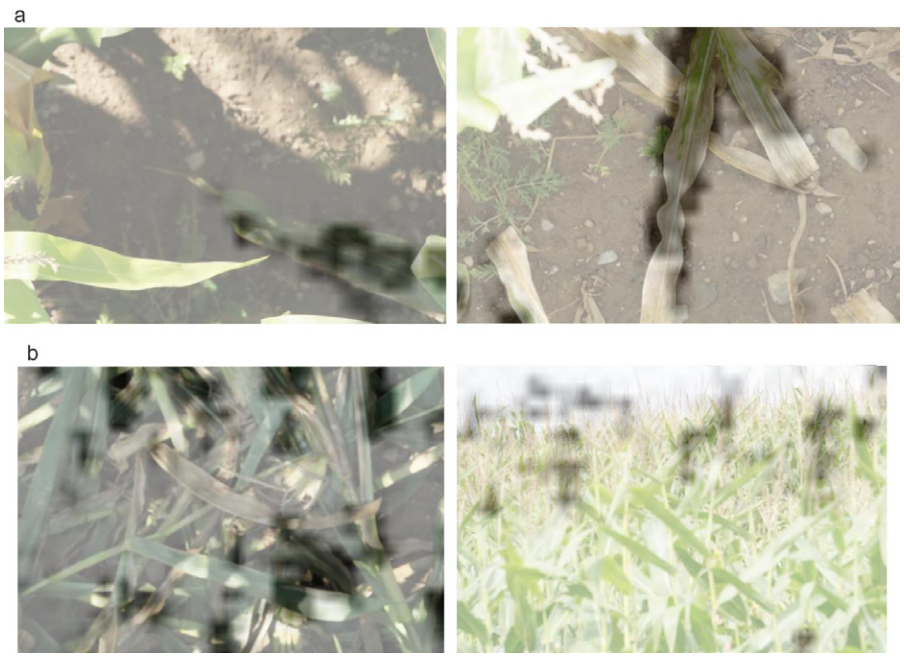


Fig. 8. (a) Two examples of the final model identifying errors in the dataset or "beating the experts," with darker areas where the model believes there to be a lesion—these areas indeed contain lesions; and (b) two examples of out-of-distribution inputs for which the final model did not achieve accurate performance, with darker areas where the model believes there to be a lesion: a mass of senesced leaves (left) and a picture of the field taken from a horizontal rather than vertical viewpoint (right).

25782703, 2019, 1, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.2135/cppj2019.03.0006, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The two potential groups of users for this technology are plant breeders and geneticists, who could use it for fast, accurate scoring of field trials, and growers, who could use it for crop scouting to make informed disease management decisions. As it stands, the technical limitations of the platform limit its utility for the former. To connect model output to an experimental unit (a plot, plant, or row), the images must either be stitched together to situate each within the entire field or georeferenced with very high accuracy. Stitching is not feasible for images taken at such high resolution, as hundreds of thousands would be needed to cover a field, and although the UAV platform was equipped with a real-time kinematic GPS system, we were not able to geo-reference each picture with the decimeter-level accuracy needed to assign it to a specific row in the experiment. Thus this system is most useful for detecting disease at many points across large areas planted with a single cultivar, i.e., a grower's field.

Conclusions

Our model was able to identify individual disease lesions with high accuracy from aerial photographs taken in the field, setting a new benchmark for field-based disease detection. Transfer learning greatly simplified the model training process, making it a promising route for others who want to deploy similar systems in any of the hundreds of other economically important pathosystems. Adding contextual information to the training process allowed us to capture the benefits of high-resolution images while retaining information on where a putative lesion was situated on the plant as a whole, increasing overall accuracy.

Data Availability

Images and annotations are available at https://osf.io/p67rz/. Code for the model is available at https://github.com/Columbia-Creative-Machines-Lab/crops.

Acknowledgments

This work was supported by the US National Science Foundation National Robotics Initiative Grant no. 1527232 (M.A. Gore, R.J. Nelson, and H. Lipson).

References

AlKhalifah, N., D.A. Campbell, C.M. Falcon, J.M. Gardiner, N.D. Miller, M.C. Romay, et al. 2018. Maize Genomes to Fields: 2014 and 2015 field season genotype, phenotype, environment, and inbred ear image datasets. BMC Res. Notes 11:452. doi:10.1186/s13104-018-3508-1

Barbedo, J.G.A. 2018. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Comput. Electron. Agric. 153:46–53. doi:10.1016/j.compag.2018.08.013

Carson, M.L. 1995. Inheritance of latent period length in maize infected with *Exserohilum turcicum*. Plant Dis. 79:581–585.

DeChant, C., T. Wiesner-Hanks, S. Chen, E.L. Stewart, J. Yosinski, M.A. Gore, et al. 2017. Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. Phytopathology 107:1426–1432. doi:10.1094/PHYTO-11-16-0417-R

Deng, J., W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL. 20–25 June 2009. IEEE, Piscataway, NJ. p. 248–255. doi:10.1109/CVPR.2009.5206848

Dewerff, R., B. Jensen, P.J. Liesch, G. Nice, M. Renz, D. Smith, and R. Werle. 2019. Pest management in Wisconsin field crops. Publ. A3646. Univ. of Wisconsin Ext., Madison.

Ferentinos, K.P. 2018. Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. 145:311–318. doi:10.1016/j.compag.2018.01.009

Goodfellow, I., Y. Bengio, and A. Courville. 2016. Deep learning. MIT Press, Cambridge, MA.

Ha, J.G., H. Moon, J.T. Kwak, S.I. Hassan, M. Dang, O.N. Lee, and H.Y. Park. 2017. Deep convolutional neural network for classifying *Fusarium* wilt of radish from unmanned aerial vehicles. J. Appl. Remote Sens. 11:042621. doi:10.1117/1.JRS.11.042621

He, K., X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. In: Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV. 26 June–1 July 2016. IEEE Comput. Soc., Los Alamitos, CA. p. 770–78. doi:10.1109/CVPR.2016.90

Hughes, D.P., and M. Salathe. 2016. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv:1511.08060 [cs.CY].

Islam, M., A. Dinh, K. Wahid, and P. Bhowmik. 2017. Detection of potato diseases using image segmentation and multiclass support vector machine. In: IEEE 30th Canadian Conference on Electrical and Computer Engineering, Windsor, ON, Canada. 30 Apr.–3 May 2017. IEEE, Piscataway, NJ. doi:10.1109/CCECE.2017.7946594

Kerkech, M., A. Hafiane, and R. Canals. 2018. Deep learning approach with colorimetric spaces and vegetation indices for vine diseases detection in UAV images. Comput. Electron. Agric. 155:237–243. doi:10.1016/j.compag.2018.10.006

Kingma, D.P., and J. Ba. 2014. Adam: A method for stochastic optimization. arXiv:1412.6980 [cs.LG].

LeCun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. Nature 521:436–444. doi:10.1038/nature14539

Lin, T.Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, et al. 2014. Microsoft COCO: Common objects in context. In: D. Fleet et al., editors, 13th European Conference on Computer Vision, Zurich, Switzerland. 6–12 Sept. 2014. Part V. Springer, Cham, Switzerland. p. 740–755.

Lu, Y., S. Yi, N. Zeng, Y. Liu, and Y. Zhang. 2017. Identification of rice diseases using deep convolutional neural networks. Neurocomputing 267:378–384. doi:10.1016/j.neucom.2017.06.023

Mohanty, S.P., D.P. Hughes, and M. Salathé. 2016. Using deep learning for image-based plant disease detection. Front. Plant Sci. 7:1419. doi:10.3389/fpls.2016.01419

Mueller, D.S., K.A. Wise, A.J. Sisson, T.W. Allen, G.C. Bergstrom, D.B. Bosley, et al. 2016. Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015. Pap. Plant Pathol. 484. Univ. of Nebraska, Lincoln. https://digitalcommons.unl.edu/plantpathpapers/484/.

Mwebaze, E., and G. Owomugisha. 2016. Machine learning for plant disease incidence and severity measurements from leaf images. In: 15th IEEE International Conference on Machine Learning and Applications, Anaheim, CA. 18–20 Dec. 2016. IEEE Comput. Soc., Los Alamitos, CA. p. 158–163.

Paszke, A., S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, et al. 2017. Automatic differentiation in PyTorch. Presented at: 31st Conference on Neural Information Processing Systems, Long Beach, CA. 4–9 Dec. 2017. https://openreview.net/pdf?id=BJJsrmfCZ.

Poland, J., and R.J. Nelson. 2011. In the eye of the beholder: The effect of rater variability and different rating scales on QTL mapping. Phytopathology 101:290–298. doi:10.1094/PHYTO-03-10-0087

Robertston, A., and D. Mueller. 2007. Fungicide application to corn at tasseling. Iowa State Univ. Ext., Ames. crops.extension.iastate.edu/fungicide-application-corn-tasseling

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al. 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115:211–252. doi:10.1007/s11263-015-0816-y

25782703, 2019, 1, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.2135/cppj2019.03.0006, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Schmidhuber, J. 2015. Deep learning in neural networks: An overview. Neural Networks 61:85–117. doi:10.1016/j.neunet.2014.09.003

Singh, A., B. Ganapathysubramanian, A.K. Singh, and S. Sarkar. 2016. Machine learning for high-throughput stress phenotyping in plants. Trends Plant Sci. 21:110–124. doi:10.1016/j.tplants.2015.10.015

Smith, J., and J. Kinsey. 1993. Latent period: A possible selection tool for *Exserohilum turcicum* resistance in corn (*Zea mays* L.). Maydica 38:205–208.

Sugiura, R., S. Tsuda, H. Tsuji, and N. Murakami. 2018. Virus-infected plant detection in potato seed production field by UAV imagery. Presented at: 2018 ASABE Annual International Meeting, Detroit, MI. 29 July –1 Aug. 2018. Am. Soc. Agric. Biol. Eng., St. Joseph, MI. doi:10.13031/aim.201800594

Tetila, E.C., B.B. Machado, N.A. de Souza Belete, D.A. Guimarães, and H. Pistori. 2017. Identification of soybean foliar diseases using unmanned aerial vehicle images. IEEE Geosci. Remote Sens. Lett. 14:2190–2194. doi:10.1109/LGRS.2017.2743715

Too, E.C., L. Yujian, S. Njuki, and L. Yingchun. 2019. A comparative study of fine-tuning deep learning models for plant disease identification. Comput. Electron. Agric. 161:272–279. doi:10.1016/j.compag.2018.03.032

Wang, G., Y. Sun, and J. Wang. 2017. Automatic image-based disease severity estimation using deep learning. Comput. Intell. Neurosci. 2017:2917536.

Wiesner-Hanks, T., and M. Brahimi. 2019. Image set for deep learning: Field images of maize annotated with disease symptoms. Open Science Foundation Repository. osf.io/p67rz.

Wiesner-Hanks, T., E.L. Stewart, N. Kaczmar, C. DeChant, H. Wu, R.J. Nelson, et al. 2018. Image set for deep learning: Field images of maize annotated with disease symptoms. BMC Res. Notes 11:440. doi:10.1186/s13104-018-3548-6

Zhang, X., Y. Qiao, F. Meng, C. Fan, and M. Zhang. 2018. Identification of maize leaf disease using improved deep convolutional neural networks. IEEE Access 6:30370–30377. doi:10.1109/ACCESS.2018.2844405

25782703, 2019, 1, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.2135/cppj2019.03.0006, Wiley Online Library on [30/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License