e-**X**tra*

Automated Identification of Northern Leaf Blight-Infected Maize Plants from Field Imagery Using Deep Learning

Chad DeChant,[†] Tyr Wiesner-Hanks, Siyuan Chen, Ethan L. Stewart, Jason Yosinski, Michael A. Gore, Rebecca J. Nelson, and Hod Lipson

First author: Department of Computer Science, Columbia University in the City of New York, 10027; second, fourth, and sixth authors: Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853; third author: Department of Mechanical Engineering, Columbia University; fifth author: Uber AI Labs, San Francisco 94103; seventh author: Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University; and eighth author: Department of Mechanical Engineering and Institute of Data Science, Columbia University.
Accepted for publication 22 June 2017.

ABSTRACT

Northern leaf blight (NLB) can cause severe yield loss in maize; however, scouting large areas to accurately diagnose the disease is time consuming and difficult. We demonstrate a system capable of automatically identifying NLB lesions in field-acquired images of maize plants with high reliability. This approach uses a computational pipeline of convolutional neural networks (CNNs) that addresses the challenges of limited data and the myriad irregularities that appear in images of field-grown plants. Several CNNs were trained to classify small regions of images as containing NLB lesions or not; their predictions were combined into separate heat maps, then fed into a final CNN trained to classify the entire image as containing diseased plants or not. The system achieved 96.7% accuracy on test set images not used in training. We suggest that such systems mounted on aerial- or ground-based vehicles can help in automated high-throughput plant phenotyping, precision breeding for disease resistance, and reduced pesticide use through targeted application across a variety of plant and disease categories.

An estimated 13% of global potential crop yield is lost to diseases each year (Oerke and Dehne 2004), with much higher losses occurring under epidemic conditions. To evaluate resistance of plant germplasm and breed for improved resistance, conventional visual assessments of disease incidence or severity are widely used. However, such assessments are prone to error through inter- and intrarater variation, which can reduce precision and accuracy of genetic inferences (Bock et al. 2009; Poland and Nelson 2011; Sherwood et al. 1983). Accurate automated high-throughput phenotyping of plant diseases has the potential to aid crop management, speed up breeding, and contribute to fundamental and applied research efforts (Pauli et al. 2016).

Northern leaf blight (NLB), also called northern corn leaf blight or turcicum blight, is a fungal foliar disease of maize caused by *Setosphaeria turcica* (anamorph: *Exserohilum turcicum*). In the United States and Ontario, NLB has been growing especially severe in recent years, with estimated yield losses rising steadily from 1.9 million metric tons in 2012 to 14 million metric tons in 2015 (Mueller et al. 2016). This estimated yield loss from NLB accounted for one-fourth of all estimated yield losses from disease in 2015, causing an estimated economic loss of $1.9 billion. In controlled yield trials, high levels of infection can cause yield losses of 18 to 62% (Perkins and Pedersen 1987; Raymundo and Hooker 1981; Shankara and Gowda 2011), as well as a decrease in silage quality and digestibility (Wang et al. 2010). Incubation period (the number of days after inoculation at which NLB lesions are present on 50% of plants in a plot) is an important component of NLB resistance that can be evaluated before the plant flowers, making it particularly useful for resistance breeding (Brewster et al. 1992; Welz and

Geiger 2000). However, scoring this by eye is very time consuming and often not feasible for large field trials (Poland and Nelson 2011).

An ideal disease phenotyping solution would be rapid, accurate, and precise. Image-based phenotyping meets these criteria. Numerous image-based plant disease phenotyping methods have been developed in recent years that show greater accuracy, precision, and repeatability than visual assessments (Bock et al. 2008; Stewart and McDonald 2014; Xie et al. 2012). However, these methods typically require destructive sampling or consistent standardized conditions in which to take the images, which is time consuming and does not allow disease progression to be tracked over time. Therefore, a system that would allow disease phenotyping using images taken under natural field conditions is desirable.

The large size of NLB lesions makes this disease an attractive candidate for image-based phenotyping. NLB causes gray-brown necrotic lesions, which can grow and coalesce over the course of a season. Within several weeks of lesion formation, they are typically at least 1 cm wide and 5 cm long. Detecting these large lesions in field photographs, which might be limited in quality or resolution, will likely be more feasible than detection of diseases that chiefly cause streaking (e.g., maize dwarf mosaic virus) or smaller lesions or pustules (e.g., common rust or northern corn leaf spot).

Machine learning techniques have been used to detect, classify, and quantify a wide variety of diseases on many crops (Arnal Barbedo 2013; Singh et al. 2016). In general, prior work on disease detection in plants has focused on analysis of individual leaves photographed under controlled, standardized conditions. Although this represents a notable achievement, the range of potential uses for such techniques is limited by the focus on single leaves and the uniform settings in which they have been photographed. It would be useful to be able to also detect disease in photographs taken in the field, which may feature many plants with highly variable perspectives and lighting conditions.

Improvements in convolutional neural networks (CNNs) in recent years have made them the state of the art among machine learning approaches for addressing computer vision problems, particularly

[†]Corresponding author: C. DeChant; E-mail: dechant@cs.columbia.edu

image classification (LeCun et al. 2015). Computer vision approaches to classification tasks have traditionally required manual selection of features that were thought to be useful in making classification decisions. By contrast, CNNs learn which features are most important. Neural networks are composed of multiple layers of linear transformations (multiplications by a "weight" matrix), each followed by a nonlinear function. The linear transformations are learned during training by making small changes to the weight matrices that progressively make the transformations more helpful to the final classification task.

The learned multilayer processing of visual input that occurs in a CNN is thought to be analogous to how the primate visual system processes information; early stages of the networks respond to basic visual elements such as lines while higher levels of the networks respond to more complicated or abstract visual concepts such as object category (Cadieu et al. 2014). Convolutional layers allow a network to efficiently learn features that are invariant to an exact location in an image by applying the same learned transformation to subsections of an entire image.

In an example of prior work in this area, Mohanty et al. (2016) trained a CNN to classify leaves as belonging to 1 of 14 species and to determine whether they had 1 of 26 diseases. They reported an accuracy of 99.4% when classifying both species and disease on held-out data. However, on images not collected under controlled conditions, their model had accuracies of 31.4 to 31.7% when classifying both species and disease and 41.1 to 54.5% when classifying disease in images of a prespecified species. In images collected from natural settings, many nuisance factors contribute to make the task difficult for a network trained on artificially controlled images, including lighting variations, shadows, and exposed soil.

Reliably distinguishing between NLB lesions and other forms of damaged or senescent leaf tissue is the first step toward estimating NLB incidence from field images. Therefore, we set out to create a system that could detect NLB lesions in diverse field images on par with human experts. To this end, we collected images of maize plants in the field that contained varying levels of NLB infection as well as differing amounts of nontarget features such as soil and sky. A three-stage image analysis pipeline based around CNNs was developed that was able to accurately detect the presence of NLB lesions in the images.

MATERIALS AND METHODS

The experiments consisted of two separate plantings of the Genomes to Fields Initiative's GxE Trial (http://www.genomes2fields.org), both grown on the Musgrave Research Farm in Aurora, NY during the 2015 growing season. Each planting consisted of two replicates, with each replicate containing 250 maize hybrids, including 5 hybrid checks adapted to New York State. One planting was inoculated with NLB as described below while the other planting received no treatment.

In the inoculated planting, each plant was inoculated using both a liquid suspension of *S. turcica* conidia and sorghum grains that were infested with *S. turcica* in a modified version of the protocol from Chung et al. (2010). Cultures of *S. turcica* for liquid suspension were grown on lactose casein hydrolysate agar for 2 to 3 weeks on a cycle of 16 h of light and 8 h of darkness at room temperature. Colonies were scraped from the plates using glass rods and sterile H_2O. Spore concentration was adjusted to 4,000 spores/ml with a hemacytometer and the final suspension was adjusted to 0.02% Tween 20. For solid inoculum, 1,000 ml of sorghum grains and 600 ml of deionized H_2O were soaked overnight in autoclavable spawn bags, sealed at the top with three seals to form an S-bend, and autoclaved for 70 min. Sterilized bags were cut open, inoculated with 1 ml of uncalibrated spore suspension scraped from 2- to 3-week-old cultures of *S. turcica* grown as described above, and resealed. Bags were massaged daily for 3 weeks to disperse fungal tissue and break up clumps. At 38 days after planting (roughly the

V5 to V6 stage), plants were inoculated by pipetting 0.5 ml of liquid spore suspension and dropping one-quarter teaspoon (approximately 1.25 ml) of colonized sorghum grains into the upper plant whorl.

Images of NLB-infected and noninfected leaves were taken with a Canon EOS Rebel or Sony a6000 camera by hand on dates ranging from 28 to 78 days postinoculation (DPI). Altogether, 1,834 images were taken over eight dates. In total, 38 images were excluded due to poor quality. The images were first classified by presence or absence of any visible lesions in the image. Following this, all visible lesions were marked with a line down the main axis of the lesion using the annotation features of the Bisque image-processing platform hosted on CyVerse (formerly iPlant) (Goff et al. 2011) (Fig. 1). Full details of how many images were taken on each date, along with the distribution of those images in the training, validation, and test sets, can be found in Supplementary Table S2.

Images of infected leaves were predominantly taken in the inoculated trial. The noninoculated trial exhibited low levels of natural infection. Therefore, images of noninfected leaves were predominantly taken in the noninoculated trial. To avoid spurious associations between the presence of lesions and trial-wide differences in plant maturity, field condition, or background objects, infected leaves were photographed in the noninoculated trial and noninfected leaves in the inoculated trial whenever possible.

The 1,028 images of infected leaves and 768 images of noninfected leaves were randomly divided such that 70% of the images were used for training (i.e., to fit the models), 15% for validation (i.e., to estimate prediction error for model and hyperparameter selection), and 15% for testing (i.e., to assess error of the final chosen model). All choices involving network architecture and all training were done without consideration of the test set, which was only used at the end to assess the performance of the final, full system.

We developed a three-stage process to analyze the images to determine whether they contained infected leaves. In stage 1, we trained several CNNs to detect the presence of lesions in small patches of the images. Those CNNs were used in the second stage to produce heat maps indicating the probability of infection of each region of the images. The third stage used those heat maps to classify the full images.

In stage 1, the images were broken up into small segments of 224 by 224 pixels. Individual lesions were almost always larger than 224 by 224 pixels, allowing for many image segments to be produced from each lesion. Image segments were generated by moving down the length of the lesion 10 pixels at a time and centering the target segment in a randomly chosen location ±15 pixels from the major axis of the lesion. A random rotation was then applied to the image before the final segmentation, which allowed six slightly different images to be produced from the same area. Nonlesion segments were produced from the noninfected images in a similar fashion from randomly rotated segments drawn from the entire image.

After training one network on the smaller image patches, that network was used to select images to augment the data set, a technique called hard negative mining. Images of noninfected plants were broken up into regions of 224 by 224 pixels using a sliding window approach with a step size of 55. Because the original training images were randomly rotated and nudged up or down, only a small fraction (fewer than 0.1%) of these newly generated segments could have been identical with the first training set. These image segments were fed into the trained neural network. Image segments that were incorrectly classified as containing lesions were then added to the training set. Versions of these images were also added to the training set after they were flipped horizontally and vertically.

Five CNNs were then trained on the newly augmented data set with variations in the architecture and hyperparameters of the networks, the balance of lesion versus nonlesion, and the balance of original nonlesion versus negative mined nonlesion images. Three

of these (classifiers A, B, and C) were used in the final system. Stage 1 classifier A was trained on a roughly equal balance of nonlesion versus lesion images, using only the hard negative mined images for the nonlesion class. Classifiers B and C were trained on both the original nonlesion images and the negatively mined ones. Because early networks produced significantly more false positives than false negatives, classifiers B and C were trained on a mix of lesion versus nonlesion images in which the proportion of nonlesion images was significantly increased; the proportion of nonlesion to lesion images was roughly 6:1.

The five CNNs trained in stage 1 were selected to create heat maps of the full images in stage 2. One set of heat maps was produced from each CNN trained to classify small patches of the images in stage 1. Using a sliding window approach with a step size of 30, sections of the larger images were fed through each of the neural networks. The output was recorded for each location in the images, which can be interpreted as the probability that each section contains a lesion. Each probability was assembled into a matrix in its appropriate place so as to represent the output of a network for each area of the image (Fig. 2).

The third stage trained CNNs to use the heat maps from the previous stage to classify each entire image as containing or not containing infected leaves. Neural networks were trained on various combinations of the heat maps produced in stage 1. The heat maps were "stacked" on top of each other, each occupying a different channel of the input. Input from three heat maps of the same underlying image,

for example, would occupy three channels just as a color image would be composed of three channels (red, green, and blue). They were flipped horizontally and vertically and rotated during training to augment their number. Various combinations of the heat map sets were used for training and a selection of three heat map sets (A, B, and C) was made based on validation set performance.

Both stages involved the use of CNNs trained using the open-source software packages Keras (Chollet 2015) and Theano (Theano Development Team et al. 2016). Theano performs automatic differentiation to allow network parameters to be learned using backpropagation; Keras is a front-end library for using Theano. We randomly initialized the weights of the neural networks following the guidance of He et al. (2015) for layers composed of rectified linear units and Glorot and Bengio (2010) for the final logistic unit layer. A rectified linear unit is equivalent to $y = x$ for $x > 0$ and $y = 0$ for $x < 0$. Hyperparameters were tuned by hand following validation set results. The convolutional layers in stage 3 have a relatively small number of filters per layer ($n = 8$) because, with any more filters, the network quickly overfit the training data. We used batch normalization (Ioffe and Szegedy 2015) in all networks and as a first stage in the stage 1 classifiers. We applied max pooling at several layers of the network; the max pooling operation takes a set of inputs, in this case a two-by-two matrix of numbers, and returns the largest number. We used the Adam optimization algorithm (Kingma and Ba 2014) in stage 1 network A and C and the stage 3 network; RMSprop (Tieleman and Hinton 2012) was used instead in network B in order to add diversity to the training methods used in the

Fig. 1. Images used to train the convolutional neural networks to detect the presence of northern leaf blight (NLB) disease symptoms. **A** and **B**, Typical leaf symptoms of NLB-infected maize plants; lesions marked during training are circled in red. **C** and **D**, Typical images of noninfected plants. Images represent the range of disease symptoms, background, and lighting conditions present.

different stage 1 networks. Training was performed with an NVIDIA Titan X GPU, taking approximately 3 days per first-stage network and 30 min for the third-stage network. At run time, it takes approximately 2 min to generate a heat map for one image (requiring 1.6 GB of memory) and less than a second to classify a set of three heat maps (requiring 800 MB of memory). Details of the CNN are shown in Supplementary Table S1. Code for the CNN is available at https://github.com/chaddech/nlb.

The image data set used for this study is available through the Bisque platform of CyVerse (https://bisque.cyverse.org/client_service/) as the public dataset nlb_annotated_public_2016. Annotation files for lesion location labels are included.

RESULTS

We analyzed 1,796 images of maize leaves, which consisted of 1,028 images of NLB-infected leaves and 768 images of noninfected leaves. There were two image sizes: 6,000 by 4,000 pixels and 5,184 by 3,456 pixels. The images of infected leaves were annotated for NLB lesions with 6,931 lines, or an average of 6.7 lines/image. This was slightly higher than the number of lesions, because lesions that curved due to leaf curvature were annotated with multiple lines. Other sources of senesced leaf tissue were present in both the inoculated and noninoculated trial. These mainly included physical damage, natural senescence of lower leaves, nitrogen deficiency, feeding from corn flea beetles, and other foliar diseases, particularly northern corn leaf spot. Lesions on inoculated plants were typical of those present in natural infections, and lesion color and shape were generally comparable between the inoculated and noninoculated plantings.

Initial attempts, before beginning on three-stage approach described in this article, consisted of training CNNs on the full images at scaled-down resolutions of 600 by 400 or 300 by 200. This approach never achieved accuracies on the validation set greater than 70%.

The final three-stage architecture described above (Fig. 3) resulted in much better performance (Table 1). The stage 1 networks selected for the final system reached accuracies of 81, 95, and 94% in classifying small image segments from the validation set. The final stage 3 network was trained on various combinations of the heat maps produced using networks that were trained in stage 1. When those heat maps were combined, the stage 3 network achieved 97.8% accuracy on the validation set.

The validation set was used to guide training by helping to make the choice of architecture of the final system as well as determining which heat maps were used and what values were chosen for various hyperparameters in the neural networks. Therefore an estimate of the final error of the system based on images in the validation set would be expected to be biased in the optimistic direction because the system was tuned to reduce validation set error (Abu-Mostafa et al. 2012). In order to have an accurate understanding of how the system would perform on new images, we kept a test set of 272 images completely unused throughout the entire training procedure. Because the test set did not influence the system in any way, an estimate of error based on it is expected to be unbiased. On this held-out test set, the network achieved an accuracy (number of correctly classified images divided by total number of images) of 96.7%, 96.8% precision (number of true positives [i.e., truly diseased] divided by the number of true positives plus false positives), 97.4% recall (number of true positives divided by the number of true positives plus the number of false negatives), and an F1 score (2 × precision × recall, all divided by precision plus recall) of 0.971. A confusion matrix, which shows a breakdown of errors on the test set, is provided in Table 2. Furthermore, the area under the receiver operating characteristic curve was found to be 0.994 (Fig. 4).

Several issues posed challenges to successfully classifying the images. The first was the small number of images to train on; successful application of deep learning techniques typically involves larger training sets, on the order of tens of thousands. Another factor contributing to the task's difficulty lay in the nature of the images themselves. Many types of dead leaf tissue, including natural senescence, can closely resemble NLB lesions to both a CNN and the untrained eye. Variation in lighting, a common issue for images taken in the field, also presented problems. Areas of shadow or, conversely, bright light appearing on a leaf were often mistaken by networks early in training as lesions; they were well represented in the false positives found during hard negative mining. Leaves in the background, dead leaf tissue on the ground, senescing leaves in the lower canopy, and insects also presented challenges (Fig. 5).

One significant benefit of the three-stage pipeline was the ability of the system to make use of the full-resolution images. Compared with scaling the images down, cropping them into smaller full-resolution sections in the first stage allowed the network to make use of the fine-grained detail that distinguishes an NLB lesion from other brown spots or dead tissue. On their own, though, the small segment predictions
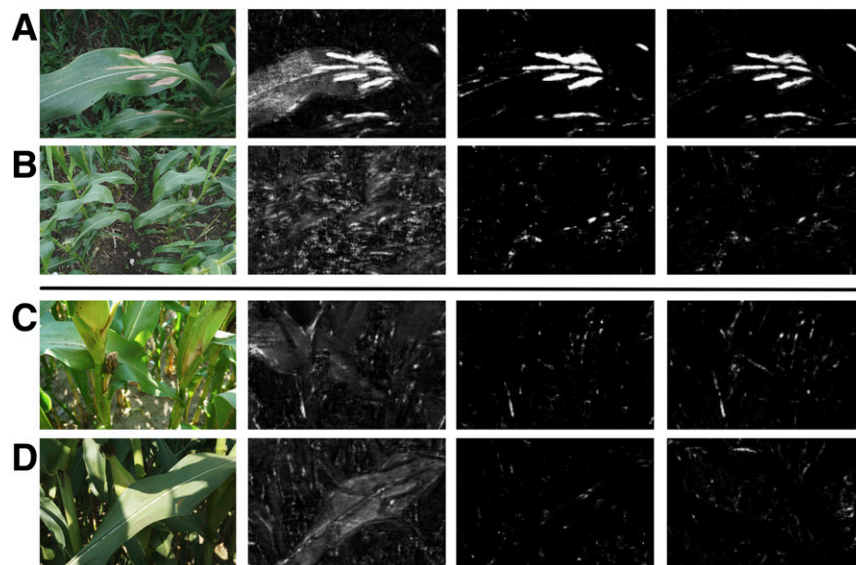

Fig. 2. A and **B**, Examples of images and their associated heat maps for images containing lesions. **C** and **D,** Images and heat maps for images not containing lesions. Heat maps are shaded to indicate the stage 1 network's output score the probability of an image segment containing a lesion. A probability of one corresponds to white while zero is represented by black, with intermediate values shown in shades of gray. From left to right columns, the heat maps are produced from stage 1 convolutional neural networks A, B, and C, respectively.
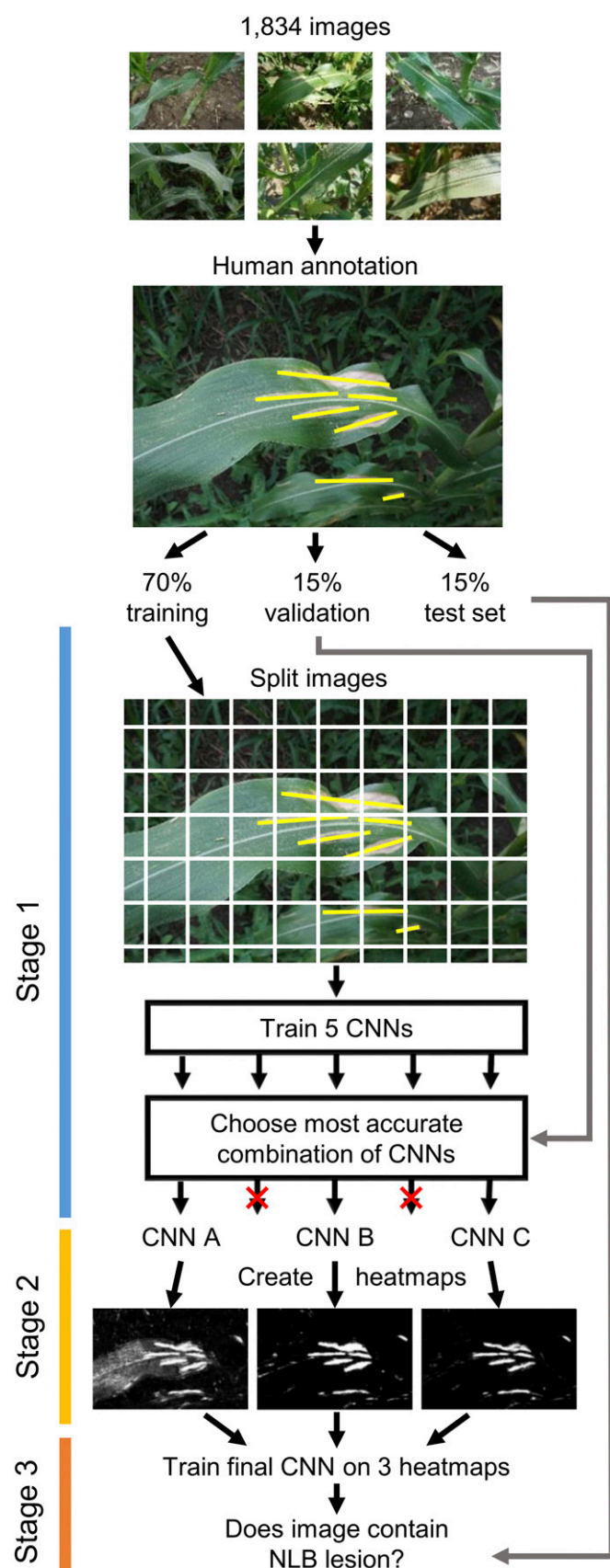
1,834 images

Human annotation

70% training   15% validation   15% test set

Split images

Train 5 CNNs

Choose most accurate combination of CNNs

CNN A   CNN B   CNN C

Create   heatmaps

Train final CNN on 3 heatmaps

Does image contain NLB lesion?

Stage 1

Stage 2

Stage 3

Fig. 3. Three stages of the classification pipeline. In stage 1, three convolutional neural networks (CNNs) are trained on subimages. In stage 2, heat maps are generated for the whole images using each of the three CNNs individually. In stage 3, a CNN is trained that takes the heat maps as input and outputs the probability of the whole image containing diseased tissue.

actually presented a problem because of their sheer number. Because the heat maps (matrices of 126 by 193 values) contained scores for 24,318 such segments, even a highly accurate classifier would have many errors in its scores for an entire image. The best stage 1 network achieved an accuracy of 94%; thus, we would expect over 1,000 incorrectly classified segments in every one of the heat maps. This is why the stage 3 classifier was necessary; it learned how to combine all of the local segment scores, including inaccurate ones, into a global classification, achieving 96.7% accuracy on whole images.

DISCUSSION

Ensemble methods that combine the output of different classifiers often show improved performance over a single classifier. We achieved the best result with a combination of three of the stage 1 networks, so that the system benefited from this effect. However, even when we used only one network's heat maps in the third stage, we still observed significant improvement over the initial, baseline network, which took as input scaled-down versions of the full images. Therefore, the three-stage system's improved performance was primarily due to other factors. Neural network performance is greatly affected by the amount of data available for training. Because we trained the first stage on small sections of the images instead of the full image, the training set size was effectively increased, at least for the networks in stage 1. One lesion might be broken up into many small regions, for example.

The need to break up the images and process them in stages arose, in part, because of memory constraints; we cannot feed reasonably sized batches of full-resolution images into a CNN with a reasonable number of convolutional filters per layer. If we could, making end-to-end training of the entire pipeline feasible from full-resolution image to final classification, it is possible that performance could be improved. It also possible, though, that the production of heat maps independent of the final classifier introduced a kind of regularization and that an end-to-end system could more easily overfit the training set. This would be a particular danger for training sets as small as this one. Future work would benefit from additional data.

On-site diagnosis will require the ability to detect the presence of disease symptoms in images that are not ideal and contain many potentially confounding factors. Because machine learning techniques typically do not perform as well on data that are significantly different from that on which they were trained, it is likely that classifiers will have to be trained on images taken in similar conditions in the field.

A significant drawback of the current method is the time needed to manually classify images required to train the CNN. However,

TABLE 1. Stage 1 and stage 3 network accuracies

| Heat maps | Network validation set accuracy while producing heat maps | |
	Stage 1	Stage 3
A	0.809	0.882
B	0.952	0.908
C	0.941	0.897
B, C	0.952, 0.941	0.959
A, B, C[a]	0.809, 0.952, 0.941	0.978

[a] Test set accuracy = 0.967.

TABLE 2. Confusion matrix

| Actual | Predicted | |
	Nonlesion	Lesion
Nonlesion	113	5
Lesion	4	150

this is a one-time investment; once the CNN is trained, it can be used on new images taken in similar conditions without further training. It does not need specialized hardware to run; thus, the network and the expertise it represents can be shared freely among different users of the technology.

The current system can detect the presence or absence of a disease in an image, information most readily used for estimating disease incidence. This may be useful for growers looking for early detection or breeders evaluating incubation period for a given disease. However, for most applications, both growers and breeders will additionally need to quantify disease severity. This information could be extracted from the heat maps used for detection but the reliability of this method remains to be seen. Factors such as distance from the camera, leaf angle, and the amount of nonplant features within the image present challenges for calculating the proportion of diseased versus healthy plant tissue within an image. Further experiments have been planned to test the replicability and accuracy of quantification, in the same way that detection has been tested here.
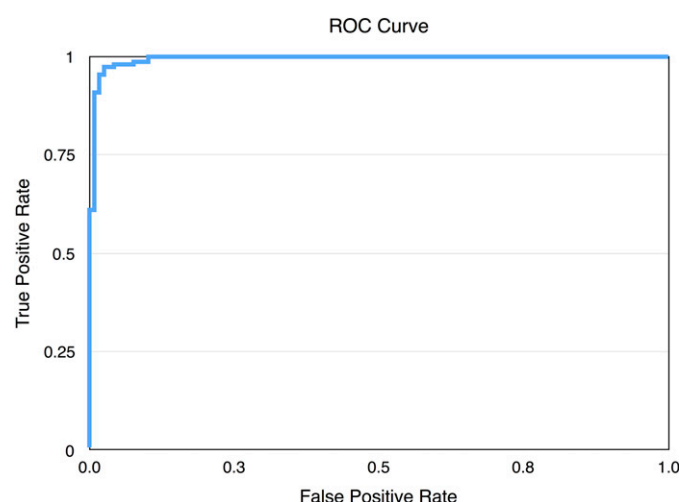

Fig. 4. Receiver operating characteristic (ROC) curve shows the relationship between the percentage of true positives (lesion images classified as containing lesions) and false positives (nonlesion images classified as containing lesions) when varying the classifier's threshold for the lesion/nonlesion decision on test set images.
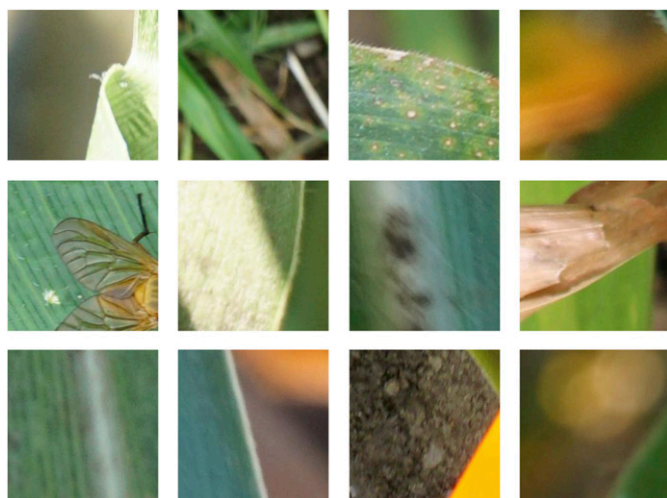

Fig. 5. Examples of image subsections that were mistakenly classified as containing lesions by the initial stage 1 classifier; these images were added to the data set along with other false positives before training the stage 1 classifiers in the final system.

We believe that this approach can be applied to other plant diseases with visible symptoms that can be recognized by human experts. As shown by Mohanty et al. (2016), when using images that were of single leaves taken on a homogenous background, CNNs could be trained to recognize many different diseases on diverse plant hosts with high accuracy. We have shown that highly accurate CNNs can be trained on field images with highly variable backgrounds, angles, fields of view, and lighting. This approach requires more human time investment to generate training data, because labeling individual lesions is more time consuming than classifying whole images as containing or not containing lesions. However, it requires less time investment during the field season, because infected leaves can be photographed from any angle without the need to place them on a homogenous background.

Training CNNs on field images of more diseases is the first step toward being able to confidently distinguish these diseases or damage in the field with an automated system. Plant breeding trials or genetic experiments that are artificially inoculated may be highly infected with only a single disease, as ours were, making misclassification of other diseases a relatively rare occurrence. For most other cases, however, accurately classifying multiple diseases will be necessary.

The images used are representative of those that could be captured from a ground vehicle as it moves through a field trial. Several ground-based phenotyping vehicles equipped with a range of different sensors have previously been developed (Andrade-Sanchez et al. 2014; Deery et al. 2014). Due to the high payload capacity of most of these vehicles, the addition of a further system trained to recognize disease is highly feasible. A further application would be to mount the system on an aerial vehicle. Small unmanned aircraft systems (sUAS) technology has progressed dramatically in recent years and shows great potential as a crop monitoring system (Chapman et al. 2014; Sankaran et al. 2015). Coupled with autonomous or semiautonomous navigation, sUAS platforms could provide measures of disease in the field with greater accuracy and the same or less human input than current visual disease monitoring. We are currently working on training CNNs to recognize NLB and other plant phenotypes from sUAS imagery. Such a system has the potential to speed up genetic improvement of disease resistance traits in plant breeding programs by allowing more lines to be screened more accurately in less time and, therefore, more frequently.

In a production setting, this system could be coupled with real-time variable-rate fungicide applicators. Such applicators feed measured crop parameters into a decision support system to gauge the required fungicide dosage. This limits fungicide application rates in areas where it is less needed, with the dual benefit of reducing fungicide usage and runoff and saving money for growers. Systems that adjust fungicide application systems to the level of canopy coverage have been developed and tested by Dammer et al. (2009) and Tackenberg et al. (2016). Both groups noted that real-time disease detection would provide the ideal data for estimating optimal application rates but that this was not yet feasible.

The application of deep learning to plant disease phenotyping has the potential to overcome the shortcomings of existing image-based phenotyping techniques while increasing the accuracy of conventional visual disease assessment. Integrating such technology into research and breeding programs has the capability to speed up fundamental research and development of plant varieties with improved disease resistance. Ultimately, deployment of the technology on autonomous vehicles could help address the current challenge placed on food production by plant diseases.

ACKNOWLEDGMENTS

for this research. The image data set used for this study is available through the Bisque platform of CyVerse (https://bisque.cyverse.org/client_service/) as the public data set 'nlb_annotated_public_2016'. This data set includes 1,017 annotated images with lesions and 768 files with no lesions.

LITERATURE CITED

Abu-Mostafa, Y. S., Magdon-Ismail, M., and Lin, H.-T. 2012. Learning from Data USA. Online publication AMLBook. http://amlbook.com/support.html

Andrade-Sanchez, P., Gore, M. A., Heun, J. T., Thorp, K. R., Carmo-Silva, A. E., French, A. N., Salvucci, M. E., and White, J. W. 2014. Development and evaluation of a field-based high-throughput phenotyping platform. Funct. Plant Biol. 41:68-79.

Arnal Barbedo, J. G. 2013. Digital image processing techniques for detecting, quantifying and classifying plant diseases. Springerplus 2:660.

Bock, C. H., Parker, P. E., Cook, A. Z., and Gottwald, T. R. 2008. Visual rating and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves. Plant Dis. 92:530-541.

Bock, C. H., Parker, P. E., Cook, A. Z., Riley, T., and Gottwald, T. R. 2009. Comparison of assessment of citrus canker foliar symptoms by experienced and inexperienced raters. Plant Dis. 93:412-424.

Brewster, V. A., Carson, M. L., and Wicks, Z. W. 1992. Mapping components of partial resistance to northern leaf blight of maize using reciprocal translocation. Phytopathology 82:225-229.

Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., and DiCarlo, J. J. 2014. Deep neural networks rival the representation of primate IT cortex for core visual object recognition. PLOS Comput. Biol. 10:e1003963.

Chapman, S. C., Merz, T., Chan, A., Jackway, P., Hrabar, S., Dreccer, M. F., Holland, E., Zheng, B., Ling, T. J., and Jimenez-Berni, J. 2014. Pheno-Copter: A low-altitude, autonomous remote-sensing robotic helicopter for high-throughput field-based phenotyping. Agronomy (Basel) 4:279-301.

Chollet, F. 2015. Keras. Online publication. https://github.com/fchollet/keras

Chung, C.-L., Longfellow, J. M., Walsh, E. K., Kerdieh, Z., Van Esbroeck, G., Balint-Kurti, P., and Nelson, R. J. 2010. Resistance loci affecting distinct stages of fungal pathogenesis: Use of introgression lines for QTL mapping and characterization in the maize-*Setosphaeria turcica* pathosystem. BMC Plant Biol. 10:103.

Dammer, K.-H., Thöle, H., Volk, T., and Hau, B. 2009. Variable-rate fungicide spraying in real time by combining a plant cover sensor and a decision support system. Precis. Agric. 10:431-442.

Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X., and Furbank, R. 2014. Proximal remote sensing buggies and potential applications for field-based phenotyping. Agronomy (Basel) 4:349-379.

Glorot, X., and Bengio, Y. 2010. Understanding the difficulty of training deep feedforward neural networks. Pages 249-256 in: Proc. 13th AISTATS.

Goff, S. A., Vaughn, M., McKay, S., Lyons, E., Stapleton, A. E., Gessler, D., Matasci, N., Wang, L., Hanlon, M., Lenards, A., Muir, A., Merchant, N., Lowry, S., Mock, S., Helmke, M., Kubach, A., Narro, M., Hopkins, N., Micklos, D., Hilgert, U., Gonzales, M., Jordan, C., Skidmore, E., Dooley, R., Cazes, J., McLay, R., Lu, Z., Pasternak, S., Koesterke, L., Piel, W., Grene, R., Noutsos, C., Gendler, K., Feng, X., Tang, C., Lent, M., Kim, S.-J., Kvilekval, K., Manjunath, B. S., Tannen, V., Stamatakis, A., Sanderson, M., Welch, S., Cranston, K., Soltis, P., Soltis, D., O'Meara, B., Ane, C., Brutnell, T., Kleibenstein, D. J., White, J. W., Leebens-Mack, J., Donoghue, M. J., Spalding, E. P., Vision, T. J., Myers, C. R., Lowenthal, D., Enquist, B. J., Boyle, B., Akoglu, A., Andrews, G., Ram, S., Ware, D., Stein, L., and Stanzione, D. 2011. The iPlant Collaborative: Cyberinfrastructure for plant biology. Front. Plant Sci. 2:34.

He, K., Xhang, X., Ren, S., and Sun, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. Online publication. arXiv:1502.01852. https://arxiv.org/abs/1502.01852

Ioffe, S., and Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Int. Conf. Machine Learning.

Kingma, D., and Ba, J. 2014. Adam: A method for stochastic optimization. Online publication. arXiv:1412.6980. https://arxiv.org/abs/1412.6980

LeCun, Y., Bengio, Y., and Hinton, G. 2015. Deep learning. Nature 521:436-444.

Mohanty, S. P., Hughes, D., and Salathe, M. 2016. Inference of plant diseases from leaf images through deep learning. Front. Plant Sci. 7:1419.

Mueller, D. S., Wise, K. A., Sisson, A. J., Allen, T. W., Bergstrom, G. C., Bosley, D. B., Bradley, C. A., Broders, K. D., Byamukama, E., Chilvers, M. I., Collins, A., Faske, T. R., Friskop, A. J., Heiniger, R. W., Hollier, C. A., Hooker, D. C., Isakeit, T., Jackson-Ziems, T. A., Jardine, D. J., Kelly, H. M., Kinzer, K., Koenning, S. R., Malvick, D. K., McMullen, M., Meyer, R. F., Paul, P. A., Robertson, A. E., Roth, G. W., Smith, D. L., Tande, C. A., Tenuta, A. U., Vincelli, P., and Warner, F. 2016. Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015. Plant Health Prog. 17:211-222.

Oerke, E.-C., and Dehne, H.-W. 2004. Safeguarding production-losses in major crops and the role of crop protection. Crop Prot. 23:275-285.

Pauli, D., Chapman, S. C., Bart, R., Topp, C. N., Lawrence-Dill, C. J., Poland, J., and Gore, M. A. 2016. The quest for understanding phenotypic variation via integrated approaches in the field environment. Plant Physiol. 172: 622-634.

Perkins, J., and Pedersen, W. 1987. Disease development and yield losses associated with northern leaf blight on corn. Plant Dis. 71:940-943.

Poland, J. A., and Nelson, R. J. 2011. In the eye of the beholder: The effect of rater variability and different rating scales on QTL mapping. Phytopathology 101:290-298.

Raymundo, A., and Hooker, A. 1981. Measuring the relationship between northern corn leaf blight and yield losses. Plant Dis. 65:325-327.

Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., Miklas, P. N., Carter, A. H., Pumphrey, M. O., Knowles, N. R., and Pavek, M. J. 2015. Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. Eur. J. Agron. 70: 112-123.

Shankara, K., and Gowda, K. 2011. Assessment of yield loss due to turcicum leaf blight of maize caused by *Exserohilum turcicum*. Mysore J. Agric. Sci. 45:691-692.

Sherwood, R. T., Berg, C. C., Hoover, M. R., and Zeiders, K. E. 1983. Illusions in visual assessment of Stagonospora leaf spot of orchardgrass. Phytopathology 73:173-177.

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

Stewart, E. L., and McDonald, B. A. 2014. Measuring quantitative virulence in the wheat pathogen *Zymoseptoria tritici* using high-throughput automated image analysis. Phytopathology 104:985-992.

Tackenberg, M., Volkmar, C., and Dammer, K.-H. 2016. Sensor-based variable-rate fungicide application in winter wheat. Pest Manage. Sci. 72: 1888-1896.

Theano Development Team, Al-Rfou, R., Alain, G., Almahairi, A., Angemueller, C., Bahdanau, D., Ballas, N., Bastien, F., Bayer, J., Belikov, A., Belopolsky, A., Bengio, Y., Bergeron, A., Bergstra, J., Bisson, V., Bleecher Snyder, J., Bouchard, N., Boulanger-Lewandowsky, N., Bouthillier, X., de Brébisson, A., Breuleux, O., Carrier, P.-L., Cho, K., Chorowski, J., Christiano, P., Cooijmans, T., Côté, M.-A., Côté, M., Courville, A., Dauphin, Y. N., Delalleau, O., Demouth, J., Desjardins, G., Dieleman, S., Dinh, L., Ducoffe, M., Dumoulin, V., Ebrahimi Kahou, S. Erhan, D., Fan, Z., Firat, O., Germain, M., Glorot, X. Theano Development Team, and 70 additional authors. 2016. Theano: A Python framework for fast computation of mathematical expressions. Online publication. arXiv:1605.02688. https://arxiv.org/abs/1605.02688

Tieleman, T., and Hinton, G. 2012. RMSProp: Divide the gradient by a running average of its recent magnitude. Online publication. Coursera: Neural Networks for Machine Learning, Lecture 6.5. https://www.coursera.org/learn/neural-networks/lecture/YQHki/rmsprop-divide-the-gradient-by-a-running-average-of-its-recent-magnitude

Wang, P., Souma, K., Kobayashi, Y., Iwabuchi, K., Sato, C., and Masuko, T. 2010. Influences of northern leaf blight on corn silage fermentation quality, nutritive value and feed intake by sheep. Anim. Sci. J. 81:487-493.

Welz, H. G., and Geiger, H. H. 2000. Genes for resistance to northern corn leaf blight in diverse maize populations. Plant Breed. 119:1-14.

Xie, W., Yu, K., Pauls, K. P., and Navabi, A. 2012. Application of image analysis in studies of quantitative disease resistance, exemplified using common bacterial blight-common bean pathosystem. Phytopathology 102: 434-442.