Received: 12 February 2024 | Accepted: 17 May 2024

DOI: 10.1002/ppj2.20110

The Plant Phenome Journal OPEN ACCESS

ORIGINAL ARTICLE

Quantifying leaf symptoms of sorghum charcoal rot in images of field-grown plants using deep neural networks

Emmanuel M. Gonzalez[1] | **Ariyan Zarei**[2] | **Sebastian Calleja**[1] | **Clay Christenson**[1] | **Bruno Rozzi**[1] | **Jeffrey Demieville**[1] | **Jiahuai Hu**[1] | **Andrea L. Eveland**[3] | **Brian Dilkes**[4] | **Kobus Barnard**[2] | **Eric Lyons**[1] | **Duke Pauli**[1,5]

[1]School of Plant Sciences, University of Arizona, Tucson, Arizona, USA

[2]Department of Computer Science, University of Arizona, Tucson, Arizona, USA

[3]Donald Danforth Plant Science Center, St. Louis, Missouri, USA

[4]Department of Biochemistry, Purdue University, West Lafayette, Indiana, USA

[5]Center for Agroecosystem Research in the Desert (ARID), Tucson, Arizona, USA

Correspondence
Duke Pauli, School of Plant Sciences, University of Arizona, Tucson, AZ, USA.
Email: dukepauli@arizona.edu

Assigned to Associate Editor Sara Tirado Tolosa.

Funding information
Cotton Incorporated, Grant/Award Numbers: 18-384, 20-720, 21-830, 23-890; Department of Energy, Biological and Environmental Research, Grant/Award Number: DE-SC0020401; National Science Foundation Plant Genome Research Program, Grant/Award Numbers: IOS–2102120, IOS–1849708, IOS–2023310; National Science Foundation Division of Biological Infrastructure, Grant/Award Numbers: DBI-1743442, DBI-2019674; Advanced Research Projects Agency - Energy, Grant/Award Number: DE-AR0001101

Abstract

Charcoal rot of sorghum (CRS) is a significant disease affecting sorghum crops, with limited genetic resistance available. The causative agent, *Macrophomina phaseolina* (Tassi) Goid, is a highly destructive fungal pathogen that targets over 500 plant species globally, including essential staple crops. Utilizing field image data for precise detection and quantification of CRS could greatly assist in the prompt identification and management of affected fields and thereby reduce yield losses. The objective of this work was to implement various machine learning algorithms to evaluate their ability to accurately detect and quantify CRS in red-green-blue images of sorghum plants exhibiting symptoms of infection. EfficientNet-B3 and a fully convolutional network emerged as the top-performing models for image classification and segmentation tasks, respectively. Among the classification models evaluated, EfficientNet-B3 demonstrated superior performance, achieving an accuracy of 86.97%, a recall rate of 0.71, and an F1 score of 0.73. Of the segmentation models tested, FCN proved to be the most effective, exhibiting a validation accuracy of 97.76%, a recall rate of 0.68, and an F1 score of 0.66. As the size of

Abbreviations: AI, artificial intelligence; AZMET, Arizona Meteorologial Network; BCE, binary cross entropy; CNN, convolutional neural network; CRF, conditional random field; CRS, charcoal rot of sorghum; EMS, ethyl methanesulfonate; FCN, fully convolutional network; FN, false negative; FP, false positive; GPU, graphics processing unit; IOU, intersection over union; ITS, internal transcribed spacer; MAC, Maricopa Agricultural Center; ML, machine learning; NLB, northern leaf blight of corn; NN, neural network; PDA, potato dextrose agar; R-CNN, region-based convolutional neural network; RGB, red-green-blue; SVWC, soil volumetric water content; TN, true negative; TP, true positive; UAV, unoccupied aerial vehicle; WL, water-limited; WW, well-watered.

Emmanuel Gonzalez and Ariyan Zarei contributed equally to this work.

the image patches increased, both models' validation scores increased linearly, and their inference time decreased exponentially. This trend could be attributed to larger patches containing more information, improving model performance, and fewer patches reducing the computational load, thus decreasing inference time. The models, in addition to being immediately useful for breeders and growers of sorghum, advance the domain of automated plant phenotyping and may serve as a foundation for drone-based or other automated field phenotyping efforts. Additionally, the models presented herein can be accessed through a web-based application where users can easily analyze their own images.

1 | INTRODUCTION

The impacts of plant biotic stresses pose a significant risk to sustainable agricultural production and threaten the availability of nutritious calories to a growing world population. Globally, plant diseases are directly responsible for yield losses ranging from 10% to 40% across major staple crops that provide ~50% of the calorie intake among humans (Savary et al., 2019). Further compounding cropping system challenges created by biotic stressors is the presence and interaction with abiotic stresses such as heat and drought (Desaint et al., 2021; Pandey et al., 2015; Ramegowda & Senthil-Kumar, 2015; H. Zhang & Sonnewald, 2017). Collectively, these factors further exacerbate the challenges facing crop production and highlight the need for resistant crop cultivars capable of mitigating these factors (Kissoudis et al., 2014). This urgent need for improved cultivars is further highlighted by the aridification of agricultural lands which will increase and intensify the effects of abiotic and biotic stresses (Overpeck & Udall, 2020); these environmental changes will likely disrupt and alter the geographic distribution and abundance of crop pathogens (Chakraborty et al., 2000; Delgado-Baquerizo et al., 2020; Newbery et al., 2016).

Traditional plant breeding methods have relied on the visual assessment and scoring of germplasm subjected to infection, either naturally or artificially inoculated, to identify genotypes with varying levels of disease resistance (Bernardo, 2014; St Clair, 2010). However, visual assessment by trained experts is subject to human biases and errors that reduce the precision, accuracy, and repeatability of disease rating resulting in decreased selection accuracy, heritability, and genetic gain (Bock et al., 2009; Poland & Nelson, 2011), which, in turn, lengthens the development of cultivars that can cope with biotic and abiotic stresses. To overcome these limitations, machine learning (ML) and artificial intelligence (AI) algorithms, in conjunction with high-throughput phenotyping, can be leveraged to conduct automated and rapid assessment of images of diseased plants/tissue to provide more accurate and reliable scoring of relevant plant germplasm (DeChant et al., 2017; Lu et al., 2021; Pauli et al., 2016; Singh et al., 2016). These data can aid in breeding by enabling direct selection or by identifying genes or alleles conferring pathogen resistance. They also allow growers to detect disease and plant stress early, facilitating prompt mitigation.

Macrophomina phaseolina (Tassi) Goid is a pluriparous plant pathogen impacting over 500 plant species in more than 100 plant families, including cereals, legumes, vegetables, and fruits throughout the world (Kunwar et al., 1986; Marquez et al., 2021). *M. phaseolina* is a necrotrophic soilborne fungus native to the Sonoran Desert soil (Mihail et al., 1989, 1992) but is also widely distributed in the United States. It survives and spreads primarily as black microsclerotia in diseased root and stem debris as well as in soil after decay of infected plant material (Bhattacharya & Samaddar, 1976). Microsclerotia infection of root tissue occurs at temperatures ranging from 20 to 40°C, affecting plants at different developmental stages, including seedling, young, and mature phases (Collins et al., 1991; Hsi, 1956). Disease development is influenced largely by drought stress and low soil moisture (Odvody & Dunkle, 1979). Under favorable environmental conditions, *M. phaseolina* invades the vascular system, disrupting the normal function of water and nutrient transport to the leaves, causing visible symptoms such as wilting and premature leaf death. Symptom onset varies with host species, cultivar, growth stage, and environment. Seedlings can be infected early, but symptoms typically appear during drought stress or physiological stress at later stages such as pollination or grain filling. Initial symptoms include wilted grayish-white leaves, reduced vigor, and scattered premature plant death. However, these symptoms are not unique to *M. phaseolina* and can be mistaken for drought stress, frost damage, early senescence, root rot, and Fusarium rot.

Sorghum [*Sorghum bicolor* (L.) Moench] is a key cereal crop that also serves as a host species for *M. phaseolina*. Globally, sorghum is the fifth most widely grown cereal crop, trailing maize, rice, wheat, and barley, and is a staple food crop for millions living in semi-arid regions (Hossain et al., 2022). Sorghum is gaining popularity given its innate ability

to produce grain on marginal land and areas that lack sufficient irrigation water needed for row crops such as maize, and because of its diverse utility as a food, feed, and biofuel crop (Ndlovu et al., 2021; Rai et al., 2016; Rooney et al., 2007; Tang et al., 2018; Yang et al., 2022). The range of sorghum's end uses combined with its innate genotypic and phenotypic diversity is helping to position this crop as a sustainable solution to agricultural production in the face of climate change (Boatwright et al., 2022; Chadalavada et al., 2021). However, for sorghum production acreage to increase, resistance to charcoal rot of sorghum (CRS, *M. phaseolina* infection in sorghum) is needed as CRS causes a variety of symptoms including root rot, soft stalk, early lodging of plants, premature drying of stalk, reduced head size, and poor filling of grain (Hsi, 1956). Charcoal rot symptoms and agronomic impacts significantly lower crop performance and profitability. Sorghum typically grows in arid regions, which are susceptible to water stress, leading to both biotic and abiotic stresses that reduce yield. These stresses cause similar visible symptoms, making it difficult to distinguish between them.

An automated tool that can distinguish between CRS and drought stress in sorghum leaves would be beneficial for researchers and crop managers in dry regions. This tool could be image based and deployable on unoccupied aerial vehicles (UAVs) or smartphones. Utilizing hardware and software advancements to create lightweight algorithms for edge devices would allow for in-field data processing. With the advent of powerful computation devices, specifically graphical processing units (GPUs), and recent advancements in ML and AI algorithms, many domain scientists have started to use neural networks (NNs) to detect, locate, and quantify many features, including disease, in various image modalities (Amsaveni & Albert Singh, 2013; Mechria et al., 2019; Siar & Teshnehlab, 2019). The class of NNs that are widely applied to this application is known as convolutional neural networks (CNNs) (LeCun, 2015). By learning useful features from image data automatically, these CNNs can perform classification or detection tasks more effectively than the older approach of manual feature extraction (Rybski et al., 2010). The CNNs have also been used for the detection, classification, and segmentation of similar foliar diseases in other crop species. In DeChant et al. (2017), the authors proposed a computational pipeline of CNNs for classifying images of field-grown maize images to determine the presence or absence of maize northern leaf blight (NLB). Expanding on this work, Stewart et al. (2019) trained a mask region-based convolutional neural network (mask R-CNN) for segmenting maize images acquired by UAVs into healthy and NLB-affected tissues. In Wu et al. (2019), the authors proposed a sliding window approach for generating heat maps highlighting regions in the aerial images of maize fields affected by the NLB. Wu et al. (2019) and Wiesner-Hanks et al. (2019) proposed a combination of CNNs and conditional random fields (CRFs) to segment maize images into normal and NLB-affected regions. For early plant disease detection, hyperspectral imaging is being increasingly utilized as an alternative to RGB images, given its ability to offer unique spectral signatures and valuable insights into plant health (Mertens et al., 2021; Yu et al., 2018; G. Zhang et al., 2022). However, its complexity and high computational demands present challenges, hindering its full implementation in agriculture (Cheshkova, 2022; Okyere et al., 2023). Meanwhile, standard RGB cameras simplify the process and broaden the application of disease detection methods from research to commercial agriculture.

Core Ideas

- Automated phenotyping tools are required for the efficient detection and quantification of charcoal rot of sorghum.
- Classification and segmentation models can distinguish between concurrent plant stresses with similar symptoms.
- Larger image patch sizes generally improve model performance and reduce processing time.

In the present study, we propose two CNN-based approaches for quantifying CRS in RGB images captured of field-grown sorghum under drought stress conditions and compare the performance of the two methods. The first approach, which involves a set of classification models, determines the presence of CRS in images by classifying small-sized patches. The second approach carries out pixel-wise classification or semantic segmentation on the images. We evaluate the performance of these two approaches to determine which is more capable of detecting and quantifying CRS as well as provide computational benchmarks for deployment by end users. We also provide a high-quality labeled dataset for classification and segmentation tasks as well as our code for the benefit of other researchers to improve upon our work. To reduce the barrier to entry, a web application is provided that allows end users to deploy all models in the present study.

2 | MATERIALS AND METHODS

2.1 | Data and field experiments

A population of ethyl methanesulfonate (EMS)-mutagenized BTx623 sorghum, the genotype used for the generation of the sorghum reference genome, consisting of 430 individuals

(Addo-Quaye et al., 2018; Paterson et al., 2009) was evaluated at the Maricopa Agricultural Center (MAC) of the University of Arizona located in Maricopa, AZ (33°04′37″ N, 111°58′26″ W, elevation 358 m) in 2020. EMS, an organosulfur compound, was used as a chemical mutagen to induce random point mutations primarily through guanine alkylation (Addo-Quaye et al., 2018; Yan et al., 2021). The population was grown under contrasting irrigation conditions representing well-watered (WW) and water-limited (WL) conditions. The trial was planted on June 17 (day 169, Julian calendar) in a partially replicated, incomplete block design with 240 of the lines replicated within each of the two irrigation treatments, while 190 lines were only observed once per irrigation treatment. The order of entries within each irrigation treatment was randomized. To reduce edge effects, the wild-type BTx623 sorghum cultivar was planted at the perimeter of each irrigation treatment. Experimental units were one-row plots, 3.5 m in length with a 0.5-m alley at the end of each plot and inter-row spacing of 0.76 m; plots were thinned to a density of five plants per plot (1 plant per 0.7 linear meter) after crop establishment, at approximately V5 growth stage. The soil type is a Casa Grande sandy loam (fine-loamy, mixed, superactive, hyperthermic Typic Natrargids). Conventional sorghum cultivation practices (fertilizer application rate/amount, weed/insect control, etc.) for the desert Southwest were employed (Ottman, 2016). Meteorological data were obtained from an automated Arizona Meteorological Network weather station (cals.arizona.edu/AZMET/06.htm) located on the premise of MAC and 738 m from the field (Brown, 1989). Crop irrigation was performed using subsurface drip irrigation with pressure compensated drip tape (DripNet PC; Netafim) buried at a depth of ∼0.15 m beneath the soil surface, directly underneath the plants. Soil volumetric water content (SVWC) was monitored on a biweekly basis using a field-calibrated neutron moisture probe (Model 503; Campbell Pacific Nuclear) with measurements taken in 0.2-m increments from a depth of 0.1– 1.9 m. Each sorghum genotype was evaluated under two irrigation treatments: 23% SVWC (WW conditions) and 15% SVWC (WL conditions). These values were based on measures of soil water-holding capacity at the study site and observation of plants, with those at 15% SVWC maintaining core function but showing stress via measures of leaf gas exchange and photosynthetic efficiency (Thorp et al., 2017; Pauli, unpublished data).

2.2 | Disease presence and description

Symptomatic disease tissue was observed in the field trial ∼5 weeks after planting with plants exhibiting typical signs of wilting and root rot (Figure 1). Initially, about 15% of research plots showed severe wilting symptoms; however, this slowly increased as the season progressed, and ultimately ∼70% of the research plots exhibited some level of infection. The presence of similar symptoms was also confirmed previously in 2019 and 2017 within the same field as used for the present work. To confirm pathogen presence and identity, symptomatic plants, with visible leaf symptoms, were dug up from the field plots ensuring that roots remained attached to the plant while avoiding damage to the roots. All sampled plants were placed in plastic bags and kept cool until they could be transported to the University of Arizona's Extension Plant Pathology lab in Tucson, AZ.

Since microsclerotia are not visible in the stem, isolation was made using putatively infected root and stem tissues (Figure 1C,D). Tissue samples, measuring 7 × 7 × 3 mm, were cut from the margin between the diseased and seemingly healthy tissues. These tissue samples were surface sterilized by soaking in 75% ethanol for 5 s, 1% sodium hypochlorite for 1 min, rinsed well with sterile distilled water, and dried on sterile filter paper in a laminar hood. Each sterile tissue sample was plated onto potato dextrose agar (PDA) plates and water agar plates. Plates were incubated at 25°C in the dark until fungal colonies were observed. Colonies were subcultured onto PDA plates and incubated at 25°C for 4 days. Hyphal tip subcultures were obtained for each isolate from the colony margin and subcultured onto fresh PDA. Morphological characteristics were observed on three isolates on 2-week-old PDA cultures. Based on the culture morphology on PDA, all three isolates were tentatively identified as *M. phaseolina* (Figure S1). To confirm its identity, genomic DNA was extracted from mycelial mats of three isolates using DNeasy Plant Pro Kit (Qiagen Inc.) according to the manufacturer's instructions. The internal transcribed spacer (ITS) region of the rRNA gene was amplified with primers ITS1/ITS4, and three nucleotide sequences were obtained. A BLASTN search revealed that sample sequences shared a 100% match with sequences of *M. phaseolina* in the NCBI GenBank Database (Sayers et al., 2022).

2.3 | Plant imaging

2.3.1 | Image capture

Images of both visibly infected and noninfected sorghum plants were taken 51 days after planting, equating to the GS-1 vegetative (germination to panicle initiation) growth stage (Roozeboom & Prasad, 2019). Images were collected from both the WW and WL irrigation treatments. A total of 1400 high-resolution, JPEG-formatted images 5184 × 3456 and 2336 × 1752 pixels were taken of sorghum plants within the field using a Canon Rebel T6 camera (Canon) and a Sony Alpha a6000 (Sony), respectively. The camera operators walked through the field, and images were taken at random to include a variety of angles, lighting conditions, background

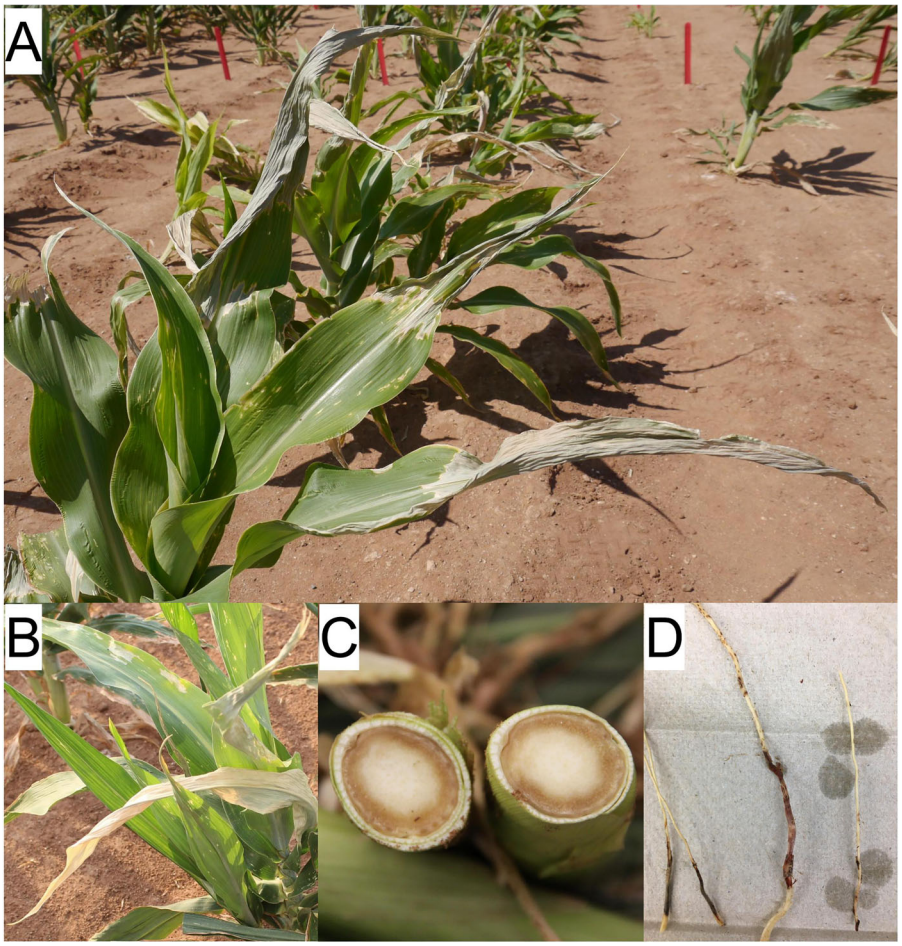

FIGURE 1 Symptoms of young sorghum plants infected with charcoal rot of sorghum (CRS), a fungal disease caused by *M. phaseolina*. (A) Grayish white appearance of young plants in a field in Maricopa, Arizona. (B) Close up of a young sorghum plant displaying symptoms of CRS, including leaf curling and hooking, chlorosis, and necrosis. (C) Discoloration of stem vascular tissue. (D) rotting of roots.

features, and zoom settings. The angle of capture was varied to encompass a broad spectrum of viewpoints, including low-angle shots, where the camera was tilted upward, and high-angle shots, with the camera angled downward. The image capture settings of both cameras were set to auto adjust (aperture, shutter speed, and ISO speed) to ensure adequate and unbiased image capture given the number of images that had to be taken. Images were captured between 10:00 and 14:00 on a single day (August 7, 2020), during which solar radiation varied from 2.42 to 3.42 MJ/m^2 (Table S1). The images captured both non-plant background elements (control) and sorghum plants displaying three distinct states: CRS infection, abiotic stress (drought), and absence of biotic and abiotic symptoms (healthy) (Figure S2). Collected images were uploaded to the CyVerse Data Store for further processing (Devisetty et al., 2016).

The 1400 CRS-impacted sorghum images were imported into the image labeling platform Labelbox (https://labelbox.com/) and annotated by researchers trained in CRS identification. Researchers were instructed to label the area exhibiting symptoms of CRS by drawing high-fidelity polygons around CRS-affected sorghum plant tissue. After annotation, images were reviewed and validated by a second set of experts and randomly split into three different sets: training (60%, 840 images), validation (20%, 280 images), and test (20%, 280 images). Images in the training set were used for model development and training, whereas images in the validation set were used to optimize model hyperparameters, and finally, images in the test set were used for reporting the final performance and accuracy. The manually labeled images represented diverse field conditions, irrigation treatments, number of plants within an image, resolution, and image collection conditions creating a diverse representation of the pathology of CRS-impacted tissue with respect to real-world conditions that would be encountered either in a sorghum production or research setting. Several pre-processing steps were carried out during model training, which included transformations of color space and generation of image patches. Using the polygons, binary masks were generated for each image and then the images alongside their corresponding masks were

separately split into smaller square image patches of different sizes (32 × 32, 64 × 64, 128 × 128, 256 × 256, and 512 × 512 pixels); this reduction in size was undertaken to test the effect of patch size on the model performance and accuracy (Figure S3). Among these patches, those that had at least one pixel annotated as CRS were then considered positive samples for the classification task. The same number of patches were randomly selected among the images without CRS to keep the three image datasets (training, validation, and test) balanced.

2.3.2 | Image analysis

Two different classes of ML models were applied to the dataset to assess and describe CRS-affected foliage: classification, which determines the presence or absence of CRS for a given patch, and segmentation, which highlights the regions affected by CRS and quantifies the amount of CRS detected in the images. Six classification models were trained and evaluated with respect to their ability to classify patches containing plants exhibiting CRS symptoms. The classification models implemented were ResNet18 (He et al., 2016); MobileNetV3 small, small custom, and large versions (Koonce, 2021b); and EfficientNet-B3 and EfficientNet-B4 (Koonce, 2021a). All of these models were initially trained and evaluated under the same conditions (same set of hyperparameters) on image patches of size 256 × 256 pixels, and the best performing model was selected for further optimization over the range of patch sizes listed above. Finally, the top image patch classification model was used to approximately quantify the area of the affected regions of each image in the test set. It is important to note that this was performed on the basis of image patches within the overall image and not on a per pixel basis. For image segmentation, three models were trained and evaluated including U-NET (Ronneberger et al., 2015), fully convolutional network (FCN) (Long et al., 2015), and DeepLabV3 (Chen et al., 2017). All models were trained and evaluated on image patch sizes of 256 × 256 pixels. Like the classification approach, the best performing model was then selected for further optimization over the range of patch sizes. Finally, the best performing models, with respect to classification and segmentation, were used to detect or quantify the regions, respectively, affected by CRS in the full-size images.

2.4 | Model training

2.4.1 | Model evaluation

To evaluate performance during the training and development of the classification models, binary cross entropy (BCE) was used as a loss function for classification (Ho & Wookey, 2020; Milletari et al., 2016; Salehi et al., 2017). The BCE calculates the negative logarithm of the prediction error to penalize the model for incorrect predictions, as defined by the following equation:

$$\mathrm{BCE}\,(p, y) = -y \times \log\,(p) - (1 - y) \times \log\,(1 - p)\,, \qquad (1)$$

where y is the ground truth, integer label associated with the image, and p is the prediction of the model. The models output a prediction between zero and one, with penalties increasing exponentially for larger prediction errors.

To evaluate the performance of the segmentation models, the Dice coefficient was used (Ho & Wookey, 2020; Milletari et al., 2016; Salehi et al., 2017). The Dice coefficient, a differentiable form of the Intersection over Union metric, was used as the loss function for semantic segmentation. It calculates the ratio of the intersecting area of the predicted mask and the ground truth to their combined area, as defined by the following equation:

$$\mathrm{Dice}\ (Y, P) = \frac{2 \times |Y\ \cap P|}{|Y\ |+|\ P|}, \qquad (2)$$

where Y is the ground truth mask and P is the predicted mask. The closer this value is to one, the more similar the prediction mask is to the ground truth.

The accuracy, precision, recall, and F1 score were computed for both classification and segmentation models to compare model performance. Accuracy, which is the proportion of correct predictions out of all predictions, is derived from the counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), as defined by the following equation:

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}. \qquad (3)$$

While accuracy is a common metric for model performance, it can be skewed by unbalanced datasets. To avoid bias toward the majority class, we also used metrics that are less sensitive to dataset composition, including precision, recall, and F1 score, as defined by the following equations:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}\ +\ \mathrm{FP}}, \qquad (4)$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}\ +\ \mathrm{FN}}, \qquad (5)$$

$$\mathrm{F1} = 2 \times \frac{\mathrm{Precision}\ \times\ \mathrm{Recall}}{\mathrm{Precision}\ +\ \mathrm{Recall}}, \qquad (6)$$

where Precision is the ratio of true positives to predicted positives, Recall is the ratio of true positives to actual positives, and F1 is the harmonic mean of these two. These metrics

collectively assess model performance, but the best performing models were chosen based on their F1 score.

Models were trained on two computing nodes: the first was an AMD EPYC 7542 32-Core processor (Advanced Micro Devices, Inc.), 1008 GB of RAM, and two NVIDIA A-100 40 GB graphical processing unit (GPU; NVIDIA Corporation); and the second node was an Intel Xeon Gold 6146 CPU @ 3.20 GHz 12-Core processor (Intel Corporation), 188 GB of RAM, and an NVIDIA GeForce RTX 2080 Ti GPU (NVIDIA Corporation). The learning rate was 1×10^{-3} and 3×10^{-4} for classification and segmentation tasks, respectively. A batch size of 32 with Adam optimizer was used for both classification and segmentation (Z. Zhang, 2018). Models were trained for 50 epochs, and early stopping was used to avoid overfitting.

The objective of the classification models is to categorize each image patch based on the presence or absence of CRS. If the model predicts the presence of CRS in a single patch, the image is labeled as CRS-affected. On the other hand, the segmentation models aim to classify each pixel within an image patch based on the presence or absence of CRS. The segmentation model quantifies the number of CRS-affected pixels per patch and reports the sum of CRS-affected pixels for each image. The algorithm generates a results data file containing the image name and prediction results. For classification models, the prediction result is a binary "True" or "False" value for the presence or absence of CRS, respectively. For segmentation models, the prediction results are the percentage of pixels classified as CRS compared to the total number of pixels in the image. For both classification and segmentation models, the total processing time for each image, which is the cumulative sum of the processing times for all individual patches within a single image, is also provided. Inference time was reported as the overall length of time from the model receiving an image patch to producing a segmentation or classification prediction.

2.4.2 | Evaluation of image patch size on accuracy and processing time

After the initial model evaluation experiments (i.e., training and evaluation of the classification and segmentation models on image patches sizes of 256 × 256 pixels), the best performing model with respect to F1 score on the validation dataset was selected from both classification and segmentation classes of models evaluated. These two models, one each from the classification and segmentation type, were then trained and evaluated on the datasets with different image patch sizes (32 × 32, 64 × 64, 128 × 128, 256 × 256, and 512 × 512 pixels) to assess the effect of image patch size on the model performance. Time benchmarking was performed on the aforementioned AMD EPYC server. In addition to the model performance on the test set, inference times were measured and reported. The code for training, inference, and evaluation of model performance can be accessed at https://github.com/phytooracle/charcoal-dryrot-quantification.

3 | RESULTS

3.1 | Model training

Nine models, including six classification and three segmentation, were trained to detect and segment the presence of CRS-affected sorghum plant tissue in images collected under field conditions (Figure 1). The two top-performing models, EfficientNet-B3 and FCN, were able to distinguish and quantify, respectively, plant tissue that was exhibiting signs of CRS compared to plant tissue that was demonstrating the effects of drought stress. The ability to distinguish between CRS and drought stress is critical given that the pathology of CRS closely resembles that of drought stress making accurate discernments between the two challenging.

3.2 | Model performance

3.2.1 | Assessing model performance and identification of top-performing models

Models were assessed on recall, accuracy, and F1 score, but F1 score was the main criterion for selection. The best performing model for the classification of image patches that contained at least one pixel of CRS-labeled plant tissue was EfficientNet-B3. This model had an accuracy of 86.97% and an F1 score of 0.73 (Table 1, Figure 2). The classification model with the lowest performance was ResNet18 with an accuracy of 82.62% and corresponding F1 score of 0.65. With respect to the other two classes of models tested, their average performance across models was as follows: MobileNet: accuracy of 84.99% and F1 score of 0.69; EfficientNet: accuracy of 86.44% and F1 Score of 0.72. EfficientNet-B3 had higher accuracy and F1 score so it was selected as the superior model (Table 1).

Given the results of the classification models and that CRS-affected tissue could be identified by ML algorithms, we next evaluated if the disease-affected regions could be segmented from the images. The expansion of our analyses to segmentation was driven by the question of whether desiccated tissue due to drought stress, which is similar in appearance to CRS-affected tissue, could reliably be distinguished. With respect to model accuracy, both DeepLabV3 and FCN performed comparably with validation F1 scores of 0.65 and 0.66, respectively (Table 1). Of the two models, FCN had higher validation recall and F1 score, so it was selected as the superior model (Table 1). U-NET gave

TABLE 1 Results of model training and testing on image patch sizes of 256 × 256 pixels for both tasks of classification and segmentation for different models in the respective categories. Values that are highlighted in bold represent the top performing models for each task, as determined by the validation F1-score.

Task/loss function	Model name	Validation accuracy (%)	Validation precision	Validation recall	Validation F1
Classification/BCE	ResNet18	82.62	0.75	0.61	0.65
	MobileNetV3 small	85.47	0.78	0.68	0.70
	MobileNetV3 small custom	84.65	0.78	0.66	0.69
	MobileNetV3 large	84.84	0.77	0.64	0.68
	EfficientNet-B3	**86.97**	**0.80**	**0.71**	**0.73**
	EfficientNet-B4	85.91	0.78	0.68	*0.71*
Segmentation/dice	U-NET	97.49	0.69	0.62	0.62
	FCN	**97.76**	**0.69**	**0.68**	**0.66**
	DeepLabV3	97.93	0.72	0.65	*0.65*

Abbreviations: BCE, binary cross entropy; FCN, fully convolutional network.

TABLE 2 Impact of image patch size on model performance, as assessed using the validation set, for EfficientNet-B3, a classification model, and fully convolutional network (FCN), a segmentation model. Values that are highlighted in bold represent the top performing patch size for each task, as determined by the validation F1-score.

Model name	Task	Patch size	Validation accuracy (%)	Validation precision	Validation recall	Validation F1
EfficientNet-B3	Classification	32	86.55	0.04	0.01	0.02
		64	92.51	0.29	0.23	0.25
		128	89.77	0.50	0.42	0.44
		256	84.97	0.79	0.63	0.68
		512	**83.76**	**0.89**	**0.83**	**0.84**
FCN	Segmentation	32	95.13	0.16	0.13	0.14
		64	95.57	0.22	0.15	0.17
		128	97.12	0.43	0.38	0.39
		256	97.80	0.71	0.65	0.65
		512	**98.48**	**0.83**	**0.77**	**0.79**

the lowest performance metrics with a validation F1 score of 0.62.

3.2.2 | Assessing the impact of image patch size on model performance

Next, we investigated the influence of image patch size on model performance to better understand how downstream applications, such as deployment on drone- or mobile-based phenotyping platforms, could be impacted by changes in image patch size. For both EfficientNet-B3 and FCN models run on the validation image set, there was a near linear increase in validation F1, recall, and precision scores as the image patch size increased from 32 × 32 pixels to the final size tested of 512 × 512 pixels (Table 2). One exception to this observed trend were the results for validation accuracy. For the FCN model, the validation accuracy results increased from 95.13% to 98.48% for image patch sizes of 32–512 pixels, respectively. However, the EfficientNet-B3 exhibited contrasting results. The highest validation accuracy occurred for image patch size of 64 × 64 pixels and was 92.51%. For the following patch sizes of 128, 256, and 512 pixels, the validation accuracy decreased to 89.77%, 84.97%, and 83.76%, respectively.

3.2.3 | Assessing the impact of image patch size on image processing time

In light of the impact that image patch size had on the performance metrics of the respective models, we next wanted to investigate how the patch sizes impacted image processing time as this is another significant factor with respect to

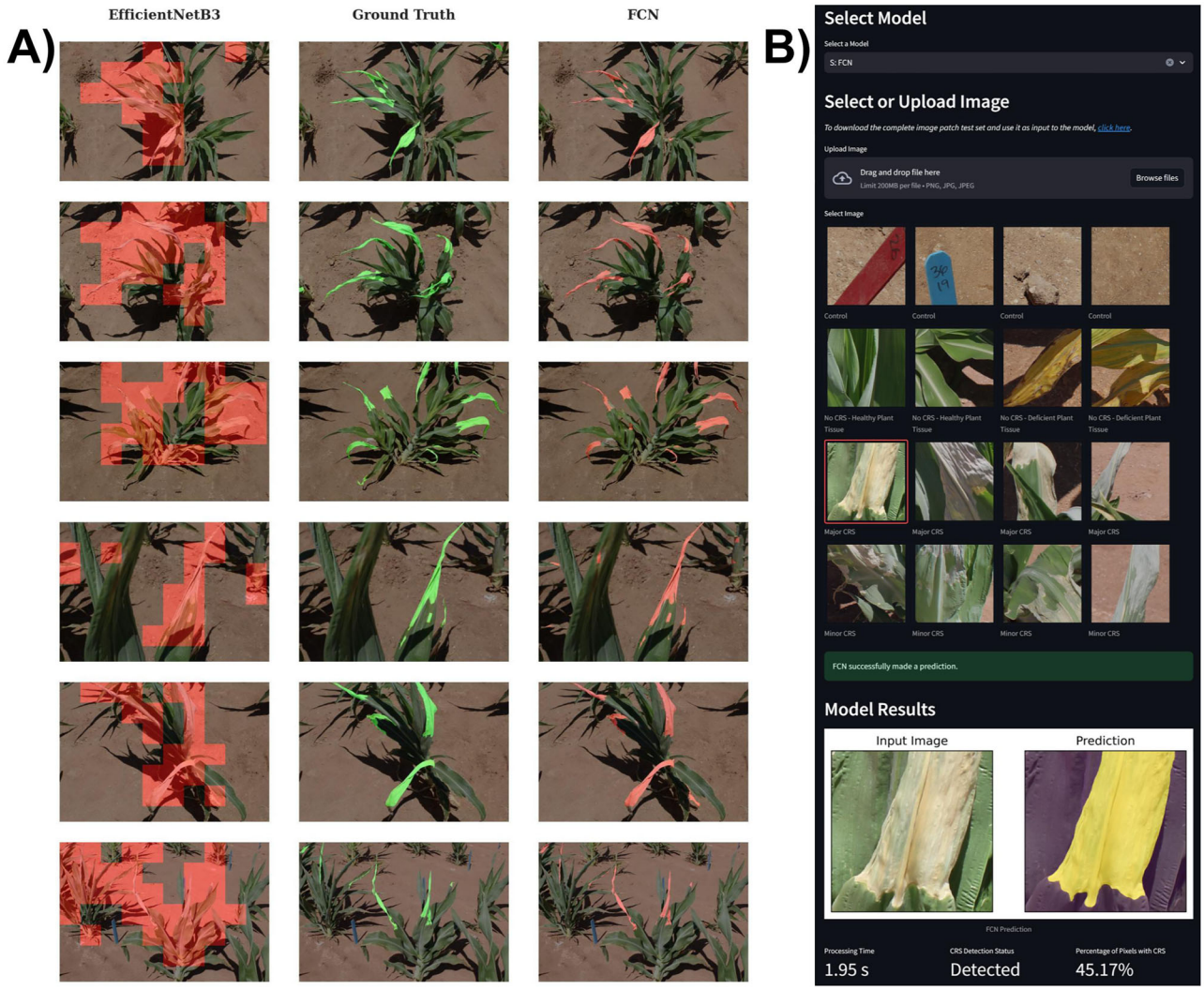

FIGURE 2 Comparison and deployment of classification and segmentation models. (A) Comparison of the identification of charcoal rot of sorghum (CRS) by the classification model EfficientNet-B3 and the segmentation model fully convolution network (FCN). Ground truth, hand-labeled images are in the center column, while the EfficientNet-B3 and FCN model outputs are on the left- and right-hand sides, respectively. For EfficientNet-B3, image patches of size 512 × 512 are highlighted since the model only classifies whether or not the image contains CRS-affected pixels. The FCN model, which performs segmentation, highlights the specific pixels that are affected with CRS closely resembling the results of the hand-labeled, ground truth data. Both models were trained on image patch sizes of 512 × 512 pixels. (B) A Streamlit application for easy deployment of all classification and segmentation models presented here. The application allows users to (i) select a model; (ii) upload an image or select one from the image gallery of test set images; and (iii) obtain model results including processing time, detection status, and percentage of pixels with CRS. The application can be accessed at https://charcoal-dryrot-quantification.streamlit.app/.

algorithm deployment. Both models exhibited an exponential decrease in mean processing time with increasing image patch size. For the 32, 64, 128, 256, and 512 patch sizes, the mean processing times were 263.34, 65.60, 24.85, 9.63, and 4.15 s and 518.09, 204.70, 70.52, 29.13, and 18.98 s for EfficientNet-B3 and FCN, respectively (Figure 3). Comparing the mean processing time within a model, we found that all processing times were significantly different ($p < 0.05$) for each image patch size level. Additionally, there was minimal variation observed for mean processing time for both models and image patch sizes; only FCN at 32 and 64 pixels exhibited any appreciable variation. With respect to the individual models themselves, it was not surprising to find that EfficientNet-B3, the classification model, was nearly twice as fast as the FCN, segmentation model, given the difference in tasks they perform.

3.2.4 | Evaluation of model performance on test set

With model training and validation completed, we next proceeded to evaluate how the respective models performed on the test data set—the holdout image set. For the classification model EfficientNet-B3, the highest F1 score obtained was 0.70 for the 256-pixel image patch size. This is in contrast to

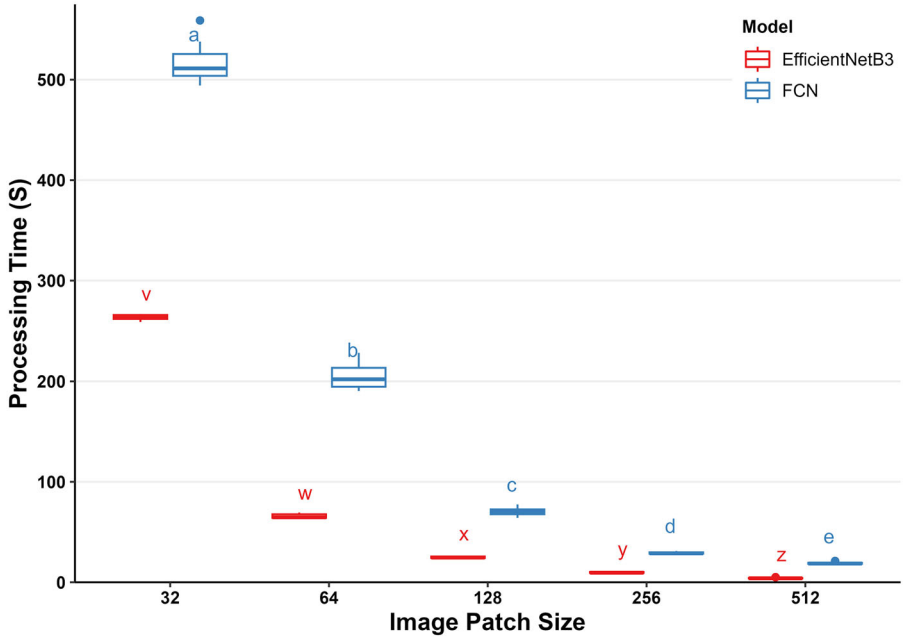

FIGURE 3 Relationship between image patch size and processing time in seconds for the EfficientNet-B3 classification model and the fully convolutional network (FCN) segmentation model. Different letters indicate significant differences among patch size processing times based on one-way analysis of variance followed by Tukey's post hoc test ($p < 0.05$). Error bars represent standard deviation.

the F1 score of 0.83 for the 512-pixel image patch size that was observed with the validation image set. With respect to accuracy, EfficientNet-B3 exhibited a near linear decrease in accuracy on the test image set with increasing patch size; however, the overall accuracy values were higher on the test image set compared to the validation image set (92.47% vs. 87.51%, respectively). For the segmentation model FCN, the highest F1 score on the validation image set was 0.77 obtained on the 512-pixel image patch size; this also corresponded to an accuracy of 98.48%. On the test image set, the FCN model also obtained the largest F1 score on the 512-pixel image patch size—an F1 of 0.80 with a corresponding accuracy of 99.43%. With respect to all FCN performance metrics, the same trend was observed for both the validation and test image sets, namely that the metrics all improved with increasing image patch size. With respect to the image processing time, the same trends were observed for the test image set as for the validation set—an approximately exponential decrease in mean processing time per image as the image patch size increased. Additionally, and not surprisingly, the FCN model's mean processing time was still much slower than the EfficientNet-B3 model. With respect to the individual models, the mean processing time for the image patches varied between the two models (Figure 3). The EfficientNet-B3 model had fairly consistent processing time per image patch with an average of 0.03 tenths of second across image patch sizes, whereas for the FCN model, the mean processing times for image patch sizes of 32, 64, 128, 256, and 512 were 0.02, 0.06, 0.08, 0.12, and 0.35 tenths of second, respectively (Table 3).

3.3 | Model deployment

The algorithm generates a data file of results, which includes the processing time (inference), the model's prediction (True or False for classification), and the percentage of pixels predicted as CRS. Alongside the results data file, visual representations of the predictions are also produced. The predictions of CRS from the classification model are depicted in each patch of the full image (Figure 2A, left). In contrast, the segmentation model's predictions are illustrated at the pixel level, emphasizing the specific regions of the leaves affected by CRS (Figure 2A, right). Users can select to run all classification and segmentation model classes tested here. In this case, the algorithm outputs a summary data file including the model's name, patch size, and processing time for each, similar to the results data file. The images for training, testing, and validation, both in their original and annotated forms, are available at https://data.cyverse.org/dav-anon/iplant/projects/phytooracle/season_11_sorghum_yr_2020/level_0/charcol_rot_sorghum/dry_rot_raw.tar.gz. The code to train models and run inference are available on GitHub at https://github.com/phytooracle/charcoal-dryrot-quantification. To enhance the usability of these tools, the models have been incorporated into a Streamlit application that is publicly accessible at https://charcoal-dryrot-quantification.streamlit.app/. The application can also be executed locally by utilizing the associated Docker container: https://hub.docker.com/r/phytooracle/charcoal-dryrot-quantification. All models presented in the current study were integrated as a module in

TABLE 3 Performance metrics for EfficientNet-B3, a classification model, and fully convolutional network (FCN), a segmentation model, with respect to the different size image patches (in pixels) and the time it took to process the whole image (seconds) as well as per image patch (tenths of second). Values that are highlighted in bold represent the top performing patch size for each task, as determined by the test F1-score.

Model name	Task	Patch size	Accuracy (%)	Precision	Recall	F1	Mean processing time (image)	Mean processing time (patch)
EfficientNet-B3	Classification	32	97.51	0.66	0.11	0.19	253.2	0.02
		64	96.38	0.52	0.77	0.62	58.42	0.02
		128	94.47	0.54	0.79	0.64	19.38	0.02
		256	**92.74**	**0.64**	**0.77**	**0.70**	**6.42**	**0.03**
		512	81.24	0.53	0.88	0.66	2.27	0.05
FCN	Segmentation	32	97.90	0.37	0.69	0.48	311.93	0.02
		64	98.81	0.58	0.56	0.57	192.15	0.06
		128	98.99	0.61	0.79	0.69	65.96	0.08
		256	99.19	0.67	0.83	0.74	25.95	0.12
		512	**99.43**	**0.77**	**0.84**	**0.80**	**17.17**	**0.35**

the PhytoOracle phenotyping workflow manager: https://github.com/phytooracle/automation/blob/main/yaml_files/other/crs_detection.yaml. The code and data are released under the open-source MIT license.

4 | DISCUSSION

4.1 | The need of phenotyping tools for detecting biotic stress

Charcoal rot of sorghum, a disease caused by the soilborne fungal pathogen *M. phaseolina*, poses a significant threat to sorghum, a critical grain crop. The pathogen, *M. phaseolina*, thrives in hot, dry conditions, which often coincide with the peak growing periods for sorghum in the southwest United States. As a result, areas with substantial sorghum production often experience high incidences of CRS, leading to considerable crop losses and economic impact (Kaur et al., 2012; Marquez et al., 2021). Effective management of CRS is therefore crucial to ensure the sustainability of sorghum production in these regions. Early detection of CRS is required for crop management systems for sorghum production and breeding improvement programs. As such, a major challenge is shifting CRS detection and quantification from manual visual assessment to automated, image-based phenotyping approaches. Handheld and drone-based RGB cameras provide low cost solutions, providing that they are accurate, precise, and efficient computer vision and AI/ML algorithms to process those data. This method not only allows for the detection and quantitative evaluation of disease severity but also supports the high-throughput screening of disease resistance variations in cultivars. Moreover, it acts as a crucial data gathering tool, potentially offering comprehensive insights into the prevalence, distribution, and effects of CRS across different environments.

To address the need for disease detection methods, two classes of algorithms were evaluated for the classification and segmentation of CRS. These algorithms were applied to RGB images of BTx623 sorghum, which were mutagenized with EMS, resulting in a phenotypically diverse population. The image dataset captured this phenotypic diversity, thereby broadening the applicability of our models to other genotypes. The algorithms comprised six and three base models of CNNs, respectively. Our objective was to create and assess models that meet the varied demands of sorghum researchers, breeders, and producers. The classification algorithm was the most efficient, requiring fewer computational resources for deployment. This enables a rapid assessment of CRS-affected tissues and is adaptable for use in the field on portable devices such as smartphones or drones. Users can efficiently identify CRS affected plants for near real-time decision-making. On the other hand, the segmentation algorithm is more computationally intensive and currently requires dedicated computing resources. Users need to send images to off-field hardware to quantify the relative amount of above-ground affected plant tissues. To address this limitation, we developed a web application that can efficiently deploy both classification and segmentation models on multiple devices, including cell phones. Tools like these will enable breeders to select genotypes with improved resistance more effectively, potentially leading to increased future yield gains. Furthermore, these tools are directly applicable to crop management, as they can assess the severity of the disease and aid in planning treatment strategies such as control point spraying.

4.2 | Model types for plant disease detection and quantification

For CRS detection in images, binary classification can be applied on image patches. Alternatively, semantic segmentation can be used to quantify the ratio of pixels identified as CRS versus non-CRS. In this study, models representing each approach were trained and tested to evaluate performance metrics and assess suitability for the task of detecting and quantifying CRS. For classification, ResNet18 (He et al., 2016), MobileNetV3 small, custom small and large (Koonce, 2021b), and EfficientNet-B3 and EfficientNet-B4 (Koonce, 2021a) were implemented; for segmentation, U-NET (Ronneberger et al., 2015), FCN (Long et al., 2015), and DeepLabV3 (Chen et al., 2017) were used. Each approach has advantages and disadvantages: classification models are generally faster, while segmentation models are often more accurate and have higher performance based on F1 score in particular.

4.2.1 | Classification models

In ResNet18, stacks of convolutional layers alongside Max Pooling (Wu & Gu, 2015), Batch Normalization (Ioffe & Szegedy, 2015), and other auxiliary layers are combined in a feed forward network. The number 18 refers to the number of learnable layers. In addition, residual or skip connections are added between these blocks to (*i*) facilitate dealing with the vanishing gradient problem and (*ii*) to help the model learn more efficiently by having the capability of skipping some of the layers if necessary. Developed by Google, MobileNetV3 models are designed for use on mobile devices and other embedded systems such as UAVs (A. G. Howard et al., 2017). These models are built out of depth-wise separable convolutions, a form of factorized convolutional layers that enables filters to be shared across channels, reducing the number of filters needed to improve computational efficiency. MobileNetV3 is the latest iteration of the MobileNet (A. Howard et al., 2019). Small and Large MobileNets have a trade-off between latency and accuracy with the final layer of the MobileNetV3 Small and Large consisting of 576 and 1280 output neurons, respectively. To investigate the impact of reducing neuron count on performance, we customized MobileNetV3 Small by modifying the number of neurons in the last block of the model. This model has extra dense layers to reduce the number of output neurons more gradually from 576–288 to 64–1 neuron. Tan and Le (2019) proposed a compound scaling method for scaling the depth, width, and resolution of neural network layers in a way to efficiently achieve better performance. They observed that while scaling up networks' width, depth, and resolution generally improves accuracy because of higher network capacity, the dimensionality of the network needs to be adjusted properly in order to maximize these improvements on metrics while minimizing overfitting to the training data. In their study, they propose a relationship between the scaling ratios for each dimension of the network which can be used to uniformly scale up the network to increase its efficiency and accuracy. Using their proposed relationship between the dimensions, they provide eight different architectures, namely EfficientNet-B0–EfficientNet-B7. In this study, we tested EfficientNet-B3 and EfficientNet-B4 based on the results presented in their study.

4.2.2 | Segmentation models

FCNs (Long et al., 2015) provide an end-to-end model for semantic segmentation of images with arbitrary sizes. Prior to FCNs, semantic segmentation used to be done by sliding a classification model over the entire image to get predictions for each pixel. U-NET is a fully convolutional neural network that has an encoder–decoder structure. It was developed by Ronneberger et al. (2015) to address the problem of semantic segmentation in biomedical images and uses skip connections. The encoder transforms the input image into a latent space with lower dimension and the decoder transforms the latent space into the output image, usually with the same dimension as the input. Our implementation varies slightly in architecture from the original paper where we used established best practices of adding padding to preserve input dimensions between downsampling and upsampling and added batch normalization between convolution layers and the Relu activation functions. DeepLabV3 (Chen et al., 2017) utilizes a novel method for enlarging the field of view of the convolution kernels to incorporate multi-scale context of images. This improves semantic segmentation results by using the atrous convolution (Holschneider et al., 1990).

4.3 | Optimizing model performance and addressing limitations

4.3.1 | Effects of image patch size on performance and processing time

Processing time and model performance are closely linked to the size of image patches. When dealing with larger patches, the results are typically faster. This is likely due to larger patches containing more information, allowing the model to make decisions based on a broader context. This can be particularly beneficial for classification tasks, where the goal is to categorize the entire image or large regions of it. For instance, in crop disease detection, larger patches could help quickly identify whether a particular disease is present or

absent in a field. On the other hand, smaller patches, while resulting in slower performance, can provide a more detailed assessment of the image. This can be especially useful when the goal is to assess the severity of a condition, such as the extent of disease in a crop field. Smaller patches allow for a more granular analysis, which can help quantify the extent of the disease and guide targeted interventions. Importantly, the choice of patch size in image analysis can significantly impact model performance, including the F1 score, a measure of a model's accuracy. In the case of EfficientNet-B3, a patch size of 512 was found to yield the highest validation F1 score (Table 3). This suggests that this patch size is optimal for balancing the trade-off between precision and recall in the model's predictions. Interestingly, this patch size also has the fastest processing time per image (Figure 3). This means that not only does it provide accurate results, but it does so efficiently, making it a good choice for applications where both accuracy and speed are important. Similarly, for the FCN, a patch size of 512 results in the highest validation and test F1 score (Table 2, Table 3). This larger patch size allows the model to capture more contextual information, which can improve the accuracy of its predictions. Moreover, this patch size also has the fastest processing time per image, making it an efficient choice for image analysis tasks (Figure 3). Our results indicate that the selection of patch size influences models, impacting accuracy and processing time. This necessitates an optimization strategy for patch size selection, guided by task-specific requirements and performance metrics.

In the current study, EfficientNet-B3 and FCN performed best for the classification and segmentation of CRS, respectively (Table 1). As expected, processing time for a single image decreased as image patch size increased due to fewer patches being needed per image, thereby reducing the computational load. While one may expect that increased patch size will result in decreased sensitivity, our results showed that the largest patch size tested for classification had the highest validation precision, recall, and F1 scores (Table 2). This is likely due to having more likelihood of dry rot being present in a larger patch. Therefore, the false positive rate would be lower compared to a smaller patch size, as we classify a prediction as a TP even if there's only a single pixel of dry rot present in the patch. It should be noted that the validation accuracy was the lowest for the largest patch size for classification. Classification models work by assigning a single label to the entire image. These models are designed to focus on the most important features in the image that are relevant to the classification task. An increase in image patch size could introduce additional details and noise, such as background soil and neighboring plants, that may interfere with the main features, potentially leading to a decrease in validation accuracy. The increase in validation accuracy as the image patch size grows could be attributed to the additional context and detail that larger patches offer for the model to assess, thereby enhancing its performance.

4.3.2 | Limitations of models

The results presented here increase the utility of these models for the classification and segmentation of CRS from image data, with the EfficientNet-B3 model taking ~6.4 s to process an image with a patch-size of 256 × 256 pixels and the FCN model taking ~17.2 s. These benchmarks utilized high-end GPUs, which are not currently deployable in the field. However, as GPU-based architectures continue to advance, the ability to run these models in the field using aerial-based or portable computing devices should be available in the near-term future, providing researchers, breeders, and growers the opportunity to detect and quantify CRS in real time using automated crop phenotyping systems (Owens et al., 2008). Although existing models can be implemented on ground-based phenotyping systems equipped with powerful computers, the deployment of these models on UAVs is not yet feasible at scale. This limitation is primarily due to the intensive computational demands of the models, which surpass the hardware currently mountable on UAVs. Specifically, these models often need powerful GPUs to operate efficiently. The GPUs can perform parallel operations, making them ideal for the complex calculations and large data volumes involved in ML. However, the physical characteristics of GPUs, such as their size, weight, and power consumption, make them currently unsuitable for mounting on UAVs, which have strict limitations on payload capacity and power supply. Given the expected progress in hardware technology, it is reasonable to anticipate that these models could be deployed on UAVs in the near future. It is therefore essential for research, like that presented here, to concentrate on the development, training, and refinement of models to prepare them for widespread deployment on UAVs in the future.

Another significant limitation of these models is the potential for ambiguity with other diseases or abiotic stresses. Models may sometimes struggle to distinguish between the disease of interest and other diseases or stress conditions with similar symptoms. This could lead to false positives or negatives, impacting the accuracy of the model. The results presented here highlight the feasibility to discriminate heterogenous symptoms, namely drought and CRS, on a single leaf. Nonetheless, additional studies are required to determine whether this is applicable to other pairs of heterogeneous symptoms.

The quantification of disease tissue predicted in a given image presents some challenges. The segmentation algorithm calculates the percentage of CRS relative to the entire image, not just the plant tissue. This approach was chosen because masking out the soil using vegetation indices could introduce errors, and, as a result, the reported result would be an amalgamation of various errors stemming from both the index calculation and the model prediction, complicating the interpretation and differentiation of these errors. The importance of having an unbiased and multi-dimensional system

for quantifying plant disease instead of relying on human scoring is also worth noting. The process of human scoring can accumulate a variety of errors, particularly when multiple individuals are involved, as it can lead to subjectivity and inconsistency. An automated, unbiased system can provide more consistent and reliable results. Such a system could process large amounts of data quickly, making it a valuable tool for large-scale disease monitoring and management.

The field design employed in our research, while standard for such research studies, may not perfectly mirror the conditions of sorghum cultivation. Nevertheless, the models developed in this study can be adapted to offer benefits to breeders and growers under their specific circumstances. It is worth noting that our planting density, which was reduced to five plants per plot post-establishment, is less than typically seen in production settings. This reduction allowed for the incorporation of a greater number of control images, such as soil, plot stakes, and other background elements, into our study. The inclusion of these control images improves the robustness of our models through increased image diversity, potentially equipping them to effectively differentiate CRS from other non-living objects and various stressors, including drought stress. Despite the discrepancy with production settings, our methodology provides tools that can be adapted to a variety of agricultural contexts. To facilitate this, labeled images are shared to enable researchers to modify or extend existing or future models, and a web application was developed for breeders and growers to assess the applicability of our models to their specific circumstances.

4.3.3 | The importance of user-friendly phenotyping tools

Increasing the accessibility and integration of trained models is essential. Platforms such as CyVerse, Streamlit, and PhytoOracle facilitate model use, encouraging adoption by non-experts and driving innovation. Open access to code and models, along with integration into web applications and workflow managers like PhytoOracle, enables efficient model deployment to large datasets (Gonzalez et al., 2023). This approach broadens user engagement and advances technological democratization.

5 | CONCLUSION

As the quantitative and qualitative results suggest, quantifying CRS in images using NNs is a difficult task because of the similarity between the symptoms of water deficit stress (i.e., drought) and CRS. However, in this study, we proposed two approaches for this task using classification and segmentation models and showed that quantifying CRS in plants exhibiting concurrent drought stress symptoms, despite being a difficult task, can be accomplished with a high level of accuracy. We showed that the segmentation models outperform the classification models in quantifying CRS. An extensive set of experiments were conducted to assess the effect of patch size on the processing time and performance of the models. We found that a patch size of 256 is suitable for classification, and a patch size of 512 yields the best results for the segmentation models. The models were integrated into existing phenomics pipelines and a web application for user-friendly deployment of trained models on various platforms.

AUTHOR CONTRIBUTIONS

Emmanuel M. Gonzalez: Conceptualization; data curation; formal analysis; investigation; methodology; resources; software; validation; visualization; writing—original draft; writing—review and editing. **Ariyan Zarei**: Conceptualization; data curation; formal analysis; investigation; methodology; resources; software; validation; visualization; writing—original draft; writing—review and editing. **Sebastian Calleja**: Data curation; Formal analysis; investigation; methodology; validation; writing—original draft; writing—review and editing. **Clay Christenson**: Data curation; formal analysis; investigation; methodology; validation; visualization; writing—original draft; writing—review and editing. **Bruno Rozzi**: Investigation; methodology; writing—original draft; writing—review and editing. **Jeffrey Demieville**: Investigation; methodology; validation; writing—original draft; writing—review and editing. **Jiahuai Hu**: Investigation; methodology; writing—original draft; writing—review and editing. **Andrea L. Eveland**: Conceptualization; funding acquisition; investigation; methodology; project administration; resources; supervision; writing—original draft; writing—review and editing. **Brian Dilkes**: Conceptualization; funding acquisition; investigation; methodology; project administration; supervision; writing—original draft; writing—review and editing. **Kobus Barnard**: Conceptualization; funding acquisition; investigation; methodology; project administration; supervision; writing—original draft; writing—review and editing. **Eric Lyons**: Conceptualization; funding acquisition; investigation; methodology; project administration; supervision; writing—original draft; writing—review and editing. **Duke Pauli**: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; supervision; validation; visualization; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

We would like to give our sincere gratitude to Cristian Salazar, Travis Simmons, Holly Ellingson, Victoria Ramsay, Brenda Esmeralda Jimenez, Hanna April Lawson, Hassan Alnamer, Jordan Pettiford, Michele Cosi, and Robert Strand for their

assistance in image annotation and assistance with the field experiment.

FUNDING INFORMATION

This project was supported by the Department of Energy Advanced Research Agency-Energy award number DE-AR0001101, Department of Energy Biological and Environmental Research award number DE-SC0020401, and National Science Foundation CyVerse project award number DBI-1743442. Support to Pauli was also provided by NSF-PGRP (Award # 2102120 and 2023310), NSF DBI (Award # 2019674 and 1743442), and Cotton Incorporated award numbers 18–384, 20–720, 21–830, and 23–890. Support to Lyons was provided by NSF-PGRP (Awards # IOS – 2023310 and IOS – 1849708).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The images used for training, validation, and testing the model can be accessed at https://data.cyverse.org/dav-anon/iplant/projects/phytooracle/season_11_sorghum_yr_2020/level_0/charcol_rot_sorghum/dry_rot_raw.tar.gz. The code for model training and inference is available at https://github.com/phytooracle/charcoal-dryrot-quantification.

The models in this study can be deployed on a Streamlit app at https://charcoal-dryrot-quantification.streamlit.app/. The Docker container to run the Streamlit app locally is available at https://hub.docker.com/r/phytooracle/charcoal-dryrot-quantification. The models were integrated as a module in the PhytoOracle phenotyping workflow manager: https://github.com/phytooracle/automation/blob/main/yaml_files/other/crs_detection.yaml.

ORCID

Emmanuel M. Gonzalez https://orcid.org/0000-0002-3021-9842
Ariyan Zarei https://orcid.org/0000-0002-3670-2472
Sebastian Calleja https://orcid.org/0000-0001-9401-4494
Clay Christenson https://orcid.org/0009-0005-4034-7433
Bruno Rozzi https://orcid.org/0009-0008-2945-2103
Jeffrey Demieville https://orcid.org/0000-0002-7725-7379
Jiahuai Hu https://orcid.org/0000-0003-1107-5414
Andrea L. Eveland https://orcid.org/0000-0003-4825-1282
Brian Dilkes https://orcid.org/0000-0003-2799-954X
Kobus Barnard https://orcid.org/0000-0002-8568-9518
Eric Lyons https://orcid.org/0000-0002-3348-8845
Duke Pauli https://orcid.org/0000-0002-8292-2388

REFERENCES

Addo-Quaye, C., Tuinstra, M., Carraro, N., Weil, C., & Dilkes, B. P. (2018). Whole-genome sequence accuracy is improved by replication in a population of mutagenized sorghum. *G3 (Bethesda)*, *8*(3), 1079–1094. https://doi.org/10.1534/g3.117.300301

Amsaveni, V., & Albert Singh, N. (2013). *Detection of brain tumor using neural network*. 2013 Fourth international conference on computing, communications and networking technologies (ICCCNT), Tiruchengode, India. https://doi.org/10.1109/icccnt.2013.6726524

Bernardo, R. (2014). *Essentials of plant breeding*. Stemma Press.

Bhattacharya, M., & Samaddar, K. R. (1976). Epidemiological studies on jute diseases. *Plant and Soil*, *44*(1), 27–36. https://doi.org/10.1007/BF00016952

Boatwright, J. L., Lucas Boatwright, J., Sapkota, S., Jin, H., Schnable, J. C., Brenton, Z., Boyles, R., & Kresovich, S. (2022). Sorghum Association Panel whole-genome sequencing establishes cornerstone resource for dissecting genomic diversity. *The Plant Journal*, *111*(3), 888–904. https://doi.org/10.1111/tpj.15853

Bock, C. H., Parker, P. E., Cook, A. Z., Riley, T., & Gottwald, T. R. (2009). Comparison of assessment of citrus canker foliar symptoms by experienced and inexperienced raters. *Plant Disease*, *93*(4), 412–424. https://doi.org/10.1094/PDIS-93-4-0412

Brown, P. W. (1989). *Accessing the Arizona meteorological network (AZMET) by computer* (Extension report No. 8733). University of Arizona. https://ag.arizona.edu/AZMET/az-data.htm

Chadalavada, K., Kumari, B. D. R., & Kumar, T. S. (2021). Sorghum mitigates climate variability and change on crop yield and quality. *Planta*, *253*(5), 113. https://doi.org/10.1007/s00425-021-03631-2

Chakraborty, S., Tiedemann, A. V., & Teng, P. S. (2000). Climate change: Potential impact on plant diseases. *Environmental Pollution*, *108*(3), 317–326. https://doi.org/10.1016/S0269-7491(99)00210-9

Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). *Rethinking atrous convolution for semantic image segmentation*. https://doi.org/10.48550/ARXIV.1706.05587

Cheshkova, A. F. (2022). A review of hyperspectral image analysis techniques for plant disease detection and identification. *Vavilovskii Zhurnal Genetiki i Selektsii*, *26*(2), 202–213.

Collins, D. J., Wyllie, T. D., & Anderson, S. H. (1991). Biological activity of *Macrophomina phaseolina* in soil. *Soil Biology & Biochemistry*, *23*(5), 495–496.

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

Delgado-Baquerizo, M., Guerra, C. A., Cano-Díaz, C., Egidi, E., Wang, J.-T., Eisenhauer, N., Singh, B. K., & Maestre, F. T. (2020). The proportion of soil-borne pathogens increases with warming at the global scale. *Nature Climate Change*, *10*(6), 550–554. https://doi.org/10.1038/s41558-020-0759-3

Desaint, H., Aoun, N., Deslandes, L., Vailleau, F., Roux, F., & Berthomé, R. (2021). Fight hard or die trying: When plants face pathogens under heat stress. *The New Phytologist*, *229*(2), 712–734. https://doi.org/10.1111/nph.16965

Devisetty, U. K., Kennedy, K., Sarando, P., Merchant, N., & Lyons, E. (2016). Bringing your tools to CyVerse discovery environment using docker. *F1000Research*, *5*, 1442. https://doi.org/10.12688/f1000research.8935.1

Gonzalez, E. M., Zarei, A., Hendler, N., Simmons, T., Zarei, A., Demieville, J., Strand, R., Rozzi, B., Calleja, S., Ellingson, H., Cosi, M., Davey, S., Lavelle, D. O., Truco, M. J., Swetnam, T. L., Merchant, N., Michelmore, R. W., Lyons, E., & Pauli, D. (2023). PhytoOracle: Scalable, modular phenomics data processing pipelines.

Frontiers in Plant Science, *14*, 1112973. https://doi.org/10.3389/fpls.2023.1112973

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. 2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, USA. https://doi.org/10.1109/cvpr.2016.90

Ho, Y., & Wookey, S. (2020). The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE Access*, *8*, 4806–4813. https://doi.org/10.1109/ACCESS.2019.2962617

Holschneider, M., Kronland-Martinet, R., Morlet, J., & Tchamitchian, P. (1990). A real-time algorithm for signal analysis with the help of the wavelet transform. In J. M. Combes, A. Grossmann, & P. Tchamitchian (Eds.), *Inverse problems and theoretical imaging* (pp. 286–297). Springer. https://doi.org/10.1007/978-3-642-75988-8_28

Hossain, M. S., Islam, M. N., Rahman, M. M., Mostofa, M. G., & Khan, M. A. R. (2022). Sorghum: A prospective crop for climatic vulnerability, food and nutritional security. *Journal of Agriculture and Food Research*, *8*, 100300. https://doi.org/10.1016/j.jafr.2022.100300

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient convolutional neural networks for mobile vision applications*. http://arxiv.org/abs/1704.04861

Howard, A., Sandler, M., & Chu, G. (2019). Searching for mobilenetv3. *IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 1314–1324). IEEE. https://doi.org/10.1109/ICCV.2019.00140

Hsi, C. H. (1956). Stalk rots of sorghum in eastern New Mexico. *Plant Disease Reporter*, *40*(5), 369–376.

Ioffe, S., & Szegedy, C. (2015). *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. https://doi.org/10.48550/ARXIV.1502.03167

Kaur, S., Dhillon, G. S., Brar, S. K., Vallad, G. E., Chand, R., & Chauhan, V. B. (2012). Emerging phytopathogen *Macrophomina phaseolina*: Biology, economic importance and current diagnostic trends. *Critical Reviews in Microbiology*, *38*(2), 136–151. https://doi.org/10.3109/1040841X.2011.640977

Kissoudis, C., van de Wiel, C., Visser, R. G. F., & van der Linden, G. (2014). Enhancing crop resilience to combined abiotic and biotic stress through the dissection of physiological and molecular crosstalk. *Frontiers in Plant Science*, *5*, 207. https://doi.org/10.3389/fpls.2014.00207

Koonce, B. (2021a). EfficientNet. *Convolutional* neural networks with swift for tensorflow (pp. 109–123). Apress. https://doi.org/10.1007/978-1-4842-6168-2_10

Koonce, B. (2021b). MobileNetV3. *Convolutional* neural networks with swift for tensorflow (pp. 125–144). Apress. https://doi.org/10.1007/978-1-4842-6168-2_11

Kunwar, I. K., Singh, T., Machado, C. C., & Sinclair, J. (1986). Histopathology of soybean seed and seedling infection by *Macrophomina phaseolina*. *Phytopathology*, *76*, 532–535. https://www.apsnet.org/publications/phytopathology/backissues/Documents/1986Articles/Phyto76n05_532.PDF

LeCun, Y. (2015). *Deep learning & convolutional networks*. 2015 IEEE hot chips 27 symposium (HCS), Cupertino, CA, USA. https://doi.org/10.1109/hotchips.2015.7477328

Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation*. 2015 IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, USA. https://doi.org/10.1109/cvpr.2015.7298965

Lu, J., Tan, L., & Jiang, H. (2021). Review on convolutional neural network (CNN) applied to plant leaf disease classification. *Collection FAO: Agriculture*, *11*(8), 707.

Marquez, N., Giachero, M. L., Declerck, S., & Ducasse, D. A. (2021). *Macrophomina phaseolina*: General characteristics of pathogenicity and methods of control. *Frontiers in Plant Science*, *12*, 634397. https://doi.org/10.3389/fpls.2021.634397

Mechria, H., Gouider, M., & Hassine, K. (2019). Breast cancer detection using deep convolutional neural network. *Proceedings of the 11th* international conference on agents and artificial intelligence (Vol. 2, pp. 655–660). SciTePress. https://doi.org/10.5220/0007386206550660

Mertens, S., Verbraeken, L., Sprenger, H., Demuynck, K., Maleux, K., Cannoot, B., De Block, J., Maere, S., Nelissen, H., Bonaventure, G., Crafts-Brandner, S. J., Vogel, J. T., Bruce, W., Inzé, D., & Wuyts, N. (2021). Proximal hyperspectral imaging detects diurnal and drought-induced changes in maize physiology. *Frontiers in Plant Science*, *12*, 640914. https://doi.org/10.3389/fpls.2021.640914

Mihail, J. D., Orum, T. V., Alcorn, S. M., & Stroehlein, J. L. (1989). *Macrophomina phaseolina* in the Sonoran Desert. *Canadian Journal of Botany. Journal Canadien de Botanique*, *67*(1), 76–82.

Mihail, J. D., Ramussen, S. L., & Turner, B. C. (1992). *Macrophomina phaseolina*: A soilborne pathogen of *Salicornia bigelovii* in a marine habitat. *Plant Disease*, *76*, 751–752. https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=5611485

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *2016 Fourth International Conference on 3D vision (3DV)* (pp. 565–571). IEEE. https://doi.org/10.1109/3DV.2016.79

Ndlovu, E., van Staden, J., & Maphosa, M. (2021). Morpho-physiological effects of moisture, heat and combined stresses on *Sorghum bicolor* [Moench (L.)] and its acclimation mechanisms. *Plant Stress*, *2*, 100018. https://doi.org/10.1016/j.stress.2021.100018

Newbery, F., Qi, A., & Fitt, B. D. (2016). Modelling impacts of climate change on arable crop diseases: Progress, challenges and applications. *Current Opinion in Plant Biology*, *32*, 101–109. https://doi.org/10.1016/j.pbi.2016.07.002

Odvody, G. N., & Dunkle, L. D. (1979). Charcoal stalk rot of sorghum: Effect of environment on host-parasite relations. *Phytopathology*, *69*, 250–254. https://www.apsnet.org/publications/phytopathology/backissues/Documents/1979Articles/Phyto69n03_250.PDF https://doi.org/10.1094/Phyto-69-250

Okyere, F. G., Cudjoe, D., Sadeghi-Tehran, P., Virlet, N., Riche, A. B., Castle, M., Greche, L., Simms, D., Mhada, M., Mohareb, F., & Hawkesford, M. J. (2023). Modeling the spatial-spectral characteristics of plants for nutrient status identification using hyperspectral data and deep learning methods. *Frontiers in Plant Science*, *14*, 1209500. https://doi.org/10.3389/fpls.2023.1209500

Ottman, M. J. (2016). *Growing grain sorghum in Arizona*. https://repository.arizona.edu/handle/10150/625542

Overpeck, J. T., & Udall, B. (2020). Climate change and the aridification of North America [Review of *Climate change and the aridification of North America*]. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(22), 11856–11858. https://doi.org/10.1073/pnas.2006323117

Owens, J. D., Houston, M., Luebke, D., Green, S., Stone, J. E., & Phillips, J. C. (2008). GPU Computing. *Proceedings of the IEEE*, *96*(5), 879–899. https://doi.org/10.1109/JPROC.2008.917757

Pandey, P., Ramegowda, V., & Senthil-Kumar, M. (2015). Shared and unique responses of plants to multiple individual stresses and stress combinations: Physiological and molecular mechanisms. *Frontiers in Plant Science*, *6*, 723. https://doi.org/10.3389/fpls.2015.00723

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A. K., Chapman, J., Feltus, F. A., Gowik, U., … Rokhsar, D. S. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature*, *457*(7229), 551–556. https://doi.org/10.1038/nature07723

Pauli, D., Chapman, S. C., Bart, R., Topp, C. N., Lawrence-Dill, C. J., Poland, J., & Gore, M. A. (2016). The quest for understanding phenotypic variation via integrated approaches in the field environment. *Plant Physiology*, *172*(2), 622–634.

Poland, J. A., & Nelson, R. J. (2011). In the eye of the beholder: The effect of rater variability and different rating scales on QTL mapping. *Phytopathology*, *101*(2), 290–298. https://doi.org/10.1094/PHYTO-03-10-0087

Rai, K. M., Thu, S. W., Balasubramanian, V. K., Cobos, C. J., Disasa, T., & Mendu, V. (2016). Identification, characterization, and expression analysis of cell wall related genes in sorghum bicolor (L.) Moench, a food, fodder, and biofuel crop. *Frontiers in Plant Science*, *7*, 1287. https://doi.org/10.3389/fpls.2016.01287

Ramegowda, V., & Senthil-Kumar, M. (2015). The interactive effects of simultaneous biotic and abiotic stresses on plants: Mechanistic understanding from drought and pathogen combination. *Journal of Plant Physiology*, *176*, 47–54. https://doi.org/10.1016/j.jplph.2014.11.008

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. Wells, & A. Frangi (Eds.), *Lecture notes in computer science* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

Rooney, W. L., Blumenthal, J., Bean, B., & Mullet, J. E. (2007). Designing sorghum as a dedicated bioenergy feedstock. *Biofuels, Bioproducts & Biorefining*, *1*(2), 147–157.

Roozeboom, K. L., & Prasad, P. V. V. (2019). Sorghum growth and development. *Sorghum* (pp. 155–172). Soil Science Society of America.

Rybski, P. E., Huber, D., Morris, D. D., & Hoffman, R. (2010). *Visual classification of coarse vehicle orientation using Histogram of Oriented Gradients features*. 2010 IEEE intelligent vehicles symposium, La Jolla, CA, USA. https://doi.org/10.1109/ivs.2010.5547996

Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Tversky loss function for image segmentation using 3D fully convolutional deep networks. In Q. Wang, Y. Shi, H. Suk, & K. Suzuki (Eds.), *Machine learning in medical imaging* (pp. 379–387). Springer. https://doi.org/10.1007/978-3-319-67389-9_44

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution*, *3*(3), 430–439.

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Siar, M., & Teshnehlab, M. (2019). *Brain tumor detection using deep neural network and machine learning algorithm*. 2019 9th international conference on computer and knowledge engineering (ICCKE), Mashhad, Iran. https://doi.org/10.1109/iccke48569.2019.8964846

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. https://doi.org/10.1016/j.tplants.2015.10.015

St Clair, D. A. (2010). Quantitative disease resistance and quantitative resistance Loci in breeding. *Annual Review of Phytopathology*, *48*, 247–268. https://doi.org/10.1146/annurev-phyto-080508-081904

Stewart, E. L., Wiesner-Hanks, T., Kaczmar, N., DeChant, C., Wu, H., Lipson, H., Nelson, R. J., & Gore, M. A. (2019). Quantitative phenotyping of northern leaf blight in UAV images using deep learning. *Remote Sensing*, *11*(19), 2209. https://doi.org/10.3390/rs11192209

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*, *97*, 6105–6114. http://proceedings.mlr.press/v97/tan19a.html

Tang, C., Li, S., Li, M., & Xie, G. H. (2018). Bioethanol potential of energy sorghum grown on marginal and arable lands. *Frontiers in Plant Science*, *9*, 440. https://doi.org/10.3389/fpls.2018.00440

Thorp, K. R., Hunsaker, D. J., Bronson, K. F., Andrade-Sanchez, P., & Barnes, E. M. (2017). Cotton irrigation scheduling using a crop growth model and FAO-56 methods: Field and simulation studies. *Transactions of the ASABE*, *60*(6), 2023–2039. https://doi.org/10.13031/trans.12323

Wiesner-Hanks, T., Wu, H., Stewart, E., DeChant, C., Kaczmar, N., Lipson, H., Gore, M. A., & Nelson, R. J. (2019). Millimeter-level plant disease detection from aerial photographs deep learning and crowdsourced data. *Frontiers in Plant Science*, *10*, 1550. https://doi.org/10.3389/fpls.2019.01550

Wu, H., & Gu, X. (2015). Max-pooling dropout for regularization of convolutional neural networks. In S. Arik, T. Huang, W. Lai, & Q. Liu (Eds.), *Neural information processing* (pp. 46–54). Springer. https://doi.org/10.1007/978-3-319-26532-2_6

Wu, H., Wiesner-Hanks, T., Stewart, E. L., DeChant, C., Kaczmar, N., Gore, M. A., Nelson, R. J., & Lipson, H. (2019). Autonomous detection of plant disease symptoms directly from aerial imagery. *The Plant Phenome Journal*, *2*(1), 1–9. https://doi.org/10.2135/tppj2019.03.0006

Yan, W., Deng, X. W., Yang, C., & Tang, X. (2021). The genome-wide EMS Mutagenesis bias correlates with sequence context and chromatin structure in rice. *Frontiers in Plant Science*, *12*, 579675. https://doi.org/10.3389/fpls.2021.579675

Yang, Q., Van Haute, M., Korth, N., Sattler, S. E., Toy, J., Rose, D. J., Schnable, J. C., & Benson, A. K. (2022). Genetic analysis of seed traits in Sorghum bicolor that affect the human gut microbiome. *Nature Communications*, *13*(1), 5641. https://doi.org/10.1038/s41467-022-33419-1

Yu, K., Anderegg, J., Mikaberidze, A., Karisto, P., Mascher, F., McDonald, B. A., Walter, A., & Hund, A. (2018). Hyperspectral canopy sensing of wheat *Septoria tritici* blotch disease. *Frontiers in Plant Science*, *9*, 1195. https://doi.org/10.3389/fpls.2018.01195

Zhang, G., Xu, T., Tian, Y., Feng, S., Zhao, D., & Guo, Z. (2022). Classification of rice leaf blast severity using hyperspectral imaging. *Scientific Reports*, *12*(1), 19757. https://doi.org/10.1038/s41598-022-22074-7

Zhang, H., & Sonnewald, U. (2017). Differences and commonalities of plant responses to single and combined stresses. *The Plant Journal:*

For Cell and Molecular Biology, *90*(5), 839–855. https://doi.org/10.1111/tpj.13557

Zhang, Z. (2018). *Improved Adam optimizer for deep neural networks* [Paper presentation]. 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, Canada. https://doi.org/10.1109/IWQoS.2018.8624183

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gonzalez, E., Zarei, A., Calleja, S., Christenson, C., Rozzi, B., Demieville, J., Hu, J., Eveland, A. L., Dilkes, B., Barnard, K., Lyons, E., & Pauli, D. (2024). Quantifying leaf symptoms of sorghum charcoal rot in images of field-grown plants using deep neural networks. *The Plant Phenome Journal*, *7*, e20110. https://doi.org/10.1002/ppj2.20110