

METHODOLOGY

Open Access



# A comparison of ImageJ and machine learning based image analysis methods to measure cassava bacterial blight disease severity

Kiona Elliott<sup>1,2</sup>, Jeffrey C. Berry<sup>1</sup>, Hobin Kim<sup>3</sup> and Rebecca S. Bart<sup>1\*</sup>

## Abstract

**Background:** Methods to accurately quantify disease severity are fundamental to plant pathogen interaction studies. Commonly used methods include visual scoring of disease symptoms, tracking pathogen growth *in planta* over time, and various assays that detect plant defense responses. Several image-based methods for phenotyping of plant disease symptoms have also been developed. Each of these methods has different advantages and limitations which should be carefully considered when choosing an approach and interpreting the results.

**Results:** In this paper, we developed two image analysis methods and tested their ability to quantify different aspects of disease lesions in the cassava-*Xanthomonas* pathosystem. The first method uses ImageJ, an open-source platform widely used in the biological sciences. The second method is a few-shot support vector machine learning tool that uses a classifier file trained with five representative infected leaf images for lesion recognition. Cassava leaves were syringe infiltrated with wildtype *Xanthomonas*, a *Xanthomonas* mutant with decreased virulence, and mock treatments. Digital images of infected leaves were captured overtime using a Raspberry Pi camera. The image analysis methods were analyzed and compared for the ability to segment the lesion from the background and accurately capture and measure differences between the treatment types.

**Conclusions:** Both image analysis methods presented in this paper allow for accurate segmentation of disease lesions from the non-infected plant. Specifically, at 4-, 6-, and 9-days post inoculation (DPI), both methods provided quantitative differences in disease symptoms between different treatment types. Thus, either method could be applied to extract information about disease severity. Strengths and weaknesses of each approach are discussed.

**Keywords:** Image analysis, Disease symptoms, ImageJ, Machine learning, Cassava bacterial blight

## Background

Annually 20–40% of crops are lost due to plant pests and disease (FAO [1]). Causal agents of plant disease such as bacteria, viruses, oomycetes, and fungi employ various strategies to promote pathogenesis and elicit

disease susceptibility in host plants. Disease susceptibility is commonly measured by the amount of *in planta* pathogen growth, reduction in crop yield/biomass, or by scaled scoring systems that use visible disease symptoms to measure severity (Strange [2], Liu [3], Guant [4], Moore [5]). Each of these methods have advantages and limitations and no single method can capture the full complexity of plant disease. For instance, it is common to introduce a small number of bacteria into a plant leaf and then quantify pathogen growth overtime (Agrios

\*Correspondence: [rbart@danforthcenter.org](mailto:rbart@danforthcenter.org)

<sup>1</sup> Donald Danforth Plant Science Center, Saint Louis, MO 63132, USA  
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

5th edition [6]). This method highly quantitative and can reveal subtle differences in virulence between related pathogen strains or mutants (Bart [7], Cohn and Bart [8], Diaz [9]). However, this assay probes only one part of the disease cycle and provides limited insight into pathogen spread, plant symptoms or defense responses. Another common method is to visually score disease symptoms on a numerical scale (Jorge and Verdier [10]). This method can be used in lab to field level experiments, is cost effective, and does not require special techniques or tools. However, accurate identification of pathogen incited symptoms can be difficult, especially in the case of multiple biotic and/or abiotic stresses. Further, disease scores may vary among different scorers and often are not sensitive enough to capture subtle changes in disease severity (Poland and Nelson [11], Strange and Scott [12]).

In recent years, there has been an increase in the use of image-based methods to analyze and measure plant health (Gehan [13], Laflamme [14], Lobet [15]). Images can be captured through many different platforms including cell phones, imaging chambers, high-throughput phenotyping facilities, drones, and satellites (Li [16], Zhang and Zhang [17]) and many analysis platforms have also been developed, for example, ImageJ (Ferreira and Rasband [18]). Image-based phenotyping tools have been successfully developed to study a broad range of plant diseases including citrus canker (Bock [19]), grapevine powdery mildew (Bierman [20]), and cereal rust disease (Gallego-Sanchez [21]). At least in some cases, image-based phenotyping can overcome some of the limitations associated with the more traditional methods described above (Mutka and Bart [22]). For example, a study investigating *Zymoseptoria tritici* infected wheat leaves found that an ImageJ analysis method provided more reliable and reproducible measures of wheat blotch disease compared to a traditional visual scoring system (Stewart [23], Stewart [24]). However, manual image analysis based on user selection of disease lesions can also be time consuming. Some image analysis methods have incorporated machine learning techniques for improved trait identification, classification, and faster analysis of plant disease symptoms (Singh [25], Tsaftaris [26]). While machine learning has enhanced the ability to process imaging data, accurate trait classification or quantification often relies on large datasets that can be expensive to acquire. Therefore, more cost effective, few-shot image analysis tools that allow for efficient segmentation and quantification of disease symptoms are needed.

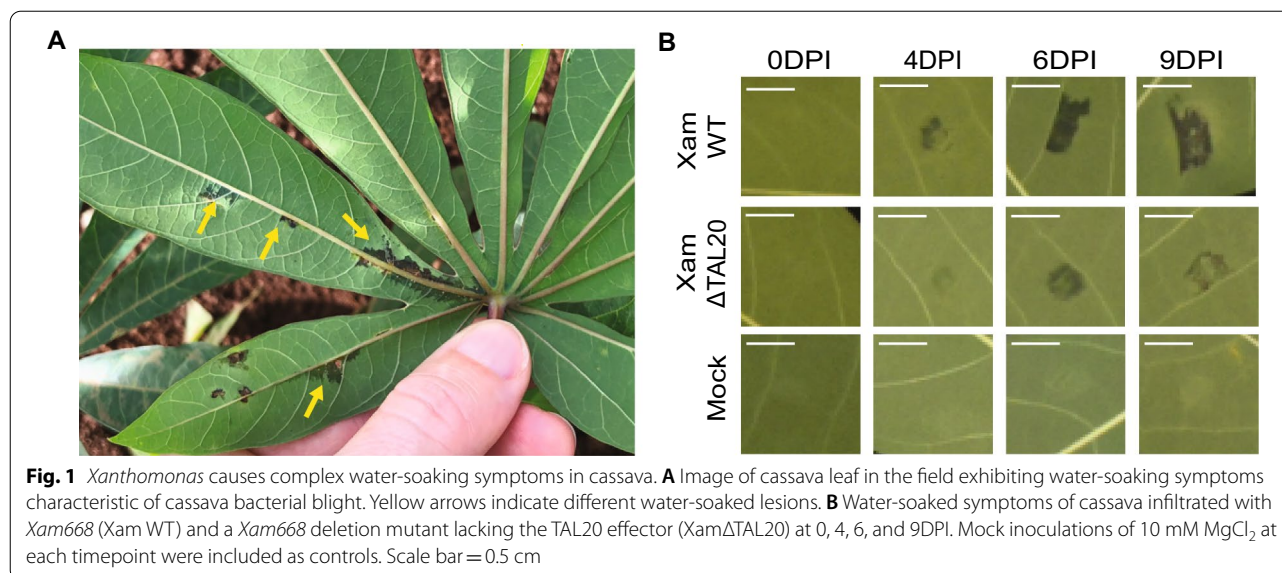
In this study, we apply image-based phenotyping to cassava (*Manihot esculenta* Crantz), a starchy storage root crop (Morgan [27]). Cassava is a hardy crop predominantly grown by smallholder farmers in South America, East Asia, and Sub-Saharan Africa (Bart and

Taylor [28], Hillock [29], El-Sharkawy [30]). Cassava production is threatened by the disease cassava bacterial blight (CBB). CBB can result in complete crop loss and is present in all cassava growing regions (Howler [31], Fanuo [32], Zárate-Chaves [33]). The causal agent of CBB is *Xanthomonas axonopodis* pv. *manihotis* also referred to as *Xanthomonas phaseoli* pv. *manihotis* (*Xam* or *Xpm*) (Constantin [34]). *Xam* infects cassava by entering through open stomata or wounds in the leaf, colonizes the surface of mesophyll cells, and spreads systemically in the plant. The first visible indicators of CBB disease are dark “water-soaked” lesions that appear on the leaf. Water-soaked lesions or spots are a common, early disease symptom of various bacterial diseases. (Aung [35]). Other CBB disease symptoms include leaf wilt, defoliation, stem browning, and eventual plant death. Like other plant pathogens, *Xam* has a repertoire of effectors that can alter the structure or function of a host cell, create a more ideal environment for pathogen colonization, and overcome plant defense mechanisms (Boch [36], Hogenhout [37]). In the *Xanthomonas* and *Ralstonia* bacterial genera, this repertoire includes specialized transcription activator-like (TAL) effectors (Bodnar [38], Van Schie and Takken [39], Koseoglou [40]). TAL effectors are secreted into the plant cell and induce expression of plant susceptibility (S) genes that enhance disease. In many pathosystems, TAL effectors target *SWEET* (Sugars Will Eventually be Exported Transporters) genes and preventing this interaction reduces disease symptoms (Li [41], Phillips [42], Cox [43]). The *Xam* strain used in this study, *Xam668*, carries the effector, TAL20, which induces ectopic expression of *MeSWEET10a* (Cohn and Bart [8]). *Xam668* mutants with loss of TAL20 (*Xam668ΔTAL20*) exhibit visibly reduced water-soaked lesions compared to wild-type *Xam*. Here, we develop and compare ImageJ and machine learning based image analysis tools that allow for segmentation and quantification of CBB induced water-soaked lesions.

## Results

### **Xam induction of water-soaked lesions in cassava**

In cassava, water-soaked lesions appear as dark angular spots at the site of infection and spread as the bacteria proliferate (Fig. 1A). To capture the progression of water-soaking in cassava, leaves were syringe-infiltrated with *Xam668*, *Xam668ΔTAL20*, or mock treatments. At 0-, 4-, 6-, and 9-days post inoculation (DPI) infected leaves were detached from the plant and imaged. Images were taken with a Raspberry Pi camera in an enclosed box to increase uniformity of imaging. An X-Rite ColorChecker Passport was included in every image for post-acquisition gray balance color correction (Berry [44]). At 4DPI, water-soaked spots began to appear in both *Xam668*



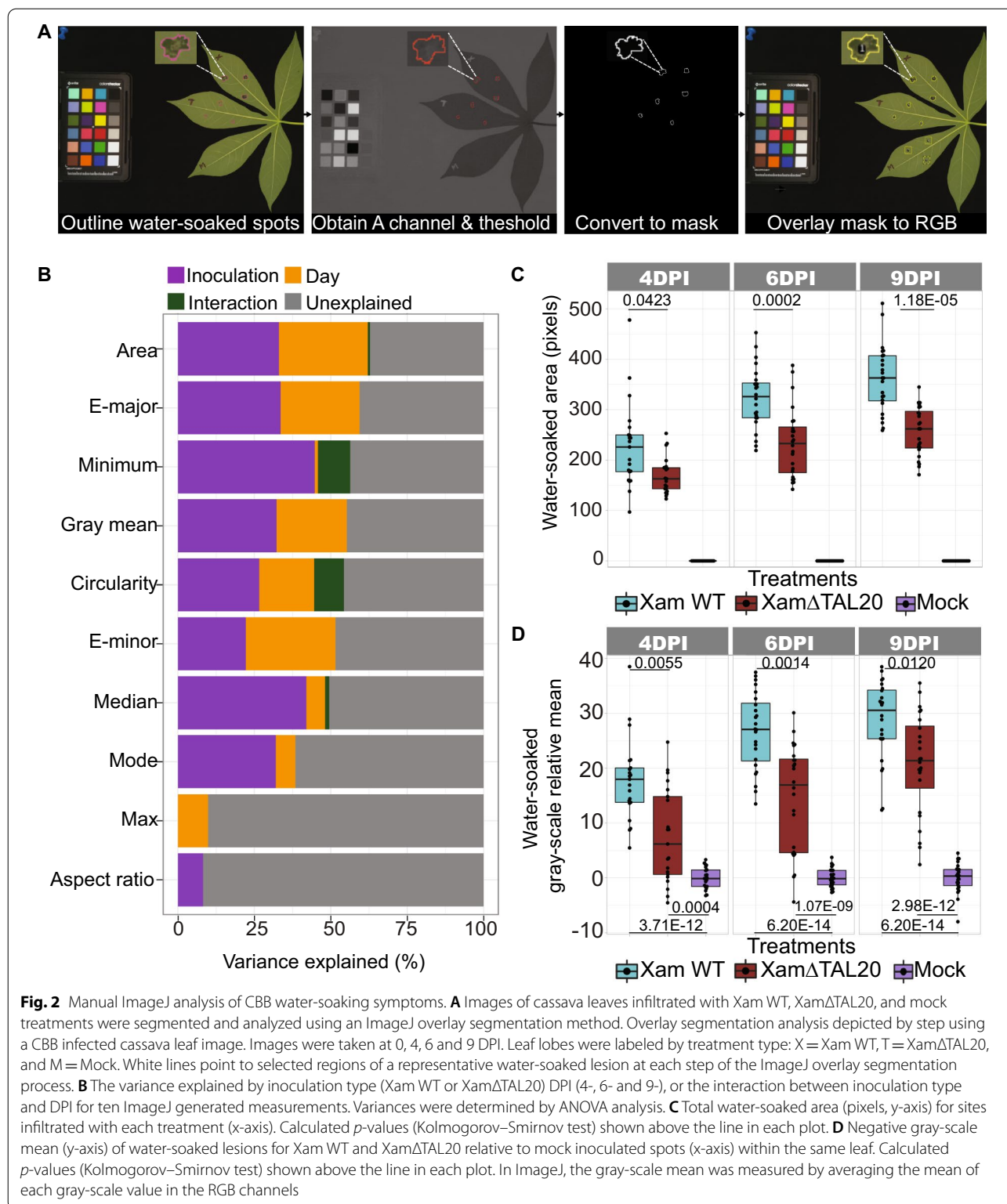
(Xam WT) and *Xam668 $\Delta$ TAL20 (Xam $\Delta$ TAL20) infiltration sites (Fig. 1B). Water-soaked lesions spread and increased in visibility at 6 and 9 DPI. However, as previously reported [8] water-soaking appeared reduced in *Xam668 $\Delta$ TAL20 infection sites as compared to wildtype *Xam668* sites. Additionally, *Xam668 $\Delta$ TAL20 infection sites appeared lighter in color compared to the darker lesions that develop at wildtype *Xam668* sites. Water-soaked lesions were not observed for any time point in mock infiltrated spots.***

#### ImageJ based quantification of water-soaked symptoms

ImageJ is regularly used for image analysis in biological studies (Ferreira and Rasband [18]). Here, we applied ImageJ based analysis to extract, quantify, and examine water-soaked lesion traits. Water-soaked lesions induced by *Xam668* and *Xam668 $\Delta$ TAL20 were segmented using a manual overlay segmentation strategy (Fig. 2A). For segmentation, color corrected images were uploaded and duplicated in ImageJ and the *Xam668* and *Xam668 $\Delta$ TAL20 lesions were outlined using the pencil tool. Outlined images were converted from RGB to the LAB color space and the “A Channel” was obtained for better separation of the outlined lesions from the leaf background. The A channel images were thresholded and converted to a binary mask. The binary masks and analyze particle tool in ImageJ were used to define the *Xam668* and *Xam668 $\Delta$ TAL20 infected sites and an overlay was created for each image. The overlays were applied to the RGB image and measurements for 27 traits were calculated. Mock sites were measured using the rectangle selection tool in the RGB image to capture information about “non-water-soaked” leaf background. ImageJ***

processing took approximately 6 min and 30 s per image. A movie example of the ImageJ based analysis method was generated as a tutorial (Additional file 1).

Ten traits were selected and further analyzed using an ANOVA analysis to determine the variance explained (VE) by three terms of interest: (1) inoculation type, (2) DPI and (3) the interaction between inoculation type and DPI (Fig. 2B). Inoculation type and DPI were selected as defining factors because we expected that water-soaking severity is dependent on these terms. Area had the highest amount of VE, with over 60% VE. We selected gray-scale mean as another trait of interest because of the color difference we observed between *Xam668* and *Xam668 $\Delta$ TAL20 water-soaked lesions. Gray-scale mean, accounted for over 50% VE. Water-soaked area (Fig. 2C) and gray-scale mean (Fig. 2D) were further analyzed as measures of CBB disease severity. The *Xam668* sites had significantly more water-soaked area compared to *Xam668 $\Delta$ TAL20 at each timepoint. We found there was noise in the gray-scale mean data due to lack of standardization across individual images despite gray balance color correction. To account for this, a linear model was applied to determine the grand mean of all gray values in each image and the *Xam668* and *Xam668 $\Delta$ TAL20 gray values were centered to mock. In each timepoint, *Xam668* treatment resulted in lesions that had a significantly larger gray-scale mean compared to *Xam668 $\Delta$ TAL20 treatment. A greater difference in gray-scale mean was observed between *Xam668* and mock treated spots compared to *Xam668 $\Delta$ TAL20 and mock spots. These results indicate that ImageJ based segmentation allowed for separation of treatment types and for the quantitative analysis of water-soaked lesions over time.*****





### Machine learning based quantification of water-soaked symptoms

While ImageJ provided sufficient segmentation of water-soaked lesions, developing an overlay mask for every individual image is time intensive. Therefore, we sought to develop a machine learning tool that would provide faster segmentation and quantification of diseased leaves. A custom workflow for machine learning disease lesion analysis was developed using the source file from PhenotyperCV, a C++ library designed for image-based phenotyping (Berry [44]). The machine learning workflow was run using the Mac terminal. Command syntax specific for each step of the machine learning tool was developed (Additional file 2). Five representative images of CBB infected leaves from different DPI were selected and combined into one graphic as a training image for the machine learning tool (Fig. 3A). A binary mask was generated from the combined leaf graphic using ImageJ. The mask was used to generate a support vector machine (SVM) learning classifier (YAML) file. The classifier file was used to process the images and eliminated the need to manually outline each lesion or make individual masks (Fig. 3B). During processing, images were color corrected and manually thresholded using a scale bar built into the program to reduce background noise and enhance segmentation of lesion pixels. Next, infiltrated spots were manually labelled and color-coded by treatment type. Output images were generated and included color corrected, pseudo-color map, and feature prediction images for every image analyzed (Fig. 3C). Machine learning processing took approximately 2 min and 30 s per image. Processing speed increased when all images were analyzed using an iteration (for loop) command in terminal allowing the machine learning tool to be executed on several images in succession. A movie example of the machine learning based analysis method was generated as a tutorial (Additional file 3). Additionally, two space separated text (TXT) files were produced with shape and color related measurements of each lesion. A list of the reported measurements is included (Additional file 4). Shape data generated by the machine learning tool includes area, hull area, height, width, etc. The color data generated by machine learning is a lightness histogram of 0–255 for each lesion which was used to calculate lesion gray-scale mean.

Twelve machine learning derived traits were selected and the ANOVA analysis was used to measure VE by each trait (Fig. 4A). Area measured by the machine learning tool had over 75% VE by the defining factors. As was determined during ImageJ analysis, area also accounts for the highest amount of VE in the machine learning analysis. The gray-scale mean had over 60% VE by the defining factors. Consistent with the ImageJ analysis, the machine

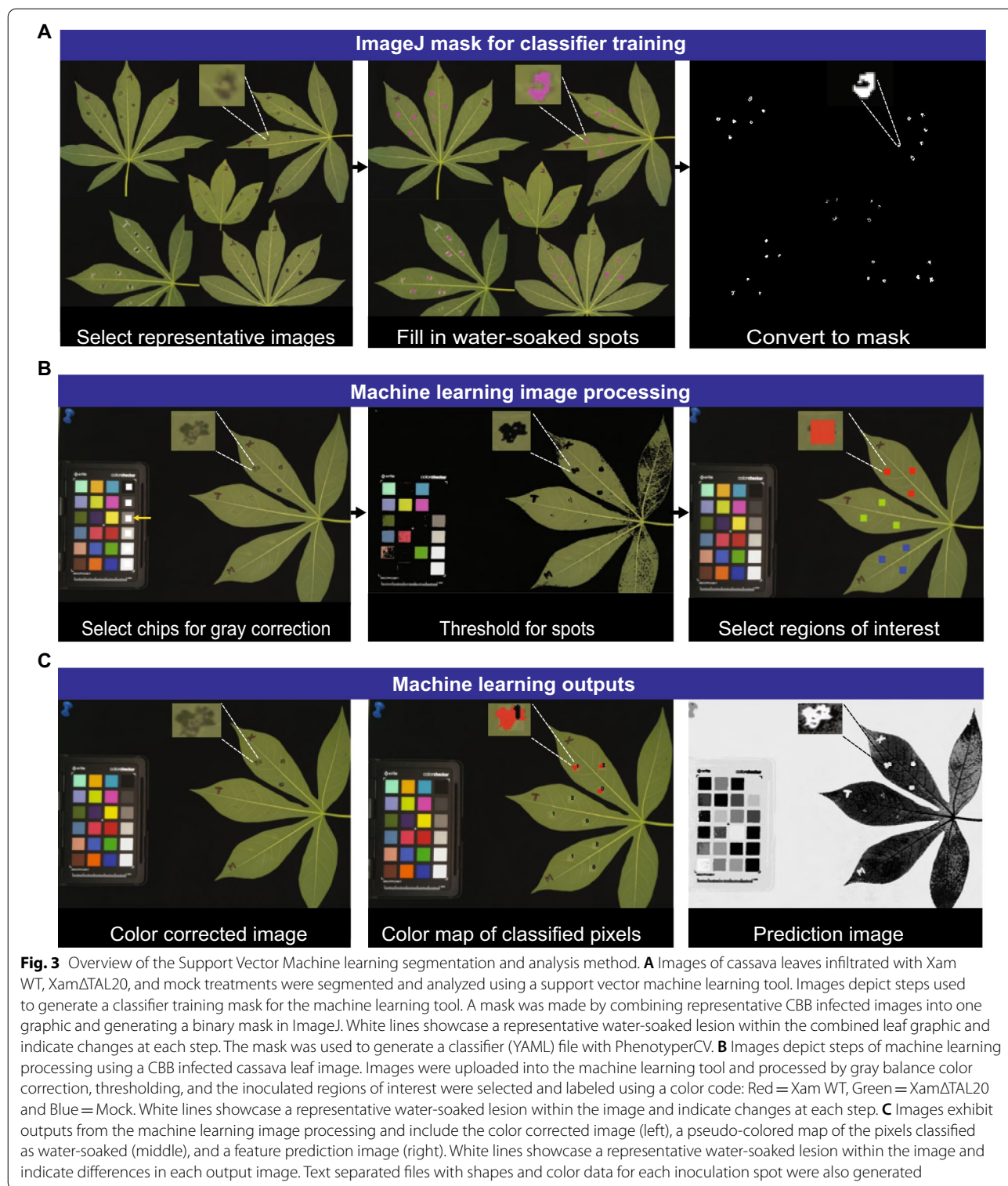
learning approach revealed that *Xam668* caused a larger water-soaked area (Fig. 4B) and relative gray-scale mean (Fig. 4C) compared to *Xam668ΔTAL20* infiltrated spots. These data suggest that the machine learning tool adequately distinguished between treatment types and provided quantitative measures of water-soaked lesions using the classifier file created from one training mask.

### Comparison of the ImageJ and Machine learning based lesion analysis methods

The ImageJ and machine learning based methods both successfully distinguished *Xam668* and *Xam668ΔTAL20* and yet the results were not equivalent. To further compare and contrast these methods, representative *Xam668* and *Xam668ΔTAL20* lesions from 4-, 6-, and 9-DPI were selected and visually inspected (Fig. 5A). We observed that machine learning was able to distinguish between water-soaked and “non-water-soaked” pixels within the lesion spot whereas in ImageJ, a boundary was put around the whole spot and could include a mix of both pixel types. This suggests that the machine learning tool is more selective in classification of water-soaked versus non-water-soaked pixels and would explain the trend of overall smaller area measurements generated by machine learning compared to ImageJ. In ImageJ, the lesion boundary is user-selected. However, to completely separate water-soaked from non-water-soaked pixels in lesions where there is a mix, smaller independent boundaries would be required. Having multiple boundaries for one lesion is not ideal as it would impact measures such as gray-scale mean and increase image processing time. The two image analysis methods were statistically compared by pairing the mock, *Xam668* and *Xam668ΔTAL20* area data and performing F-statistic variance tests on each respective treatment type (Fig. 5B). At each time-point, there was no significant difference in the variance observed between ImageJ and machine learning data suggesting the two methods have equal variation within each treatment type.

### Discussion

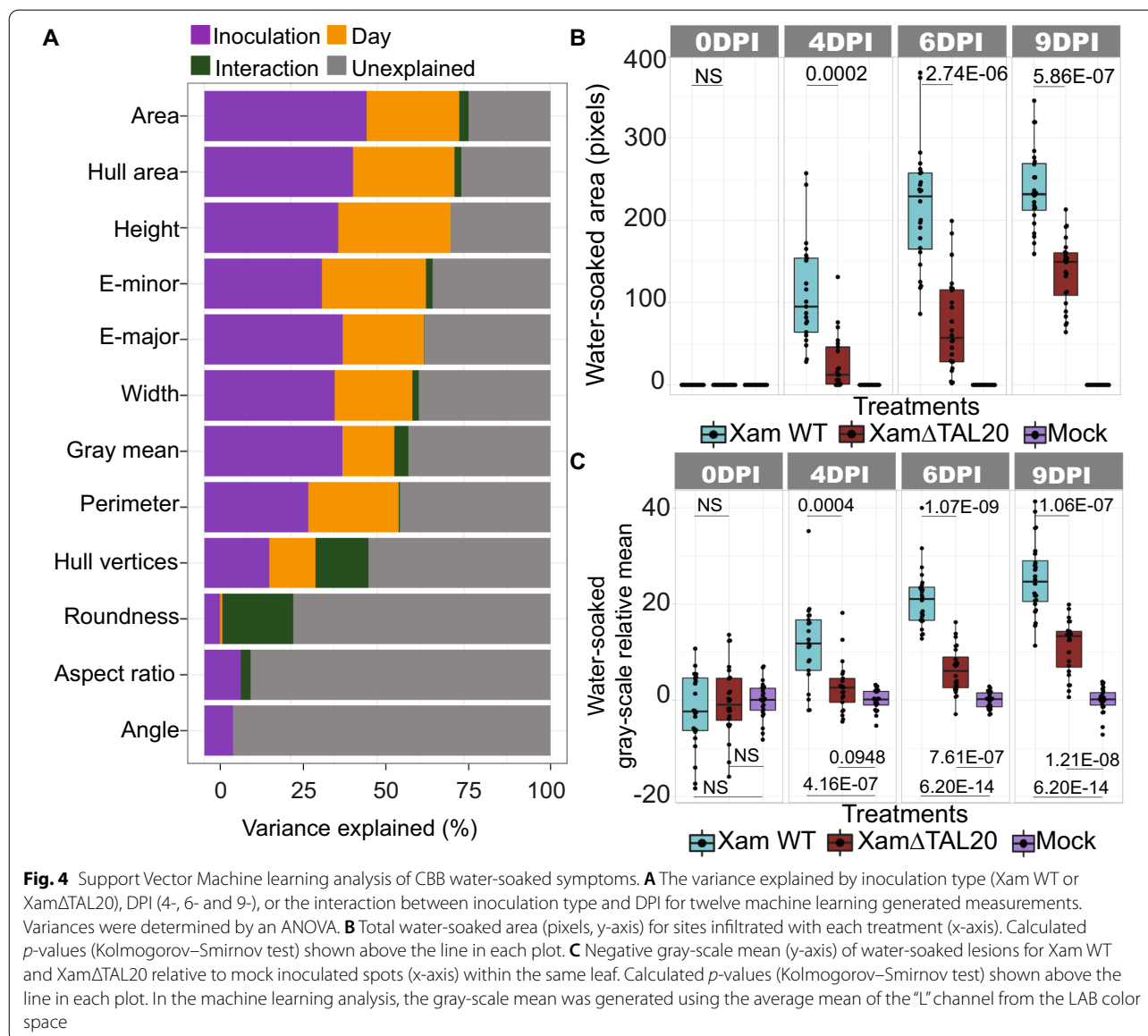
To quantify CBB, we developed and compared ImageJ and machine learning image analysis methods for accurate segmentation and quantification of water-soaked lesion symptoms. We found that an ImageJ overlay segmentation method allowed for adequate separation between cassava infected with mock, *Xam668* and *Xam668ΔTAL20* treatments based area and gray-scale mean values of disease lesions. However, the ImageJ analysis was time-consuming because an individual mask had to be made for every image analyzed. Other ImageJ analysis methods tested with this data set such as non-segmentation and color-threshold based segmentation



of water-soaked lesions failed to accurately capture the water-soaking phenotype.

Machine learning has previously been applied to detect and measure several cassava diseases including

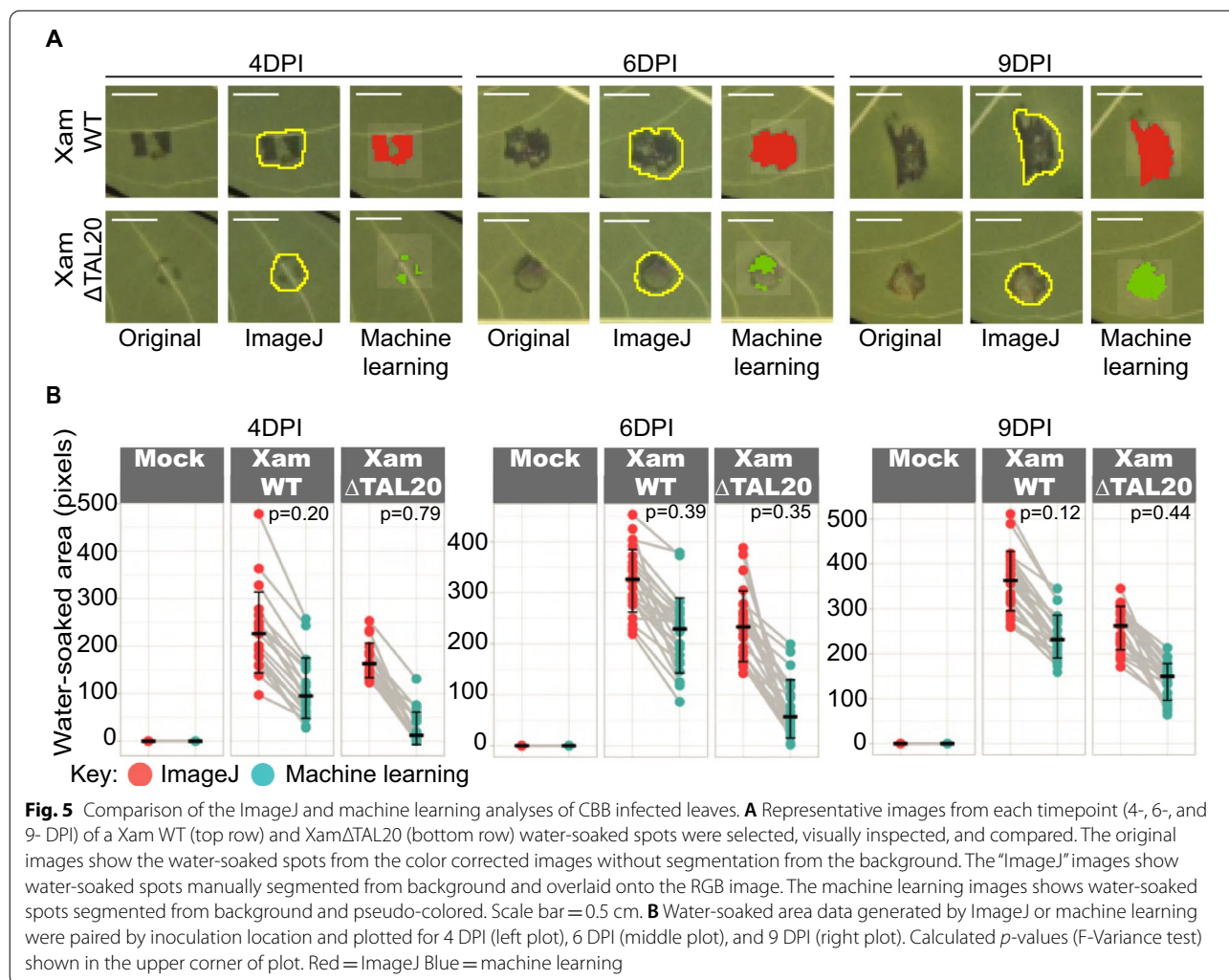
bacterial blight, brown streak and mosaic disease (Sangbamrung [45], Ramcharan [46]). However, these tools rely on hundreds to thousands of images for classifier training. Any machine learning tool is heavily



reliant on its classifier file for adequate segmentation and measure of an object of interest. If a classifier file does not adequately capture the range of traits for an object of interest, classification of that object will fail. To determine if a classifier file would work accurately for our data set, we tested its predictive capability by spot checking analysis accuracy in a subset of images and visually inspecting classification of pixels defined as water-soaked. We initially developed classifier files based on a single representative CBB infected leaf image and found it could not reliably predict features of interest for all images. However, by combining representative images of cassava infected with three replicates each of mock, *Xam*668, and *Xam*668Δ*TAL20* treatments across different timepoints into one training

graphic, we developed a classifier that better predicted water-soaked lesions. The accuracy of the combined leaf graphic was tested by again spot-checking a subset of color map images and inspecting classification of pixels defined as “water-soaked”. Similarly, our classifier file was developed using one genotype of cassava, TME419. In future studies, if this approach were to be applied to datasets derived from multiple genotypes or a breeding program, the classifier file would need to be updated with representative images to capture any additional variability in leaf traits.

Another important consideration for classifier file development is the machine learning algorithm used. The machine learning workflow presented here functions with either support vector machine (SVM) or Naïve



Bayes learning algorithms. During testing of classifier files, we found that SVM training files predicted water-soaked lesion features in our system more accurately than Naïve Bayes. Similarly, a previous study tested three machine learning methods and reported that SVM had high performance in predicting and classifying cassava diseases (Ramcharan [47]).

Despite the limitations, we found that the few-shot machine learning based image analysis tool presented here offered a fast and accurate approach to segment water-soaked lesions. Processing for the machine learning tool took less than half the time of ImageJ based analysis for each image. The machine learning tool worked as well as the ImageJ overlay segmentation method for separating lesions by treatment type and extracting quantifiable data. Due to the time needed to validate a classifier file, we suggest that a machine learning approach for image-based lesion analysis is appropriate when there is a large number of images to

be processed. If the data set is small, ImageJ could be a faster approach as the accuracy of the method does not rely on a classifier file. Moreover, manual thresholding is still required for segmentation of the lesions in each image and may be slightly variable within the data set. Thresholding performed within either the machine learning or ImageJ methods requires user decision to determine the threshold cut-off. In the case of the machine learning tool, it is important to inspect the color maps generated for each image analyzed to ensure proper classification of water-soaked lesions. In some cases, we found it necessary to re-process images in the machine learning tool and adjust the threshold for more precise capture of a lesion.

While improvement is still needed in image-based phenotyping, there are several potential uses for the machine learning and ImageJ analyses presented in this study. Image based phenotyping has become increasingly popular for examining the link between disease symptoms



and genetics in plant science (Casto [48]). The tools presented here provide a new resource for experiments investigating CBB disease susceptibility. Additionally, the general framework of the machine learning workflow can be applied to other plant species and disease symptoms using classifier files representative of the disease of interest.

## Conclusions

To quantify CBB, we developed and compared ImageJ and machine learning image analysis methods for accurate segmentation and quantification of water-soaked lesion symptoms. Both the ImageJ and machine learning image analysis methods are described in detail, along with video tutorials and we hope these resources will help other researchers use these tools and/or design similar tools that can be applied to other pathosystems. We found that both methods accurately distinguished between and quantified different water-soaked lesion types in the cassava-*Xanthomonas* pathosystem. The ImageJ method is best used from smaller datasets as it relies on the user developing a mask for every image. The machine learning based tool is best used for larger datasets as it is more time efficient to develop a single classifier file to process many images. Many machine learning tools rely on thousands of training images for accurate function. However, the machine learning tool presented here is few-shot learning based and functions as well as ImageJ for disease segmentation and measurement.

## Methods

### Plant materials and growing conditions

Cassava plants from the cultivar TME419 were kept in greenhouse conditions set to 28 °C; 50% humidity; 16 h light/8 h dark and 1000 W light fixtures that supplemented natural light levels below 400 W/m<sup>2</sup>. Cuttings were taken from the woody stem of mature plants and propagated to 4-inch pots of Berger45 soil. 4–5-week-old propagated plants that were well established were used for infection experiments. During infection experiments, plants were kept in a post-inoculation room set to 50% humidity, ambient room temperature, 12 h light/12 h dark and 32 W light fixtures.

### Bacterial inoculations

*Xanthomonas* strains were struck from glycerol stocks onto NYG agar plates containing appropriate antibiotics. The strains used for this study were *Xam668* (rifampicin 50 µg/ml) and *Xam668ΔTAL20* (suicide vector knockout (Cohn and Bart [8]) tetracycline 5 µg/ml, rifampicin 50 µg/ml). *Xanthomonas* strains were grown in a 30 °C incubator for 2–3 days. Inoculum for each strain was made by transferring bacteria from plates into 10 mM

MgCl<sub>2</sub> using inoculation loops and brought up to a concentration of OD<sub>600</sub>=0.01. Leaves from 4–5-week-old cassava plants were inoculated using a 1.0 mL needleless syringe. For each replicate assay, two cassava plants were used for inoculations and four leaves were inoculated on each plant. One bacterial strain was inoculated per leaf lobe with three injection sites. Mock inoculations of 10 mM MgCl<sub>2</sub> alone were included resulting in nine infiltrated sites per leaf. Four replicate rounds of inoculations were done in total.

### Imaging

Cassava leaves were detached and imaged at 0-, 4-, 6-, and 9-days post inoculation (DPI). One leaf from each cassava plant was collected and imaged for a total of two leaves per timepoint. In all, thirty-two leaves were imaged and analyzed across four replicate rounds of inoculations. Leaves were imaged from above using a Raspberry Pi Sony IMX219 camera in an enclosed box with an overhead light. To account for setting inconsistencies between images, images were color-corrected by gray balancing using a X-Rite ColorChecker Passport color card. Images were uploaded to the machine learning workflow and six gray color chips (black-white) were manually selected using a selection tool built into the program. Saturation of each chip was estimated and the brightness of each image was adjusted accordingly. The gray corrected images were then used for water-soaking analysis. Analytical standardization of the gray values post-image-processing by ImageJ and machine learning was performed separately by estimating the grand mean of all gray values within each image and centering those values to the grand mean across all images. This is achieved by creating a linear model with a single fixed effect term accounting for each image and extracting model residuals.

### ImageJ image analysis

Gray corrected images were uploaded to ImageJ version FIJI (Schindelin [49]) and duplicated. Water-soaked lesions were manually outlined on the duplicate image using the pencil tool (color: #ff00b6 and size 2). The outlined images were converted from RGB to LAB and split to obtain the A color channel. The A channel images were thresholded, converted to a mask and the mask for each spot was added to the ROI manager using the analyze particle tool. The ROI masks were applied to the original RGB gray corrected images. Mock infiltrated spots (no water-soaking, plant background data) were added to the ROI manager using an arbitrarily sized rectangle selection tool consistently set to a W=26 and H=30. Area, gray-scale mean, and eight other measurement data were obtained for each infiltrated spot using the FIJI

measure tool. The measurements were saved as a comma separated value (CSV) file. The variance explained by ten image J derived traits were calculated and plotted in the software program R using a custom partial correlations script. Area and gray-scale mean data for all lesions were compared across different treatment types and timepoints using a Kolmogorov–Smirnov (KS) statistical test in R. All plots were generated in R with a dpi = 300, width = 8.66, and height = 6.86.

### Machine learning image analysis

Five images of *Xanthomonas* inoculated cassava leaves from different timepoints were selected as representatives to make a classifier file for the machine learning image analysis tool. The images were combined into one graphic, uploaded to ImageJ, and water-soaked spots were outlined and filled in using the pencil tool (color: #ff00b6). The outlined combined leaf image was converted to a binary mask and referred to as the “labeled image”. The machine learning image analysis tool is part of PhenotyperCV, a C++11 header-only library designed for image-based plant phenotyping. The machine learning workflow and software download instructions are available on GitHub.

([https://github.com/jberry47/ddpsc\\_phenotypercv/wiki/Machine-Learning-Workflow](https://github.com/jberry47/ddpsc_phenotypercv/wiki/Machine-Learning-Workflow)).

All steps of the machine learning workflow were run on the Mac terminal command line. The labeled leaf mask image and original combined leaf graphic were used to create a support vector machine learning classifier or YAML file. Individual images of inoculated cassava leaves were processed in the machine learning tool by uploading the images and gray correcting. The images were thresholded using a scale bar built into the program to set a cut-off for pixels that can be classified as water-soaked. The inoculated sites were manually selected with a color-coded region of interest (ROI) selector (mouse right click-red, left click-green, and middle click-blue). The ROI selector tool size ranges from 0 to 20. The ROI size was consistently set to 11 for this study. The ROI selector does not restrict the size of the object identified as a water-soaked lesion. If a part of the object defined as a lesion is included in the ROI selection, then the entire object will be labelled and color-coded. For this study, we designated red as *Xam668*, green as *Xam668ΔTAL20*, and blue as mock inoculation spots. If color-code separation is not required for other studies using the machine learning tool, one click/color type can be used for all lesion selections. Outputs from the workflow include a color corrected image (also used in the ImageJ analysis), a prediction image of what could be captured as pixels of interest, and a pseudo-colored map image showing

what was captured as pixels of interest. Additionally, two space separated text files were generated with measurement data about the shape and color of each lesion. The shape file includes nineteen trait measures such as area, height, circularity, etc. The color file includes is a lightness histogram of 0–255 for each lesion. The text files were uploaded into R and processed using a custom script designed to read and format the data and create a comma separated value (CSV) file. For the color file, the histogram data were used to calculate lesion gray-scale mean. The variance explained by twelve machine learning derived traits were calculated and plotted in R using a custom partial correlations script. Area and gray-scale mean data for all lesions were compared across different treatment types using a Kolmogorov–Smirnov (KS) statistical test in R. All plots were generated in R with a dpi = 300, width = 8.66, and height = 6.86.

### Abbreviations

CBB: Cassava Bacterial Blight; Xam: *Xanthomonas axonopodis* pv. *manihotis*; Xpm: *Xanthomonas phaseoli* pv. *manihotis*; PV: Pathovar; DPI: Days Post Inoculation; TAL: Transcription Activator-Like; VE: Variance Explained; SVM: Support Vector Machine learning; WT: Wildtype; SWEET: Sugars Will Eventually be Exported Transporters; ROI: Region Of Interest; KS: Kilmogrov-Smirnov; CSV: Comma Separated Values plain text file.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-022-00906-x>.

**Additional file 1: Movie S1.** Movie example of ImageJ based analysis method.

**Additional file 2: Table S1.** Machine learning tool commands. A table of the command syntax, function, and description of inputs and outputs for each command.

**Additional file 3: Movie S2.** Movie example of machine learning based analysis method.

**Additional file 4: Table S2.** Machine learning measurement types. A table of measurements generated from the machine learning tool and their descriptions.

### Acknowledgements

We acknowledge the Bart Lab members who provided insightful discussion and feedback on this project, especially Dr. Kira Veley, Dr. Qi Wang, Dr. Ben Mansfeld, and Taylor Harris.

### Author contributions

KE and RSB designed the study. KE completed bacterial inoculations and collected images at each timepoint. HK and KE completed ImageJ analysis and interpreted the results. JB set up the Raspberry Pi camera system and developed the machine learning workflow in PhenotyperCV. JB and KE tested the machine learning tool performance. KE processed all images analyzed through the machine learning tool and interpreted results. JB developed the gray correction and image effects color correction methods for image analysis. JB provided expertise on statistical analyses and developed initial R scripts used to run statistical tests. KE wrote the original manuscript draft, completed statistical analysis, and generated figures. KE, JB, and RSB assisted in manuscript and figure review and editing. RSB provided supervision over the project. All authors read and approved the final manuscript.

**Authors' information**

Kiona Elliott: PhD candidate at Washington University in St. Louis and the Donald Danforth Plant Science Center.  
 Jeffrey C. Berry: Sr. Data Scientist at the Donald Danforth Plant Science Center.  
 Hobin Kim: Donald Danforth Plant Science Center high school summer intern from the Army and Navy Academy.  
 Rebecca S. Bart: Associate member and principal investigator at the Donald Danforth Plant Science Center.

**Funding**

National Science Foundation GRFP DGE-2139839 and DGE-1745038 (KE). Bill and Melinda Gates Foundation OPP1125410 (RSB).

**Availability of data and materials**

The datasets and custom R scripts generated and/or analyzed in this study are available in the figshare repository, <https://doi.org/10.6084/m9.figshare.17334407>.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Donald Danforth Plant Science Center, Saint Louis, MO 63132, USA. <sup>2</sup>Division of Biological and Biomedical Sciences, Washington University in Saint Louis, St. Louis, MO 63110, USA. <sup>3</sup>Army and Navy Academy, Carlsbad, CA 92008, USA.

Received: 4 January 2022 Accepted: 16 May 2022

Published online: 21 June 2022

**References**

- Access to food in 2020. Results of twenty national surveys using the Food Insecurity Experience Scale (FIES). FAO. 2021. <https://doi.org/10.4060/cb5623en>.
- Strange RN. Introduction to plant pathology. New York: Wiley; 2003.
- Liu X, Sun Y, Körner CJ, Du X, Vollmer ME, Pajerowska-Mukhtar KM. Bacterial leaf infiltration assay for fine characterization of plant defense responses using the *Arabidopsis thaliana*-*Pseudomonas syringae* pathosystem. *J Vis Exp*. 2015. <https://doi.org/10.3791/53364>.
- Gaunt RE. The relationship between plant disease severity and yield. *Annu Rev Phytopathol*. 1995;33:119–44. <https://doi.org/10.1146/annurev.py.33.090195.001003>.
- Moore WC. The measurement of plant diseases in the field: Preliminary report of a sub-committee of the Society's Plant Pathology Committee. United Kingdom: Chartered Institute Of Horticulture; 1949.
- Plant Pathology—5th Edition n.d. <https://www.elsevier.com/books/plant-pathology/agrios/978-0-08-047378-9>. Accessed 30 Mar 2022.
- Bart R, Cohn M, Kassen A, McCallum EJ, Shybut M, Petriello A, et al. High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc Natl Acad Sci USA*. 2012;109:E1972–1979. <https://doi.org/10.1073/pnas.1208031109>.
- Cohn M, Bart RS, Shybut M, Dahlbeck D, Gomez M, Morbitzer R, et al. *Xanthomonas axonopodis* virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar transporter in cassava. *Mol Plant Microbe Interact*. 2014;27:1186–98. <https://doi.org/10.1094/MPMI-06-14-0161-R>.
- Díaz Tatis PA, Herrera Corzo M, Ochoa Cabezas JC, Medina Cipagauta A, Prías MA, Verdier V, et al. The overexpression of RXam1, a cassava gene coding for an RLK, confers disease resistance to *Xanthomonas axonopodis* pv. *manihotis*. *Planta*. 2018;247:1031–42. <https://doi.org/10.1007/s00425-018-2863-4>.
- Jorge V, Verdier V. Qualitative and quantitative evaluation of cassava bacterial blight resistance in F1 progeny of a cross between elite cassava clones. *Euphytica*. 2002. <https://doi.org/10.1023/A:1014400823817>.
- Poland JA, Nelson RJ. In the eye of the beholder: the effect of rater variability and different rating scales on QTL mapping. *Phytopathology*. 2011;101:290–8. <https://doi.org/10.1094/PHYTO-03-10-0087>.
- Strange RN, Scott PR. Plant disease: a threat to global food security. *Annu Rev Phytopathol*. 2005;43:83–116. <https://doi.org/10.1146/annurev.phyto.43.113004.133839>.
- Gehan MA, Fahlgren N, Abbasi A, Berry JC, Callen ST, Chavez L, et al. PlantCV v2: image analysis software for high-throughput plant phenotyping. *PeerJ*. 2017;5: e4088. <https://doi.org/10.7717/peerj.4088>.
- Laflamme B, Middleton M, Lo T, Desveaux D, Guttman DS. Image-based quantification of plant immunity and disease. *MPMI*. 2016;29:919–24. <https://doi.org/10.1094/MPMI-07-16-0129-TA>.
- Lobet G. Image analysis in plant sciences: publish then perish. *Trends Plant Sci*. 2017;22:559–66. <https://doi.org/10.1016/j.tplants.2017.05.002>.
- Li L, Zhang Q, Huang D. A review of imaging techniques for plant phenotyping. *Sensors*. 2014;14:20078–111. <https://doi.org/10.3390/s141120078>.
- Zhang Y, Zhang N. Imaging technologies for plant high-throughput phenotyping: a review. *Front Agr Sci Eng*. 2018;5:406–19. <https://doi.org/10.15302/J-FASE-2018242>.
- Ferreira T, Rasband W. ImageJ user guide. Madison: University of Wisconsin; 2012.
- Bock CH, Parker PE, Cook AZ, Gottwald TR. Visual rating and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves. *Plant Dis*. 2008;92:530–41. <https://doi.org/10.1094/PDIS-92-4-0530>.
- Bierman A, LaPlumm T, Cadle-Davidson L, Gadoury D, Martinez D, Sapkota S, et al. A high-throughput phenotyping system using machine vision to quantify severity of grapevine powdery mildew. *Plant Phenomics*. 2019;2019:9209727. <https://doi.org/10.34133/2019/9209727>.
- Gallejo-Sánchez LM, Canales FJ, Montilla-Bascón G, Prats E. RUST: a robust, user-friendly script tool for rapid measurement of rust disease on cereal leaves. *Plants*. 2020;9:1182. <https://doi.org/10.3390/plants9091182>.
- Mutka AM, Bart RS. Image-based phenotyping of plant disease symptoms. *Front Plant Sci*. 2015;5:734. <https://doi.org/10.3389/fpls.2014.00734>.
- Stewart EL, McDonald BA. Measuring quantitative virulence in the wheat pathogen *Zymoseptoria tritici* using high-throughput automated image analysis. *Phytopathology*. 2014;104:985–92. <https://doi.org/10.1094/PHYTO-11-13-0328-R>.
- Stewart EL, Hagerly CH, Mikaberidze A, Mundt CC, Zhong Z, McDonald BA. An improved method for measuring quantitative resistance to the wheat pathogen *Zymoseptoria tritici* using high-throughput automated image analysis. *Phytopathology*. 2016;106:782–8. <https://doi.org/10.1094/PHYTO-01-16-0018-R>.
- Singh A, Ganapathysubramanian B, Singh AK, Sarkar S. Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci*. 2016;21:110–24. <https://doi.org/10.1016/j.tplants.2015.10.015>.
- Tsaftaris SA, Minervini M, Scharf H. Machine learning for plant phenotyping needs image processing. *Trends Plant Sci*. 2016;21:989–91. <https://doi.org/10.1016/j.tplants.2016.10.002>.
- Morgan NK, Choct M. Cassava: nutrient composition and nutritive value in poultry diets. *Animal Nutrition*. 2016;2:253–61. <https://doi.org/10.1016/j.aninu.2016.08.010>.
- Bart RS, Taylor NJ. New opportunities and challenges to engineer disease resistance in cassava, a staple food of African small-holder farmers. *PLoS Pathog*. 2017;13: e1006287. <https://doi.org/10.1371/journal.ppat.1006287>.
- Hillocks RJ, Thresh JM, Bellotti A. Cassava: biology, production and utilization. Wallingford: CABI; 2002.
- El-Sharkawy MA. Cassava biology and physiology. *Plant Mol Biol*. 2003;53:621–41. <https://doi.org/10.1023/B:PLAN.0000019109.01740.c6>.
- Howeler RH, Lutaladio N, Thomas G. Save and grow: cassava: a guide to sustainable production intensification. Rome: Food and Agriculture Organization of the United Nations; 2013.
- Fanou AA, Zinsou VA, Wydra K. Cassava bacterial blight: a devastating disease of cassava. *Cassava*. 2017. <https://doi.org/10.5772/intechopen.71527>.

33. Zárte-Chaves CA, Gómez de la Cruz D, Verdier V, López CE, Bernal A, Szurek B. Cassava diseases caused by *Xanthomonas phaseoli* pv. *manihotis* and *Xanthomonas cassavae*. *Mol Plant Pathol*. 2021;22:1520–37. <https://doi.org/10.1111/mpp.13094>.
34. Constantin EC, Cleenwerck I, Maes M, Baeyen S, Van Malderghem C, De Vos P, et al. Genetic characterization of strains named as *Xanthomonas axonopodis* pv. *dieffenbachiae* leads to a taxonomic revision of the *X. axonopodis* species complex. *Plant Pathol*. 2016;65:792–806. <https://doi.org/10.1111/ppa.12461>.
35. Aung K, Jiang Y, He SY. The role of water in plant–microbe interactions. *Plant J*. 2018;93:771–80. <https://doi.org/10.1111/tpj.13795>.
36. Boch J, Bonas U. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol*. 2010;48:419–36. <https://doi.org/10.1146/annurev-phyto-080508-081936>.
37. Hogenhout SA, Van der Hoorn RAL, Terauchi R, Kamoun S. Emerging concepts in effector biology of plant-associated organisms. *Mol Plant Microbe Interact*. 2009;22:115–22. <https://doi.org/10.1094/MPMI-22-2-0115>.
38. Muñoz Bodnar A, Bernal A, Szurek B, López CE. Tell me a tale of TALEs. *Mol Biotechnol*. 2013;53:228–35. <https://doi.org/10.1007/s12033-012-9619-3>.
39. van Schie CCN, Takken FLW. Susceptibility genes 101: how to be a good host. *Annu Rev Phytopathol*. 2014;52:551–81. <https://doi.org/10.1146/annurev-phyto-102313-045854>.
40. Koseoglou E, van der Wolf JM, Visser RGF, Bai Y. Susceptibility reversed: modified plant susceptibility genes for resistance to bacteria. *Trends Plant Sci*. 2021. <https://doi.org/10.1016/j.tplants.2021.07.018>.
41. Li T, Liu B, Spalding M, Weeks D, Yang B. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol*. 2012;30:390–2. <https://doi.org/10.1038/nbt.2199>.
42. Phillips AZ, Berry JC, Wilson MC, Vijayaraghavan A, Burke J, Bunn JI, et al. Genomics-enabled analysis of the emergent disease cotton bacterial blight. *PLoS Genet*. 2017;13: e1007003. <https://doi.org/10.1371/journal.pgen.1007003>.
43. Cox KL, Meng F, Wilkins KE, Li F, Wang P, Booher NJ, et al. TAL effector driven induction of a SWEET gene confers susceptibility to bacterial blight of cotton. *Nat Commun*. 2017;8:1–14. <https://doi.org/10.1038/ncomms15588>.
44. Berry JC, Fahlgren N, Pokorny AA, Bart RS, Veley KM. An automated, high-throughput method for standardizing image color profiles to improve image-based plant phenotyping. *PeerJ*. 2018. <https://doi.org/10.7717/peerj.5727>.
45. Sangbamrung I, Praneetpholkrang P, Kanjanawattana S. A novel automatic method for cassava disease classification using deep learning. *JAIT*. 2020;11:241–8. <https://doi.org/10.12720/jait.11.4.241-248>.
46. Ramcharan A, McCloskey P, Baranowski K, Mbilinyi N, Mrisho L, Ndalawha M, et al. A mobile-based deep learning model for cassava disease diagnosis. *Front Plant Sci*. 2019;10:272. <https://doi.org/10.3389/fpls.2019.00272>.
47. Ramcharan A, Baranowski K, McCloskey P, Ahmed B, Legg J, Hughes DP. Deep learning for image-based cassava disease detection. *Front Plant Sci*. 2017. <https://doi.org/10.3389/fpls.2017.01852>.
48. Casto L. Picturing the future of food. *Plant Phenome J*. 2021. <https://doi.org/10.1002/ppj2.20014?af=R>.
49. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9:676–82. <https://doi.org/10.1038/nmeth.2019>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

