Detecting Plant Diseases in Browser Using Efficient Deep Learning

Griffin Davis
The University of Texas at Dallas
800 W Campbell Rd, Richardson, TX 75080
gcd@utdallas.edu

Canton Zhou
The University of Texas at Dallas
800 W Campbell Rd, Richardson, TX 75080
canton.zhou@utdallas.edu

Arshia Elahi
The University of Texas at Dallas
800 W Campbell Rd, Richardson, TX 75080
arshia.elahi@utdallas.edu

1. Problem Statement

Costing the global economy an estimated $220 million yearly [4], plant disease has remarkable effects on the lives of both farmers and consumers. Plant disease management is a critical aspect of ensuring food production can meet the demands of a growing human population [3]. On a smaller scale, people who own houseplants or small-scale gardeners may not know why their plants are declining. The goal of this project is to implement a classifier to identify plant diseases using efficient computer vision and deep learning techniques, allowing the model to be deployed on a web application with all computation done in the browser. With this type of model and application environment, users such as gardeners or plant owners can upload pictures of their plants to the web application to get a predicted class of disease that their plant might have, helping them to find treatment for the plants.

While plant disease detection has been widely studied, including through the application of convolutional neural networks, these methods often focus on individual plant types and large models that require significant computational resources. Mohanty et al. expand their study to 14 crop species and 26 diseases [10], but the authors rely on AlexNet [8] and GoogleLeNet [14], which are not designed for mobile or browser environments. Agarwal et al. and Mishra et al. each focus on a single crop species, focusing on tomato and corn plants respectively [1, 9].

This project will be application-oriented, applying well-known techniques of image classification to the subject area of plant disease detection across a variety of species. We will also impose additional evaluation metrics to build a model that is efficient and lightweight.

2. Approach

To expand the size of the training and testing data, we will merge multiple publicly available plant leaf disease data sets and apply various data augmentation strategies. These strategies include both data warping, which modifies the images themselves, and oversampling, which produces synthetic images to increase the data size [12]. By combining multiple data sets and expanding the data with image augmentation, we can avoid biases that may be present in individual datasets [11] and improve generalization performance [12].

This data will then be used to train convolutional neural networks designed for usage in environments with low computational capacity, including MobileNet [6] and EfficientNet [15]. We will test both the accuracy of these models as well as the performance inside internet browsers on multiple device types to determine the optimal model for this usage. These results will be compared to traditional large-scale image recognition models. In future work, the application can allow for user input regarding the correctness of the model prediction to add new training data.

3. Data

We will use a combination of publicly available data sets that have both annotated plant diseases and healthy leaves across a range of different plant species. By using images from different sources, the model will ideally generalize better to different conditions and plant diseases not seen at training time.

The data sources used for this project will include the PlantVillage data set, introduced by Hughes and Salathé [7]; Identification of Plant Leaf Diseases data from Arun and Geetharamani [2]; the PlantDoc data set, introduced by Singh et al. [13]; and a plant disease data set by Frabotta [5].

4. Evaluation

The performance of the model will be evaluated based on both its accuracy and computational efficiency. For the accuracy of the model, we will compute metrics such as precision, recall, and F_1 score. This will be done by comparing true positives, true negatives with incorrectly classified examples on a testing data set. Efficiency will be measured by speed required to train the model initially, and then the amount of time it takes for the deployed model to output a result. A successful model is one that has both high accuracy and can be deployed reasonably quickly in a browser on commodity hardware.

References

[1] Mohit Agarwal, Abhishek Singh, Siddhartha Arjaria, Amit Sinha, and Suneet Gupta. ToLeD: Tomato leaf disease detection using convolution neural network. In *International Conference on Computational Intelligence and Data Science*, volume 167, pages 293–301, 2020. 1

[2] J Arun Pandian and Gopal Geetharamani. Data for: Identification of plant leaf diseases using a 9-layer deep convolutional neural network. Type: dataset. 1

[3] Jeremy J Burdon, Luke G Barrett, Li-Na yang, Dun-Chun He, and Jiasui Zhan. Maximizing world food production through disease control. *BioScience*, 70(2):126–128, 2019. 1

[4] Food and Agriculture Organization of the United Nations. New standards to curb the global spread of plant pests and diseases. 1

[5] Jessica Frabotta. Plant diseases dataset with augmentation. 1

[6] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. MobileNets: Efficient convolutional neural networks for mobile vision applications. 1

[7] David P Hughes and Marcel Salathé. An open access repository of images on plant health to enable the development of mobile disease diagnostics, 2015. 1

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. 1

[9] Sumita Mishra, Rishabh Sachan, and Diksha Rajpal. Deep convolutional neural network based detection system for real-time corn plant disease recognition. In *International Conference on Computational Intelligence and Data Science*, volume 167, pages 2003–2010, 2020. 1

[10] Sharada P. Mohanty, David P. Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 2016. 1

[11] Mehmet Alican Noyan. Uncovering bias in the PlantVillage dataset. 1

[12] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019. 1

[13] Davinder Singh, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawat, and Nipun Batra. PlantDoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, pages 249–253, 2020. 1

[14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9. IEEE, 2015. 1

[15] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking model scaling for convolutional neural networks. 1