Accessible Large-Scale Plant Pathology Recognition

Marcos V. Conde* **Dmitry Gordeev**

H2O.ai
Mountain View, CA
{marcos.conde,dmitry.gordeev}@h2o.ai

Abstract

Plant diseases are costly and threaten agricultural production and food security worldwide. Climate change is increasing the frequency and severity of plant diseases and pests. Therefore, detection and early remediation can have a significant impact, especially in developing countries. However, AI solutions are yet far from being in production. The current process for plant disease diagnostic consists of manual identification and scoring by humans, which is time-consuming, low-supply, and expensive. Although computer vision models have shown promise for efficient and automated plant disease identification, there are limitations for real-world applications: a notable variation in visual symptoms of a single disease, different light and weather conditions, and the complexity of the models. In this work, we study the performance of efficient classification models and training "tricks" to solve this problem. Our analysis represents a plausible solution for these ecological disasters and might help to assist producers worldwide. More information available at: https://github.com/mv-lab/mlplants

1 Introduction

Plant diseases affect crop production in developed as well as developing countries. Their rapid identification remains difficult, particularly in parts of the world that lack the necessary resources. This can sometimes trigger serious **environmental disasters** that severely affect the economy.

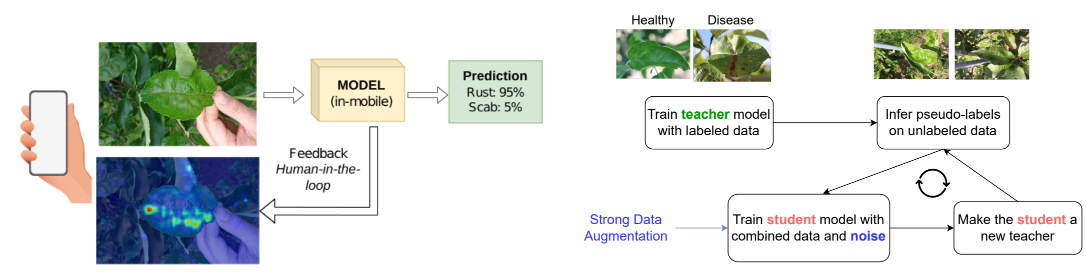

Figure 1: (Left) We aim to process images captured in-situ in the field using a smartphone, due to the efficiency of the studied models. (Right) Semi-supervised noisy student workflow [25].

Cassava (*Manihot esculenta*) is one of the key staples and food security crops in Africa. At least 80% of household farms in Sub-Saharan Africa grow this starchy root. It is robust to adverse weather conditions and is the second-largest provider of carbohydrates in Africa. Several diseases plague the crop and cause annual yield losses valued at an estimated 20 million dollars [17].

*MC is also with University of Würzburg, CAIDAS. Supported by The Alexander von Humboldt Foundation.

Accurate disease diagnostics requires agricultural experts from government agencies to visually inspect the plants. This makes it difficult to effectively monitor and treat disease progression.

In terms of value, Apple (*Malus domestica* Borkh. 1803) is one of the most economically important fruit crops in the world; in the US $15 billion annually [19]. Apple production is heavily affected by a wide range of pests and diseases including mites, aphids, stink bug, european sawfly, etc. Early and accurate disease detection is critical for management in orchards, not least because pesticide (over-)use has negative environmental impact.

Recent advances in machine learning and computer vision (CV) paired with universal and global access to smartphones lead a clear way towards low-cost plant disease diagnosis in the wild. The potential impact of an efficient smartphone implementation of such technology is immeasurable. Although CV models have shown promising results for efficient plant disease detection [23, 19, 12], there are many factors which complicate the design of such a tool *e.g.,* multiple diseases are present in an image, small datasets, and noisy inputs (light and shade variations, non-uniform backgrounds).

Our contributions aim to tackle these challenges in the wild. We explore *state-of-the-art* models and regularization techniques to achieve such general and robust application, as illustrated in Figure 1.

2 Related work

In recent years, multiple large-scale annotated datasets [17, 23, 19, 12] for this topic have been proposed. We will discuss the datasets in Section A. Kumar *et al.* [15] presented the first mobile app ("Leafsnap") for identifying plant species using automatic visual recognition in 2012. Since 2016, deep learning and computer vision approaches have been used in many agriculture problems such as automated plant species identification and phenotyping [2] and plant pathology recognition [1], among others. These approaches use Convolutional Neural Networks (CNNs) [8, 22]. Mohanty *et al.* [16] introduced the use of deep learning and CNNs for image-based plant disease detection using the "PlantVillage" dataset [12]. More recently, Keaton *et al.* [14] proposed a model for plant species identification using object detection as a form of attention.

Existing computer vision research in categorization struggles with fine-grained attributes recognition due to the inherently high intra-class variance and low inter-class variance [11]. The large variation in visual symptoms of a single disease class across different samples present the major challenge for deep learning models. This variation arises, for example, from differences in natural and image capturing environments, leaf color and morphology. Hence, models must be robust to environmental variation (*i.e.* viewpoints, light, background), and, due to the nature of the public large-scale datasets, the models should learn from noisy labels [18] .

2.1 Datasets

Plant pathology researchers have collected real world datasets with enough variability and size to train deep learning models. Images of disease symptoms on leaves were captured using smartphones at different distances and angles from the leaves, with different focus and light conditions to represent real world scenarios. We use the following public datasets from different FGVC-CVPR editions:

1. **iCassava 2019** [17] consists of labeled and unlabeled images of healthy and diseased cassava plant leaves showing 4 types of diseases.
2. **Cassava Leaf Disease Classification 2021 Dataset** an extended training dataset of 21397 labeled images, and 15000 test images.
3. **FGVC Plant Pathology 2020** [23, 19]. This is a large, real-life disease pilot dataset, which contains approximately 23000 high-quality expert-annotated images of apple foliar diseases.
4. **Our dataset.** We pre-process the images, unify and extend the previous datasets. We perform a 70/30% stratified split, obtaining 29295 training images, and 12555 test images.

We aim to reflect real-world scenarios: (1) Imbalanced classes, (2) Different backgrounds, light, angles, and noise conditions. (3) Different physiological age of the plants, (4) Co-occurrence of multiple diseases on the same plant, and (5) Different focus of the images. We do not use curated datasets captured in-lab under non-realistic scenarios, *i.e.* Plant Village Dataset [12]. More information about the datasets can be consulted in the Appendix A.

a b c d

Figure 2: We can see (a) sample from the Plant Village Dataset [12] with uniform background, focus and illumination; we consider this non-realistic. In contrast, (b) (c) (d) represent real scenes in the field, where leaves are surrounded and "mixed", we define these as noisy labels (or hard samples).

3 Learning with Noisy Labels

In Section 2, we discuss some challenges when developing computer vision models for this task: (1) fine-grained attributes and great intra-class variability, (2) noisy labels. We focus on the second point, as fine-grained visual categorization is a well-studied topic [11].

We define two different types of noisy labels, both are represented in Figure 2.

1. Misclassified images: The provided annotations from non-experts and farmers for these images are apparently wrong. We estimate these to be a small portion of the dataset.
2. Images with noisy labels: Multiple plants (or leaves) appear in the same image, each one might have a different label. For instance, in Figure 2, we can see healthy leaves surrounded by diseased leaves, and these with different type of diseases.

In this Section, we introduce our experimental setup. In Section 3.3 we explain how to transfer learning from models pre-trained on a generic task, to few-shot fine-grained multi-class classification. We also propose a simple approach for automated annotation and semi-supervised learning.

3.1 Models

We use SOTA Deep Convolutional Neural Networks: ResNet [8], ResNeSt [26], and EfficientNet [22]. We also explore compact models designed for mobile devices: GhostNet [7] and MobileNetV2 [21]. Note that we focus on efficient light-weight models that can fit into a smartphone, our main objective is to outperform complex deep models using them, while being robust - as we can see in Table 1 -.

3.2 Implementation details

In Figure 3 we show our framework for training and evaluating deep learning models: **Hydrogen Torch** [6]. Our framework integrates well-known tricks for image classification [9], such as: mixed precision training, cosine annealing LR scheduler, gradient accumulation, and advanced augmentations (Cutout [4] and Mixup [27]) amongst other state-of-the-art techniques. The implementation is done using Pytorch. We use model architectures from the `timm` package [24]. The image size is set to 512px. Experiments are performed using a single NVIDIA RTX-3090 GPU. We refer the reader to our repository and documentation for more details.

3.3 Transfer Knowledge and Distill

Deep learning models are data-hungry and require large amounts of annotated data. In Section 3 we have introduced our main experimental setup. Initially we train different SOTA models for the task of classifying if a crop is healthy or not, we consider the cost of annotating healthy/disease (binary class) "cheaper" than producing fine-grained annotations per disease. In this Section we explain how to **transfer the learning** from such models, trained on a general task using cheap and noisy labels, to perform fine-grained classification [11] of the disease using few-shot learning. In Table 1 we show the results from our experiments using different light-weight models and training techniques. Few-shot (FS) models were **pre-trained** using our dataset (29295 training images), and in this stage, they are fine-tuned using 550 images with labels of five classes: healthy, rust, scab multiple diseases.

Model	img. size	ROC AUC	Mobile
1st place (SEResNeXt [10])	512	0.984	✗
2nd place (ResneSt101 [26])	545	0.981	✗
3rd place (EffNet B7 [22])	768	0.980	✗
ResNet34 [8]	512	0.969	✗
Ours EfficientNet B0 [21]	512	0.975	✓
Ours GhostNet [7]	512	0.970	✓
Ours MobileNetV2 [21]	512	0.967	✓
Ours **FS** MobileNetV2 [21]	512	0.910	✓
Ours **FS** GhostNet [7]	512	0.900	✓
Ours **SS** MobileNetV2 [21]	512	0.973	✓
Ours **AL** MobileNetV2 [21]	512	0.950	✓

Model	F1-score	F1-score TTA	Mobile	N. params (M)
ResNet-50 [8]	0.9486	0.9500	✗	23.5
ResNeSt-50 [26]	0.9526	0.9530	✗	25.4
GhostNet-100 [7]	0.9559	0.9570	✓	3.9
FBNet-V3 [3]	0.9590	0.9601	✓	3.6
EfficientNet B0 [22]	0.9583	0.9595	✓	4.9
MobileNetV2 [21]	0.9540	0.9552	✓	3.4
ViT [5]	0.9185	0.9240	✗	86.6

Table 1: (Left) Leaderboard Plant Pathology Challenge, (right) Helthy-Diseased Classification using our dataset. The metric is the mean class-wise ROC AUC. We use our "bag of tricks" to train lightweight models suitable for mobile devices and achieve competitive performance even when trained with Few-Shot (FS) data or Active Learning (AL).

For comparison, we also train the proposed light-weight models from scratch using the entire competition dataset and our *bag-of-tricks*. We can see that the proposed models, even when trained with few data, achieve competitive results, but most important, they can generalize while being extremely small in comparison with other solutions (*i.e.* $10\times$ smaller than 1st and 2nd place solutions, in terms of parameters).

Semi-supervised (SS) Learning extends the idea of self-training and distillation [25]. We train a teacher model EfficientNet B0 [22] on the fine-grained classification task using the few-data from the challenge (pre-trained on our dataset). Next, we use this model to auto-label unseen data (test set) and generate pseudo-labels, these samples are used to expand the training set. In Table 1 we show the improvement due to this technique (SS), and we illustrate this process in Figure 1 (right).

Using **Active Learning (AL)** [20], one can keep optimising the labelling strategy as the models improve. We use this technique in a experimental setup that illustrates a real scenario. We use the Plant Pathology 2020 dataset [23] and start from a subset of 384 labeled images (20% of the training set), we apply AL to train iteratively MobileNetV2 as proposed in the original algorithm [20]. This model, started training with only 20% of labels, yet, it ended achieving competitive results as we show in Table 1 ("Ours AL MobileNetV2"). Because we train with few data, this is trained using classical augmentations (flips, color shifts, etc.) and MixUp [27]. More applications in Appendix B.

4 Conclusion

In this work we study different computer vision models for automated plant pathology recognition in the wild. We focus on efficient and compact models that can be used on smartphones in real-time for fine-grained plant disease categorization. Moreover, we provide some empirical tricks to improve model's generalization and robustness to different light and weather conditions, noisy image background, or multiple instances in the same image. We hope this work serves as a spotlight to tackle one of the major threats to food security in developing countries.

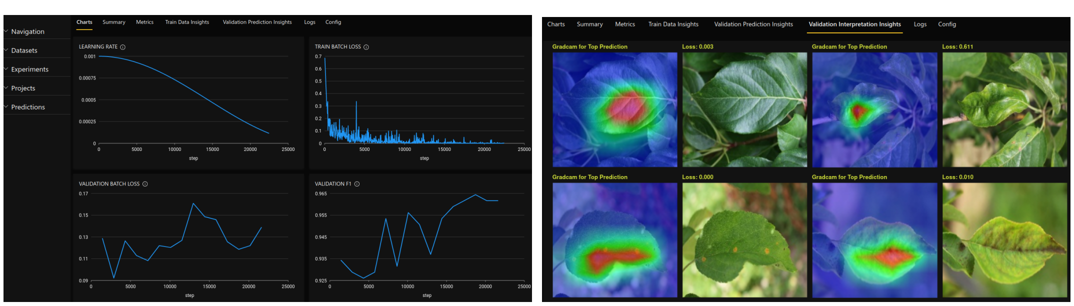

Figure 3: Our framework for fast experimentation, suitable for non-experts in machine learning [6]. We train the different studied models using the state-of-the-art image classification "tricks" [9].

A Datasets

To encourage the development of computer vision algorithms, plant pathology researchers and experts generated real world datasets with enough variability and size to train deep learning models. Images of disease symptoms on leaves were captured using mainly smartphone cameras at different distances from the leaves, from different angles, with different focus and light conditions to represent real world scenarios. We use the following public datasets from different FGVC CVPR Workshop editions:

1. **iCassava 2019** [17] consists of labeled and unlabeled images of healthy and diseased cassava plant leaves showing 4 types of diseases. The annotations hence consist of the following 5 classes: Healthy, Cassava Mosaic Disease (CMD), Cassava Brown Streak Disease (CBSD), Cassava Bacterial Blight (CBB), Cassava Green Mite (CGM). All the collected images were manually labelled by experts from the National Crops Resources Research Institute (NaCRRI) in Uganda, who scored each of the images for disease incidence and severity.
2. **Cassava Leaf Disease Classification 2021 Dataset** is an extension of the previous dataset. Authors propose a challenge[2] where participants train models on a public training dataset of 21397 labeled images, and test them on 15000 unknown test images.
3. **FGVC Plant Pathology 2020** [23]. This is a large, high-quality, real-life disease pilot dataset of multiple apple foliar diseases captured during the 2019 growing season. Photos were taken using a DSLR camera and smartphones under various conditions. The dataset contains a total of 3651 images of leaves with 1200 apple scab, 1399 cedar apple rust, 187 complex disease, and 865 healthy leaves, respectively. The dataset was randomly split into training and stratified test set of 50% and 50%, respectively, or 1821 training images and 1821 test images. The challenge[3] associated to this dataset provided all the images to the participants, but only training data ground-truth.
4. **Our dataset.** We unify and extend the previous datasets. We unify the different diseases into a unique disease class, therefore we only consider two classes (healthy and diseased). This dataset contains 41850 images, we holdout using 70/30% stratified split, 29295 train and 12555 test images. This dataset represents a realistic scenario where fine-grained annotations are not available (due to the time-consuming and expensive process), but it is straight-forward to determine, whether a plant is healthy or not.

B Application Context

In previous sections and Figure 1 we already studied ML models to help farmers and producers (especially in developing countries) to deal with plant pathologies and plagues, which represent a critical ecological disaster. We have explained the technical challenges (see Sec. 3.3), and studied lightweight models suitable for real-time applications on smartphones [21, 13]. Another potential example of combining our ideas and Active Learning[20] with noisy labels is as follows:

1. Capture 1000 images and label 80% (set A) using general annotations (*i.e.* healthy, diseased). This process is cheaper and faster than producing fine-grained multi-class annotations, we only produce such high-quality annotations only for 20% of the images (set B).
2. Train a general (teacher) model M_a (see Sec. 3) on the set A.
3. Transfer knowledge from M_a to a light-weight compact M_b trained on set B for fine-grained disease recognition. Our experiments show that these models even when trained with few noisy data, achieve competitive results (see Sec. 3.3).
4. We can deploy M_b easily on a smartphone (*i.e.* using AI Benchmark app [13]).
5. Using models M_a and M_b we can automatically annotate more images in the future and apply AL. These annotations or *pseudo-labels* allow us to expand the original dataset and re-train models efficiently [25] as we propose in Sec. 3.3.

This process allows to annotate data and train models in a dynamic way, while getting real feedback from farmers and producers. We also designed a framework for training our models, tracking results and do interpretability analysis in a interactive manner. This framework is illustrated in Figure 3.

[2] `https://www.kaggle.com/c/cassava-leaf-disease-classification/`

[3] `https://www.kaggle.com/c/plant-pathology-2020-fgvc7/`

References

[1] Andre S Abade, Paulo Afonso Ferreira, and Flavio de Barros Vidal. Plant diseases recognition on images using convolutional neural networks: A systematic review. *arXiv preprint arXiv:2009.04365*, 2020.

[2] Akshay L Chandra, Sai Vikas Desai, Wei Guo, and Vineeth N Balasubramanian. Computer vision with deep learning for plant phenotyping in agriculture: A survey. *arXiv preprint arXiv:2006.11391*, 2020.

[3] Xiaoliang Dai, Alvin Wan, Peizhao Zhang, Bichen Wu, Zijian He, Zhen Wei, Kan Chen, Yuandong Tian, Matthew Yu, Peter Vajda, et al. Fbnetv3: Joint architecture-recipe search using predictor pretraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16276–16285, 2021.

[4] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017.

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[6] H2O.ai. H2O Hydrogen Torch, 2022. URL `https://docs.h2o.ai/h2o-hydrogen-torch/v1.2.0/`.

[7] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1580–1589, 2020.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[9] Tong He, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. Bag of tricks for image classification with convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 558–567, 2019.

[10] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[11] Tao Hu, Honggang Qi, Qingming Huang, and Yan Lu. See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification, 2019.

[12] David Hughes, Marcel Salathé, et al. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*, 2015.

[13] Andrey Ignatov, Radu Timofte, William Chou, Ke Wang, Max Wu, Tim Hartley, and Luc Van Gool. Ai benchmark: Running deep neural networks on android smartphones. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.

[14] Matthew R Keaton, Ram J Zaveri, Meghana Kovur, Cole Henderson, Donald A Adjeroh, and Gianfranco Doretto. Fine-grained visual classification of plant species in the wild: Object detection as a reinforced means of attention. *arXiv preprint arXiv:2106.02141*, 2021.

[15] Neeraj Kumar, Peter N. Belhumeur, Arijit Biswas, David W. Jacobs, W. John Kress, Ida C. Lopez, and João V. B. Soares. Leafsnap: A computer vision system for automatic plant species identification. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision – ECCV 2012*, pages 502–516, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-33709-3.

[16] Sharada P. Mohanty, David P. Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 2016. ISSN 1664-462X. doi: 10.3389/fpls.2016.01419. URL `https://www.frontiersin.org/article/10.3389/fpls.2016.01419`.

[17] Ernest Mwebaze, Timnit Gebru, Andrea Frome, Solomon Nsumba, and Jeremy Tusubira. icassava 2019 fine-grained visual categorization challenge. *arXiv preprint arXiv:1908.02900*, 2019.

[18] Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with noisy labels. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL `https://proceedings.neurips.cc/paper/2013/file/3871bd64012152bfb53fdf04b401193f-Paper.pdf`.

[19] Thapa Ranjita, Wang Qianqian, Snavely Noah, Belongie Serge, and Awais Khan. The plant pathology 2021 challenge dataset to classify foliar disease of apples, 2021.

[20] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B Gupta, Xiaojiang Chen, and Xin Wang. A survey of deep active learning. *ACM Computing Surveys (CSUR)*, 54(9):1–40, 2021.

[21] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[22] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 09–15 Jun 2019. URL `http://proceedings.mlr.press/v97/tan19a.html`.

[23] Ranjita Thapa, Kai Zhang, Noah Snavely, Serge Belongie, and Awais Khan. The plant pathology challenge 2020 data set to classify foliar disease of apples. *Applications in Plant Sciences*, 8(9):e11390, 2020. doi: https://doi.org/10.1002/aps3.11390. URL `https://bsapubs.onlinelibrary.wiley.com/doi/abs/10.1002/aps3.11390`.

[24] Ross Wightman. Pytorch image models. `https://github.com/rwightman/pytorch-image-models`, 2019.

[25] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10687–10698, 2020.

[26] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Haibin Lin, Zhi Zhang, Yue Sun, Tong He, Jonas Mueller, R Manmatha, et al. Resnest: Split-attention networks. *arXiv preprint arXiv:2004.08955*, 2020.

[27] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.