Deep Learning approach to Tomato Plant Diseases Recognition on PlantVillage dataset and related Sensitivity Analysis
DL-IC 2018 Project

Alessandro Erba Mirco Manzoni
Giuseppe Mascellaro
Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy
{alessandro2.erba, mirco.manzoni, giuseppe.mascellaro}@mail.polimi.it

Abstract

The ABSTRACT should be self contained and explain what the paper is about. Usually abstracts are no longer than 300 words. You should state what are the main contributions of your work and tempt the reader to continue to read your paper. A good abstract briefly describes your problem, approach, and key results. This document is based on the CVPR submission template, and it has been adapted to submit a technical report of a project of the Deep Learning and Image Classification course.

1. Introduction

The advances of science and technology in history have given the possibility to produce enough food to meet the demand of more than 7 billion people. However, food provisioning is threatened by a number of factors such as climate change [1], the decline in pollinators [10], plant diseases [20], and others. Plant diseases are not only a threat to food security at the global scale, but can also have disastrous consequences for smallholder farmers whose livelihoods depend on healthy crops. In the developing world, more than 80 percent of the agricultural production is generated by smallholder farmers [26], and reports of yield loss of more than 50% due to pests and diseases are common [4]. Various efforts have been developed to prevent crop loss due to diseases based on pesticide usage. Independent of the approach, identifying a disease correctly when it first appears is a crucial step for effective and efficient disease management [13].

Historically, diseases identification has been supported by agricultural experts directly on the field. In recently times, these efforts have been additionally supported by leveraging the increasing of Internet penetration worldwide with on-line diagnoses and the tools based on mobile phones, taking advantage of the rapid uptake of mobile phones technology in all parts of the world [11]. These factors, together with advances in computer vision and machine learning, lead to a situation where disease diagnosis based on automated image classification, if technically feasible, can be made available at an unprecedented scale and cost-effectiveness.

On this line, our work focuses first on a Deep Learning approach to disease identification task of 10 tomato plant classes (1 healthy and 9 diseases) using PlantVillage [19] dataset (leaves images) and secondly on the targeted sensitivity analysis of the dataset which has been, in fact, used in state-of-the-art related works [2, 23]. The rationale behind the choice of a particular specie of plants (i.e., tomato plants) is that farmers do know what their plantations are about, hence we exploit this fact as prior knowledge.

As first step we reproduce the experiments of [2] improving their performance results. Then, we show how the learned models perform against variations of input images. This analysis is further supported by the usage of visualization techniques. Finally, on that basis, we conduct a sensitivity analysis of the dataset by building ad hoc variations of it. The analysis of these variations gives insights on actual robustness of the dataset. Indeed, the experiments are pursued keeping in mind realistic deployment environments in which the prediction phase will be performed (i.e., images taken from plantations, greenhouses, and so on). On this assumption we show that the dataset has some not negligible limitations. In light of this we propose a reasonable image augmentation choice that lets the dataset be more robust to various deployment environments.

2. Related work

Several works have been proposed in the literature to plant diseases identification. The classical approach given by the expertise support directly on the field has offered

diverse solutions that show outstanding performance, nonetheless they do not provide yet a scalable and cost-effective solution [15, 17, 18].

After analysis of their work and investigation presented by the authors of [16, 21], it has been decided to employ the image processing approach among other laboratory-based approaches. Several handcrafted feature-based methods have been widely applied specifically for image processing. These methods are usually combined with classifiers from machine learning (e.g., Support Vector Machines (SVMs) [5], K-Nearest Neighbors [6], Random Forests [3], and so on). In [22], the authors have presented a survey of well-known conventional methods for handcrafted feature extraction.

The main drawback of these methods regards feature engineering, that is a complex and time-consuming process which needs to be revisited every time according to the problem at hand. Thus, the performance of classifiers depends heavily on the underlying features. For these reason, deep learning has allowed researchers to consider and design systems as a unified and automated process with no *handcrafted* features [14]. In particular, Convolutional Neural Networks (CNNs), first introduced in [28], have showed, in fact, how to bind together feature extraction to classification in image recognition by means of LeNet architecture. Tremendous achievements have been made by CNNs in the past few years in image classification and benchmarked, for instance, against ImageNet dataset [12].

The principles of CNNs have spread also to plant diseases identification. The authors of [2], have proposed a comparison between shallow models with combined handcrafted features and deep models for the identification of 9 tomato plant diseases. They have analyzed different models using PlantVillage dataset containing 14,828 images of cropped leaves put on a table. The best shallow model has achieved 95.47% accuracy using Random Forests, while the best deep model has achieved 99.19% accuracy using GoogLeNet architecture (pre-trained on ImageNet). Furthermore, they have used Occlusion [29] as model visualization method and on its basis they claim that backgrounds in images do not influence the prediction of their best model. In other related works [8, 23] and in our, we argue that this is not always true.

The work proposed in [23] have focused on developing deep models for the identification of 38 classes (i.e., crop and disease information) using PlantVillage dataset (54,306 images of leaves). They have compared several models having various hyper-parameters and dataset variations (RGB, gray-scale, segmented). Their best result, 99.34% accuracy, has been achieved by GoogLeNet (pre-trained on ImageNet) employing RGB images. They noticed that their best model, when tested on a set of images (derived from trusted Internet sources) taken under conditions different from the images employed for training, determines a substantial accuracy reduction, to just above 31%. According to the authors, this limitation is caused by homogeneous background in the images.

A more robust approach has been pursued in [8] where the authors proposed a robust deep-learning-based detector for real-time tomato diseases and pests recognition. They have built their own dataset of 5,000 images taken under different conditions and scenarios divided in 9 classes (and the backgrounds class). Then, with the support of experts, they manually annotated the areas of every image containing the disease or pest with a bounding box and class. Finally, they have analyzed several models for object detection and achieving outstanding performances. Interestingly, they criticize PlantVillage dataset since the images it contains have been previously cropped in the field and captured by a camera in the laboratory causing image recognition and object detection on realistic environments images unfeasible. Instead, their work has aimed at dealing with background variations mainly caused by the surrounding areas of plants or the place itself (i.e., images taken from plantations, greenhouses, and so on).

According to the complains that have been made about PlantVillage dataset, our work provides a reasonable heuristic solution to overcome its limitations for what backgrounds are concerned.

3. Proposed approach

Hereby, we present how we tackle the challenges set up in the introduction.

3.1. Workflow overview

First, we inspect the dataset by manually visualizing the images and by counting them for each class. Then, as showed in Figure 1, we split the dataset (70% train and 30% test) so that we can train the model and assess its final performance on independent data.

As second step, we enhance train data by applying augmentation techniques. By doing this we let the model learn also slight modification of available data allowing for better generalization capabilities. Optionally, we can make further transformations according to a previous sensitivity analysis of the dataset. Important to notice is that, at this stage, all data splits are normalized.

After that, we set up the cross-validation framework by randomly dividing the data into train and validation sets (since we have a large number of samples, we can assume that each class has a sufficient number of representative samples, i.e., stratification). Cross-validation method is used to estimate the performance of the model on unseen samples and, hence, its generalization capabilities. At this point, once hyper-parameters are chosen, we can run the training phase. Hyper-parameters choice is of tremendous

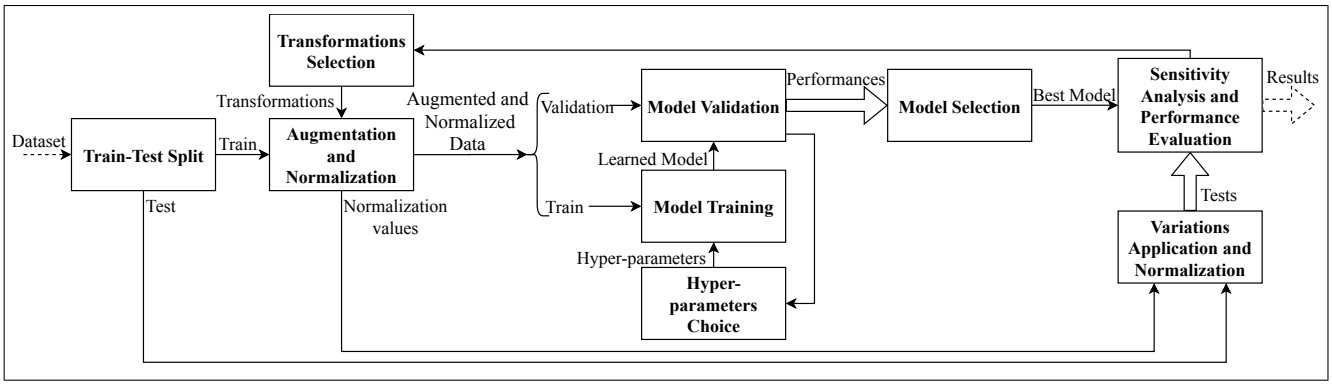

Figure 1. Workflow overview schema.

importance as it largely affects the resulting model in both qualitative (the model may learn different representations) and computing time terms. The main ones involve the architecture of the model, the learning rate and its decay speed, the batch size, the optimizer and its arguments choice, regularization type and its value choice, the number of epochs. In our experiments we search the space of hyper-parameters guided by the model validation performances.

Once a satisfactory number of trials is made, we move on to the selection of the best model according to the performances' estimates made by model validation. After that, as a further model performance evaluation, we test it against test data coming from the original dataset and possibly with the application of several variations so to better assess its generalization capabilities. This is further supported by model visualization on test datasets. At this stage, according to sensitivity analysis and performance evaluation, we may decide to make ad hoc modifications of the transformations we apply to the data in earlier stages, and then go through the workflow again.

3.2. The classification task

On the basis of the set up framework, we conduct several classification experiments. For each experiment we take note of the confusion matrices on both validation phase – for model selection – and tests phase – for sensitivity analysis and performance evaluation. In order to assess the quality of the results and to summarize them, we use metrics of performance commonly employed within classification problems [25]. Measures for multi-class classification are based on a generalization of binary classification measures for C classes. M index represent macro-averaging over the C classes.

3.3. Model visualization

Sensitivity analysis and performance evaluation are not only dealt with metrics of performance but also by visualizing the model using the following approaches.

- **Kernels visualization** of the first (it is the most interpretable) layer of a CNN is useful because well-trained models usually display nice and smooth filters without any noisy patterns. Noisy patterns can be an indicator of a network that has not been trained for long enough, or possibly a very low regularization strength that may have led to overfitting.
- **Gradient-weighted Class Activation Mapping (GradCAM)** uses the gradients of any target concept, flowing into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image for predicting the concept. Among the other properties, this method helps in achieving generalization by identifying dataset bias [24].
- **Occlusion method** acts by systematically occluding different portions of the input image, and monitoring the output. Therefore, if the model is correctly localizing the relevant objects, the probability of the correct class drops significantly when the objects are occluded [29].

4. Experiments

Our experiments are about two main tasks: binary and multi-class classification. In this section, first we describe the dataset and the variations created to perform sensitivity analysis. Then we go trough the experiments setup where we describe how we have conducted the experiments to accomplish the two task and finally we show the results.

4.1. Datasets

We analyze 18,158 images of tomato plant leaves, which are split in 10 classes, 9 diseases and 1 healthy. The scope is to predict the correct class of a leaf image. All the images are sized 256x256 pixels.

Given the original dataset we have conducted the training phase on different variations of the original dataset, we

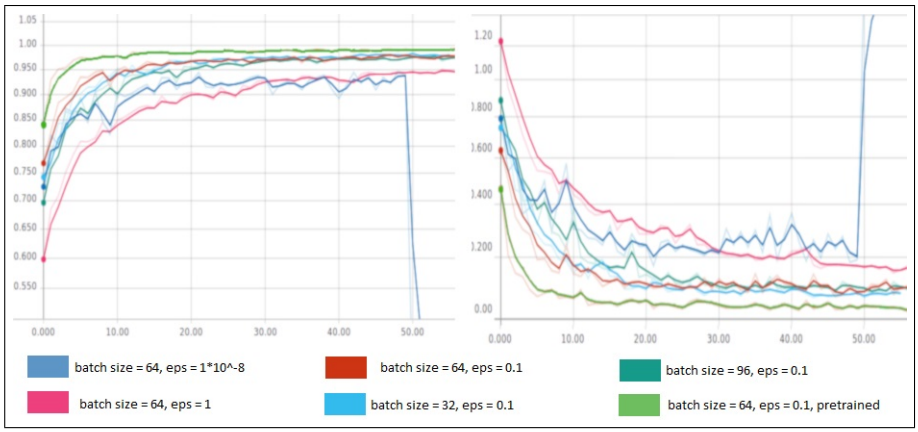

Figure 2. Tensorboard graph of experiments, from left to right: **Validation Accurancy over epochs**, **Valdiation error over epochs**.

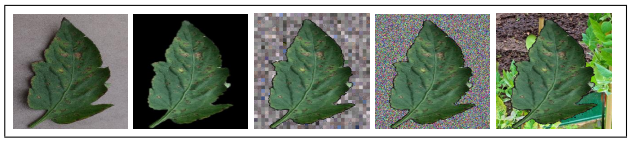

Figure 3. Samples of dataset images, from left to right: **Original PlantVillage**, **Segmented PlantVillage**, **Random Background Crop**, **Random Noise Background** and **Random Background Image**.

have worked on the backgrounds in order to find a way to overcome the limits highlighted in [8, 23] and confirmed by us. Especially, we have found out a dependence between the original backgrounds and their belonging class. In Figure 3 we can see all the images and the different variations of the dataset that we have tested.

- **Original PlantVillage dataset.** It consists of images of tomato leaves laying over a gray table.
- **Segmented PlantVillage dataset.** It consists of images of tomato leaves with no background, this variation has been used also in [23]. In our work we have re-segmented the images since in the available segmented dataset they were incomplete.
- **Random Background Crop dataset.** It consists of the leaf images coming from PlantVillage with a background composed of crops of background coming from the Original PlantVillage dataset.
- **Random Noise Background dataset.** It consists of leaf images coming from PlantVillage with a background composed of random colored pixels.
- **Random Background Image dataset.** It consists of leaf images coming from PlantVillage dataset with a background taken from a set of 4,176 images we have built by looking for tomato plantation pictures on the Internet. The objective of this set was not only to decouple the background influence on the classifier but also try to take the leaf in a situation similar to a realistic one, where the leaf image is taken directly in the plantation.

4.2. Experiments setup

4.2.1 Data augmentation and normalization

We have performed data augmentation applying to the original sample two randomly selected transformations from this pool: Flip Top Bottom, Flip Left Right, Rotate 90°, Rotate 180°, Rotate 270°, Flip Top Bottom and Rotate 90°, Flip Top Bottom and Rotate 270°.
Another key aspect of our setup is the normalization of the input images, indeed, the chosen normalization changes dramatically the model performance. In our work we have considered the following normalization approaches:
Dataset Normalization. Given the chosen variation of the dataset we have computed the mean and the standard deviation of the train dataset.
Per-leaf Normalization. Starting from the segmented train dataset we have computed the mean and standard deviation of the non-black pixels.

4.2.2 Binary classification setup

LeNet architecture. This network proposed in [28] is the first and simplest deep model available. We have adapted it to perform our binary task.
Weights initialization. For this network we have started using training from scratch approach. The weights in the network are initialized using Xavier initialization ([27]). This helps the training since we start from a normally distributed set of weights and we avoid that certain weighs vanish through epochs.

Activation function. The layer activation functions is the rectified linear unit (ReLU) function since it diminishes the likelihood of a gradient to vanish.
Loss function. The chosen loss function for the binary classification is the binary cross-entropy since it is properly designed for binary tasks.
Validation. For this task we have performed 10-fold cross validation. This kind of validation has been used to estimate the hyper-parameters.

4.2.3 Multi-class classification setup

AlexNet architecture. For this task we have chosen a more complex model since we have first tested LeNet architecture achieving results very far from the state-of-the-art [2].
Weights initialization. For this network we have tried both training from scratch approach and transfer learning. The weights in the first case are initialized using Xavier initialization ([27]). In the transfer learning approach we have used the weights coming from the training on ImageNet [9].
Activation function. Layers activation function is again the rectified linear unit (ReLU) function.
Loss function. In this case we have used cross-entropy as loss function since it is largely used for multi-class classification tasks. ==========
Validation. in the case of Multi-class Classification we opt to proceed with cross validation. We have chosen a different approach from binary classification. In fact k-fold cross validation applied on AlexNet requires a lot of computing time that is not justified in our dataset since there is small amount of variance in our pictures and consequently in each possible fold.
Optimizer: As optimizer we have choosen Adam [7]. An important annotation should be done about this optimizer. In fact trough the training process we noticed that accuracy after a bunch of epochs drops. This was caused by the ε coefficient, this coefficient used to avoid division by zero when the gradient is almost zero in parameters update. By default is set to $1 * 10^{-8}$ and it causes large weights updates. So we set it to 0.1 to avoid this problem.

4.3. Results and discussion

4.3.1 Binary Classification

After a training phase using 10-fold cross validation, we have selected as best hyperparameters for our binary classification problem, the batch size equal to 32 and the total number of epochs equal to 15. With this configuration, after retraining the net on the global train set, we have obtained the results on the test set shown in the table 1. Beeing a very unbalanced dataset because of the number of healthy leaves with respect to the non healthy ones, we have expected to have an healthy accurancy lower than non healthy. The result shows that our assumptions was correct. In fact LeNet accurancy for healthy aren't good as for non healthy one. By the way the average performances of our net are good.

Measure	Results
$AverageAccuracy$	0.999184
$Precision_M$	0.993377
$Recall_M$	0.996678
F_1score_M	0.995025

Table 1. Performance of binary classification.

4.3.2 Multi-class Classification

As we have done for binary classification, we have tried different run using different hyperparameters and we test our result on the validation set. As figure 2 shows, it's possible to understand that the best configuration it has been given from Alexnet pretrained, with batch size equal to 64, eps equal to 0.1 for 57 epochs. The results obtained from this configuration are described in the first row of the table 2. AlexNet pretrained has a very good performance for all the classes and obtained also better result than the state of the art [2]. We have expected that classes with less samples, e.g. Tomato Mosaic Virus, would have been missclassified compared to classes with an higher number of samples, e.g. Tomato Yellow Leaf Curl. This does not happen in the reality, because our classifier missclassified only some samples that aren't related to the size of the class predicted.

4.3.3 Sensitivity Analisys

After having reached a good level of accuracy for our multi-class classification task, we have moved our attention to analise the robustness of our dataset by computing gradcam and occlusion methods to analise which parts of the leaf trigger the classifier. The table 2 shows all the different approach we have tried. It's possible to see that the preformances of the results obtained are very far from the what we have done in the previous task. Sometimes also happens that some classes have never be predicted at all. Then we have tried a different way, applying both normalization and backgrounds moving to something similar to what a predictor will see when it will applied on the field. Using per-leaf normalization and random background images we obtain the result shows in figure 4. It is also possible to see in figure 5 how the system is more robust to some background variation, activating only on the leaf surface.

Train	Test	Performance
Original	Original	$AverageAccuracy = 0.998560$ $Precision_M = 0.998339$ $Recall_M = 0.998173$ $F_1score_M = 0.998254$
	Segmented	$AverageAccuracy = 0.390169$ $Precision_M = NaN$ $Recall_M = 0.285877$ $F_1score_M = Nan$
Segmented	Original	$AverageAccuracy = 0.617933$ $Precision_M = 0.730197$ $Recall_M = 0.494125$ $F_1score_M = 0.486940$
	Segmented	$AverageAccuracy = 0.966511$ $Precision_M = 0.959517$ $Recall_M = 0.956894$ $F_1score_M = 0.958019$
Random Background Crop	Original	$AverageAccuracy = 0.769175$ $Precision_M = 0.794344$ $Recall_M = 0.673085$ $F_1score_M = 0.671527$
	Segmented	$AverageAccuracy = 0.848037$ $Precision_M = 0.885544$ $Recall_M = 0.737432$ $F_1score_M = 0.739969$
Random Noise Background	Original	$AverageAccuracy = 0.578862$ $Precision_M = 0.666754$ $Recall_M = 0.490099$ $F_1score_M = 0.499304$
	Segmented	$AverageAccuracy = 0.748469$ $Precision_M = 0.820938$ $Recall_M = 0.709873$ $F_1score_M = 0.697417$

Table 2. Performance of multi-class classification with difference training and test sets. We use a NaN value for the class that are never predicted to avoid division by zero in precision, recall and F1 score.

5. Conclusion and Future Works

In this study we have proposed deep models approach to build classifiers for disease classification and the related sensitivity analysis on plant village dataset. Our results confirms what suggested in [2] that deep models performs very good on this task. Furthermore we provide an improved dataset to overcome the problem of sensitivity analysis as we highlined during the work.

A final remark about occlusion methods purposed [2]. We have tried to see the outcomes of the same occlusion method but with our trained models, we are not able at all to come up with the same conclusion about independence of the prediction with respect to the background.

In order to develop further works on the Random Background Image we have created a new dataset containing 256 images found on Internet. We have used

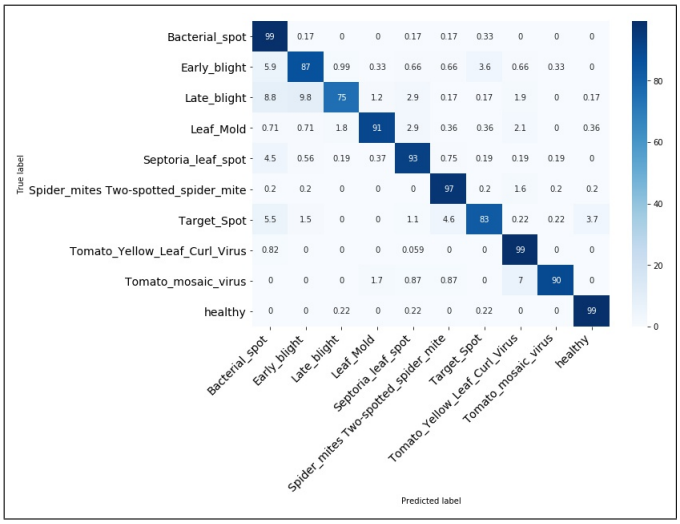

Figure 4. Confusion matrix of AlexNet pretrained in multi-class classification problem in percentage using per-leaf normalization and random background images.

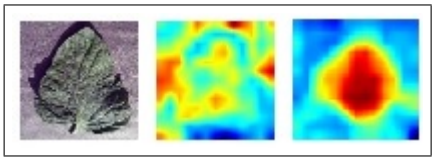

Figure 5. From left to right: **Original image, class: Target Spot**, **GradCAM of AlexNet pretrained on Original dataset**, **GradCAM of AlexNet pretrained on random background images**. It is possible to see the difference on the activation of GradCAM, one is distributed both on the leaf and on the background, the other is locate only on the leaf disease.

tested these images against the AlexNet model trained on Random Background Image dataset. The accuracy on this test set is 29.8%. The reason of this result is in line with the results on internet images purposed in [23]. Moreover we have tried to train a vgg11 network on Random Background Image dataset and tested it against the images coming from internet; the accuracy reached 35.8% We think this result is so poor if compared with the results on Normal Test because no one validated this test set. In order to improve this work we think that two ways should be followed:

- build a dataset of images coming from plantations that is validated by a domain expert.
- use more complex deep learning models as the experiment with vgg11 suggests

Acknowledgements

aaa

A. Supplementary Material

References

[1] A. P. Tai, M. V. Martin, and C. L. Heald. Threat to future global food security from climate change and ozone air pollution. *Nat. Clim. Change*, 4(9):817–821, 2014. https://doi.org/10.1038/nclimate2317. 1

[2] M. Brahimi, B. Kamel, and A. Moussaoui. Deep Learning for Tomato Diseases: Classification and Symptoms Visualization. *Applied Artificial Intelligence*, 31(4):299–315, Apr. 2017. https://doi.org/10.1080/08839514.2017.1315516. 1, 2, 5, 6

[3] L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, Oct. 2001. https://doi.org/10.1023/A:1010933404324. 2

[4] C. A. Harvey et al. Extreme vulnerability of smallholder farmers to agricultural risks and climate change in Madagascar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1639):20130089–20130089, Feb. 2014. https://doi.org/10.1098/rstb.2013.0089. 1

[5] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, Sep. 1995. https://doi.org/10.1007/BF00994018. 2

[6] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, Jan. 1967. https://doi.org/10.1109/TIT.1967.1053964. 2

[7] J. L. B. Diederik P. Kingma. Adam: A method for stochastic optimization. *international Conference on Learning Representations*, 2015. https://arxiv.org/pdf/1412.6980.pdf. 5

[8] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park. A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. *Sensors*, 17(9), 2017. https://doi.org/10.3390/s17092022. 2, 4

[9] ImageNet. ImageNet dataset. www.image-net.org/. 5

[10] Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services Fourth session. *Report of the Plenary of the Intergovernmental Science-PolicyPlatform on Biodiversity and Ecosystem Services on the work of its fourth session*, Kuala Lampur, 2016. https://www.ipbes.net/event/ipbes-4-plenary. 1

[11] International Telecommunication Union (ITU). *ICT Facts and Figures – the World in 2015*, Geneva, 2015. https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf. 1

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira and C. J. C. Burges and L. Bottou and K. Q. Weinberger, editor, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf. 2

[13] L. E. Ehler. Integrated pest management (IPM): definition, historical development and implementation, and the other IPM. *Pest Management Science*, 62(9):787–789, 2006. https://doi.org/10.1002/ps.1247. 1

[14] Y. LeCun, Y. Bengio, and G. Hinton. Deep Learning. *Nature*, 521:436–444, 2015. https://doi.org/10.1038/nature14539. 2

[15] O. W. Liew, P. C. J. Chong, B. Li, and A. K. Asundi. Signature Optical Cues: Emerging Technologies for Monitoring Plant Health. *Sensors*, 8(5):3205–3239, 2008. https://doi.org/10.3390/s8053205. 2

[16] A.-K. Mahlein, E.-C. Oerke, U. Steiner, and H.-W. Dehne. Recent advances in sensing plant diseases. *European Journal of Plant Pathology*, 133, May 2012. https://doi.org/10.1007/s10658-011-9878-z. 2

[17] M. Mazarei, I. Teplova, M. R. Hajimorad, and C. N. Stewart. Pathogen Phytosensing: Plants to Report Plant Pathogens. *Sensors*, 8(4):2628–2641, 2008. https://doi.org/10.3390/s8042628. 2

[18] M. Meroni, M. Rossini, V. Picchi, C. Panigada, S. Cogliati, C. Nali, and R. Colombo. Assessing Steady-state Fluorescence and PRI from Hyperspectral Proximal Sensing as Early Indicators of Plant Stress: The Case of Ozone Exposure. *Sensors*, 8(3):1740–1754, 2008. https://doi.org/10.3390/s8031740. 2

[19] PlantVillage. PlantVillage Dataset. https://github.com/spMohanty/PlantVillage-Dataset. 1

[20] R. N. Strange and P. R. Scott. Plant Disease: A Threat to Global Food Security. *Annual Review of Phytopathology*, 43(1):83–116, Sep. 2005. https://doi.org/10.1146/annurev.phyto.43.113004.133839. 1

[21] P. R. Reddy, S. Divya, and M. R. Vijayalakshmi. Plant disease detection techniquetool-a theoretical approach. *IJITR*, pages 91–93, 2015. 2

[22] T. Reed and J. Dubuf. A Review of Recent Texture Segmentation and Feature Extraction Techniques. *CVGIP: Image Understanding*, 57(3):359–372, 1993. https://doi.org/10.1006/ciun.1993.1024. 2

[23] S. P. Mohanty, D. P. Hughes, and M. Salathé. Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science*, 7:1419, Sep. 2016. https://doi.org/10.3389/fpls.2016.01419. 1, 2, 4, 6

[24] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *CoRR*, abs/1610.02391, 2016. http://arxiv.org/abs/1610.02391. 3

[25] M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009. https://doi.org/10.1016/j.ipm.2009.03.002. 3

[26] UNEP, International Fund for Agricultural Development (IFAD). *Smallholders, Food Security, and the Environment*, Rome, 2013. https://www.ifad.org/documents/10180/666cac24-14b6-43c2-876d-9c2d1f01d5dd. 1

[27] X. Glorot, Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.

http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf. 4, 5

[28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324", Nov. 1998. https://doi.org/10.1109/5.726791. 2, 4

[29] M. D. Zeiler and R. Fergus. Visualizing and Understanding Convolutional Networks. *CoRR*, abs/1311.2901, 2013. http://arxiv.org/abs/1311.2901. 2, 3