Hierarchical Deep Learning Approach
for Plant Disease Detection

Joana Costa[1,2(✉)], Catarina Silva[1,2], and Bernardete Ribeiro[2]

[1] School of Technology and Management, Polytechnic Institute of Leiria,
Leiria, Portugal
[2] CISUC - Center for Informatics and Systems, University of Coimbra,
Coimbra, Portugal
{joanamc,catarina,bribeiro}@dei.uc.pt

Abstract. In this paper we propose a hierarchical deep learning approach for plant disease detection. The detection of diseases in plants using deep image approaches is attracting researchers as a way of taking advantage of cutting-edge learning techniques in scenarios where major benefits can be achieved for mankind. In this work, we focus on diseases of three major different agricultural crops: apple, peach and tomato. Using a real-world dataset composed of nearly 24,000 images, including healthy examples, we propose a hierarchical deep learning approach for plant disease detection and compare it with the standard deep learning approaches. Results permit the conclusion that hierarchical approaches can overcome standard approaches in terms of detection performance.

1 Introduction

One of the greatest technological advances that humans have ever made has been the domestication of plants during the agricultural revolution 8 to 12,000 years ago at multiple sites around the world [5]. These and the following breakthroughs allowed for the evolution of human population to the current remarkable 7.7 billion people[1].

Such evolution has put colossal pressure on agriculture technology, that is nowadays under immense challenges due to infectious diseases and pests spread by the globalization and compounded by climate change [1].

Hence, major benefits can arise from more accurate and faster detection of plant diseases, which is being increasingly enabled by sensors that obtain real time information from crops plantation sites.

Such approaches, usually coined as smart farming since they constitute an evolution of traditional techniques, are also important to tackle the challenges of agricultural production in terms of productivity, environmental impact, food security and sustainability [7].

---

[1] http://www.worldometers.info/world-population/.

To tackle these challenges, more information on agricultural ecosystems is needed, which is being obtained through new technologies achieved by monitoring, measuring and continuously analyzing various physical aspects and phenomena. Such scenarios have become possible through the development of Internet of Things (IoT), allowing that remote and often autonomous sensors provide information, whether sensory as humidity and temperature, or imaging as snapshots taken by fixed cameras or drone-based imaging.

This setup permits the gathering of huge amounts of data over large geographical areas, thus originating problems that can be framed into big data scenarios, which have not yet been widely applied in agriculture [6].

A recent technique that can deal with such challenges is Deep Learning (DL) [10] that by using "deeper" neural networks better represents data by means of various convolutions, allowing larger learning capabilities and higher accuracy [7].

In this paper we propose a two-level deep learning hierarchical approach for plant disease detection. The first level starts to determine the crop in the image and the second level focuses on learning the specific diseases of that specific crop. Results show the overall advantages of the approach and allow for the application in real scenarios in which the crop is a given.

The rest of the paper is organized as follows. In the next section we will present the current state of the art approaches to plant disease detection with deep learning, including a brief introduction to existing deep learning implementation strategies. In Sect. 3 we will introduce the hierarchical approach for plant disease detection starting from a standard baseline approach. Section 4 will detail on the experimental setup, namely on the dataset used, the deep learning model setup and the performance metrics. Section 5 will present and discuss the results obtained, and in Sect. 6 final conclusions and future lines of research are pointed out.

2   Plant Disease Detection with Deep Learning

2.1   Deep Learning Strategies

Deep learning is extremely good for visual feature extraction from images, audio signals, or text, which makes it very attractive to be used today. Given that specific datasets of plant disease images can be easily processed by deep neural networks without the need to extract manually hand-crafted features, we posit that higher performance than traditional methods can be obtained from these models. In particular, Convolutional Neural Networks (CNN) (and its variants) are becoming the current state of art in many image processing problems.

Deep neural network architectures fall into the four main types (for a recent survey see [12]):

1. Stacked Auto-Encoder (SAE)
2. Convolutional Neural Networks (CNN)
3. Resctricted-Boltzmann Machines (RBM)

4. Deep Belief Networks (DBN)

Despite the diversity of models, improved learning algorithms and new application players are exponentially increasing.
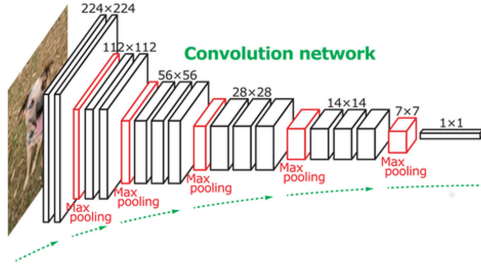

Fig. 1. Convolutional neural network [16]

CNN are depicted in Fig. 1. Being biologically inspired variants of Multilayer perceptrons. From Hubel and Wiesel's early work on the cat's visual cortex [3,4], we know that the visual cortex contains a complex disposition of cells, called receptive fields, that are sensitive to small regions of the visual field. These simple cells act as filters that respond to edge-like patterns and form complex hierarchies that are invariant to the position of the pattern. Being the visual cortex the most powerful visual processing system known, its behavior has inspired many models, for instance, the CNN.

2.2 Image-Based Approaches

Convolutional Neural Networks (CNN) are multilayer neural networks designed to recognize visual patterns directly from image pixels. The work by [9,11] has been pioneer for the current CNN that are researched today. As a consequence, CNN have been catapulted to the center of object recognition research. The rekindled interest in CNN is largely attributed to [8] CNN model, that showed significantly higher image classification accuracy on the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Their success resulted from a model inspired by LeCun's previous work and a few twists that enabled training with 1.2 million labeled images (*e.g.* GPU programming, $\max(x,0)$ rectifying non-linearities and dropout regularization). Likewise, the emergence of the giant processing power from parallel programming models developed by large HPC (High Performance Computing) teams working in the industry also contributed by a large amount to their success.

2.3 Disease Detection

Although it is still a developing research field, different successful approaches have been pursued as we will show in this section.

In [2] a deep-learning-based approach is presented to detect diseases and pests in tomato plants using images captured in-place. Three main families of DL detectors were considered and combined with additional information to effectively recognize nine different types of diseases and pests, with the ability to deal with complex scenarios from a plant's surrounding area.

In [14] DL is applied to Cassava disease detection. Cassava is the third largest source of carbohydrates for human food in the world, hence any improvement in the disease control can prove precious specially in less developed countries. In this work, the authors have applied transfer learning to train a deep convolutional neural network to identify three diseases and two types of pest damage, achieving accuracy of up to 98%.

In [13] the authors take advantage of increasing global smartphone penetration and recent advances in computer vision made possible by deep learning to propose smartphone-assisted disease diagnosis. Using a deep convolutional neural network to identify 14 crop species and 26 diseases, the trained model achieves an accuracy of 99.35% on a held-out test set, demonstrating the feasibility of this approach.

In [7] it is possible to find a survey on deep learning in agriculture, giving a broad view of the wide range of applications and methods that are being pursued.

3 Hierarchical Approach for Plant Disease Detection

To introduce the proposed approach, we will firstly define the problem of plant disease detection using images. It can be stated as follows.

Consider a set I of N images of C agricultural crops. For each crop c there are different target classes, i.e it is a multiclass problem. Those classes include healthy crops and specific diseased crops.

The goal can then be stated as, given a set of images I,

$$I = \{i_1, i_2, \ldots, i_N\}, \tag{1}$$

identify, for each image i_n the target class T_d^c, i.e. make the correspondence of each image i_n to a specific disease d of crop c or just the correspondence to a healthy state.

3.1 Standard Baseline Approach

The standard baseline approach is depicted in Fig. 2 and consists of building a model that uses all images and detects for each input simultaneously the crop and the disease. This approach is the typical deep learning approach in which all preprocessing efforts are supported by the neural network and has usually good results.

However, in plant disease detection, specially in approaches based in the crop/disease pair, difficulties can arise for instance due to the same disease in different crops, or different diseases producing similar damaged patterns but in different crops and hence a more structured approach may become appropriate.
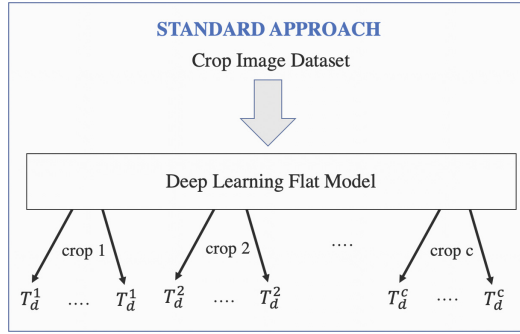
Fig. 2. Standard baseline deep learning flat approach.

3.2   Hierarchical Approach

Figure 3 shows the proposed two-step hierarchical deep learning approach.

The rationale is to divide the problem of crop disease detection in two subproblems. The first (STEP 1) consists in identifying the crop and the second (STEP 2) consists in determining for a specific crop which, if any, disease in the dataset is identified/recognized. Such division may prove reasonable for various reasons and applications.

It also allows for the reuse of the models in STEP 2 in scenarios where we have plantations of only one crop and the goal is to detect diseases in that specific crop.
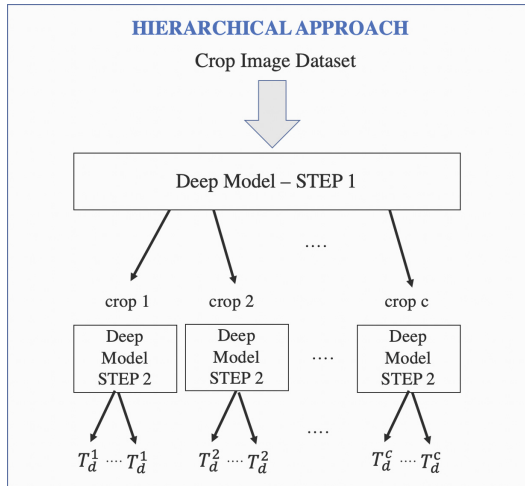

Fig. 3. Hierarchical deep learning approach.

4   Experimental Setup

4.1   Dataset

PlantVillage is an online platform dedicated to crop health and crop diseases (http://www.plantvillage.org) with more than 2 million site visits reported [5]. It started as an online crowdsourced forum where users could put questions and receive answers from other users. Additionally, a library of open access information of images of over 150 crops and over 1,800 diseases was also constructed and curated by plant pathology experts.
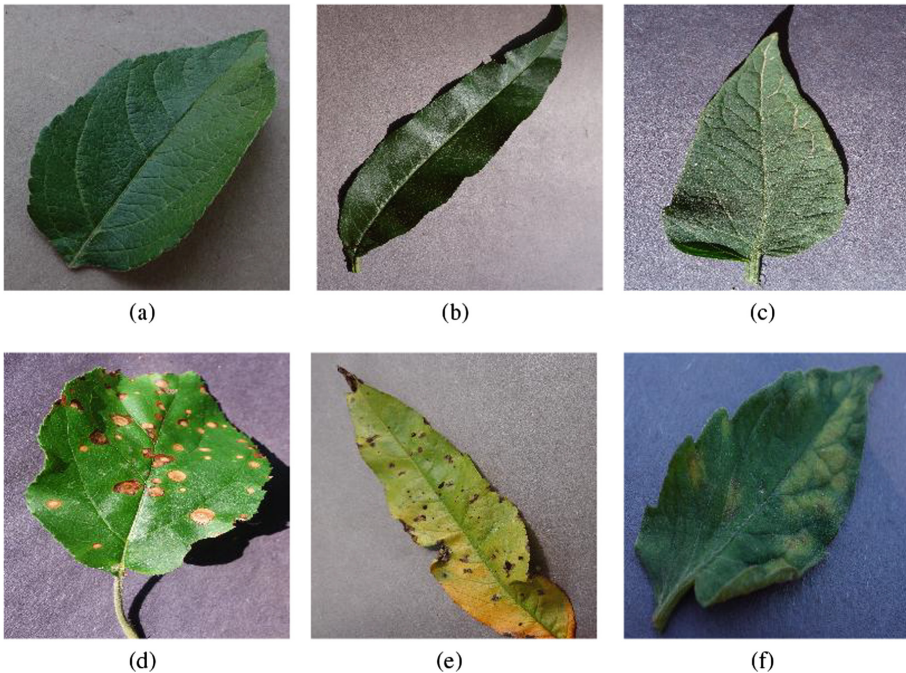

Fig. 4. Example images of some classes in the dataset: (a) Apple healthy; (b) Peach healthy; (c) Tomato healthy; (d) Apple black rot; (e) Peach bacterial spot; and, (f) Tomato leaf mold.

In this work, we have constructed a dataset using the PlantVillage images to test the proposed hierarchical approach for plant disease detection. We have focused on three major different agricultural crops: apple, peach and tomato. Figure 4 shows example images of the classes in the dataset, namely three examples of healthy apple, peach, and tomato crops, and three examples of diseases from the same crops.

Table 1 presents the classes associated with these three crops, consisting in 16 classes and nearly 24,000 image examples. We have used 70% of each class

Table 1. Dataset classes and training/testing sets

Class	Number of samples	Train samples	Test samples
Apple___Apple_scab	630	441	189
Apple___Black_rot	621	435	186
Apple___Cedar_apple_rust	275	193	82
Apple___healthy	1645	1152	493
Peach___Bacterial_spot	2297	1608	689
Peach___healthy	360	252	108
Tomato___Bacterial_spot	2127	1489	638
Tomato___Early_blight	1000	700	300
Tomato___healthy	1591	1114	477
Tomato___Late_blight	1909	1336	573
Tomato___Leaf_Mold	952	666	286
Tomato___Septoria_leaf_spot	1771	1240	531
Tomato___Spider_mites Two-spotted_spider_mite	1676	1173	503
Tomato___Target_Spot	1404	983	421
Tomato___Tomato_mosaic_virus	373	261	112
Tomato___Tomato_Yellow_Leaf_Curl_Virus	5357	3750	1607
Total	23988	16793	7195

for training purposes and 30% for testing. Notice that both classes per crop and examples per class are heterogeneous.

4.2   Deep Learning Models

The evaluation of our approach was performed with a fine-tuned InceptionV3 CNN proposed in [17]. In an Inception CNN, the input is split into a few lower-dimensional embeddings, transformed by a set of specialized filters, and merged by concatenation. The solution space of this architecture is a strict subspace of the solution space of a single large layer operating on a high-dimensional embedding, and this it is expected the representational power of large and dense layers, but at a considerably lower computational complexity [18].

We have used the implementation of the InceptionV3 distributed in Keras[2] with the TensorFlow[3] backend. The InceptionV3 was initialized with weights pre-trained on the ImageNet dataset, and fine-tuned to transfer learning.

We have empirically defined our experimental configurations, and thus all CNN run for 30 epochs and 50 steps each. They consistently converged to a better accuracy performance.

The main goal was to compare the performance of the standard and hierarchical approaches using identical models as base classifiers, not targeting improvements in the classification performance of the models.

---

[2] http://keras.io.
[3] http://tensorflow.org.

4.3   Evaluation Metrics

In this setup a multiclass single label problem is defined, i.e., each image belongs to just one of the specified classes. One of the most common evaluation strategies is to calculate the accuracy of a classifier, i.e., the percentage of image instances that are correctly assigned to their class.

We can also simplify our problem into a binary decision problem, as each image can be classified as being in a given class, or not. In order to evaluate the binary decision task, a contingency matrix can be defined to represent the possible outcomes of the classification, as shown in Table 2.

Table 2. Contingency table for binary classification

	Class positive	Class negative
Assigned Positive	a	b
	(True Positives)	(False Positives)
Assigned Negative	c	d
	(False Negatives)	(True Negatives)

Traditional measures can be defined based on this contingency table, such as error rate ($\frac{b+c}{a+b+c+d}$) and the above-mentioned accuracy ($\frac{a+d}{a+b+c+d}$). However, for unbalanced problems, i.e., problems where the number of positive examples is rather different among classes, or in the case of a binary problem, the number of positive examples is rather different from the negative examples more specific measures should be defined to capture the performance of each model. Typical examples include recall ($R = \frac{a}{a+c}$) and precision ($P = \frac{a}{a+b}$). Additionally, combined measures that give a more holistic view of performance in just one value, like the van Rijsbergen F_1 measure [15], which combines recall and precision:

$$F_1 = \frac{2 \times P \times R}{P + R}. \tag{2}$$

Two conventional methods are widely used in multiclass scenarios, namely macro-averaging and micro-averaging. Macro-averaged performance scores are obtained by computing the scores for each class and then averaging these scores to obtain the global means. Differently, micro-averaged performance scores are computed by summing all the previously introduced contingency matrix values (a, b, c and d), and then use the sum of these values to compute a single micro-averaged performance score that represents the global score.

5   Experimental Results and Analysis

Table 3 summarises the performance results obtained by classifying the testing set, with the macro-averaged performance measures, namely recall, precision, F1,

Table 3. Standard and hierarchical results for each crop/disease pair

Class	Standard approach				Hierarchical approach			
	Recall	Precision	F1	Accuracy	Recall	Precision	F1	Accuracy
Apple_Apple_scab	91.53%	77.93%	84.18%	99.10%	97.88%	72.55%	83.33%	98.97%
Apple_Black_rot	91.40%	92.92%	92.64%	99.62%	98.39%	91.96%	95.06%	99.74%
Apple_Cedar_apple_rust	71.95%	95.16%	81.94%	99.64%	95.12%	95.12%	95.12%	99.89%
Apple_healthy	93.31%	99.78%	96.44%	99.53%	99.80%	77.48%	87.23%	98.00%
Peach_Bacterial_spot	99.98%	96.88%	97.92%	99.60%	95.50%	95.50%	95.50%	99.14%
Peach_healthy	99.07%	96.40%	97.72%	99.93%	80.56%	95.60%	87.44%	99.65%
Tomato_Bacterial_spot	65.05%	93.47%	76.71%	95.50%	72.41%	98.93%	83.62%	97.48%
Tomato_Early_blight	84.67%	51.31%	63.90%	96.01%	78.67%	65.92%	71.73%	97.41%
Tomato_healthy	57.44%	92.57%	70.89%	96.87%	98.32%	89.85%	93.89%	99.15%
Tomato_Late_blight	72.60%	97.20%	83.12%	97.65%	95.99%	90.76%	93.30%	98.90%
Tomato_Leaf_Mold	80.77%	84.62%	82.65%	98.65%	87.06%	96.51%	91.54%	99.36%
Tomato_Septoria_leaf_spot	96.42%	76.42%	85.26%	97.54%	88.51%	93.81%	91.09%	98.72%
Tomato_Spider_mites 2-spotted_spider_mite	41.15%	84.15%	55.27%	95.34%	94.63%	90.84%	92.70%	98.96%
Tomato_Target_Spot	72.21%	80.21%	76.00%	97.33%	74.11%	86.91%	80.00%	99.83%
Tomato_Tomato_mosaic_virus	57.14%	25.10%	34.88%	96.68%	91.07%	86.44%	88.70%	99.64%
Tomato_Tomato_Yellow_Leaf_Curl_Virus	98.88%	80.74%	88.90%	94.48%	93.96%	98.63%	96.24%	98.36%
Average	79.54%	82.87%	79.28%	97.78%	90.12%	89.18%	89.16%	98.83%

and accuracy. Each row refers to one of the 16 classes considered. Results are presented both for the standard flat approach and for the hierarchical approach.

A global analysis of macro-average shows the potential benefits of the proposed method. Specifically in classes where the flat model underperforms in terms of recall and precision, the hierarchical approach shows significant improvement. Notice that the accuracy often used as the only metric, in fact offers a rather biased notion of the performance obtained.

Table 4 shows the macro-averaged performance measures for the three crops in STEP 1: Apple, Peach, and Tomato. This remarkably easier problem exhibits an almost perfect result. This achievement allows for the improvement that STEP 2 presents over the flat approach.

Table 4. Hierarchical approach STEP 1 results

Class	Hierarchical STEP 1			
	Recall	Precision	F1	Accuracy
Apple	100.00%	81.13%	89.58%	96.93%
Peach	94.98%	97.05%	96.01%	99.12%
Tomato	96.26%	100.00%	98.09%	97.16%
Average	97.08%	92.73%	94.56%	97.74%

6   Conclusions and Future Work

In this paper we have proposed a two-level deep learning hierarchical approach for plant disease detection. The first level starts to determine the crop in the image and the second level focuses on learning the specific disease of that specific crop. Results show the overall advantages of the approach and are expected to be used in real scenarios applications in which the crop is given.

Results show the effectiveness of the proposed hierarchical approach in terms of overall performance. When targeting specific crop diseases such approach can bring significant advantages, when considering other measures than accuracy that only provides a partial representation of the resulting performance.

Future work is foreseen on further exploring the hierarchical nature of the proposal and additional research on different deep architectures.

References

1. Bourne, J.E.: The End of Plenty: The Race to Feed a Crowded World. W. W. Norton & Company, New York (2015)
2. Fuentes, A., Yoon, S., Kim, S.C., Park, D.S.: A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors **17**(9), 2022 (2017)
3. Hubel, D.H., Wiesel, T.N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. **160**, 106–154 (1962)
4. Hubel, M., Wiesel, T.N.: Brain and Visual Perception. Oxford Univeristy Press, Oxford (2005)
5. Hughes, D.P., Salathé, M.: An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. CoRR (2015)
6. Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F.: A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. **143**(1), 23–37 (2017)
7. Kamilaris, A., Prenafeta-Boldú, F.X.: Deep learning in agriculture: a survey. Comput. Electron. Agric. **147**, 70–90 (2018)
8. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
9. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Hubbard, R.E.H.W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Comput. **1**(4), 541–555 (1989)
10. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
11. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
12. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E.: A survey of deep neural network architectures and their applications. Neurocomputing **234**, 11–26 (2017)
13. Mohanty, S.P., Hughes, D.P., Salathe, M.: Using deep learning for image-based plant disease detection. Frontiers Plant Sci. **7**, 1419 (2016)

14. Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., Hughes, D.P.: Deep learning for image-based cassava disease detection. Frontiers Plant Sci. **8**, 1852 (2017)
15. van Rijsbergen, C.: Information Retrieval. Butterworths, London (1979)
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556 (2014)
17. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
18. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5987–5995. IEEE (2017)