Machine Learning for detection of crop diseases

Giovanni Belval

1 Introduction

The last few years , the world population has been growing faster and faster. This high rate of growth implies bigger and bigger ressources in order to feed humans. However , recent studies have shown that the crop yields only improve from 1.3% each years , instead of the 2.4% required [1]. Moreover , crops are continuously threatened by various diseases. it is estimated that between 20% and 40% or yearly crops are lost each year due to diseases and insect pests , costing a global loss of $200 billions and $70 billions respectively [2]. In order to strop this trend , we could try to use more chemical products , but this is probably not the best solution. Increasing the amount of chemicals product will ultimately lead to more pollution of the soil , greater risk or humain health due to crops contaminations and finally , bacteria and insect will develop higher resistance , forcing us to use stronger and stronger chemical products. in order to solve these problems , one can think about using Machine learning methods to detect and forecast crops diseases.
In this paper , I will try to develop some machine learning algorithms in order to detect crops diseases (mainly on tomatoes , potatoes and peppers) using the **PlantVillage** dataset available on Kaggle [3].

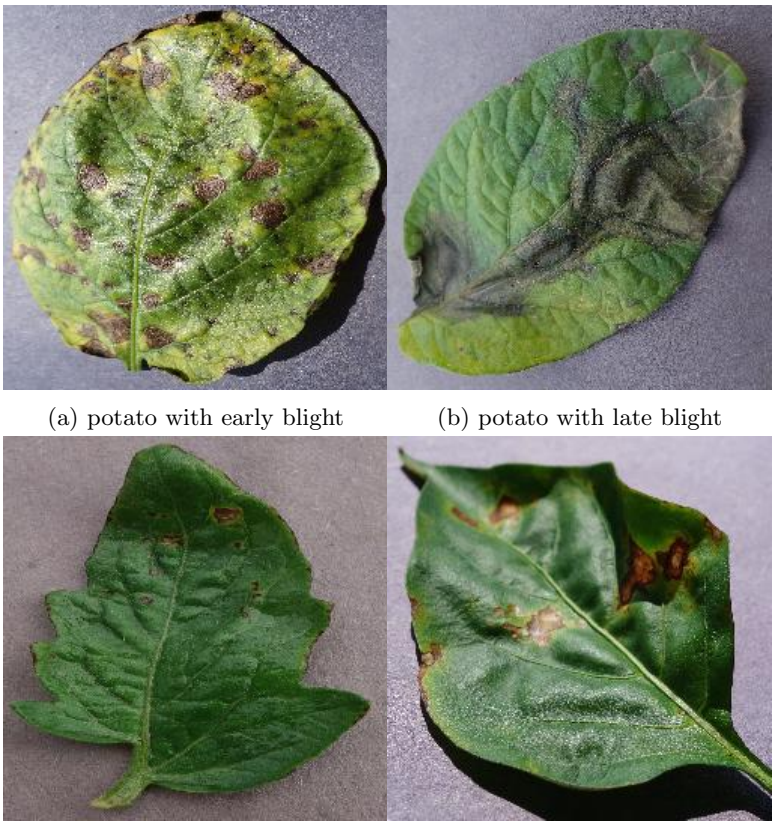
2 Data Exploration

The **PlantVillage** Dataset gives us three categories of crop , **potatoes** , **tomatoes** (mainly) and **peppers**. Our dataset contains pictures of tomatoes infected with **9** diseases/insects , potatoes with **2** , and peppers with only one disease , these information are listed in **Figure 1**. The entire Dataset is about 20000. these images are already cleaned , so we don't have to do clean them ourselves.

	Tomatoes	**Potatoes**	**Peppers**
Diseases	Target spot	Early blight	Bacterial spot
	Mosaic virus	Late blight	
	YellowLeaf curl virus		
	Bacterial spot		
	Early blight		
	Spider mites		
	Late blight		
	Leaf Mold		
	Septoria leaf spot		

Figure 1: Diseases per crops available in the dataset

having given an eye to the dataset , I found the dataset too "perfect" , images were always well delimited , oriented north etc. I was worried that a model trained on this dataset wouldn't be robust enough on real exemple (if I use my model on the real field , thus without clinical adjustments) so I made some data augmentation. Data added were randomly rotated , zoomed , or contrasted. Some images of the initial dataset can be visualize in the **Figure 2** bellow.



(a) potato with early blight (b) potato with late blight

(c) tomato leaf with bacterial spot (d) pepper leaf with bacterial spot

Figure 2: Random images from PlanVillage dataset

3 Model creation

I created three different models , one for each type of crop. at the end , I also tried to create a model that can predict any categories from any crops. All the model and the code are available on **Github** [4]. The conception of the models are similar and not very hard. the main idea was to use convolutional layers followed by Maxpooling Layer and finally connect it to fully connected layers.

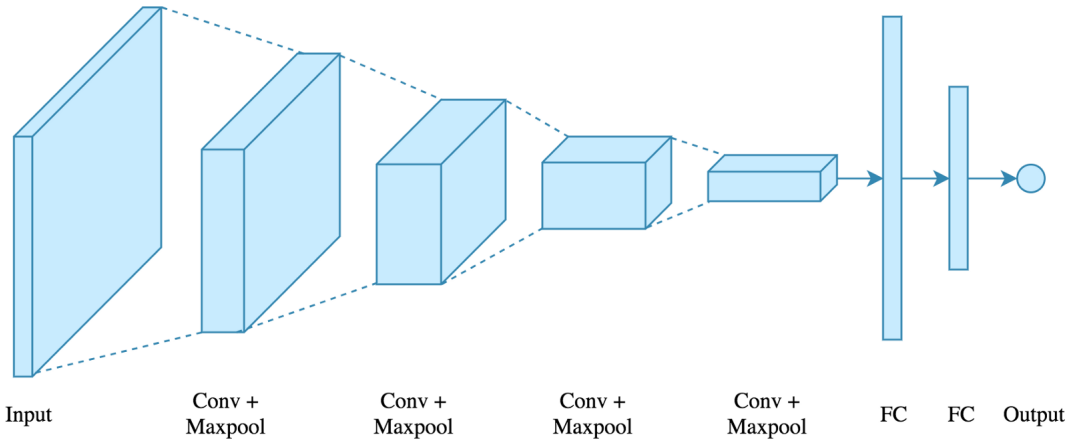


Figure 3: Blue print of the models architecture

4 Results

The results are really promising , each model achieve a great performance on the test set as you can see in the **Figure 4**. there is no real differences between the final accuracy of each model (around 91%).

Models	Accuracy
potatoes model	0.92
peppers model	0.99
tomatoes model	0.92
global model	0.91

Figure 4: Accuracies on the test sets

in y opinion , these results could be far better , especially for the global model. the main issus is that I used a very simple model containing only CNN layers and FC layers. One can add Batch Normalization layers and

adding more layers in general. you can take a look at each model's structure on the report files available on my **Github** [4].

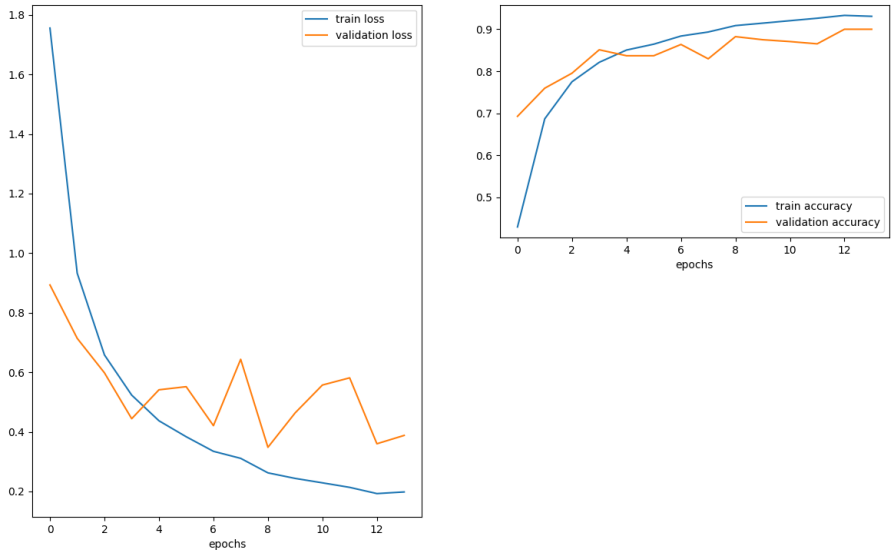


Figure 5: Global model's loss and accuracy curves

5 references

1. Roser, M. Future population growth. In Our World in Data; University of Oxford: Oxford, UK, 2013.
2. https://www.agrivi.com/blog/yield-losses-due-to-pests
3. https://www.kaggle.com/datasets/emmarex/plantdisease
4. Github