Minor AI January 2022

Machine Learning Project

Nassim Bekkali, Rhea Groenenberg, Jeanice Koorndijk

Afbeelding met tekst, boom, plant

Automatisch gegenereerde beschrijving

Contents

[REPORT INTRODUCTION 2](#_Toc92811325)

[Apple Tree Disease Project description 2](#_Toc92811326)

[Data section 2](#_Toc92811327)

Milestone 1

*Deadline: Thursday January 13th at 17:59*

*This is the first model you create for your project. The report introduction should include a description of the problem and data section might include some basic pre-processing. The model itself should be very simple, but show to learning something from the training data, i.e. make a prediction that is (a little) better then randomly selecting an output.*

*Meeting 1: Wednesday January 12th*

*Meeting 2: Friday January 14th*

# 1. REPORT INTRODUCTION

## 1.1 Apple Tree Disease Project description

The overall productivity and quality of apple orchards can be negatively affected by foliar (leaf) diseases. Current disease diagnosis based on human scouting is time-consuming and expensive.[[1]](#footnote-1) Computer-vision based models may be able to increase the efficiency with which diseases are detected.

The difficulty of machine learning algorithms to account for variations in symptoms due to aspects like age of infected tissues, genetic variations, and light conditions within trees can be hurdles to efficient and accurate detection of different diseases.[[2]](#footnote-2) This machine learning project sets out to overcome these hurdles by developping a machine learning-based model to accurately classify a given leaf image from the test dataset to a particular disease category, and to identify an individual disease from multiple disease symptoms on a single leaf image.

## 1.2 Data section

**1.2.1 What the dataset looks like**

Our project utilises the dataset from the Kaggle competition ‘Plant Pathology 2020 Challenge’. Details about the competition and the dataset were published as a peer-reviewed research article:

[Thapa, Ranjita; Zhang, Kai; Snavely, Noah; Belongie, Serge; Khan, Awais. The Plant Pathology Challenge 2020 data set to classify foliar disease of apples. Applications in Plant Sciences, 8 (9), 2020.](https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/aps3.11390)

The provided dataset has X images of appletree leafs, with X possible labels.

The training set metadata consists of:

the image ID and labels. The Labels are the target classes, a space delimited list of all diseases found in the image. An image can have multiple labels, indicating that multiple diseases have been identified in for that image. Unhealthy leaves with too many diseases to classify visually will have the complex class, and may also have a subset of the diseases identified.

Note that, the competition had a hidden test set: only three images were provided as testing data as samples while the remaining 5,000 images would be made available to the notebooks of contenders once it they had submitted their work.

The dataset is not balanced. (wat betekent dit ookalweer?)

**1.2.2 How the data has been processed**

The data has been subjected to some basic pre-processing.

….. (fill in tmrw.)

1. Thapa, R., K. Zhang, N. Snavely, S. Belongie, and A. Khan. 2020. The Plant Pathology Challenge 2020 data set to classify foliar disease of apples. Applications in Plant Sciences 8(9): e11390. [↑](#footnote-ref-1)
2. Thapa, R., K. Zhang, N. Snavely, S. Belongie, and A. Khan. 2020. The Plant Pathology Challenge 2020 data set to classify foliar disease of apples. Applications in Plant Sciences 8(9): e11390. [↑](#footnote-ref-2)