

Machine Learning Engineer Nanodegree

Capstone Proposal

Rairo Mukamuri June 5th, 2020

Proposal

Plant Disease Detection using Pytorch and Fastai

Domain Background

As the world grows in population every year the demand for food is also rising, however a growing threat in the form of plant diseases threatens to worsen the already strained food supply.

Pest and disease induced crop losses have devastated farms across the world but thanks to advances in machine learning farmers can detect diseases early through the use of computer vision in the browser or on a mobile device. Convolutional Neural Networks which a class of deep neural networks for analysis visual imagery have been widely used in image classification problems and their applications have been promising.

Problem Statement

The goal of this project is to fit a Convolutional Neural Network on the data using the Fastai library and Pytorch. I am going to use transfer learning with ResNet, a deep residual learning framework developed at Microsoft in order to achieve accurate classification of the plant images. Transfer learning focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. Resnets have been trained on the imagenet dataset which contains millions of images so their knowledge of those features will be very help in this case.

Datasets and Inputs

The PlantVillage dataset consists of 54303 healthy and unhealthy leaf images divided into 38 categories by species and disease.

The images span 14 crop species: Apple, Blueberry, Cherry, Grape, Orange, Peach, Bell Pepper, Potato, Raspberry, Soybean, Squash, Strawberry, and Tomato. It contains images of 17 fungal diseases, 4 bacterial diseases, 2 mold (oomycete) diseases, 2 viral diseases, and 1 disease caused by a mite. 12 crop species also have images of healthy leaves that are not visibly affected by a disease.

Solution Statement

The solution will be a fastai classification model capable of correctly identifying whether an image contains a specific plant disease or not with a high accuracy. The model will also be deployed in a web application for farmers to use.

Benchmark Model

For this problem, the benchmark model will be a resnet-152 pytorch classifier developed by Viridiana Romero Martinez as referenced in her medium article <https://medium.com/datadriveninvestor/creating-an-ai-app-that-detect-diseases-in-plants-using-facebooks-deep-learning-platform-pytorch-15faeb6bec3>. The model produced a high accuracy score, 96.1% and the goal is to see if fastai and another resnet can improve on that.

Evaluation Metrics

The model will evaluated on accuracy, precision and recall in classifying the images correctly. A confusion matrix will be tabulated to show more details.

Project Design

The project is expected to follow the steps below:

- Data exploration and preprocessing
- Model selection
- Training and validation
- Summarisation of the results

References

- Original crowdai dataset <https://github.com/SpikerJG/PlantDiseaseData.git>
- Resnet, <https://neurohive.io/en/popular-networks/resnet/>
- David Hughes paper, <https://arxiv.org/ftp/arxiv/papers/1604/1604.03169.pdf>
- Viridiana Romero Martinez's medium article <https://medium.com/datadriveninvestor/creating-an-ai-app-that-detect-diseases-in-plants-using-facebooks-deep-learning-platform-pytorch-15faeb6bec3>