New Plant Disease Detection

CSML1020 Course Project

Jerry Khidaroo  
 CSML1020 – Group 3  
York University  
Toronto, ON Canada  
 email@email.com

Paul Doucet  
 CSML1020 – Group 3  
York University  
Toronto, ON Canada  
 [email@email.com](mailto:email@email.com)

ABSTRACT

This paper will explore the classification of plant images to identify new plant diseases using a machine learning model.

The dataset was obtained from Kaggle and consists of over 87,000 rgb images of healthy and diseased crop leaves labeled by plant and disease type in 38 different classes.

CCS CONCEPTS

• Artificial Intelligence • Machine Learning   • Image Classification

1 Introduction

The problem we will examine is a supervised multi-class image classification problem. The goal is to investigate which supervised machine learning models will give the best results in classifying the images from our dataset in the predefined categories.

2 Existing Work

|  |  |
| --- | --- |
| Plant Desease Classifictaion-VGG16  https://www.kaggle.com/wiwidsetiawan/plant-desease-classifictaion-vgg16 | Example of classification using the VGG16 pre trained model |
| Fork of Plant Diseases Classification Using incep3  https://www.kaggle.com/vimaladit/fork-of-plant-diseases-classification-using-incep3 | Example of classification using the Inception Version 3 pre trained model |
|  |  |
|  |  |

3 Methodology

3.1 Data Preparation

The dataset was downloaded from the Kaggle website: <https://www.kaggle.com/vipoooool/new-plant-diseases-dataset>

The dataset did not need any manipulation as it was previously divided into a useable directory structure for training and validation as well as several test images.

Table 1: Dataset Parsed from Category Folder Names



3.2 Data Exploration

The dataset consists of images of plant leaves in various conditions. The following graphs, show the distribution of this data.

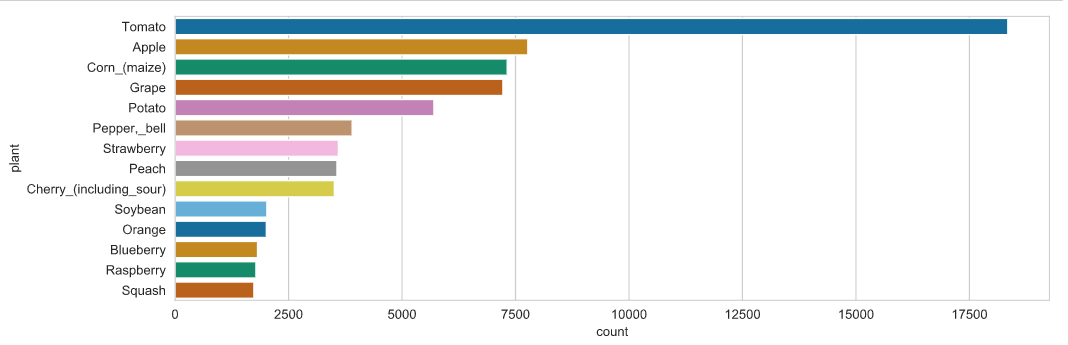


Figure 1: Number of images by plant

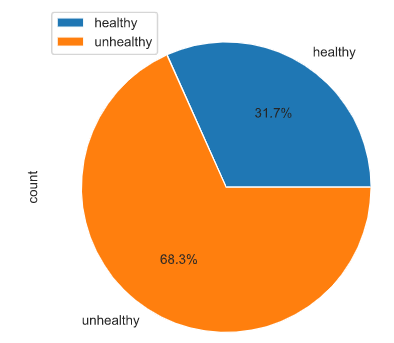


Figure 2: Relative image percentages by health status

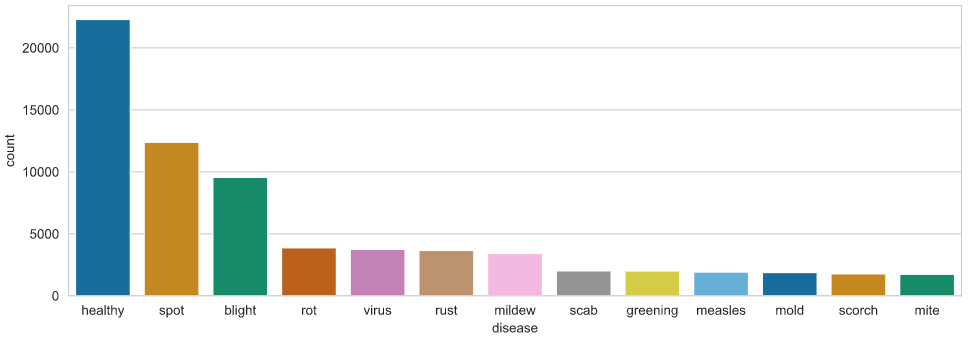


Figure 3: Number of images by disease

3.3 Data Preprocessing

The following data preprocessing methods will be evaluated to determine the best data augmentation for our input layer.

Data Augmentation

* Random Horizontal Shift
* Random Vertical Shift
* Random Horizontal Flip
* Random Vertical Flip
* Random Rotation
* Random Brightness
* Random Zoom

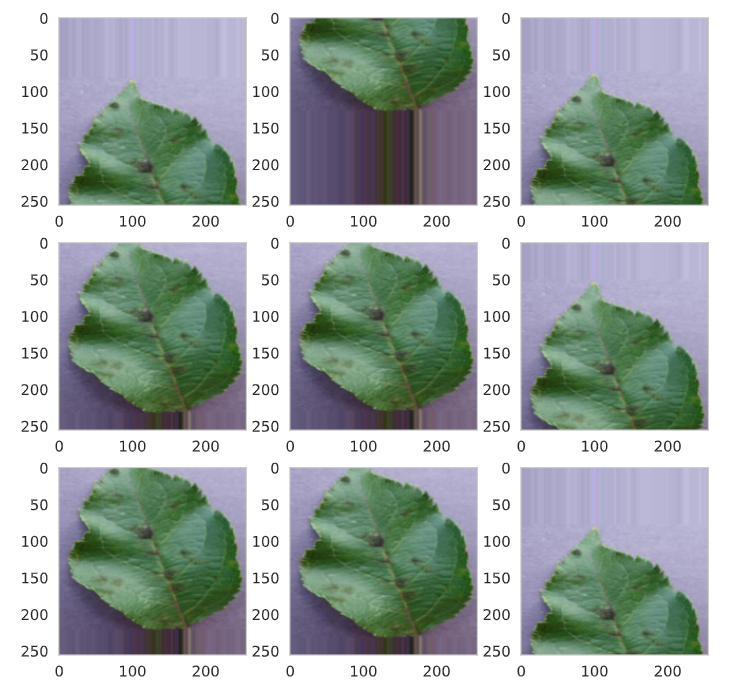
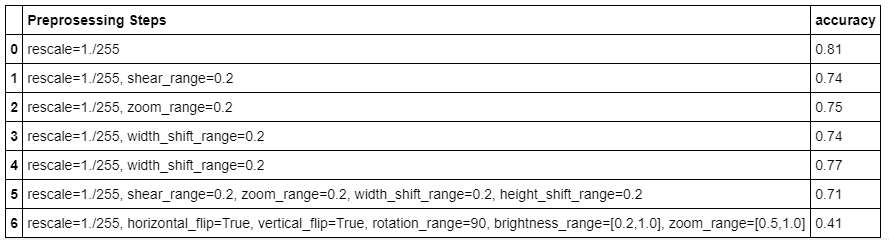


Figure 4: Data Augmentation Visualization for Random Vertical Shift



Figure 5: Code Snippet For Data Preprocessing Steps

Table 2: Preprocessing Results Based on Baseline VGG16 Model with Transfer Learning



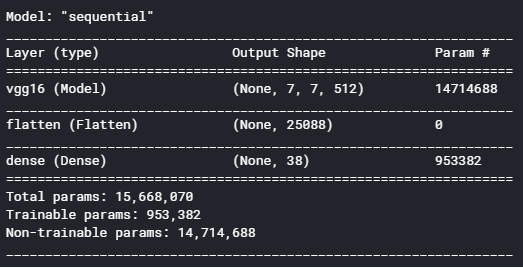
3.5 Model Evaluation & Selection

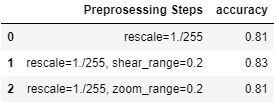
For the model evaluations, we will be running baselines (From scratch, with transfer learning, …), Improving

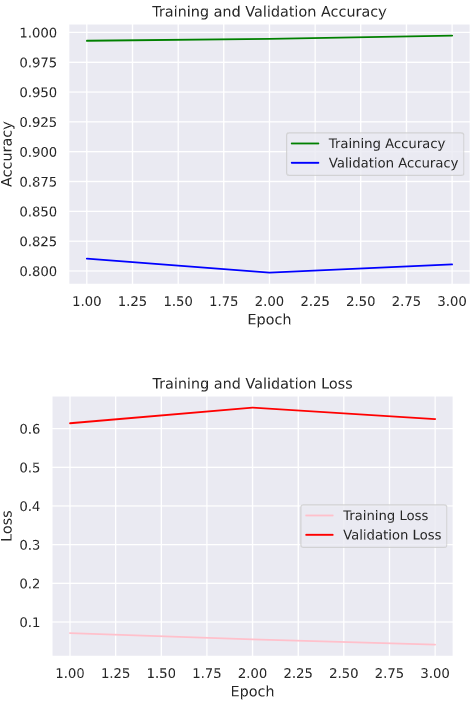
3.5.1 Baseline Models

Several base models with transfer learning were run, using a sample of the complete dataset. The accuracy and loss were evaluated for each of the three base models: VGG16; ResNet50 and InceptionV3.

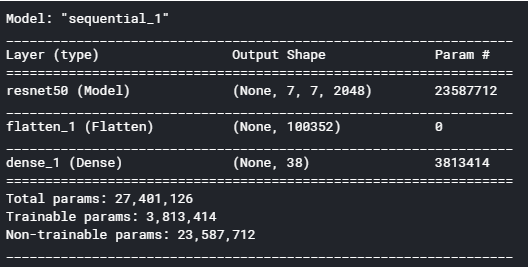
3.5.1.1 VGG16 Base Model with Transfer Learning

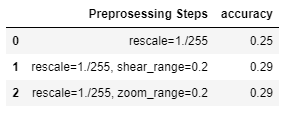






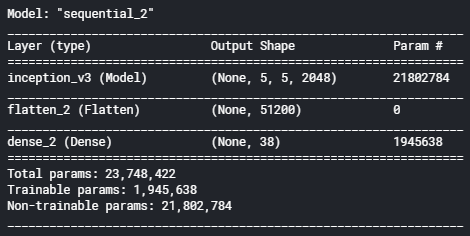
3.5.1.2 ResNet50 Base Model with Transfer Learning

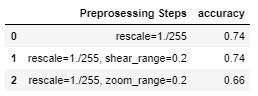


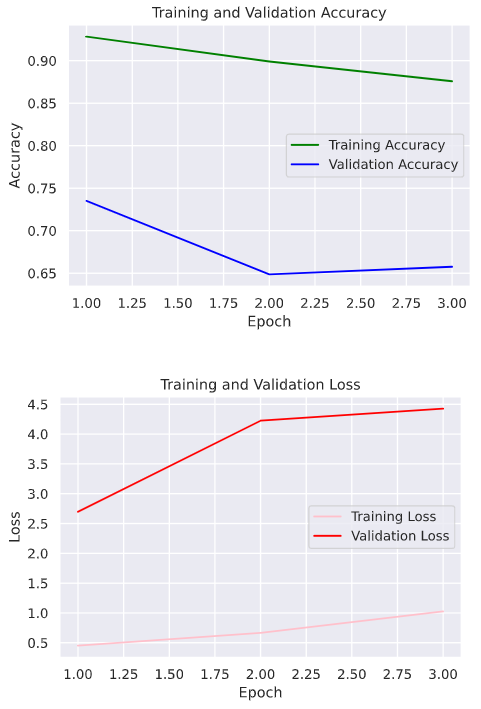




3.5.1.3 InceptionV3 Base Model with Transfer Learning







3.5.2 Improved Models

Hyperparameter Tuning



Best Paramaters: {'learning\_rate': 0.0001, 'epochs': 3, 'batch\_size': 32, 'activation': 'softmax'}

3.5.3 Final Model and Predictions

To Complete

4 Results

To Complete

5 Discussion

To Complete

6 Conclusion

To Complete

ACM Reference format:

FirstName Surname, FirstName Surname and FirstName Surname. 2018. Insert Your Title Here: Insert Subtitle Here. In *Proceedings of ACM Woodstock conference (WOODSTOCK’18). ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/1234567890

ACKNOWLEDGMENTS

Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

REFERENCES

[1] Wiwid Setiawan, 2020. T Plant Desease Classifictaion-VGG16 <https://www.kaggle.com/wiwidsetiawan/plant-desease-classifictaion-vgg16>

[2] Vimal Adit. 2019. Fork of Plant Diseases Classification Using incep3 <https://www.kaggle.com/vimaladit/fork-of-plant-diseases-classification-using-incep3>

[3]

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00