Voltron_2021_29

Detecting grape vine diseases with A.I

Geoffroy Huck
Ryan Headley
Simon Cormier
4 July 2021

Abstract

In this paper, we develop an algorithm that can determine the health status of a grape vine from a picture of its leaves. A dataset of ____ images was provided to not only classify a healthy leaf, but also identify four different types of diseases that commonly affect this plant : black rot, leaf light, black measles and **mildew.** The classifier chosen was the convolutional neural network (CNN), which has grown to achieve the highest results in the computer vision world. After building and training, the CNN model was able to achieve a 94.4% accuracy.

Keywords: Vineyard, plant disease, vineyard, machine learning, convolutional neural network

Introduction

The alcohol industry is one of the largest in the world, with wine being at the forefront of the several options available. Creating high quality wine is created through a process of preparation, chemistry and presentation. The first step and one of the most important, is the agricultural aspect of the preparation : without grapes, you cannot make wine. All wine producers must not only grow, but also maintain the size and quality of their crops, which directly dictate their potential to profit or be swept into bankruptcy.

While there are many factors that negatively affect the vineyards, plant diseases can rapidly destroy entire crops without notification; often these diseases are allowed to grow to a size that significantly reduces the output of the crop, without being noticed. For this reason, a solution must be provided to notify companies at the beginning of the onset of a plant disease, so that it can be eradicated while it is small and easier to handle. This is exactly what is proposed in this paper.

Material

Dataset

The dataset used consist of a total of ___ colored leaf images (jpeg) of thousands of different grape vines. These images were divided by their classes and later into training and validation data for the training of the model. Afterwards, a section of the data was used as testing data to validate the efficacy of the model.

Fig 1: Grape leaf with black rot

Fig 1: Healthy grape leaf

Tools

Since this project is founded in artificial intelligence with the goal of being completed and shared as quickly as possible, a jupyter notebook was created. This option not only facilitates collaborative work, but it takes advantage of the most commonly used language in data science: python.

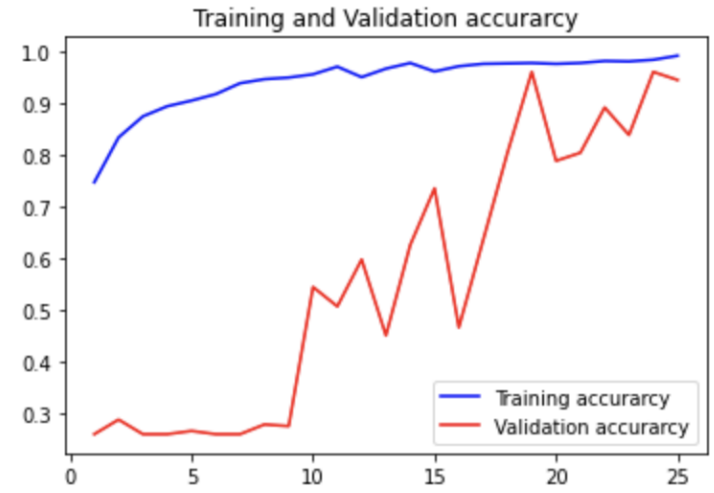
Alongside python, the numpy, panda, sklearn and tensorflow libraries were incorporated to build on top of pre-built machine learning algorithms.

Results

The performance of the model was determined to greatly increase with more epochs, to the point where running the algorithm on less than twenty epochs results in unsatisfactory data.

The two plots below denote the accuracy and loss for the training and validation data.

Plot 1: Training and validation loss



Plot 2: Training and validation accuracy

After 25 epochs, the training accuracy arrived at 99.1% and the validation accuracy at 94.4%. Analyzing the other loss metrics, it was clear that the results were well balanced between the classes and the different types of precision. These results are shown in the classification report below.

	precision	recall	f1-score	support
0	0.93	0.99	0.96	76
1	1.00	0.94	0.97	69
2	0.88	0.99	0.93	92
3	1.00	0.86	0.92	83
accuracy			0.94	320
macro avg	0.95	0.94	0.95	320
weighted avg	0.95	0.94	0.94	320

Fig 3: Classification report

For further analysis of the breakdown of the classes, a confusion matrix was plotted to determine whether there were large errors associated with a specific class. This is shown in the figure below.

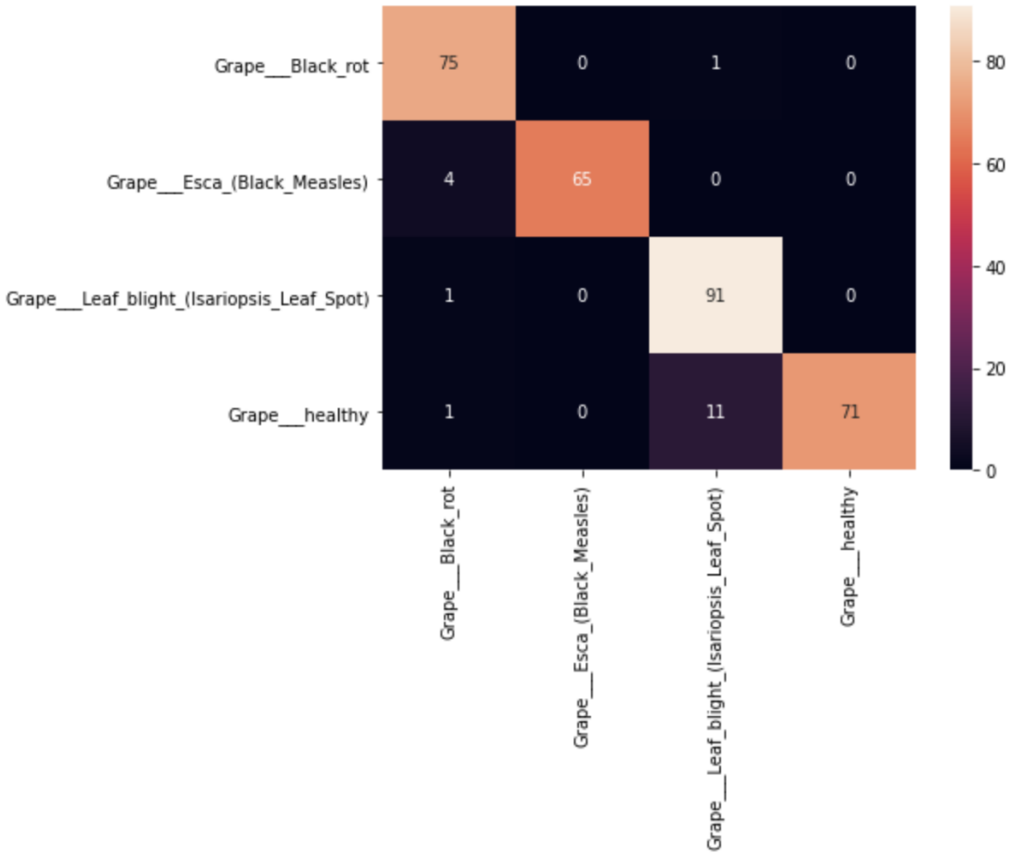


Fig 4: Confusion matrix

As we can see, most of the classes were very accurately predicted with only an outlying error of a healthy leaf being classified as a leaf bright disease.

In the end, the model was evaluated on the test data and provided an accuracy of 94.4%, which matches our validation accuracy at the end of training.

Discussion

The goal of this paper was to create a model capable of accurately classifying the health status of a vineyard by analyzing photos of the leaves. While this model has only been trained on

Conclusion

Method