Plant Pathology

Disease Identification of Apple Leaves

Alex & Brian

Topics We'll Cover

Agenda

Project Intro

The Dataset

What is Object Recognition?

Classification Model

Localization Model

Conclusions

Project Intro

What is our project about?

- **Kaggle Competition (Plant Pathology 2020 - FGVC7)**
- **Image Classification task**
- **Assessing the health of apple leaves**

Health Categories

Healthy

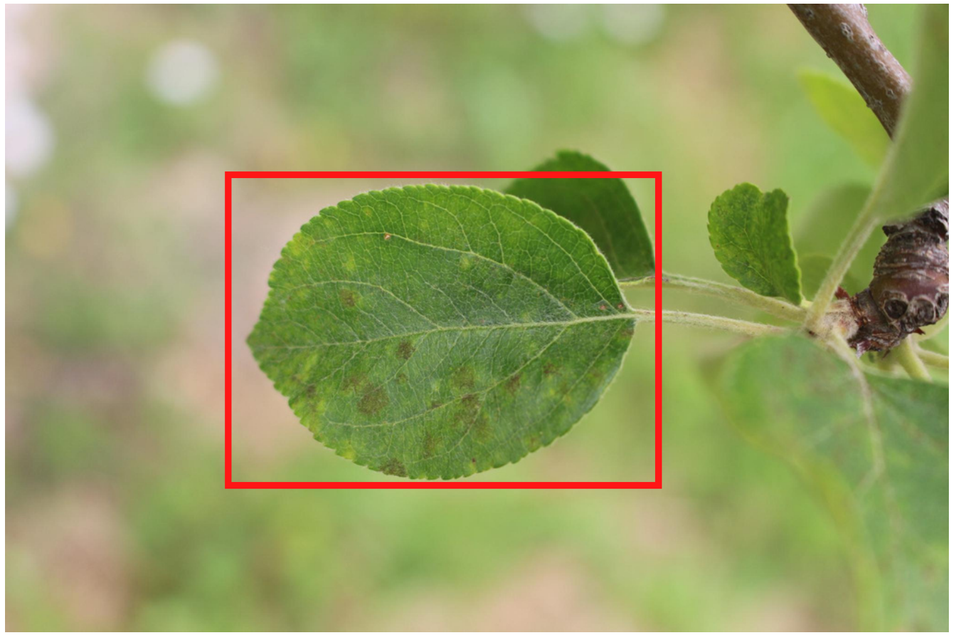
Scab disease

Rust disease

Multiple diseases

Extended Project Task

- Extended task: Object Localization
- Draw a box around the location of the leaf in the photo



The Dataset

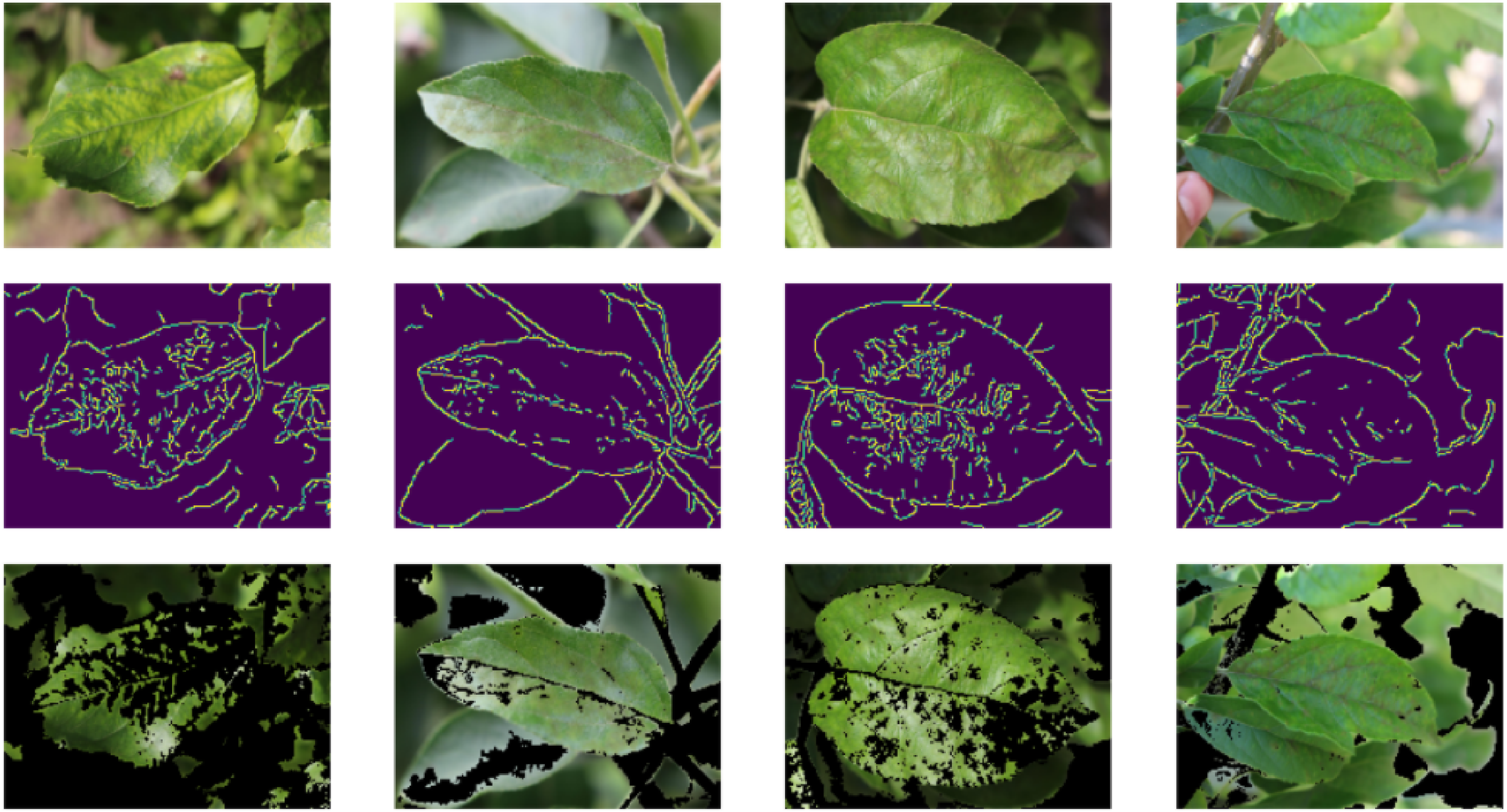
Analysis of the leaf images

- 1821 images with labels
- All images are 2048 x 1365px
- We split into 80% training and 20% validation

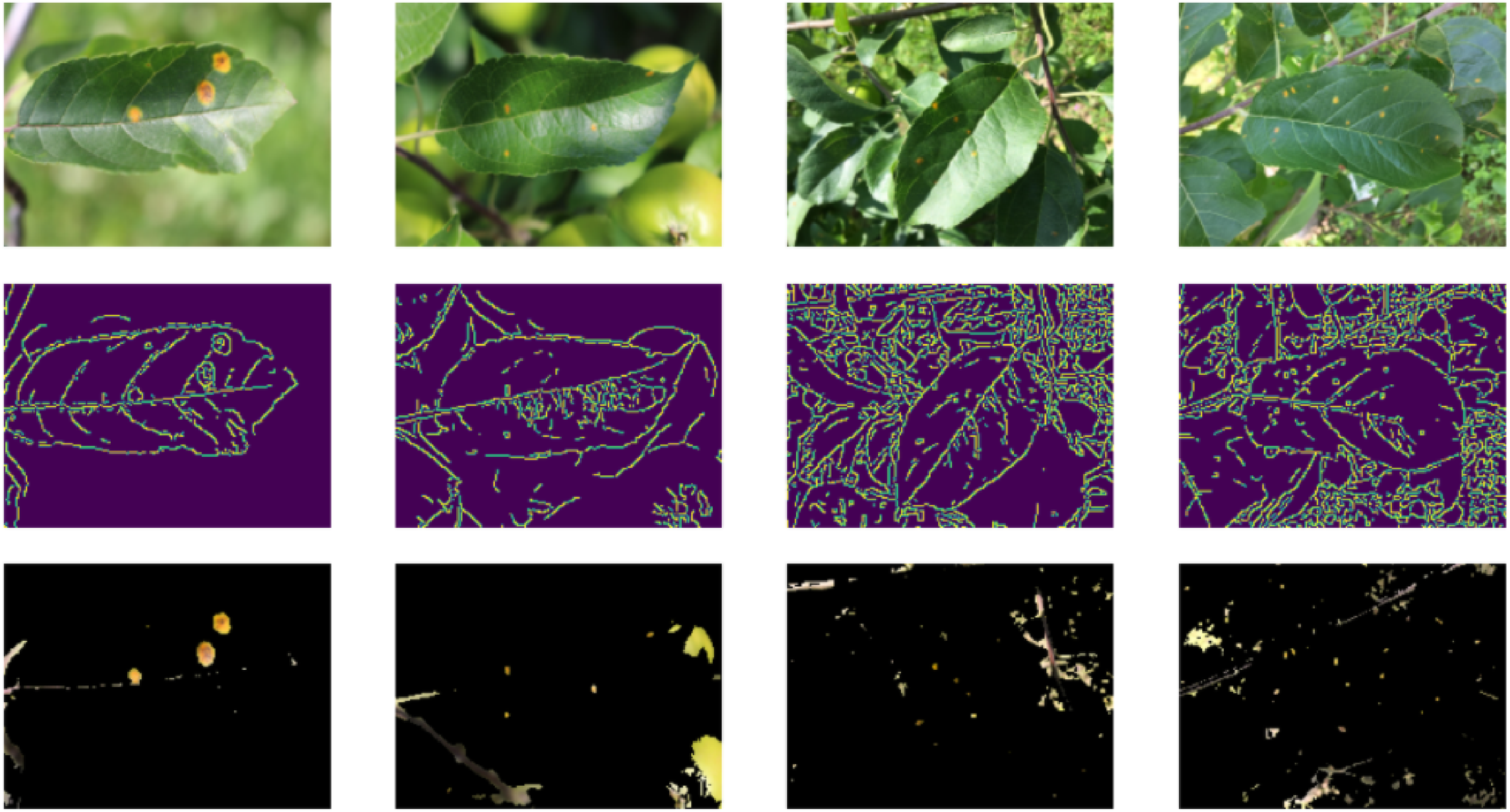
- 1821 test images with no labels (for competition submission)

Healthy leaves

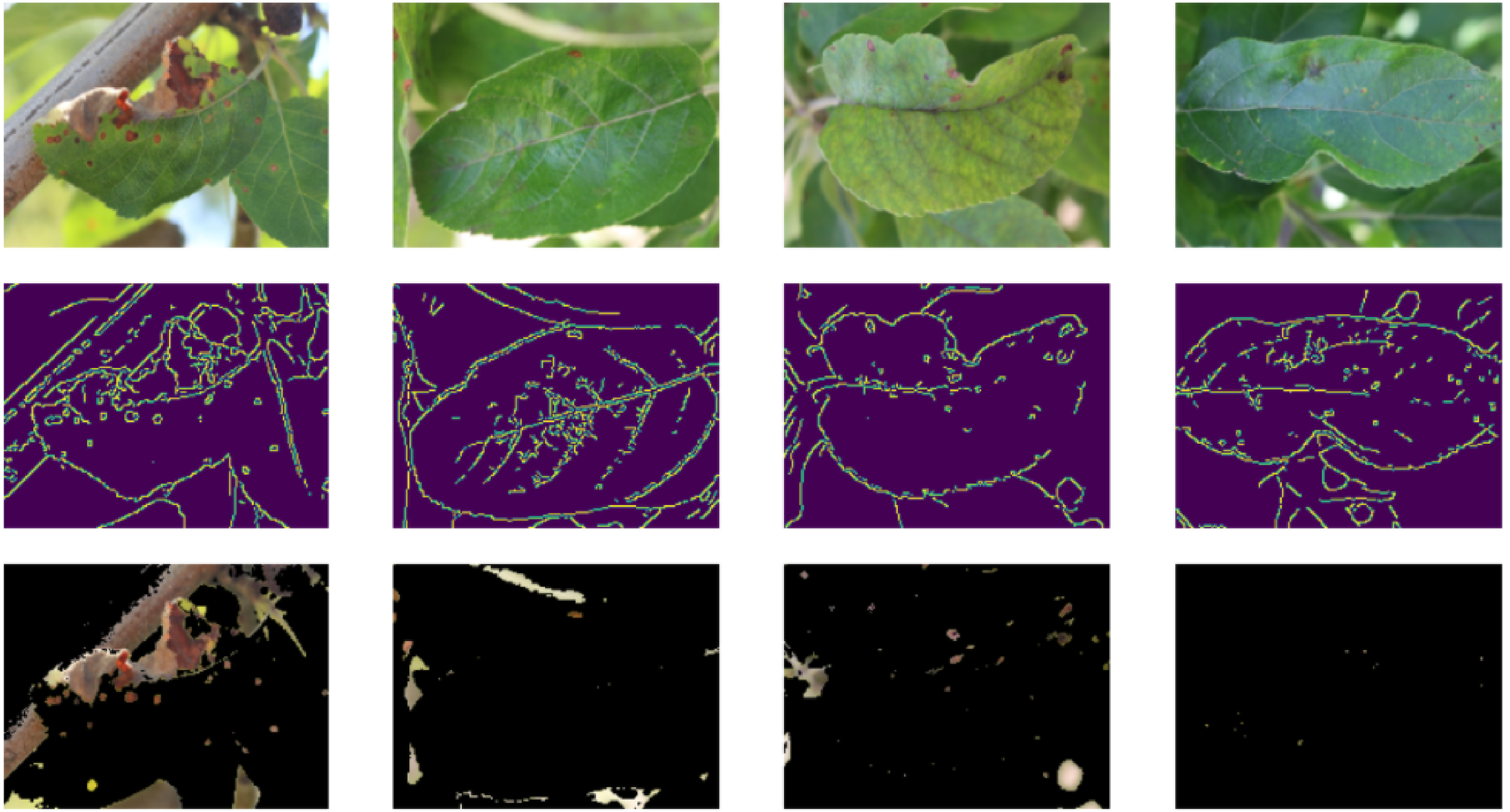
Leaves with scab disease



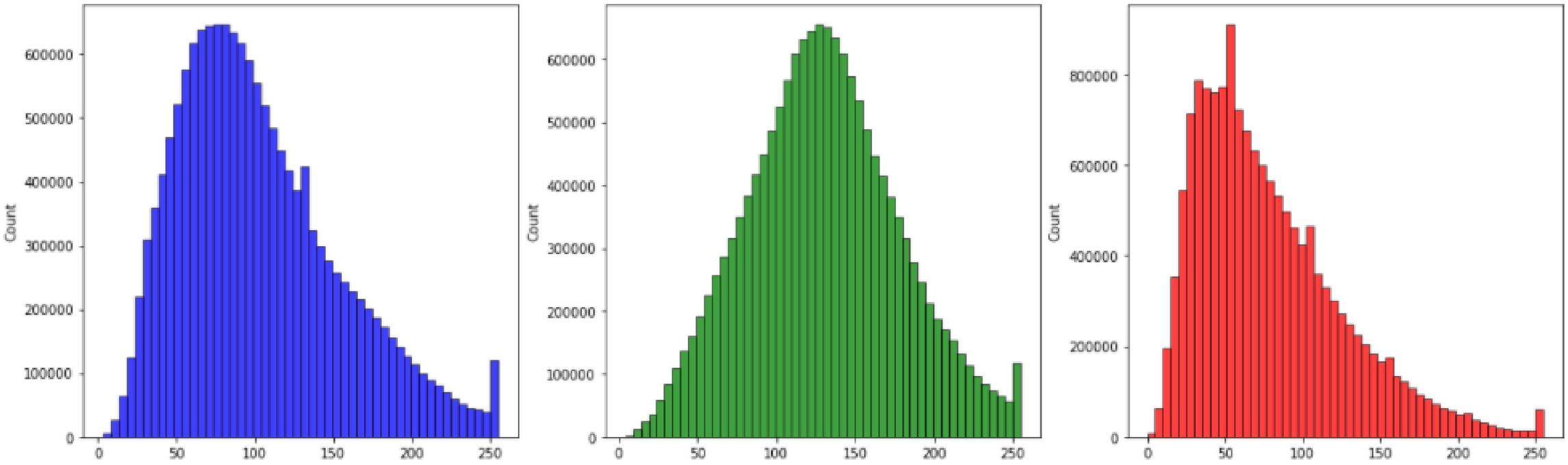
Leaves with rust disease



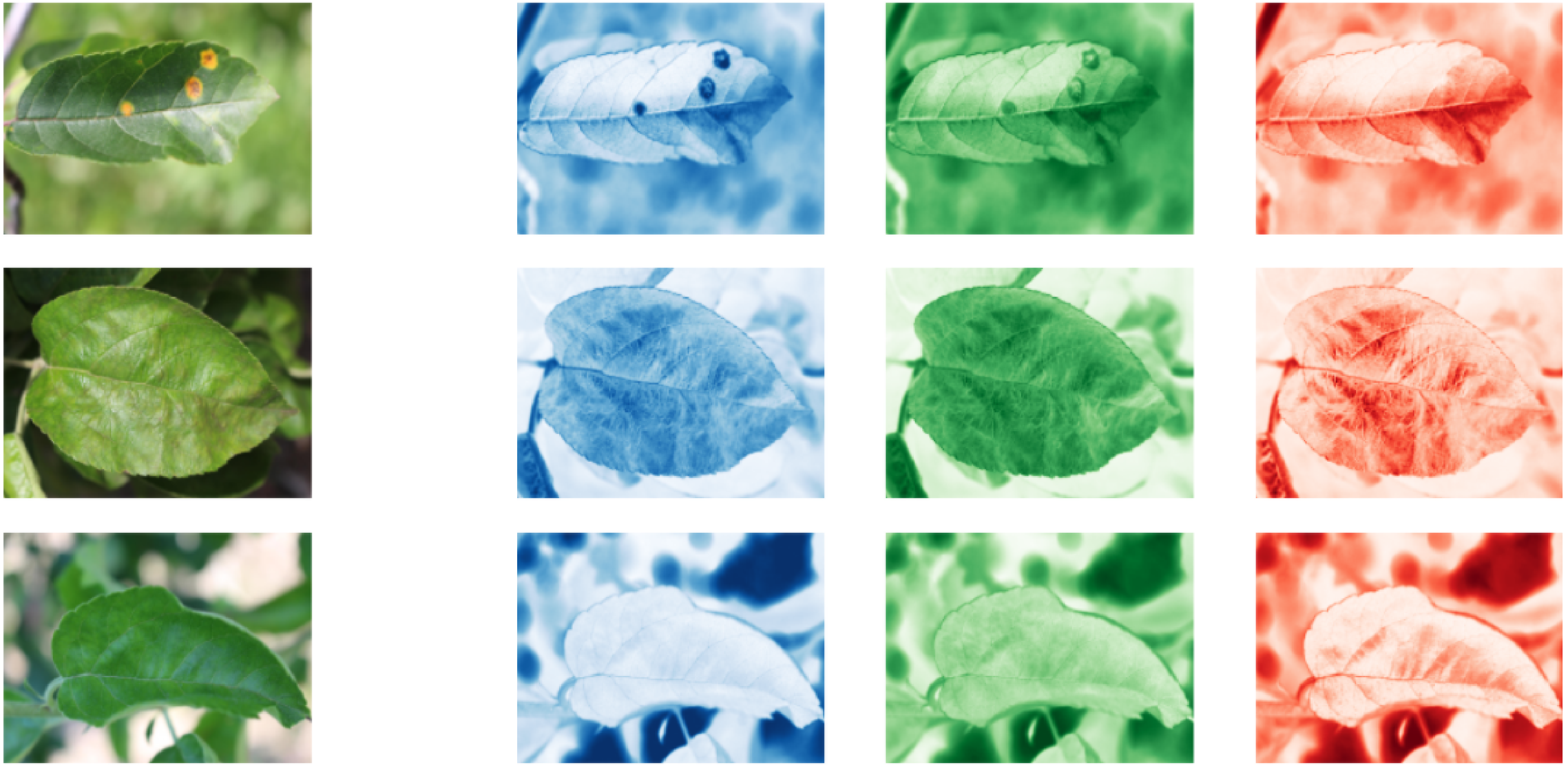
Leaves with multiple diseases



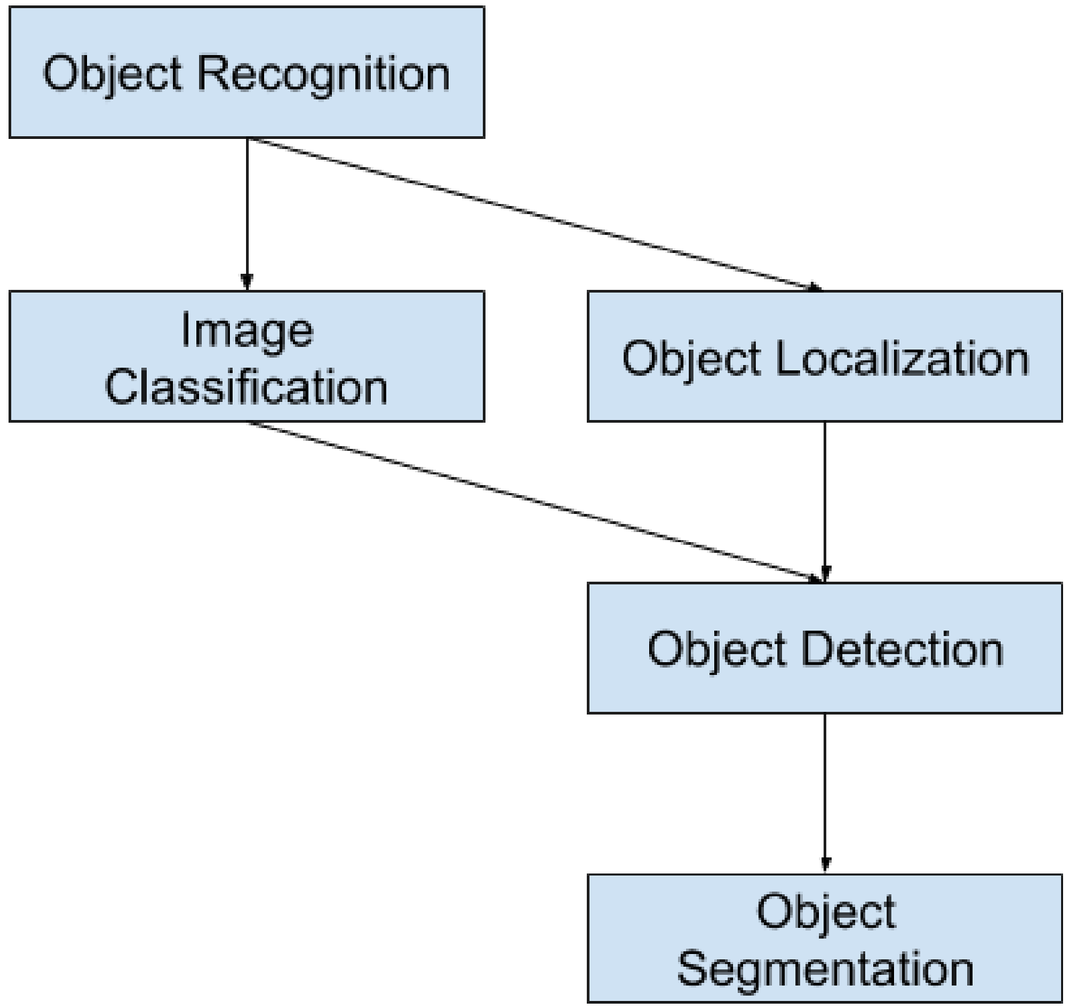
RGB distribution across a sample of 500 images



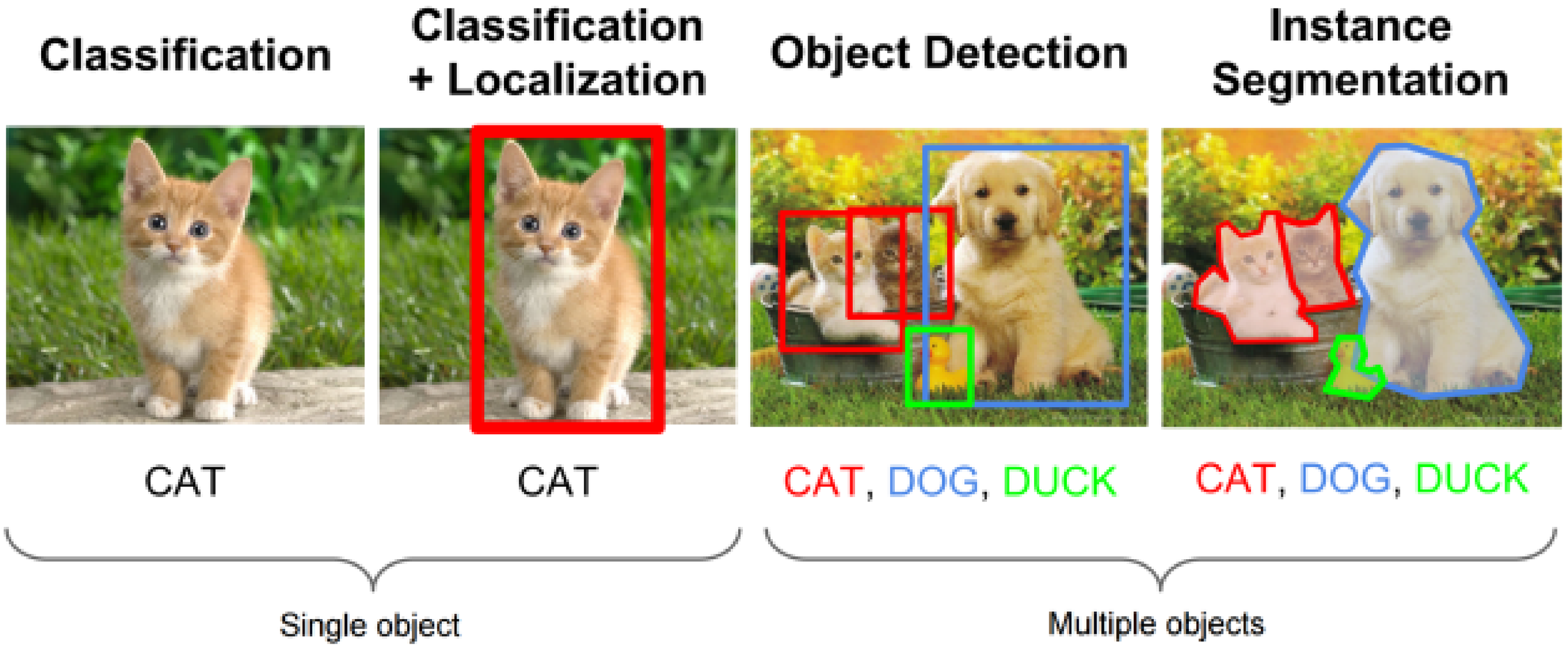
RGB channels of three sample images



What is Object Recognition?

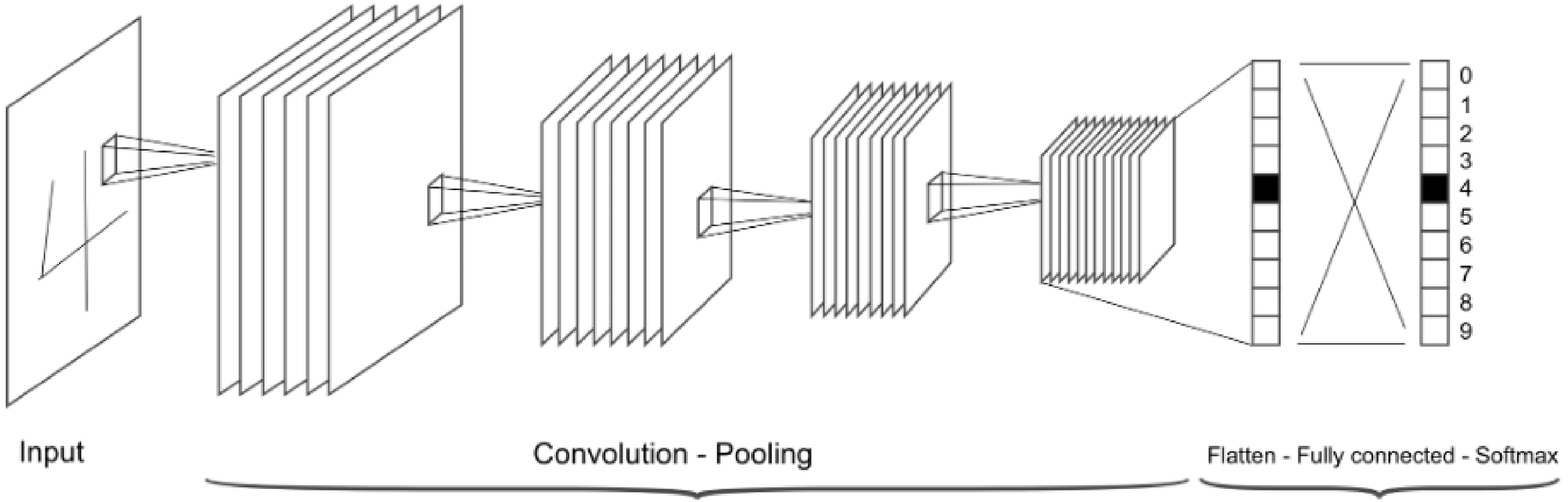


Jason Brownlee, machinelearningmastery.com

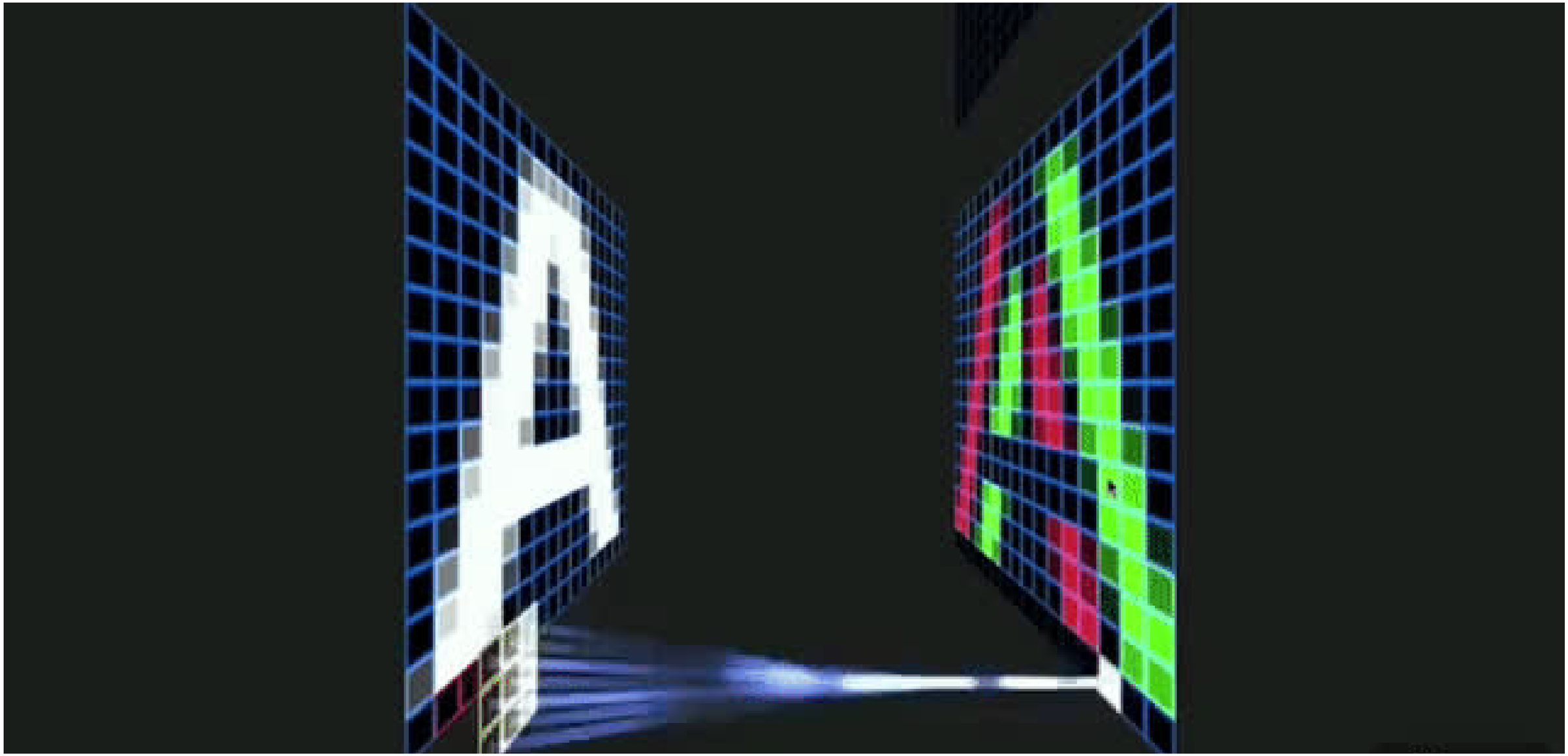


Michal Maj, kdnuggets.com

Convolutional Neural Networks



Diego Junco, mobiquity.com

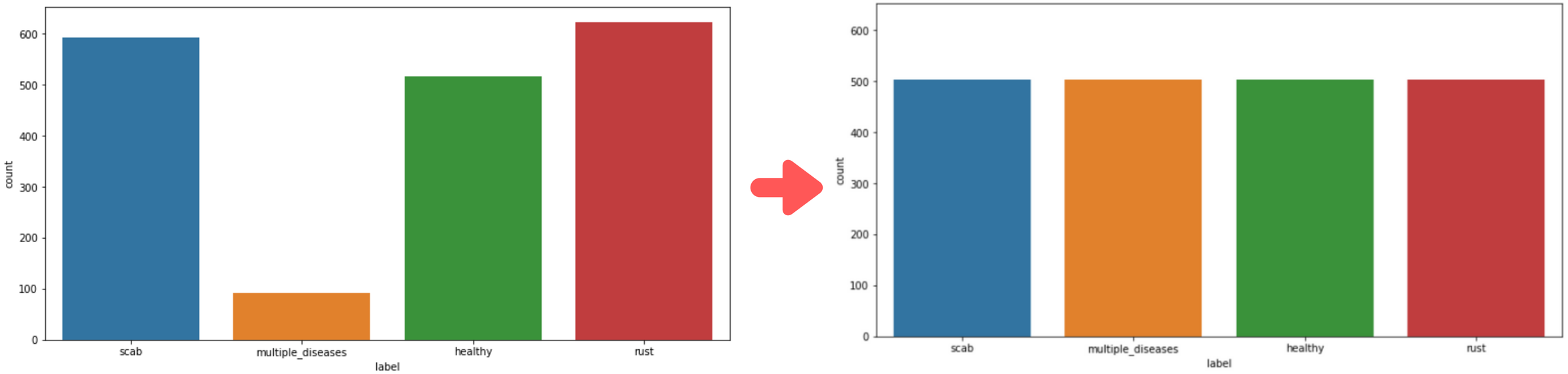


Octavio Good, https://www.youtube.com/watch?v=f0t-OCG79-U

Our Classification Model

Which disease does this leaf have?

Class Imbalances in the Training Dataset



Oversampling using SMOTE (Synthetic Minority Oversampling Technique)

Image Augmentation (Random Transformations)

Scaling the Images (dividing pixel values by 255)

109	46	209	211	76
234	67	34	45	23
234	56	102	109	105
80	123	68	46	45
91	127	238	254	34

0.43	0.18	0.82	0.83	0.30
0.92	0.26	0.13	0.18	0.09
0.92	0.22	0.40	0.43	0.41
0.31	0.48	0.27	0.18	0.18
0.36	0.50	0.93	0.99	0.13

Flowchart for our Classification Model

Convolutional layer
Max Pooling layer
Fully-connected layer
Flatten layer

INPUT

200 x 150px

32 128

100 x 75px

32 128

50 x 38px

32 128

25 x 19px

32 128

13 x 9px

16640

512

256

128

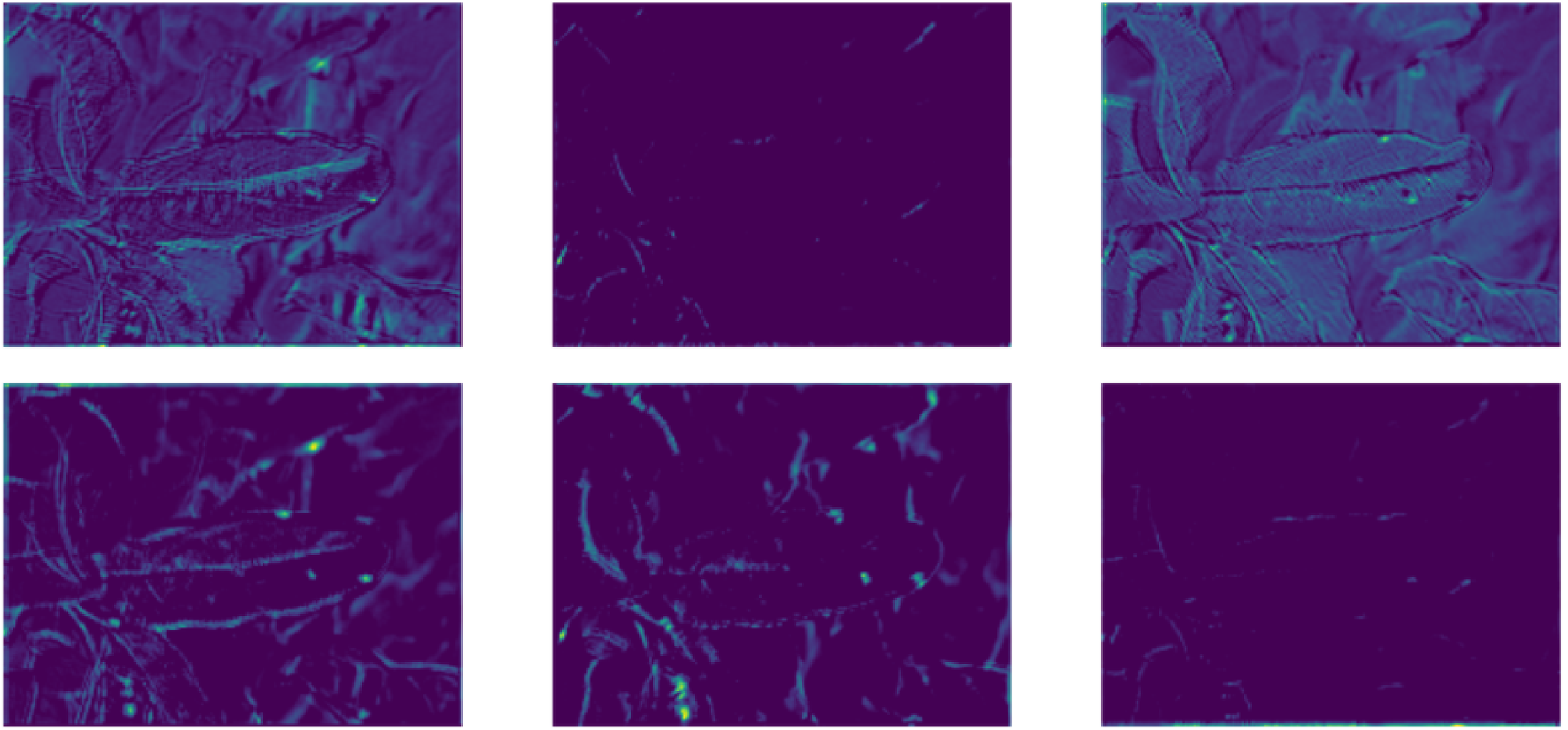
OUTPUT (Softmax)

- healthy? **1%**
- multiple? **4%**
- rust? **1%**
- scab? **94%**

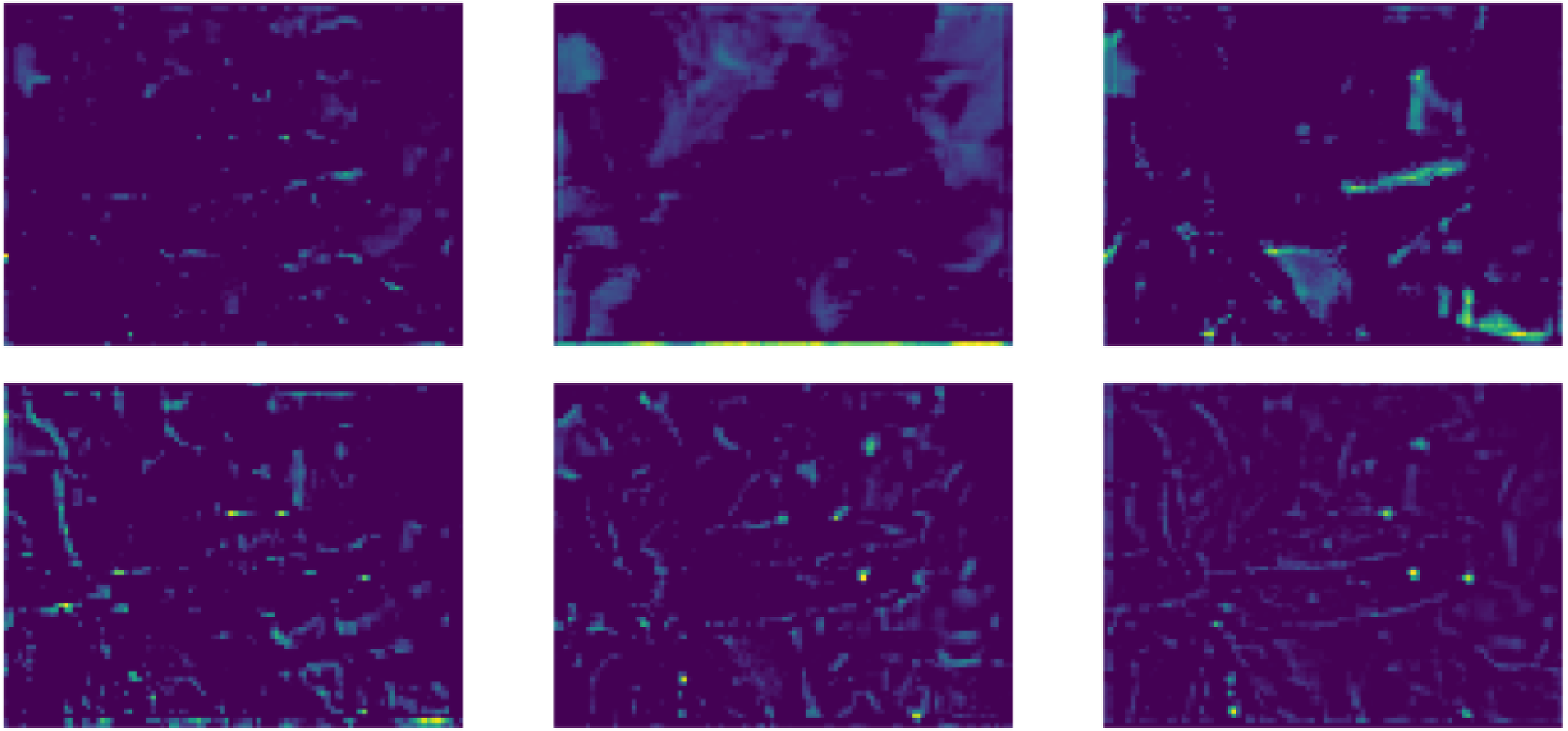
Prediction: scab

Example Input Image

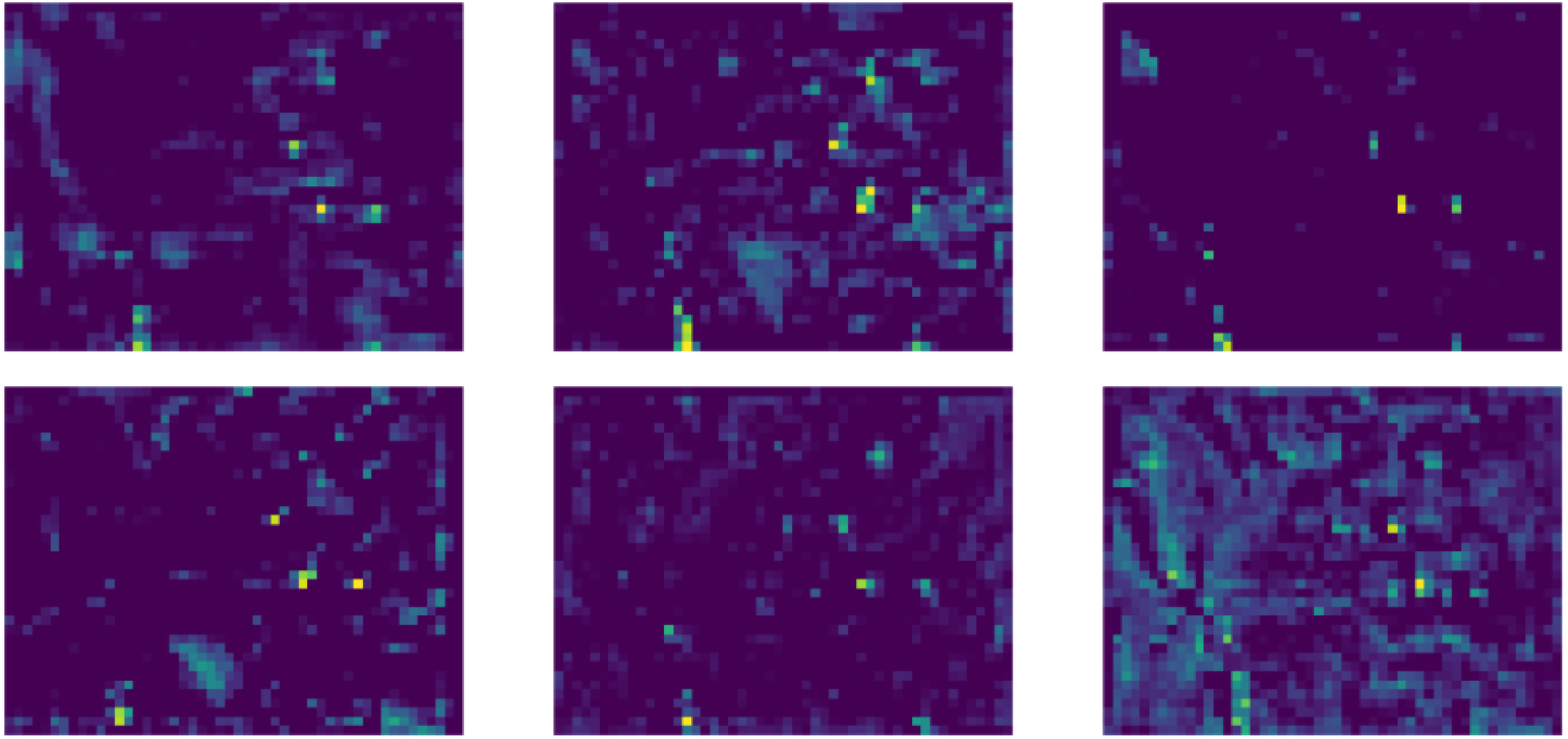
1st Convolutional Layer



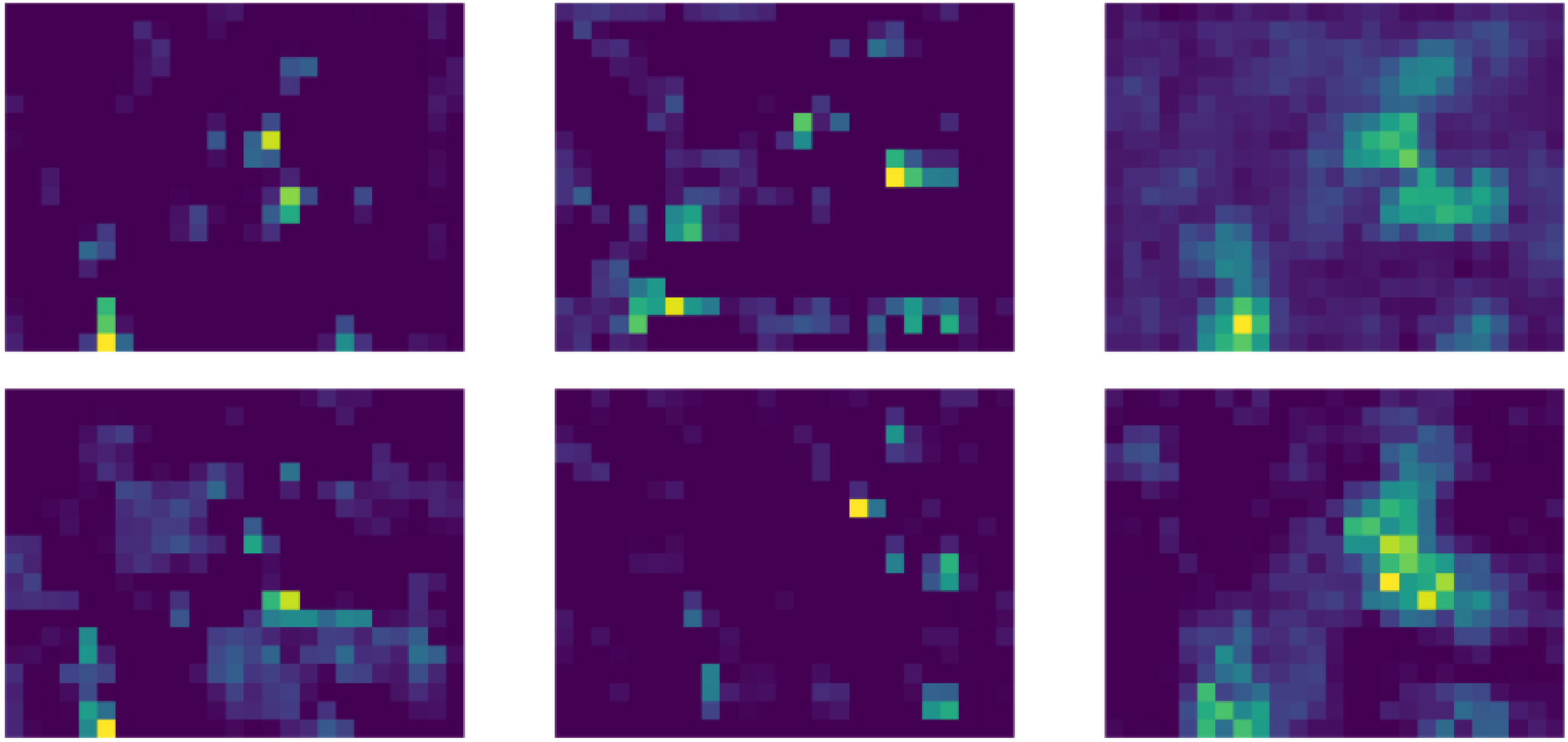
3rd Convolutional Layer



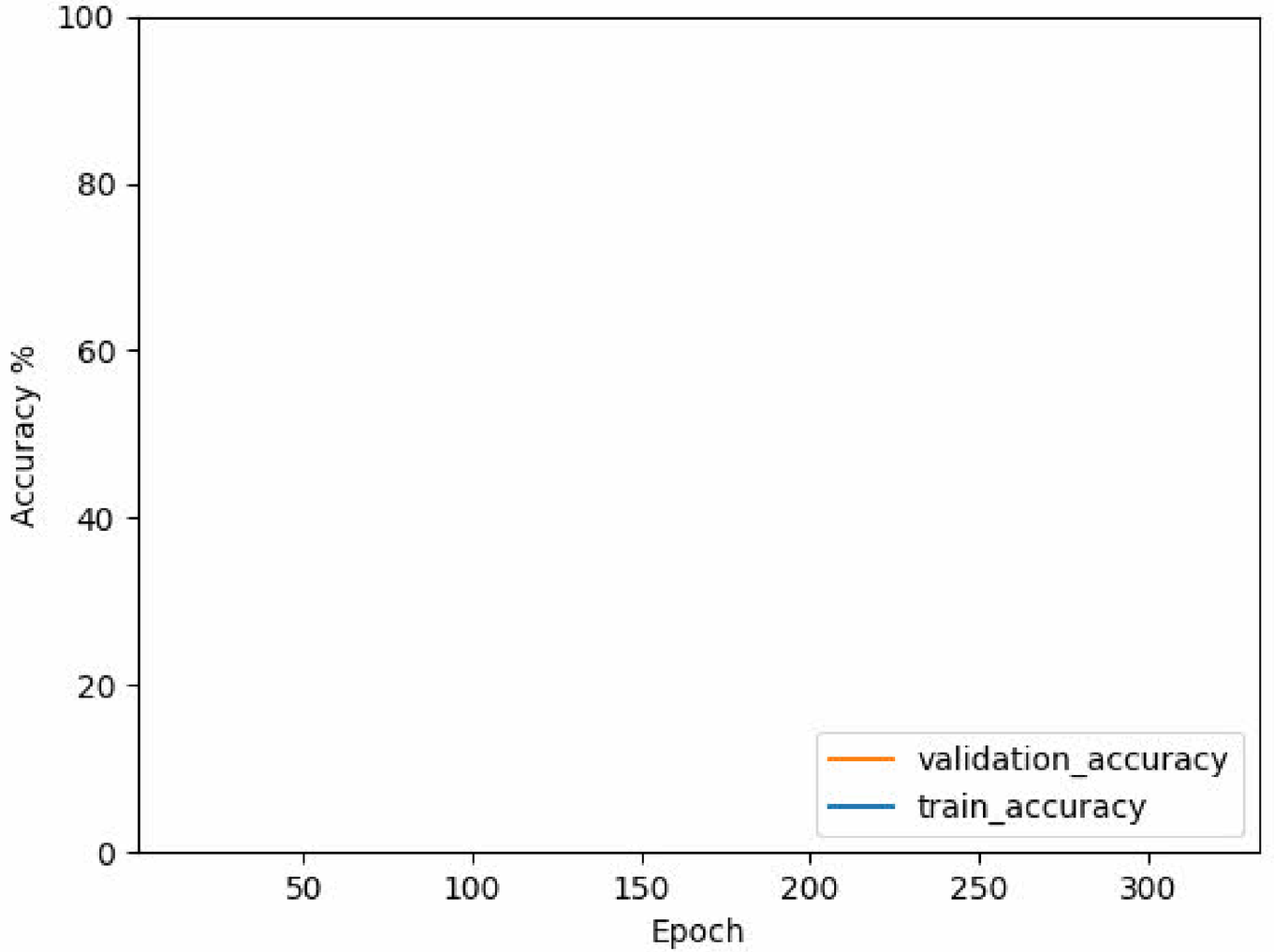
5th Convolutional Layer



7th Convolutional Layer



Classification Model Training

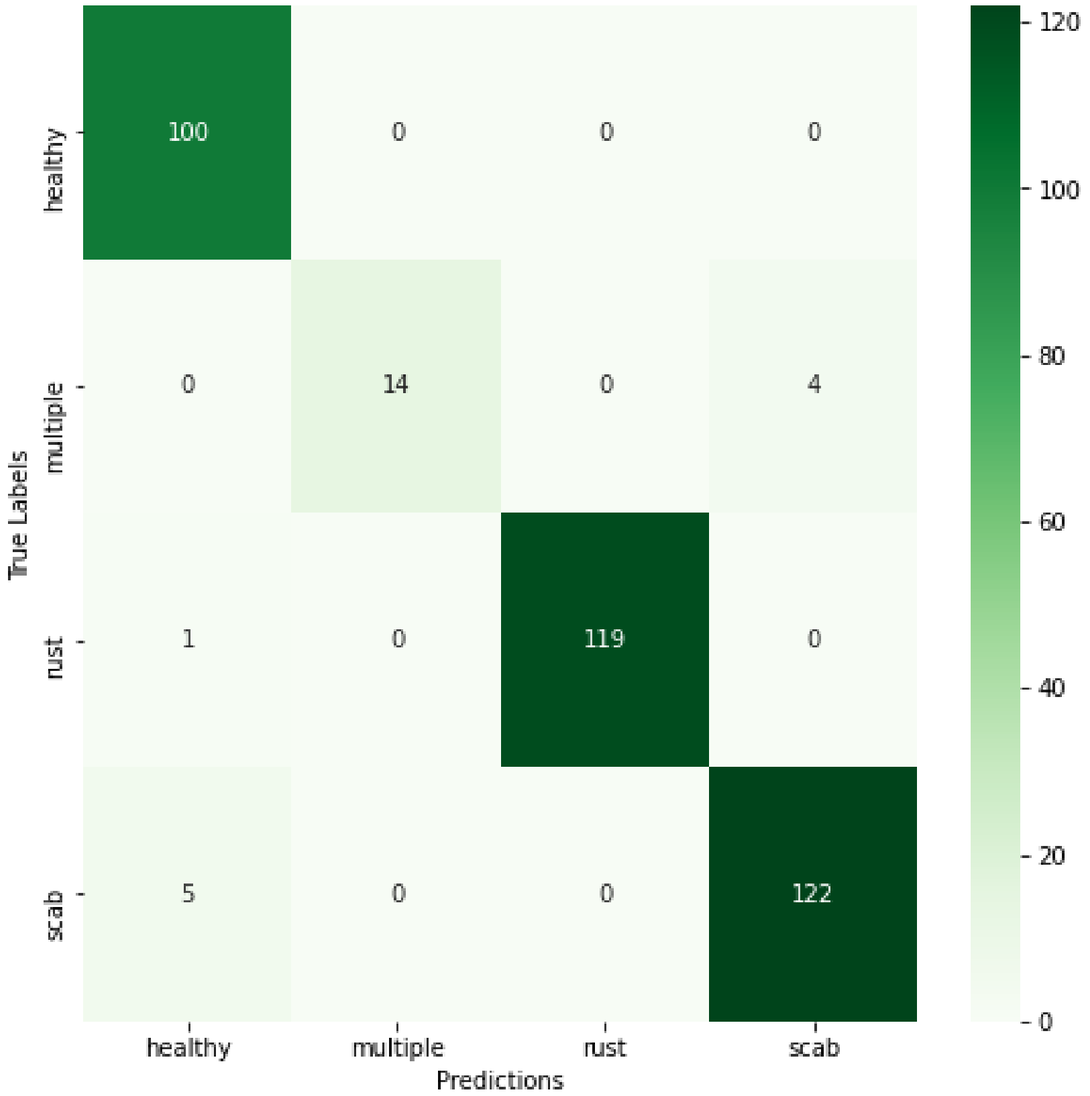


Training accuracy: **99.6%**

Validation accuracy: **97.3%**

Epochs run: **332**

Confusion Matrix for Validation Set Predictions



(365 images in validation set)

Healthy accuracy: **100%**

Multiple accuracy: **77.8%**

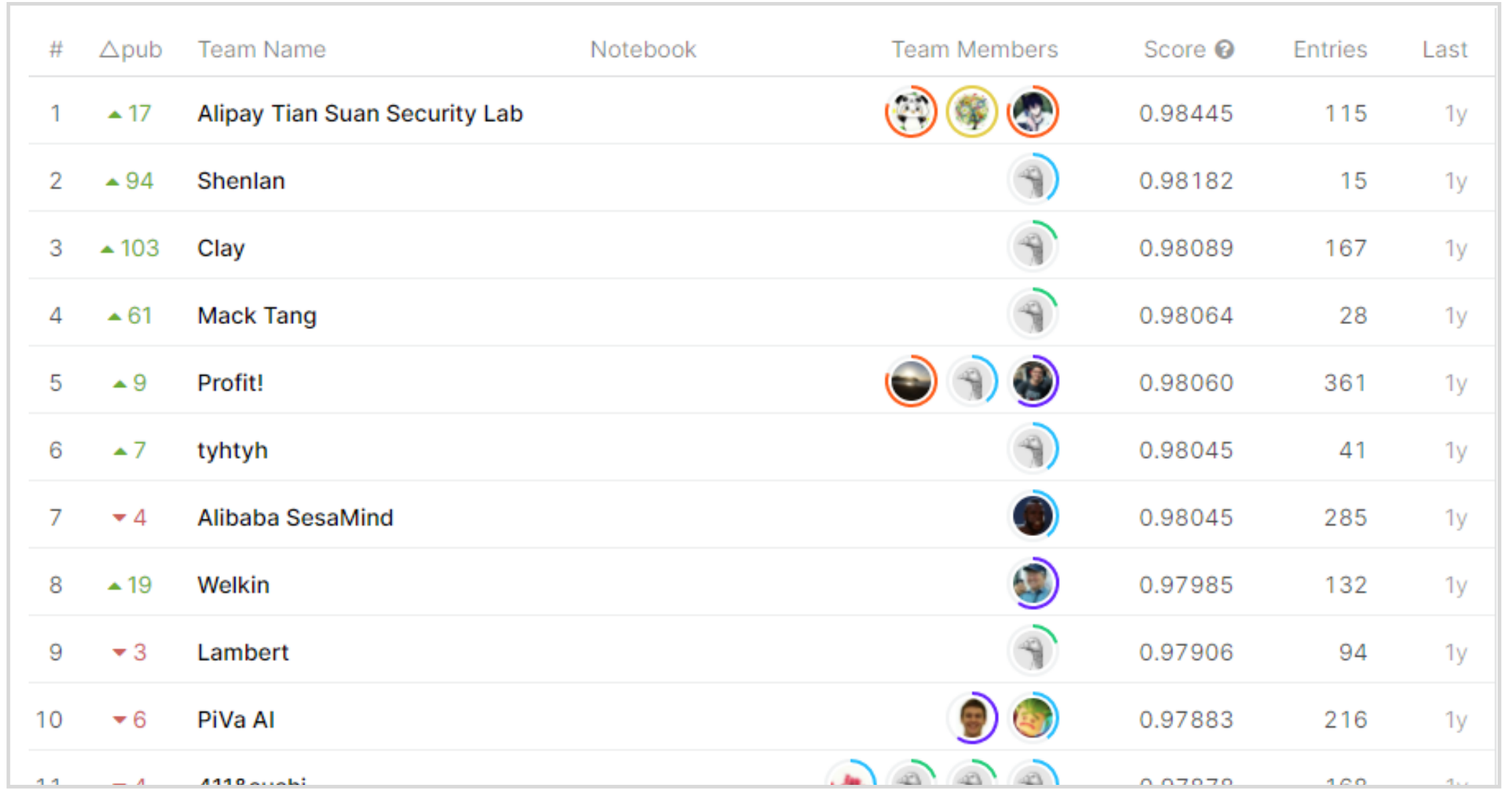
Rust accuracy: **99.2%**

Scab accuracy: **96.1%**

Our Competition Submission

kaggle

- 1317 Teams
- Top 75% of teams had an accuracy score of > 0.9 (90% accuracy)



#	△pub	Team Name	Notebook	Team Members	Score	Entries	Last
1	▲ 17	Alipay Tian Suan Security Lab			0.98445	115	1y
2	▲ 94	Shenlan			0.98182	15	1y
3	▲ 103	Clay			0.98089	167	1y
4	▲ 61	Mack Tang			0.98064	28	1y
5	▲ 9	Profit!			0.98060	361	1y
6	▲ 7	tyhtyh			0.98045	41	1y
7	▼ 4	Alibaba SesaMind			0.98045	285	1y
8	▲ 19	Welkin			0.97985	132	1y
9	▼ 3	Lambert			0.97906	94	1y
10	▼ 6	PiVa AI			0.97883	216	1y

Our submission score:

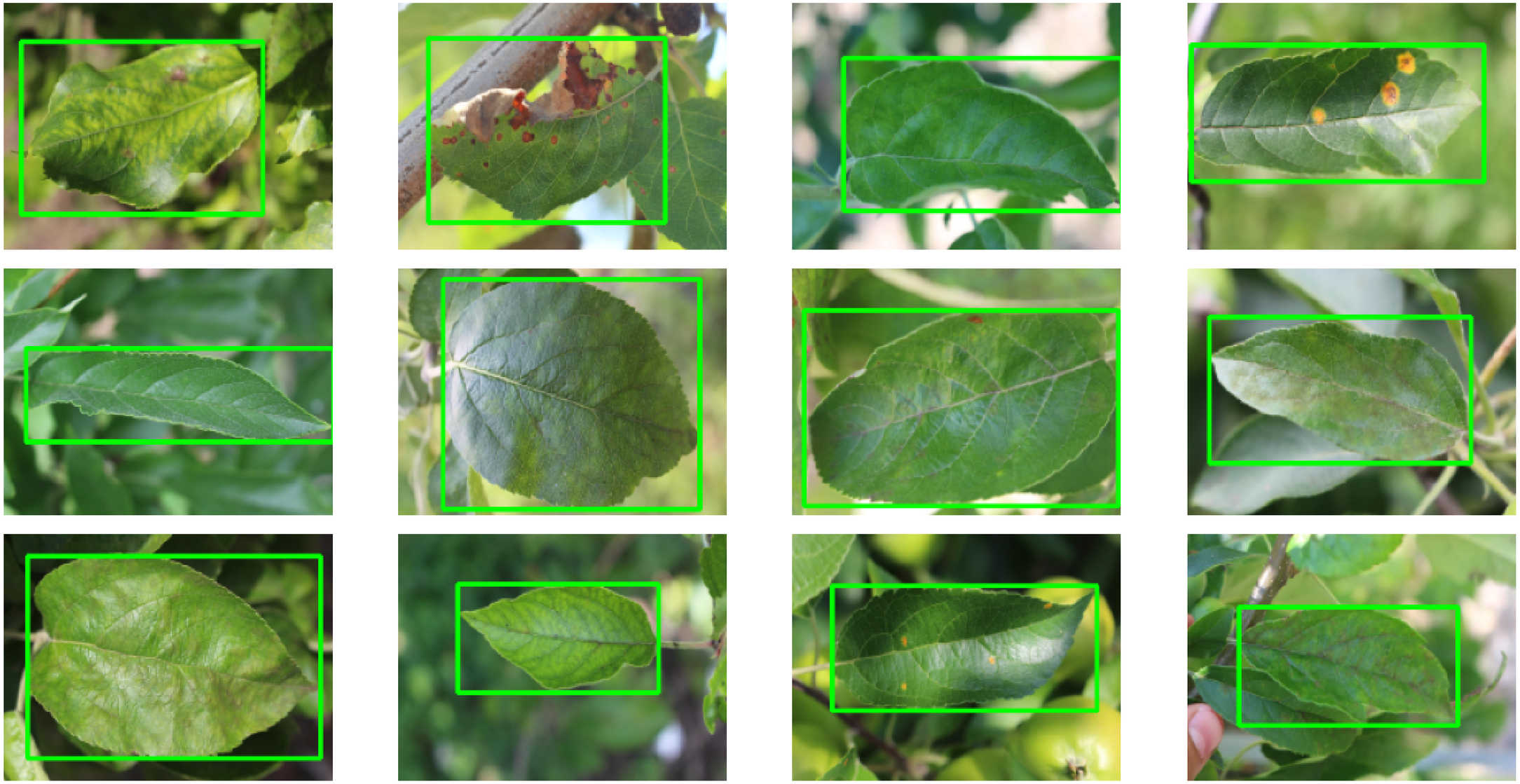
0.95576 (95.6%)

which would have put us at 540th place, in the Top 40% of teams.

Our Localization Model

Where is the leaf in this image?

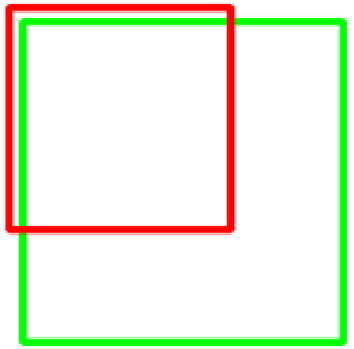
Labelling our own Bounding Boxes for 500 Images



Evaluation Metric for Bounding Boxes

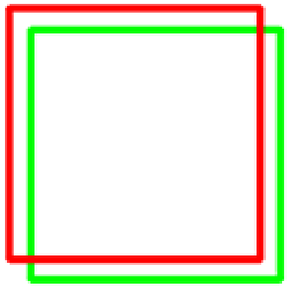
$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

IoU: 0.4034



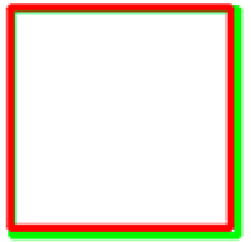
Poor

IoU: 0.7330



Good

IoU: 0.9264



Excellent

Adrian Rosebrock, pyimagesearch.com

Flowchart for our Localization Model

Convolutional layer
Max Pooling layer
Fully-connected layer
Flatten layer

INPUT

200 x 150px

32 128

100 x 75px

32 128

50 x 38px

32 128

25 x 19px

32 128

13 x 9px

16640

512

256

128

OUTPUT

x0

y0

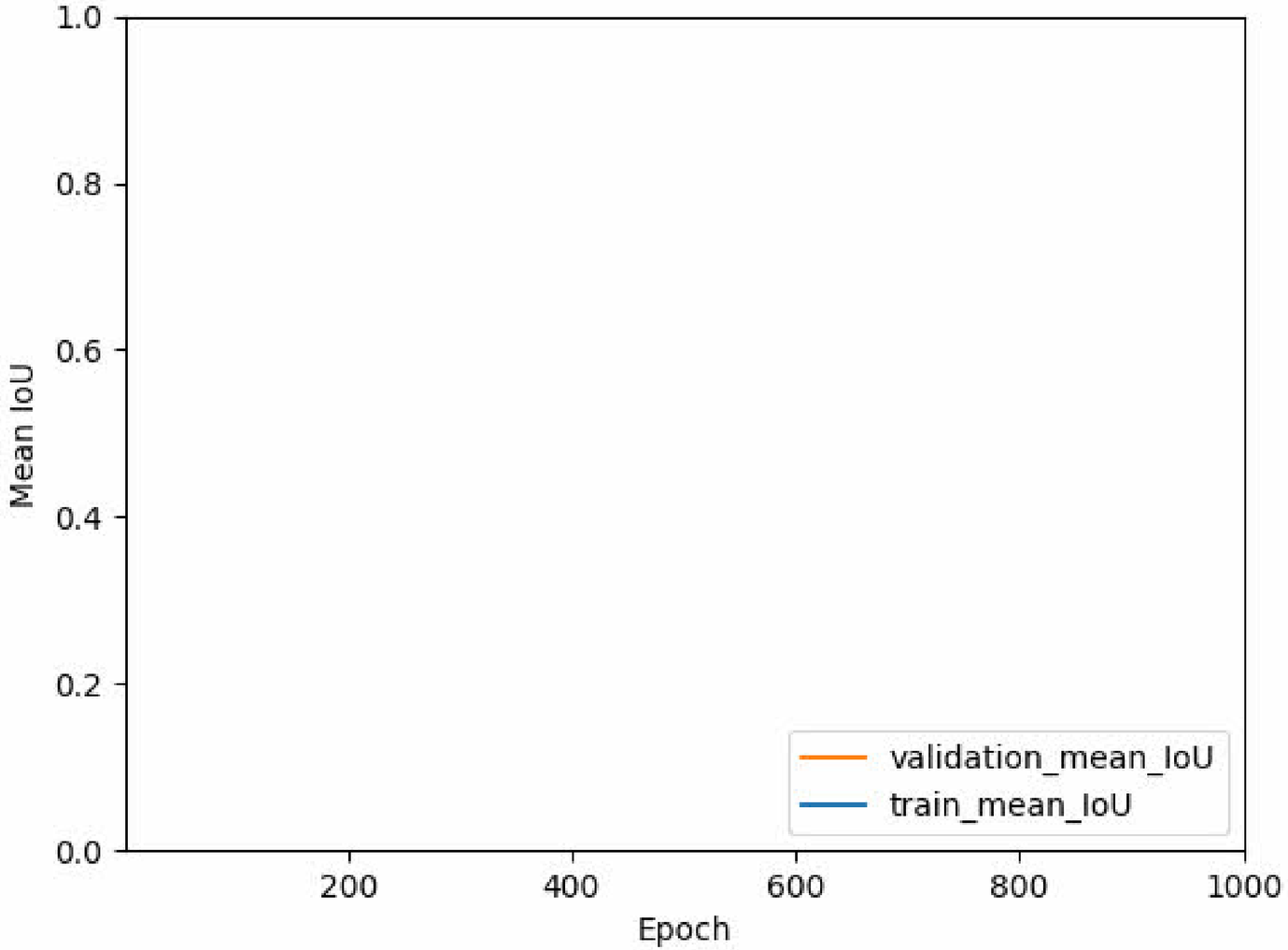
width

height

(x0, y0) width

height

Object Localization Model

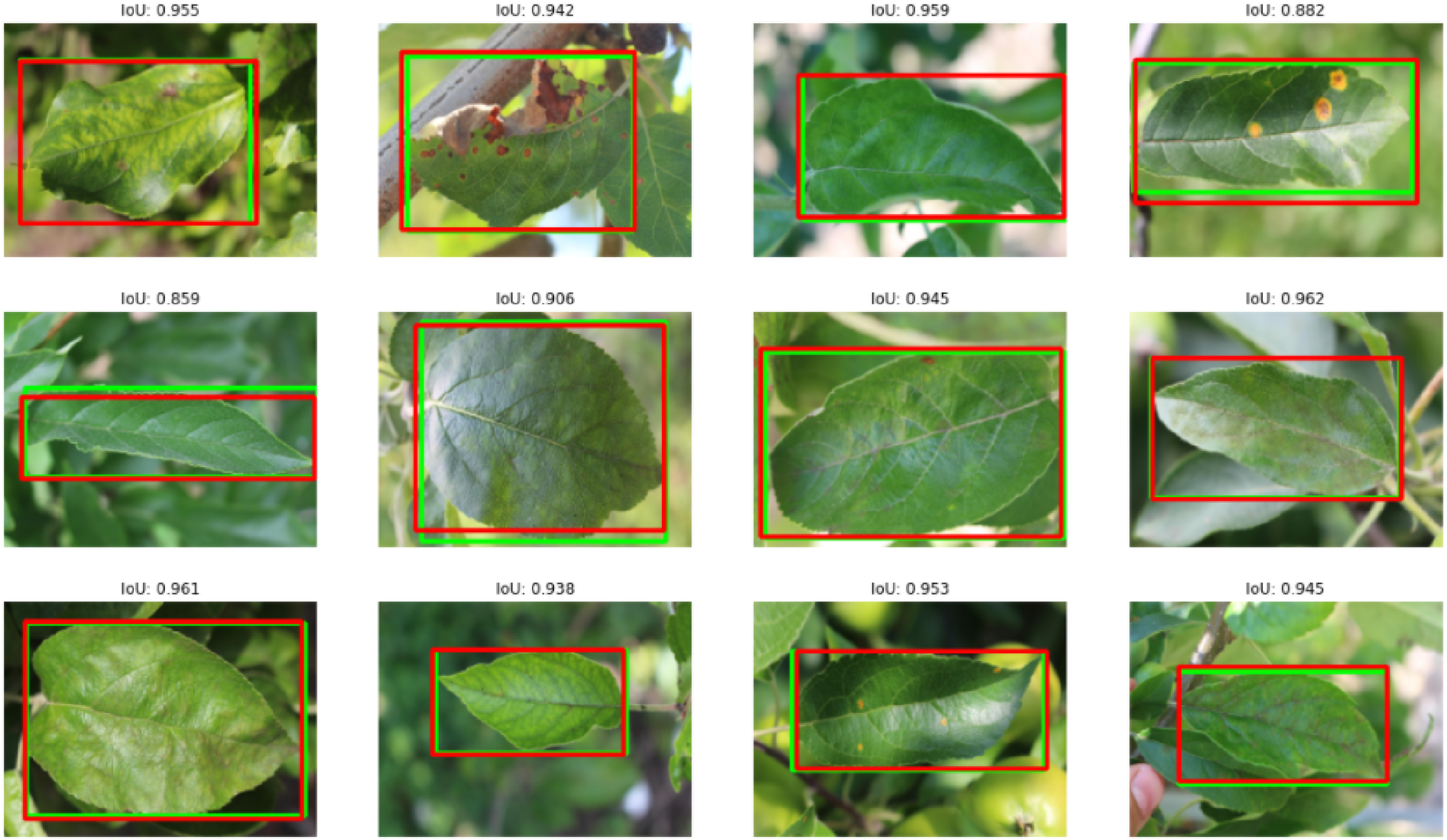


Training mean IoU: **0.929**

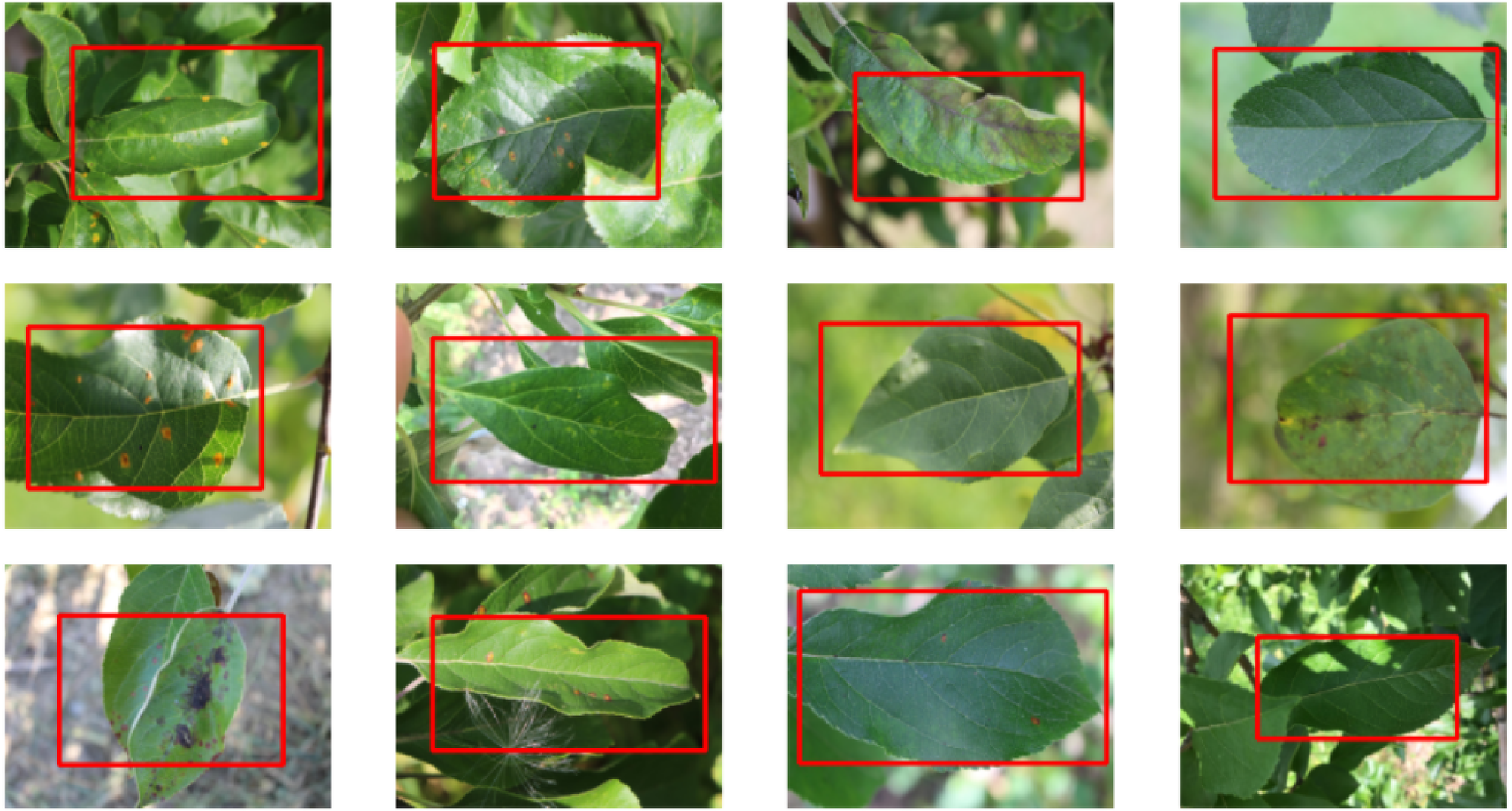
Validation mean IoU: **0.782**

Epochs run: **1000**

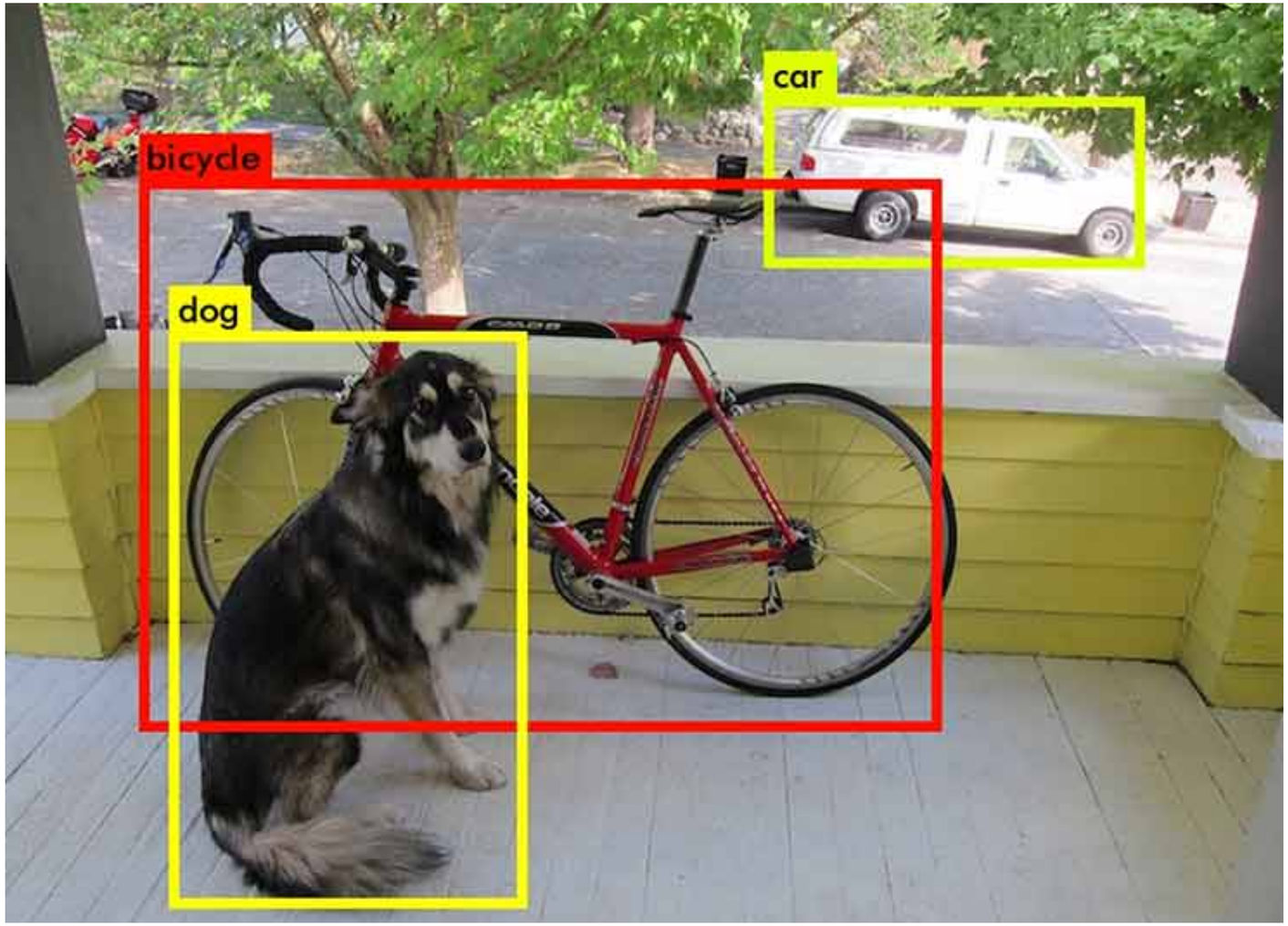
Example Bounding Box Predictions (Training Images)



Example Bounding Box Predictions (Test Images)

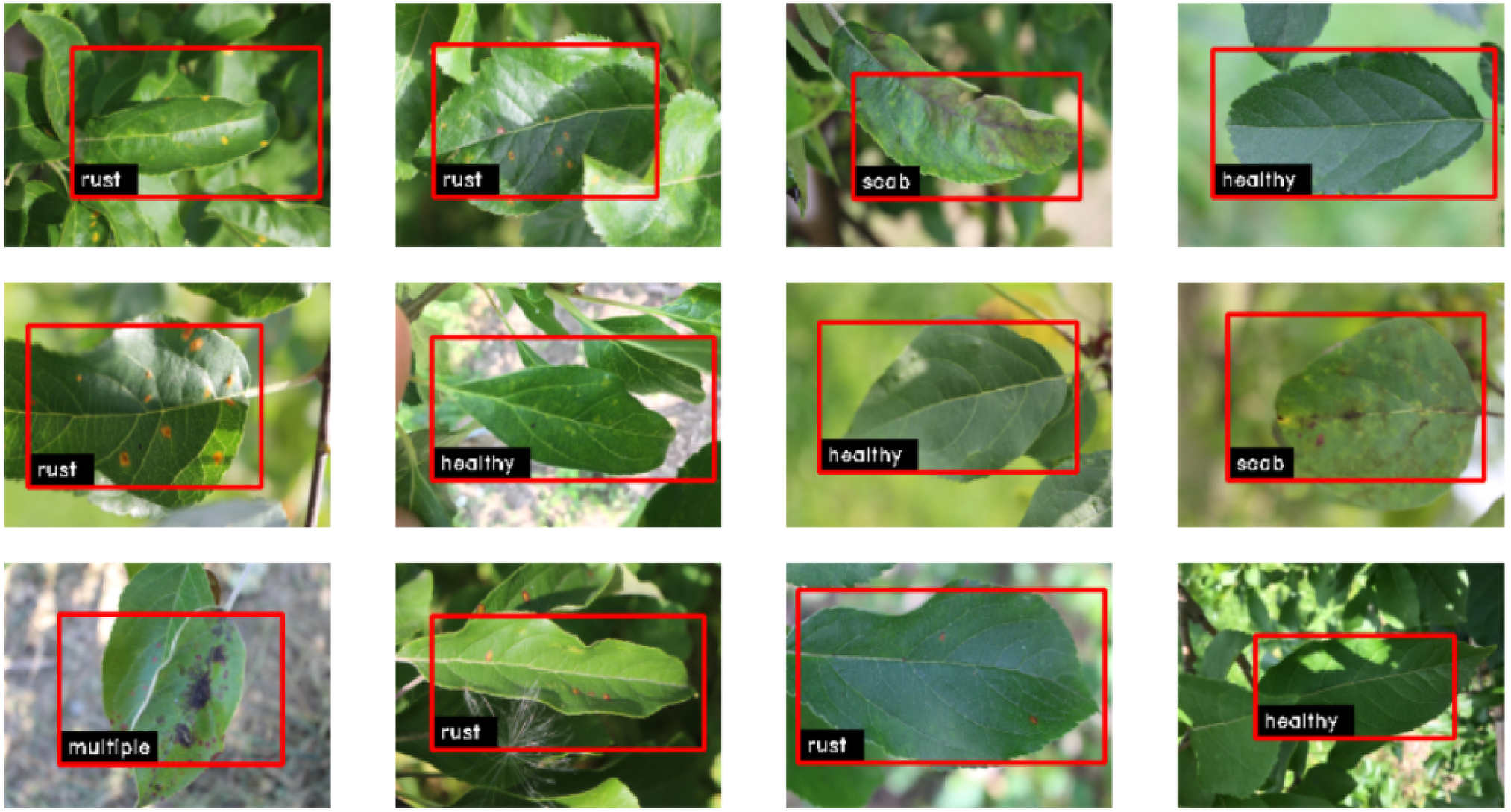


Let's combine our models to create something like this!



Gaurav Sinha, towardsdatascience.com

Combining our Classification + Localization Model (Test Images)



Conclusions

So far...

- Model that classifies leaf diseases
- Another model that performs object localisation

Next steps

- Spend more time to train the models to achieve a higher position in the competition
- Label more training images with bounding boxes
- Model that detects all leaves in the image and classifies their health status

THANK YOU

ANY QUESTIONS?