Plant Disease Classification Using Leaf Images

Josh Slizinov

Agenda

- Problem Statement
- Solution Proposition
- Gathering Data & EDA
- Modeling
- Model Performance
- Improvements
- Recommendation
- Questions

Problem Statement

- Horticulture is great until your plants get sick
- Plant disease can significantly affect yield and sometimes kill/stunt plants altogether
- Diseases are difficult to identify because some look very similar to others and the grower's domain knowledge can impede their diagnosis
- Can a machine learning be used to assist in the classification/identification of diseases in tomato plants?

Solution Proposition

Develop a convolutional neural network to classify images to the correct disease class.

Background

- Commercial vertical hydro farming
- Increase accuracy and speed of/speed to identification

- Consumer hydro farming
- Possibly limited knowledge

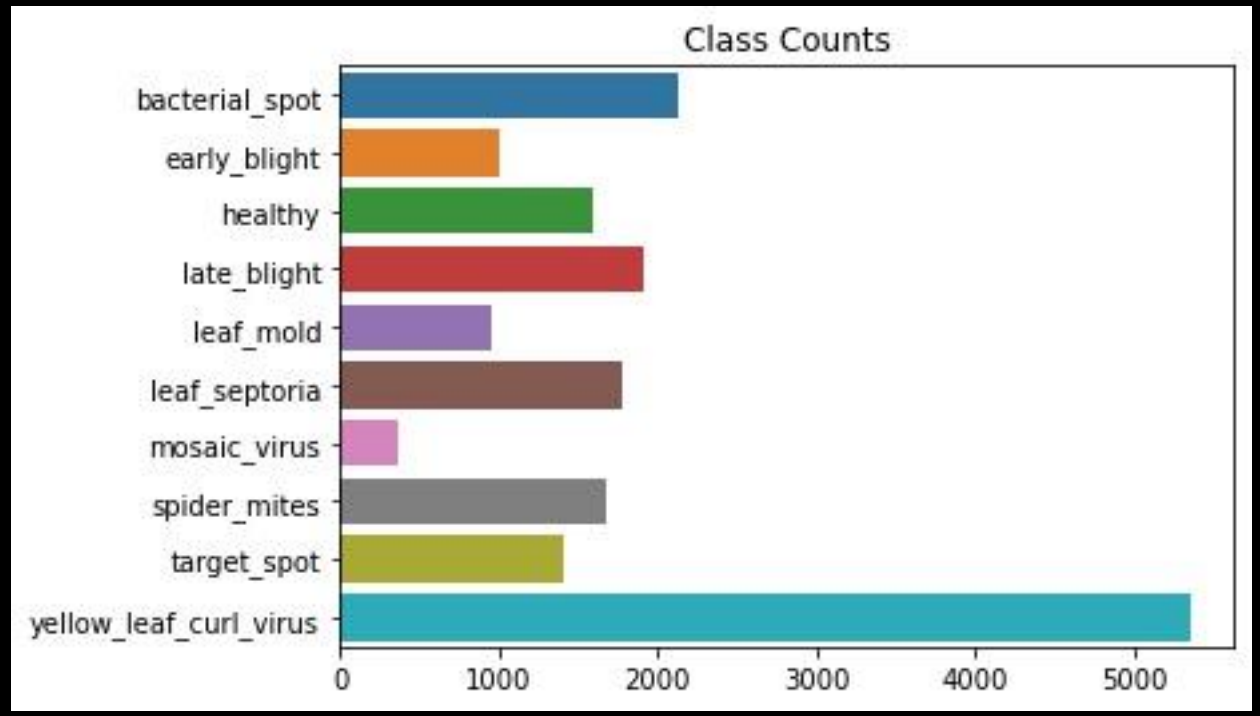
Data Gathering

- Easy process
- All pre-prepared
- 18,160 images

Samples of Data




Class Imbalance



10 classes

Baseline: 0.295

Class Balancing

- Data Augmentation
  - Random zoom
  - Random rotation
  - Random shift
- 3,000 images per class

Baseline : 0.1

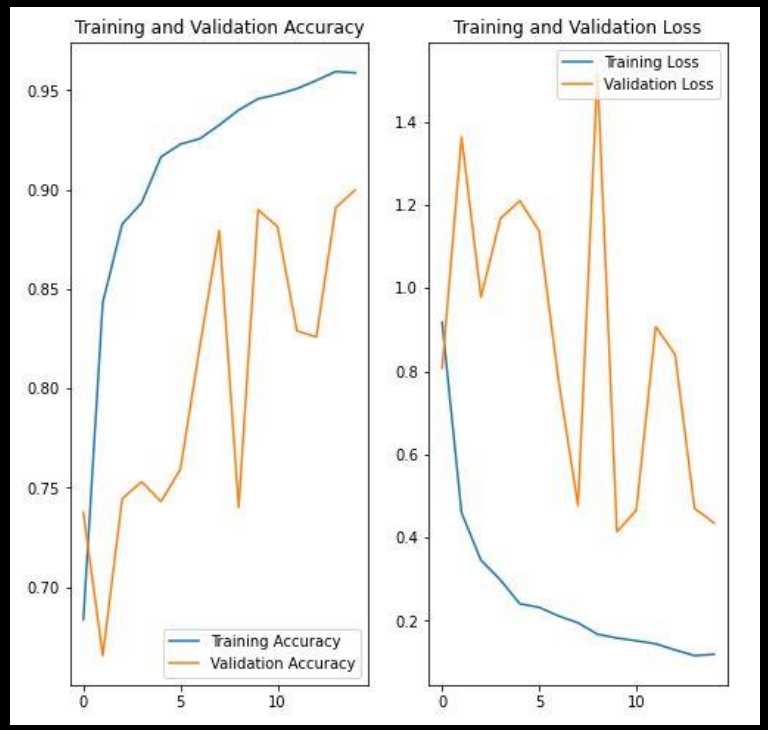
Samples of Augmented Data



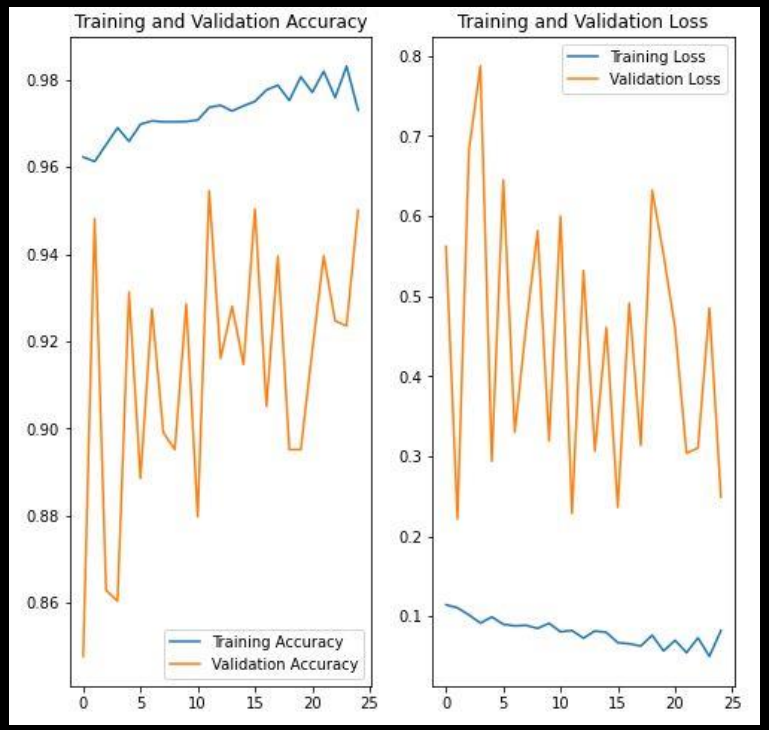
Modeling

- Convolutional neural network
  - Suffered from overfitting
    - Mitigated with:
      - Data augmentation layer
      - Kernel regularization (L1 0.0001)
      - Dropout (0.2)
- After overfitting correction
  - Sporadic loss and accuracy
    - Decrease learning rate

Model Performance



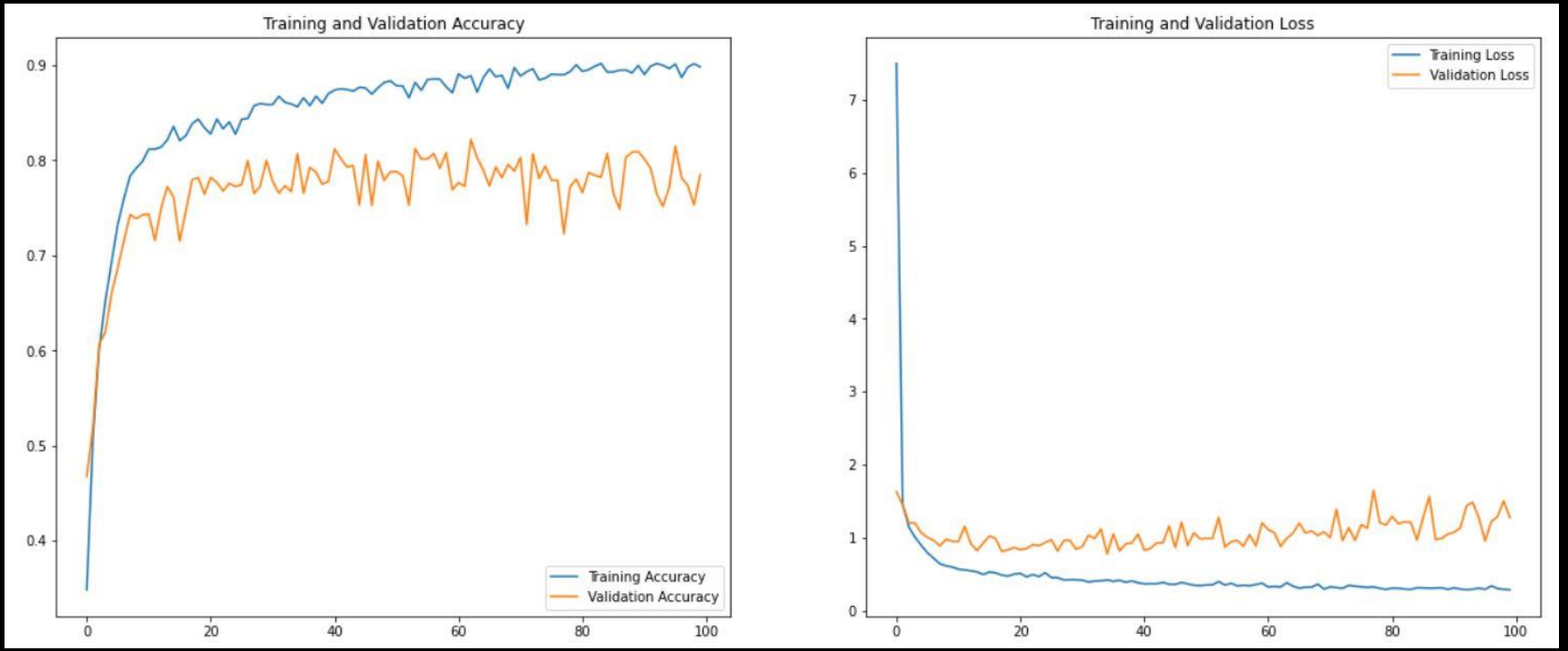
Overfitting Mitigation Model



Decreased Learning Rate

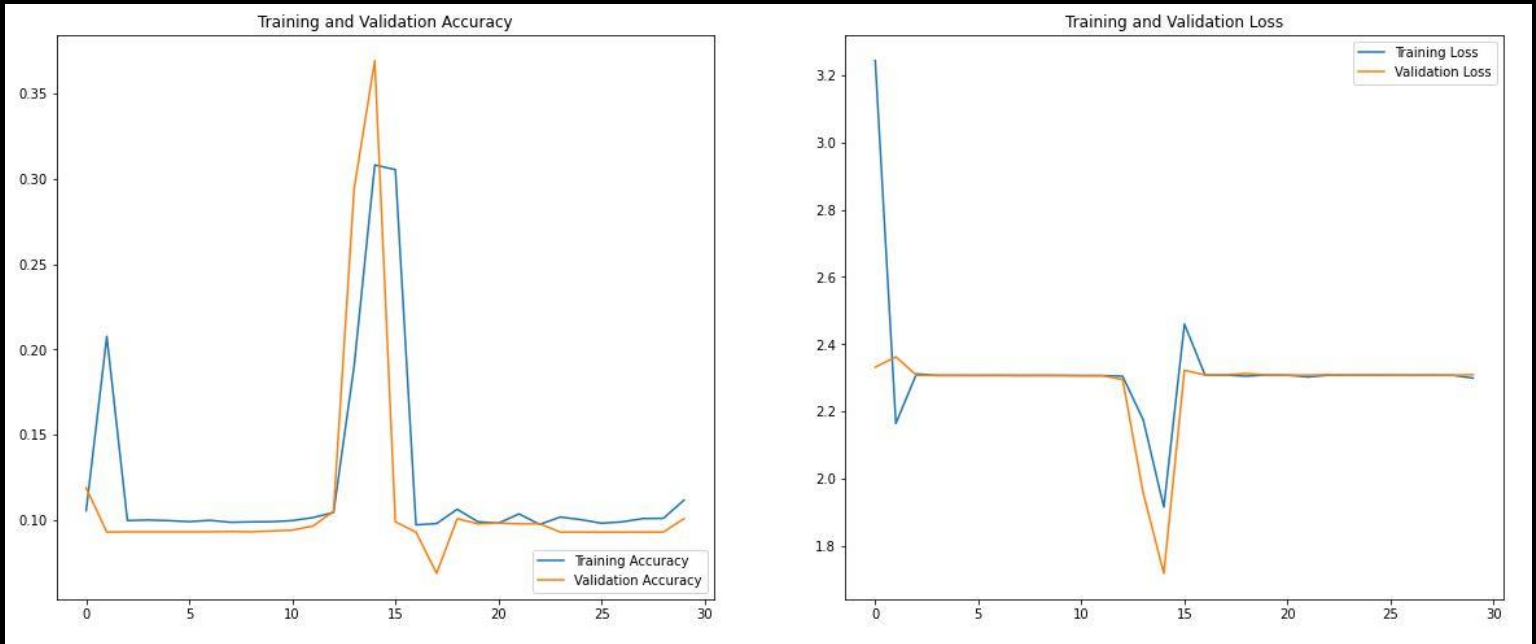
- High accuracy
- Very sporadic
  - Suspicious
- Random loss
- Beats baseline

Least Sporadic



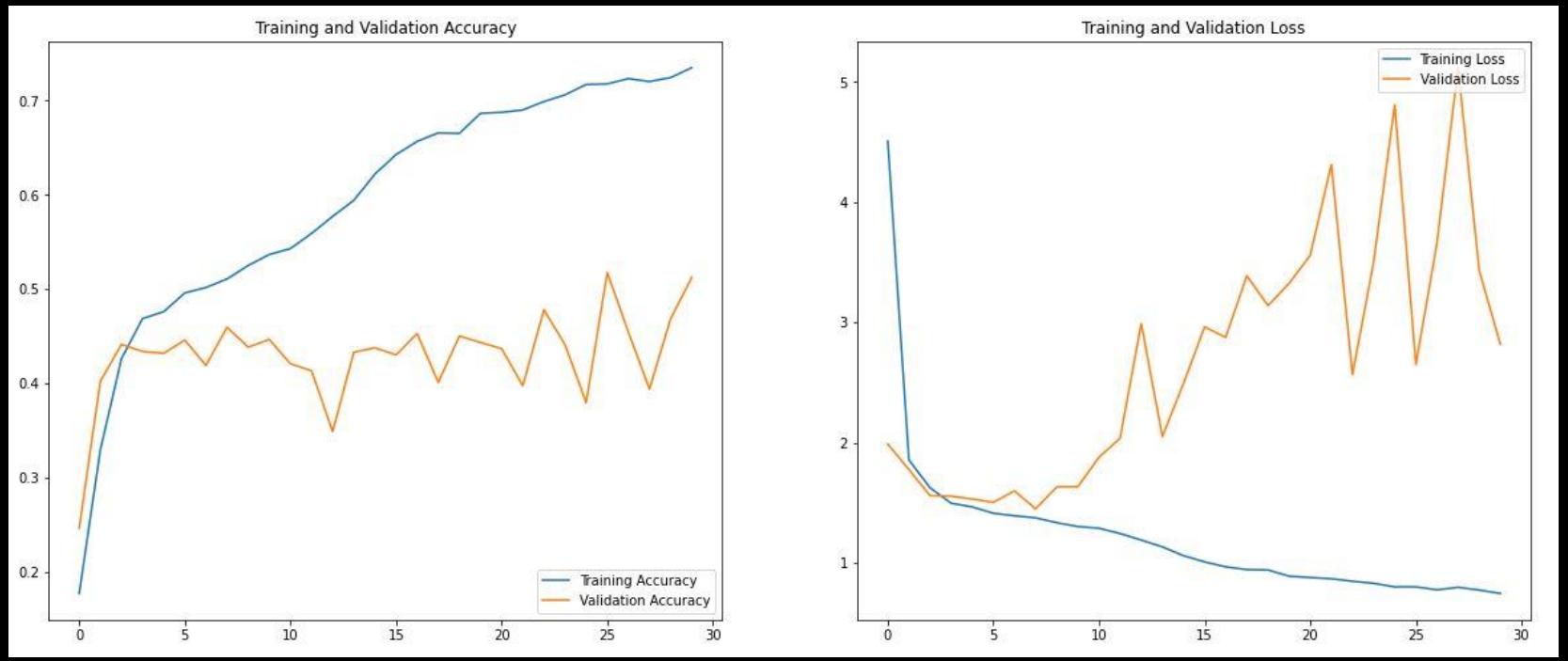
- Still sporadic, but less
- Peaks at ~0.8 accuracy
- Beats baseline

Balanced Class Performance



- Just looks suspicious
- Beats baseline, but it's iffy

Balanced Class Performance Cont.



Still beats baseline, but looks poor, visually

Loss or Accuracy?

- High accuracy, low loss would be nice
- Accuracy may be more important because the correct classification is what matters in the end

Improvements

- Better processing power
  - Huge obstacle
  - Hyperparameter tuning
  - More epochs
- More knowledge and research into CNNs
- More time
  - More epochs
- Additional data augmentation methods

Recommendation

- Almost all the models can technically be used because they beat the baseline
  - Would be better if accuracy was higher
    - It's possible to achieve higher accuracy

Questions?