Crop Disease Image Classification

A Convolutional Neural Network Analysis

by Ben Geissel

Why is Plant Disease Important?

- **80%** of human diet
- Up to **40%** of losses in global food production
- Food demand set to **double by 2050**
- **50%** of land used for agriculture

Data Source

- PlantVillage image data from David Hughes
  - **54K images**
  - **38 classes**
  - Not all crops have healthy and infected classes

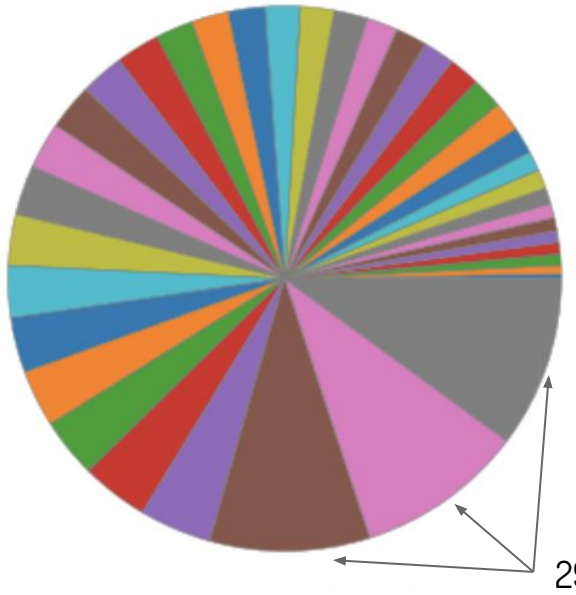
Potato: Healthy

Potato: Late Blight

Machine Learning Models & Limitations

- Typical Machine Learning models are **fast**, **BUT...**
- Flatten image arrays → HUGE data size
- Unable to perform SMOTE, Principal Component Analysis, Support Vector Machine, XGBoost
- Lower accuracy than desired
- Unable to handle class imbalances

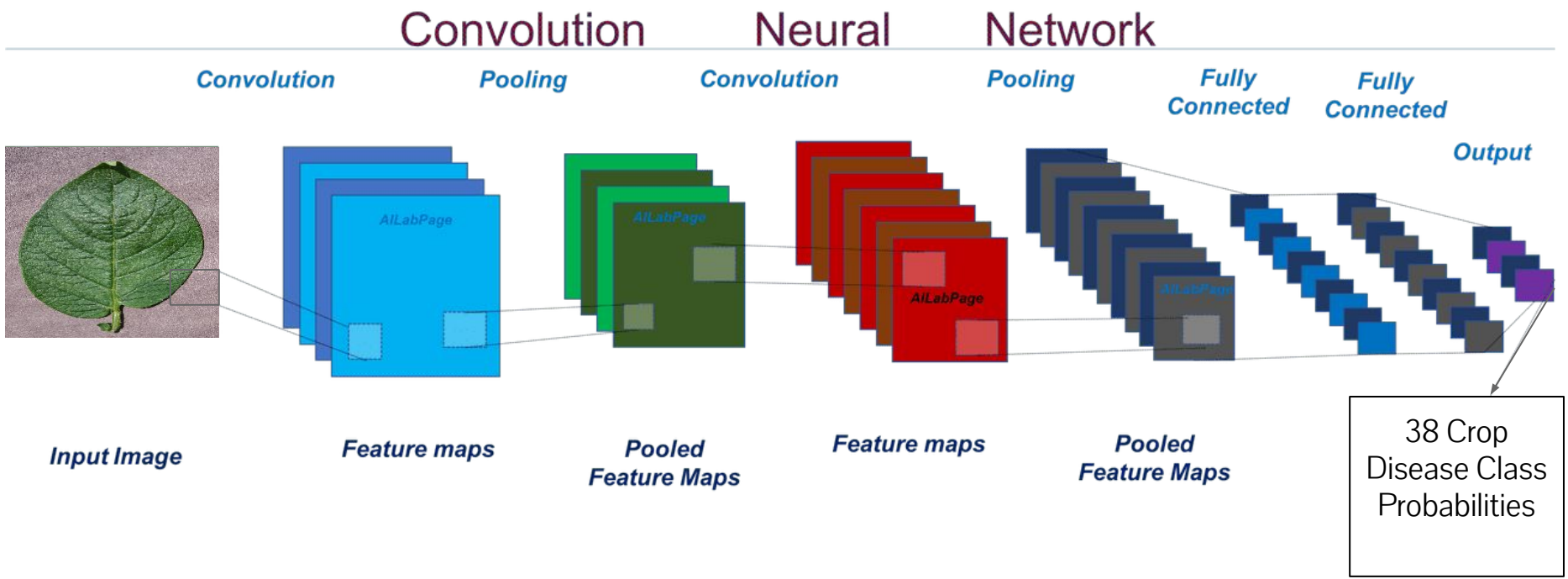
Class Imbalance



Multinomial Naive Bayes - **35% Accuracy**

Random Forest - **64% Accuracy**

Convolutional Neural Networks



Deep Learning - Convolutional Neural Network

- Standardize pixel values of each image
- **Image Data Generator**
- 17 Layers
- 25 Epochs
- 3 hr 20 min run time

97%
Accuracy

Conclusions

- Machine Learning Models: Fast, but inaccurate
- CNN: Slow, but very accurate
- Model can be used to develop application to help farmers identify crop disease quickly
- Less Diseases → Greater Food Stability
- Application Prototype


Thank You