Creating an Image Classifier to Predict Wheat Blast Disease

Introduction and Problem Identification

- Wheat blast-fungal disease-harm wheat yield around the world
- Image classification machine learning (ML) could help diagnose?
- If successful, could replace expense of professional botanist
- Success=model with high predictive power
- Ideally low computational load

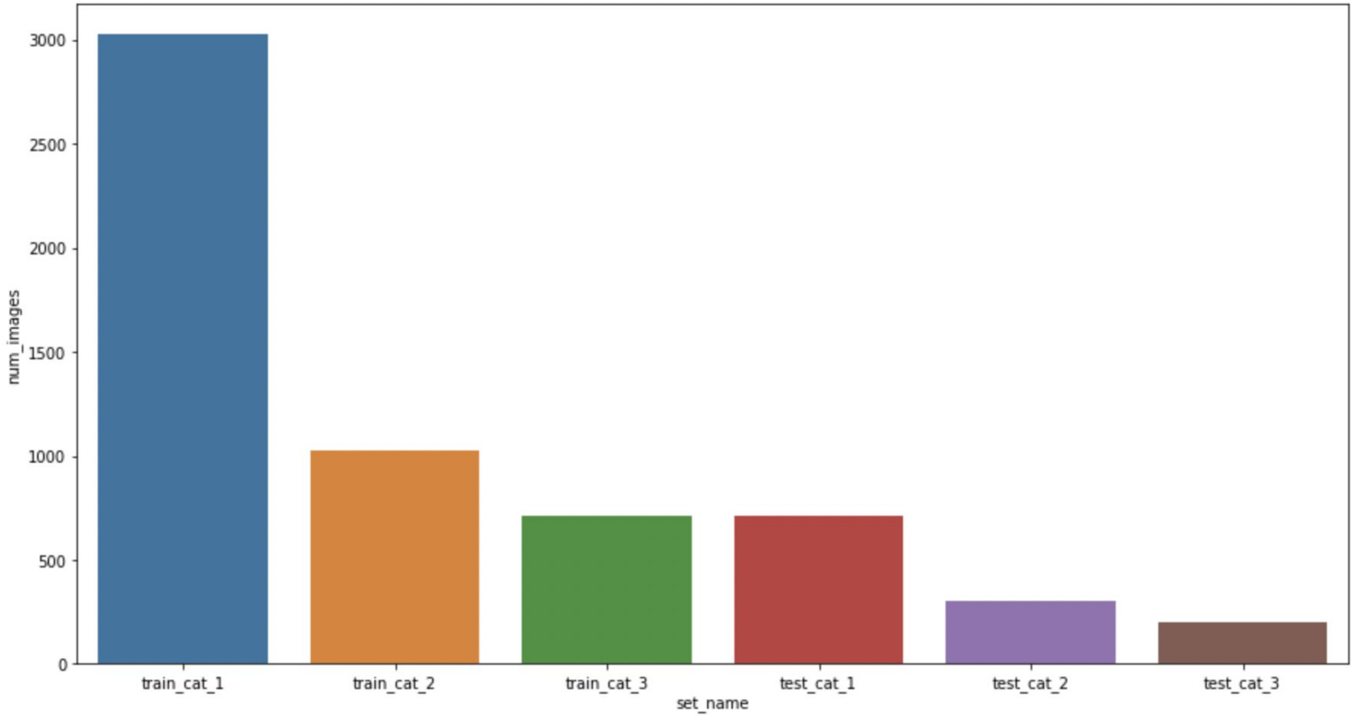
Introduction and Problem Identification 2

- Stakeholders-shareholders of a company selling a predictive app
- App could be mobile or online
- Constraint-this dataset contains photographs taken in controlled environment
- Greater number/variation of photos could increase predictive power
- Constraint-not enough time to fully fine tune model

Dataset Description

- Publicly available dataset from Fernandez-Campos et al.
- ~6,000 wheat spike images
- All images are stated to be unique, relevant, and high quality
- Photographed w high quality camera, blank background, standardized distance
- Pre-split into training and test set, test size 0.2.
- Pre-split into categories of blast severity, as marked by plant pathologist.

A Visualization of the Dataset:



Data Wrangling

- Data was downloaded from Fernandez-Campos et al website onto hard disk
- Zip file ~3 GB
- Uploaded google drive, drive mounted in google colab
- Data briefly inspected, determined clean
- Data reported clean-Fernandez-Campos et al

Exploratory Data Analysis (EDA)

- Displayed images, inspected
- Plotted image sizes, aspect ratio
- Visualized number images each subset
- Image Feature Analysis
  - Canny edge, fast features, graphical representation

Data Preprocessing

- Resizing-Keras’ Image Data Generator
- Assigned train/test-Keras’ fit_generator()

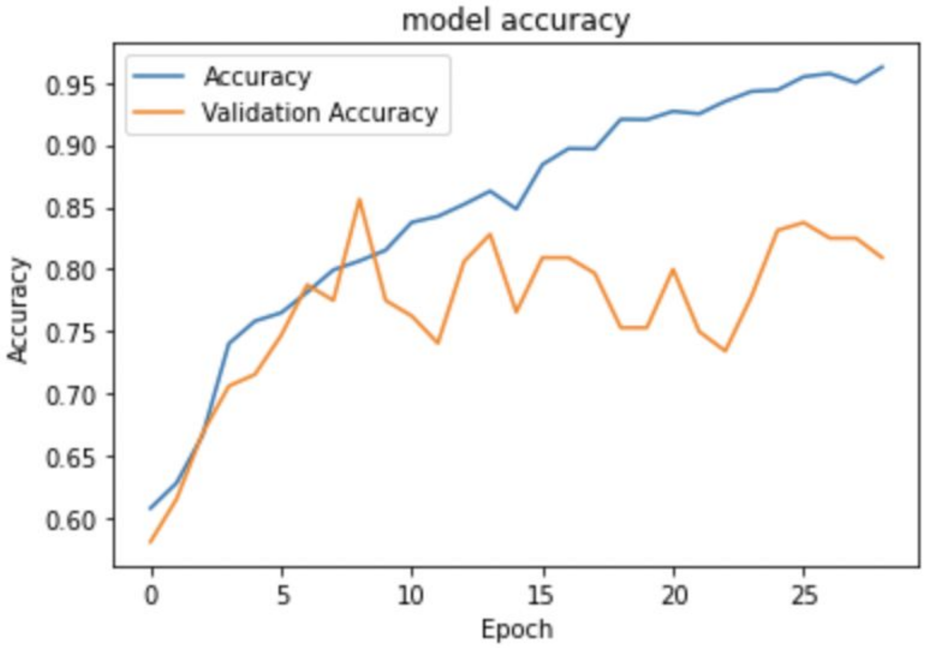
Modeling

- 2 models created
- Model 1-pretrained VGG16 model
- Model 2-smaller CNN-from scratch
- Chosen because CNNs-tend have strong predictive power images
- Model 2 allows us see if simpler model can be as effective

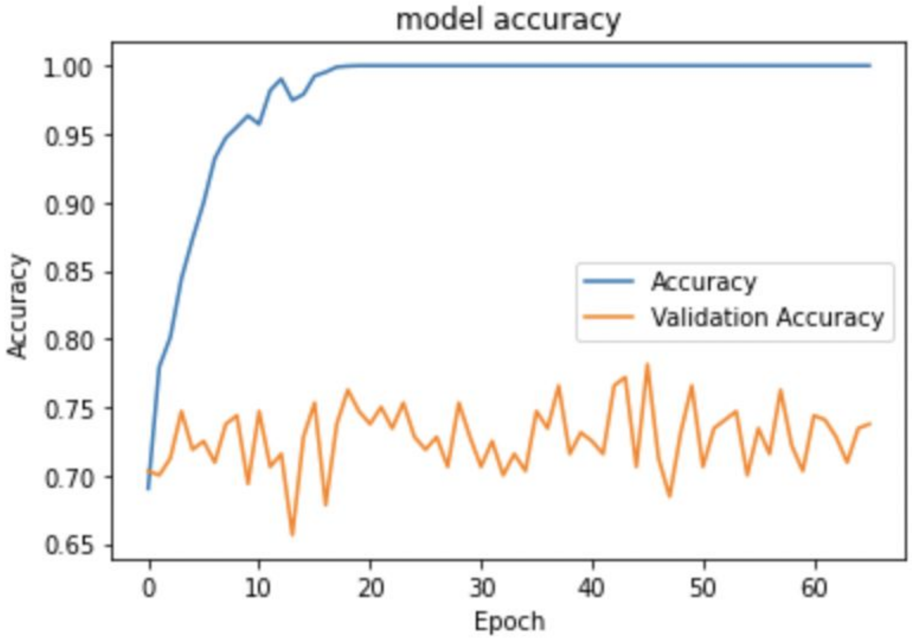
Results

- Model 1 (pretrained, VGG16) highest validation accuracy across epochs: 0.85625
- Model 2 (simple, from scratch) highest validation accuracy across epochs: 0.78125

Results: Model 1 (pretrained, VGG16)



Results: Model 2 (simpler, from scratch)



Conclusions

- Model 1 is better if purely performance is concerned
- Model 2 may be better if a combination of performance and simplicity is preferred

Recommendations

- If a model performs better than expert, make into app
- Market to farmers who want catch blast early
- Market to farmers-want identify blast resistant cultivars
- Market to agriculturalists-trying create blast resistant cultivars
- Liaise with botanists and ML engineers to try boost model performance even more

Future Steps

- Try other pretrained models
- Fine tune hyperparameters
- k-fold cross validation
- Data augmentation
- Additional datasets
- Change to a binary prediction

Acknowledgements

I would like to thank my mentor Branko Kovac for providing me with constructive feedback on this project.

Works Cited

- Fernandez-Campos, Mariela, et al. "Wheat spike blast image classification using deep convolutional neural networks." *Frontiers in Plant Science* 12 (2021): 1054.

- https://www.kaggle.com/keras/vgg16
- https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c
- https://purr.purdue.edu/publications/3772/1
- Sitaula, Chiranjibi, and Mohammad Belayet Hossain. "Attention-based VGG-16 model for COVID-19 chest X-ray image classification." *Applied Intelligence* 51.5 (2021): 2850-2863.

Images Cited:

1. https://www.google.com/search?q=wheat+blast&source=lnms&tbm=isch&sa=X&ved=2ahUKEwiN1dmIx_XyAhXaFVkFHZqOAQEQ_AUoA3oECAEQBQ&biw=1431&bih=727#imgrc=shxXvjkv7TswtM