Predicting Diseases in Chinese Cabbage

With an increase in sustainable growth, Singapore has started to expand on urban farms and community gardens. By providing ML solutions to hobbyist and volunteers with little to no experience, this project aims to assist them to identify common diseases in chinese cabbage.

DSIF-2 Capstone Project

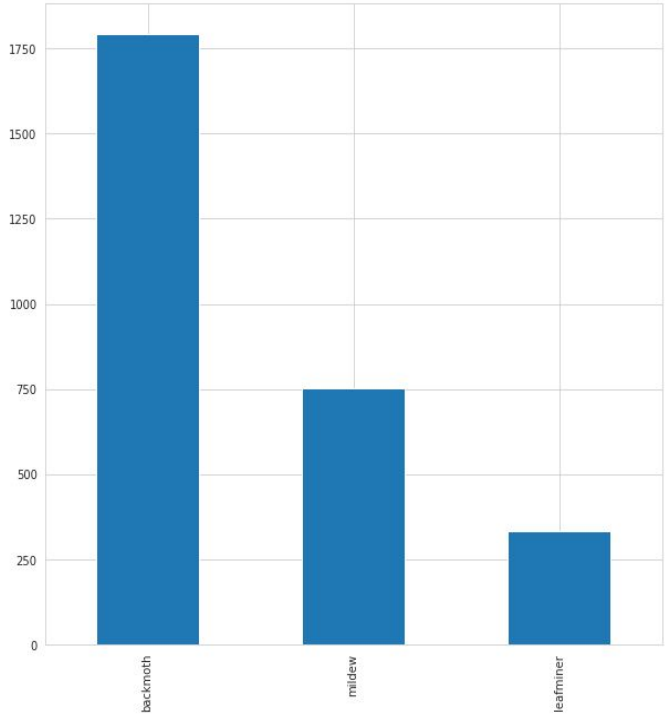
Melvin Chandra

The Data

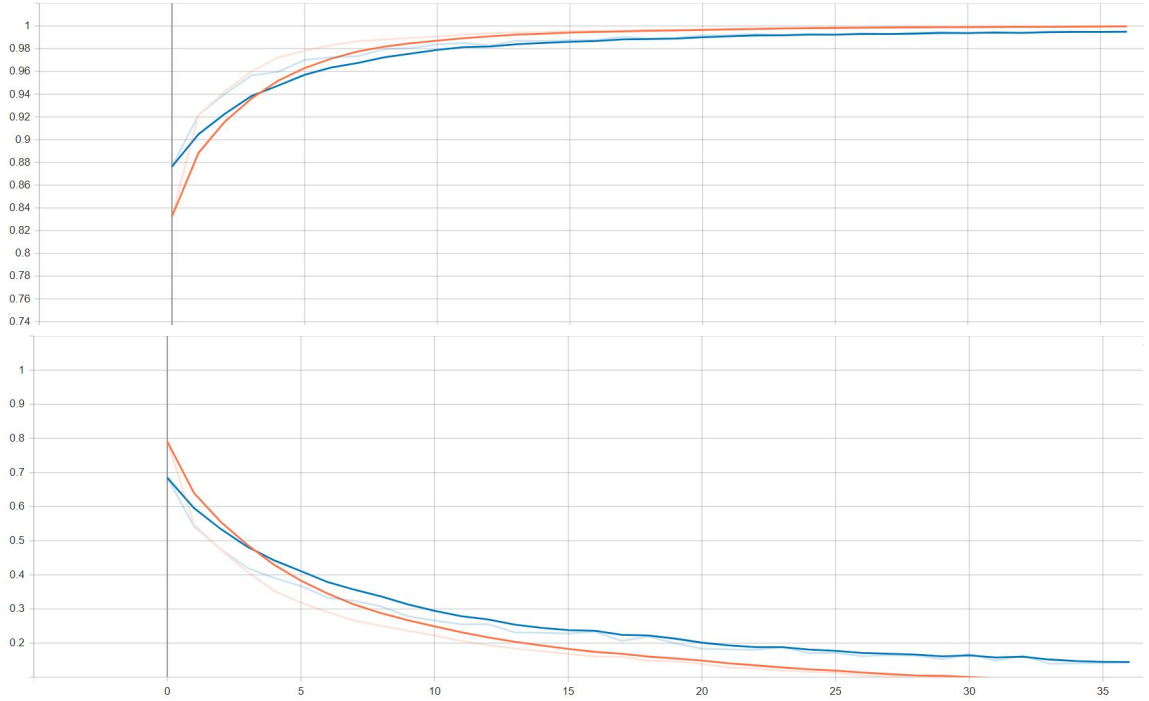
Backmoth

Leafminer

Mildew



MobileNetV2



Model: MobileNetV2

Compile Metrics: AUC

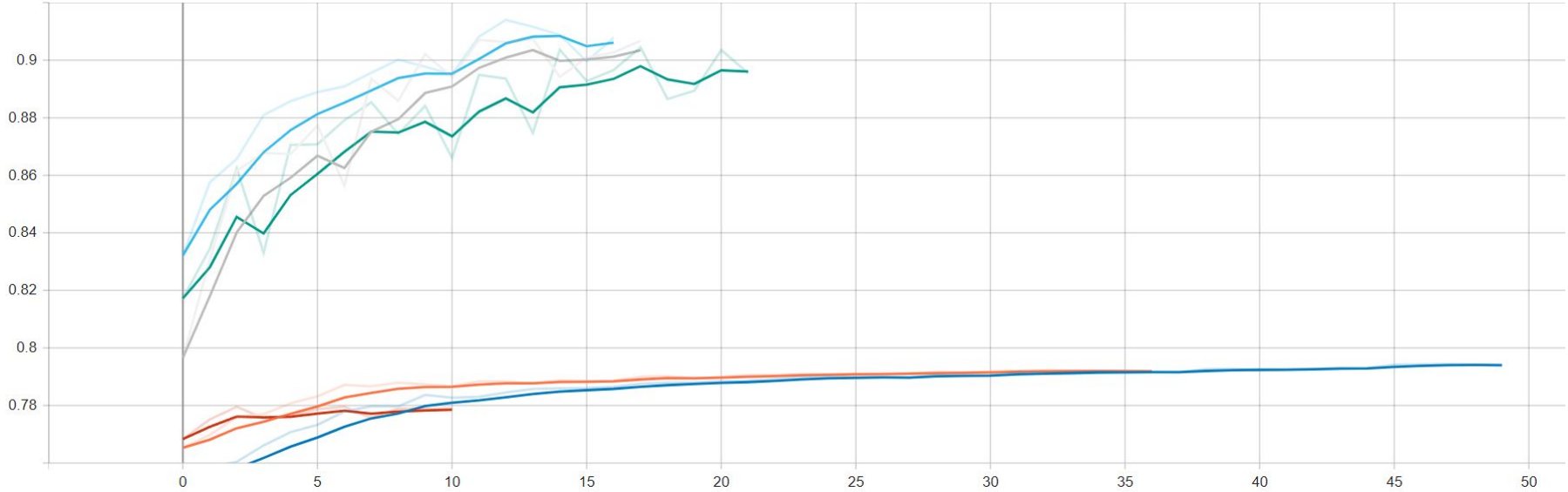
Loss: Categorical Cross Entropy

Metric	Value
Train AUC	0.9997
Validation AUC	0.9951
Train Loss	0.0756
Validation Loss	0.1435

What went wrong?

- Problem:
  - Data Leakage in train test split
- Solution:
  - Manual splitting
  - Data Augmentation
  - Try with other models

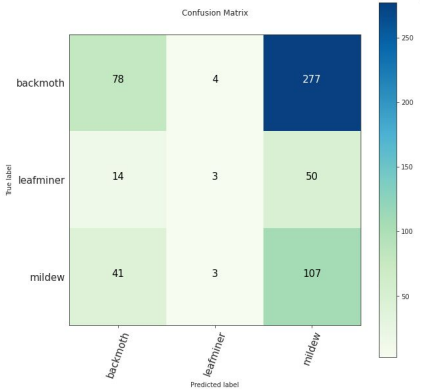
Musings from multiple models



Model	MobileNetV2	EfficientNetB0	MobileNetV3	InceptionV3	EfficientNetB5	EfficientNetB6
Val AUC	0.9067	0.794	0.9078	0.8955	0.7918	0.7789
Early Stop Epoch	17	49	16	21	36	10

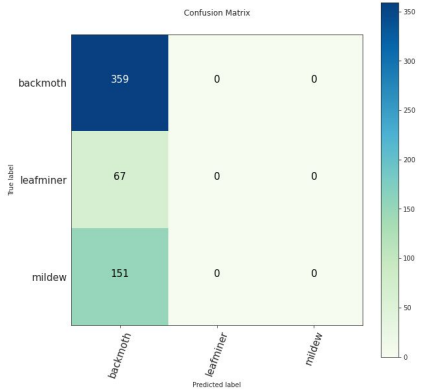
MobileNetV2 vs EfficientNetB6

MobileNetV2



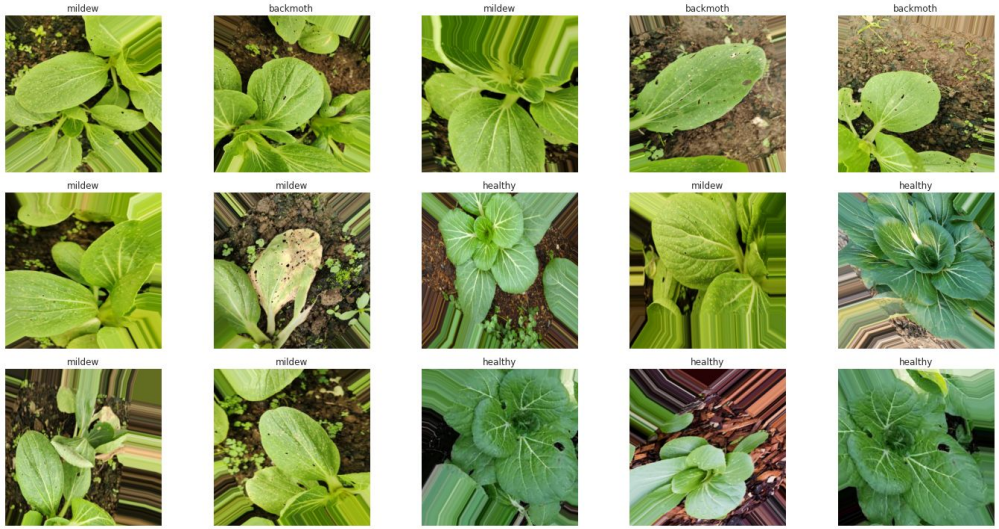
	precision	recall	f1-score	support
0	0.59	0.22	0.32	359
1	0.30	0.04	0.08	67
2	0.25	0.71	0.37	151
accuracy			0.33	577
macro avg	0.38	0.32	0.25	577
weighted avg	0.46	0.33	0.30	577

EfficientNetB6



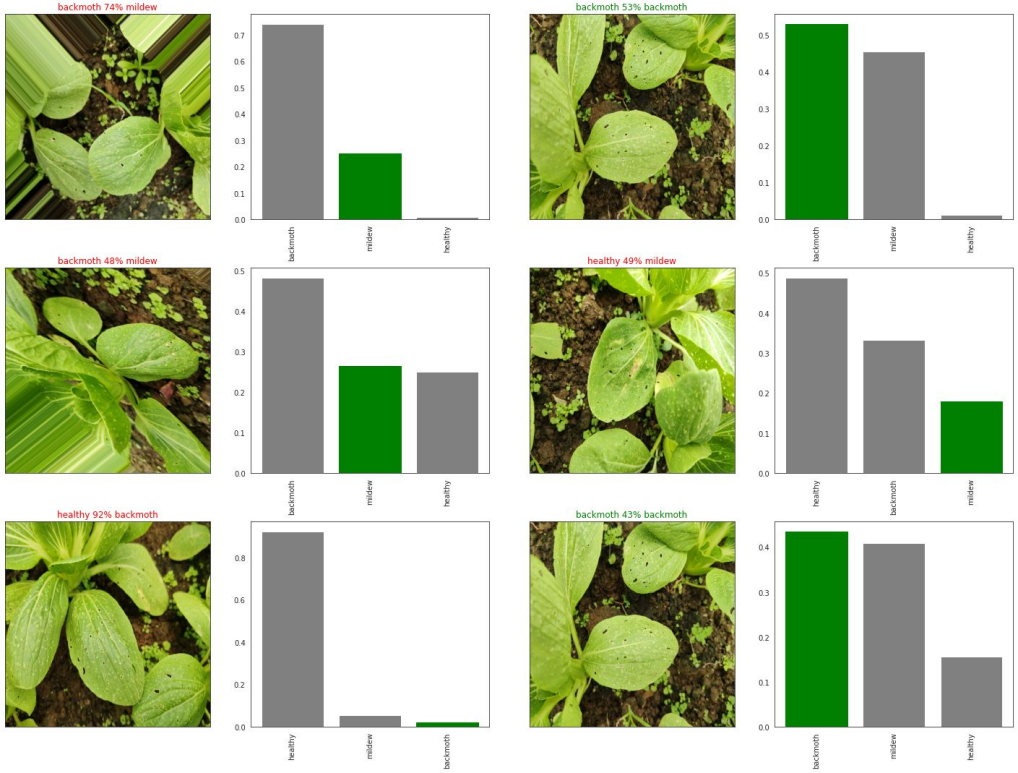
	precision	recall	f1-score	support
0	0.62	1.00	0.77	359
1	0.00	0.00	0.00	67
2	0.00	0.00	0.00	151
accuracy			0.62	577
macro avg	0.21	0.33	0.26	577
weighted avg	0.39	0.62	0.48	577

The Remedy

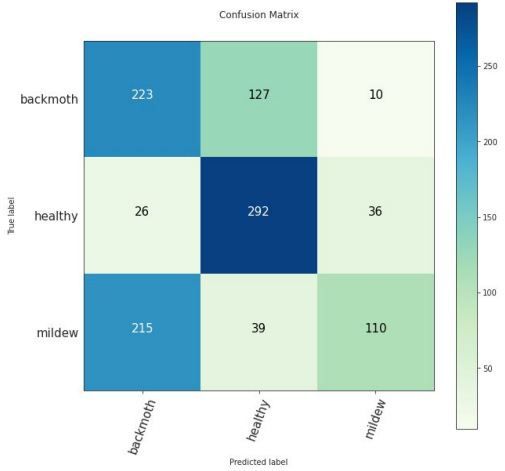


- Removed leafminer class
- added healthy samples
- Oversampling with ImageDataGenerator
- Increased more parameter variance in preprocessing
- Remove Early Stopping
- Run lots and lots more epoch

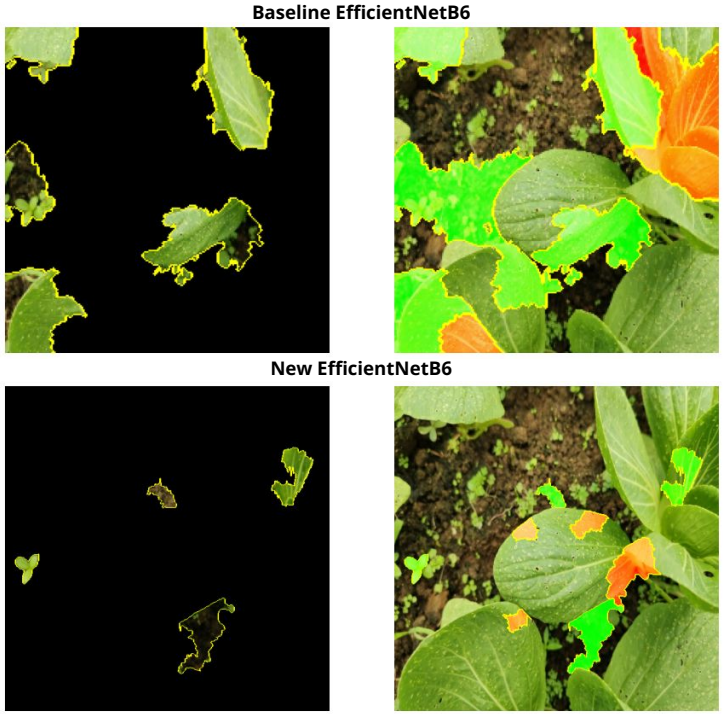
300 Epochs Later...



	precision	recall	f1-score	support
0	0.48	0.62	0.54	360
1	0.64	0.82	0.72	354
2	0.71	0.30	0.42	364
accuracy			0.58	1078
macro avg	0.61	0.58	0.56	1078
weighted avg	0.61	0.58	0.56	1078



Conclusions



- Forms smaller clusters of features
- Still picks up noise
- Still has problems isolating the main subject of the image

Recommendations

- Better data
  - Sourced from multiple farms, unaugmented
  - More distinct blights
- Modelling
  - Bounding boxes
  - More fine-tuning (unfreeze & retrain)