Ana Cira and Anna

- Dataset from Kaggle (Playground)
- Pictures of Seedlings (12 sorts)
- Important to recognise the good seedlings from the “bad” weed seedlings
- Goal: Train a network to recognise seedlings

Cleavers

Common Chickwead

Maize

Charlock

Common wheat

Shepperd’s Purse

Blackgrass

Fat Hen

Loose SilckyBent

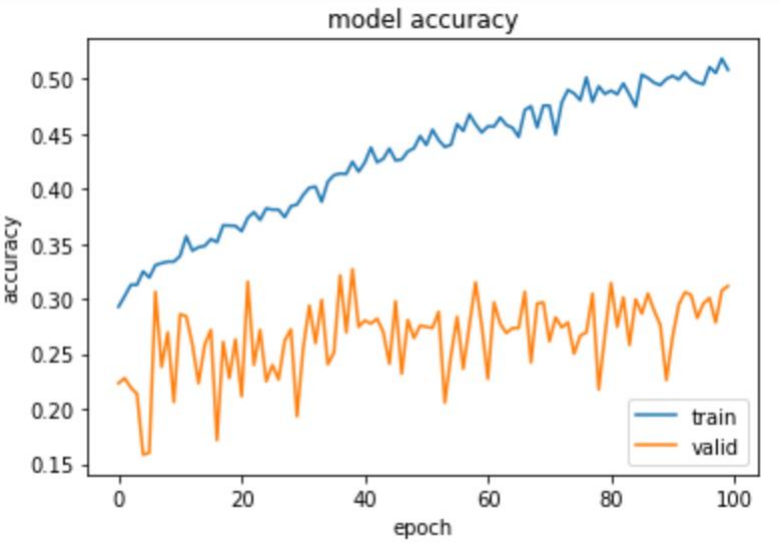
Sentless Mayweed

Small-flowered Cranesbill

Sugar Beet

- Read the images (one class could not be read → 11 kinds of seedlings)
- Shuffle the Pictures and Labels
- Create Test, Training, Validationsets

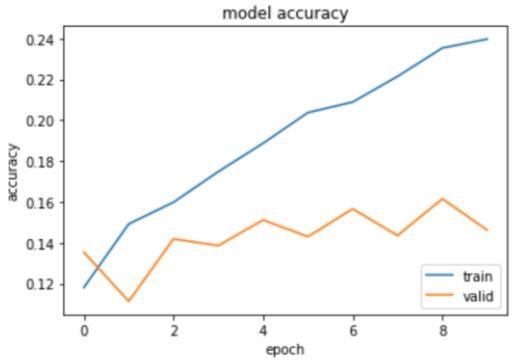
- Convolutional Neural Network with 3 Layers and BatchNormalization
- Data read in as RGB, reshaped but
- not further processed
- **Unbalanced dataset?**
- **Not enough data?**

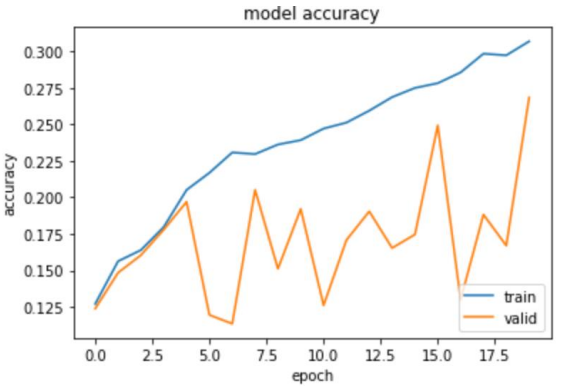


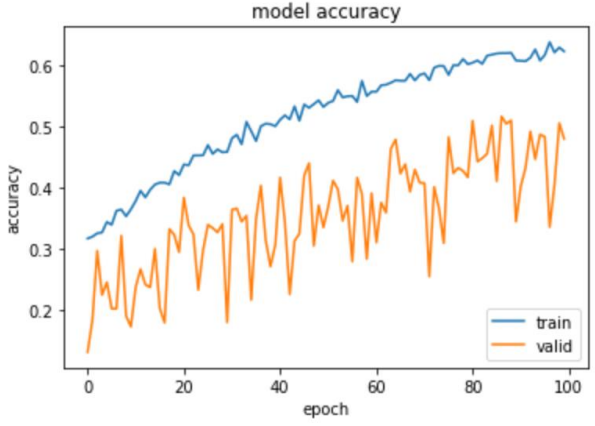
- Check whether data is balanced →
  - Not really
- Rebalance the dataset

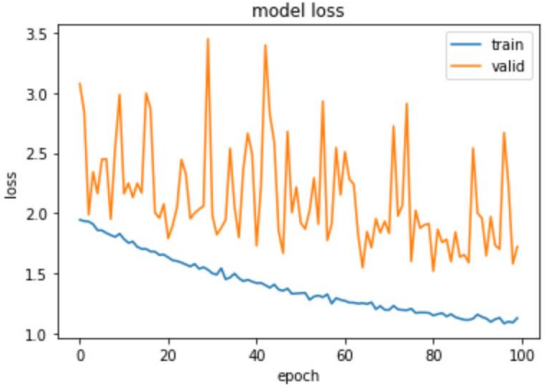
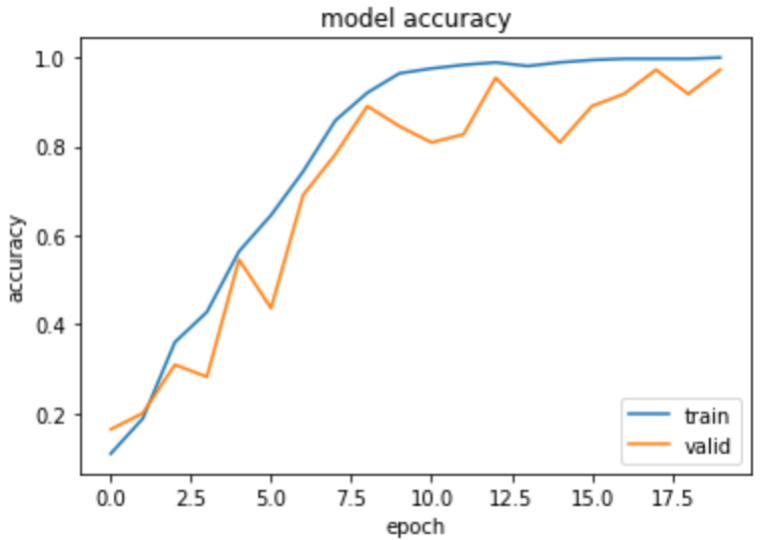
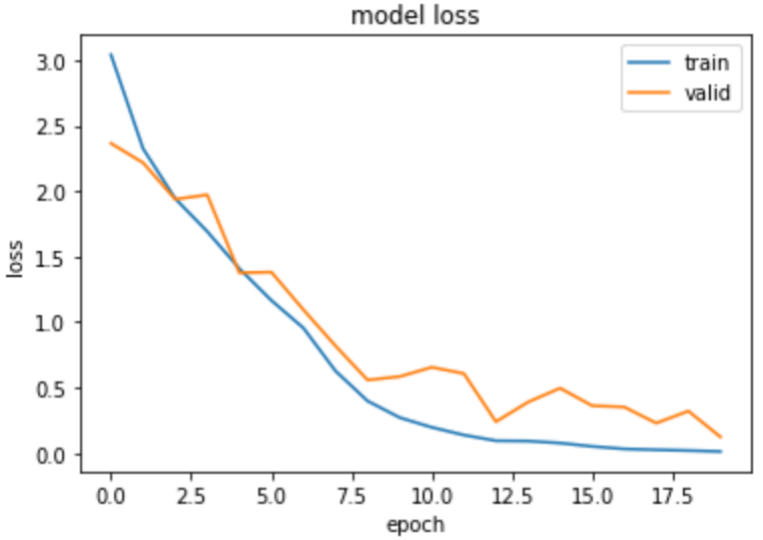
```
In [37]: ## check whether data is balanced
         pd.DataFrame(labels)[0].value_counts()

Out[37]: Loose Silky-bent              762
         Common Chickweed              713
         Scentless Mayweed             607
         Small-flowered Cranesbill     576
         Fat Hen                       538
         Sugar beet                    463
         Charlock                      452
         Cleavers                      335
         Black-grass                   309
         Maize                         257
         Common wheat                  253
         Name: 0, dtype: int64
```


model accuracy
accuracy
epoch
train
valid


model accuracy
accuracy
epoch
train
valid


model accuracy
accuracy
epoch
train
valid


model loss
loss
epoch
train
valid

- Make model simpler
- !! instead of validation set, the test set was given for validation!!
- Make dataset smaller → Result: