ETH zürich

Classification of a Melanoma with Deep Learning

Silvan Hostettler, Urs Mayr, Vera Jäggi

1 Introduction – Classes

Goal to classify three types of skin growth**.**

Class	**Nevus**	**Seborrheic keratosis**	**Melanoma**
Appears at	Birth (congenital)	Increased age	Increased age
Risk	Benign (harmless)	Benign (harmless)	malign (harmfull cancer)
Characet ristics	- hyperpigmented - mostly single colored - mostly arised skin **Can turn into a Melanoma**!	- Flat/raised papule /plaque - mostly one single - smooth, waxy or warty surface Can look similar to a Melanoma	ABCDE Rule (vs. Nevus) - **Asymmetry** - **Border** irregular in shape. - **Color** multicolored - **Diameter** > 6mm in diameter, - **Evolution** change characteristics over time.

2 Data

Melanoma Detection Dataset from Kaggle.
- *Unbalanced data*

	# total images
test	600
train	2000
valid	150

3 Models and Methods

1. *Baseline Model (with RandomForest [RF])*
2. *Pretrained Neuronal Network [NN]*
3. *Train a ConvolutionalNN from scratch*

RF and pretrained NN used the weights of the CNN VGG:

```
## Load base model weights ----
base_model <- application_vgg16(weights = "imagenet",
                                include_top = FALSE,
                                input_shape = c(image_size, image_size, 3L))
summary(base_model)

# i.e. freeze all convolutional VGG16 layers
freeze_weights(base_model)
```

All models were complied the same:

```
## compile model
transfer_model %>% compile(
  optimizer = "adam",
  loss = "categorical_crossentropy",
  metrics = c("accuracy")
)
```

4 Results and Discussion

Model	**Loss**	**Accuracy**
Baseline model with RF	NA	0.70
Pretrained VGG	1.52	0.63
CNN from Scratch	0.80	0.68

1. Baseline model with RF

The Baseline model using the Random Forest as classification lead to the highest accuracy with 70 %.

RF uses no distribution assumption and can handle well imbalanced data sets (weighted random forest, balanced random forest).

2. Pretrained VGG with

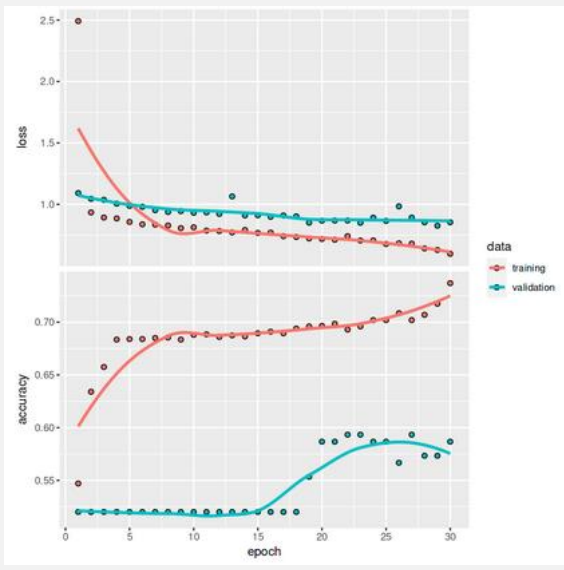
The pretrained VGG has the highest loss function and the lowest accuracy of all the trained models. This is surprising as the model provides 4096 features for the final densely connected classifier.

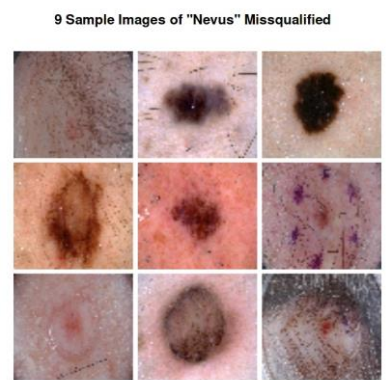
[…] 4 Results and Discussion

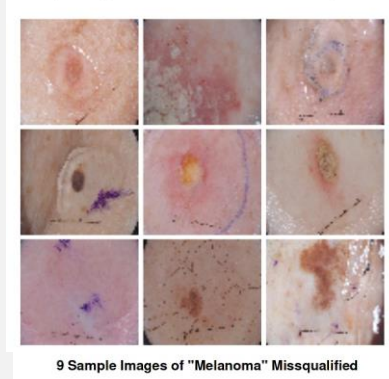
3. CNN from Scratch:

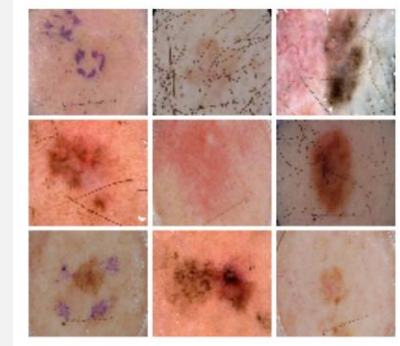
Overfitting (accuracy of Train data > test data)

Solution: data augmentation, neuronal dropout









4 Conclusion

Data difficulties:

- Different pixel size/coloring of images
- Imbalanced data
- Cumbersome data load -> writing of new data import function

Language/Versioning difficulties:

- Tensorflow/Keras better documented for Python than for R
- Image data augmentation seems to depend on the version of keras and tensorflow. -> data augmentation skipped

Findings:

- Baseline model with RF best classification accuracy
- From scratch trained CNN: Minimization of loss and the accuracy vary among the batch_sizes.
  - -> Effect is not completely understood.
  - *Hypothesis: model optimization algorithm ADAM aims for solution with larger weights -> lower test accuracy*

5 Base script

Skript: https://github.com/hosts1/ethz-deepl-melanoma

References

1. https://www.kaggle.com/datasets/wanderdust/skin-lesion-analysis-toward-melanoma-detection/code.
2. https://tensorchiefs.github.io/dl_rcourse_2022/.
3. https://tensorflow.rstudio.com/.
4. https://tensorflow.rstudio.com/guides/keras/preprocessing_layers.
5. https://medium.com/mini-distill/effect-of-batch-size-on-training-dynamics-21c14f7a716e

10. October 2022