Prostate Cancer Recognition in MR-images with Keras Deep Learning

Research Project

Larissa Triess

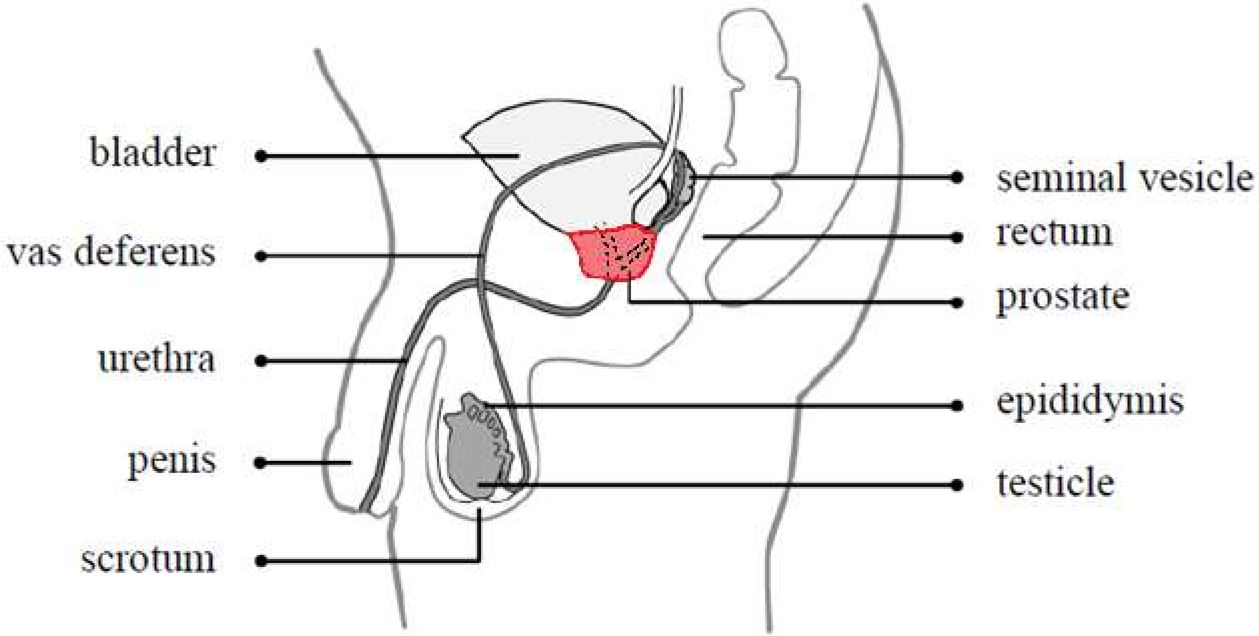
12/04/2018

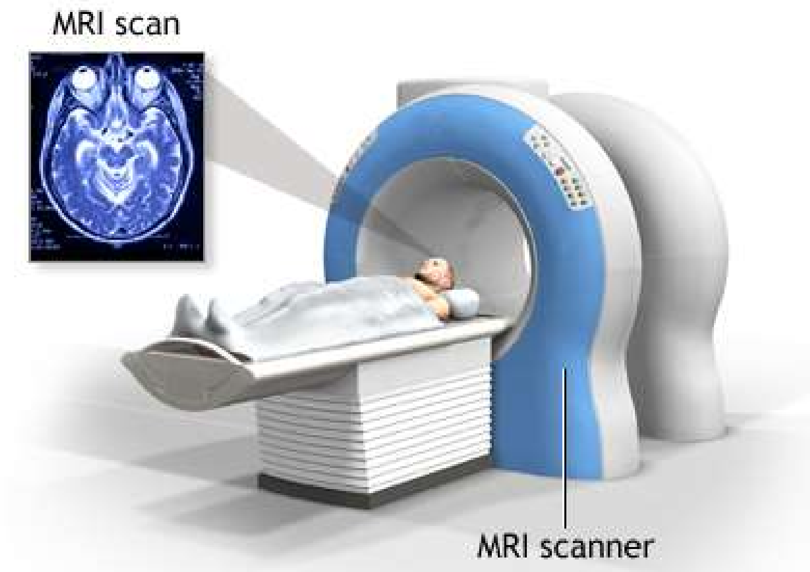
Contents

INTRODUCTION

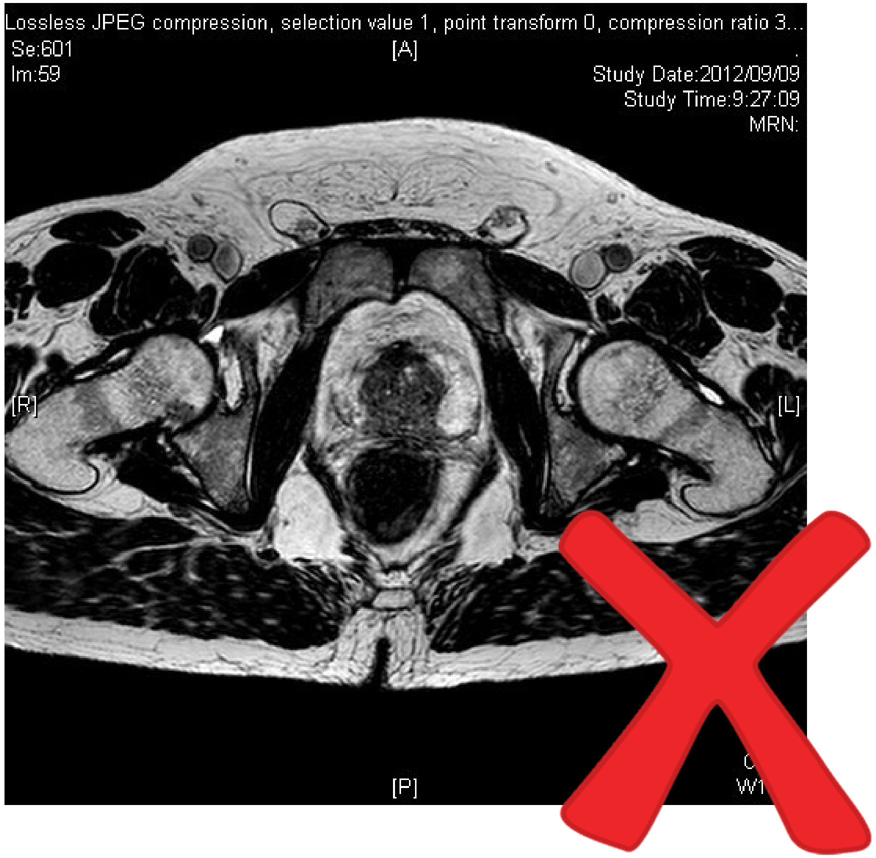
Motivation

- Prostate cancer is the second most common cancer in men
- Early detection is essential for successful treatment
- State-of-the-art imaging technique:
  Multi Resonance Imaging (MRI)
- Requires expertise of experienced radiologists
- Computer aided diagnosis can help agreement among doctors





Aim



- Classification into cancer and healty patient scans
- Usage of a simple neural network

Image Database

- 1465 images of 218 cancer patients
- 3460 images of 128 healthy patients
- Cancer annotations included

- Training: 65% ≙ 3216, validation: 15% ≙ 741, testing: 20% ≙ 968

Keras and Tensorflow

Keras

- Open-source neural network library
- Written in Python
- Runs on top of Tensorflow, Theano or CNTK
- Enables fast experimentation with deep neural networks

Tensorflow

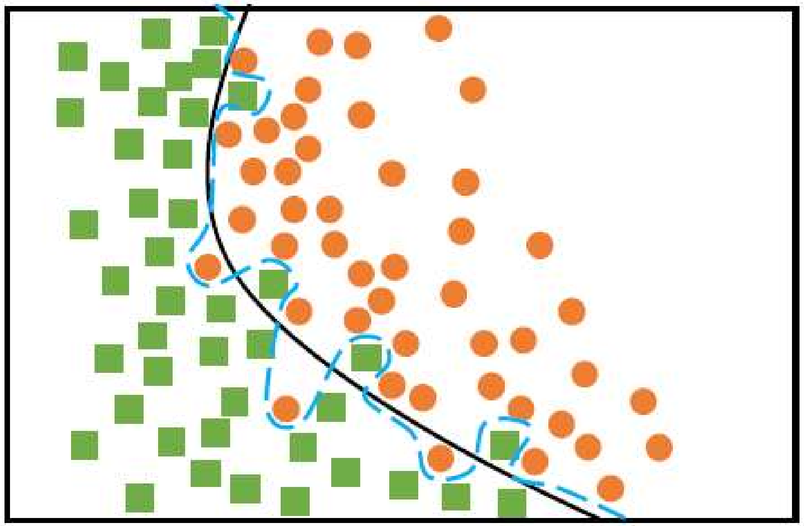
- Open-source software library developed by Google
- Can run on multiple CPUs and GPUs with optional CUDA extentions

MAIN PART

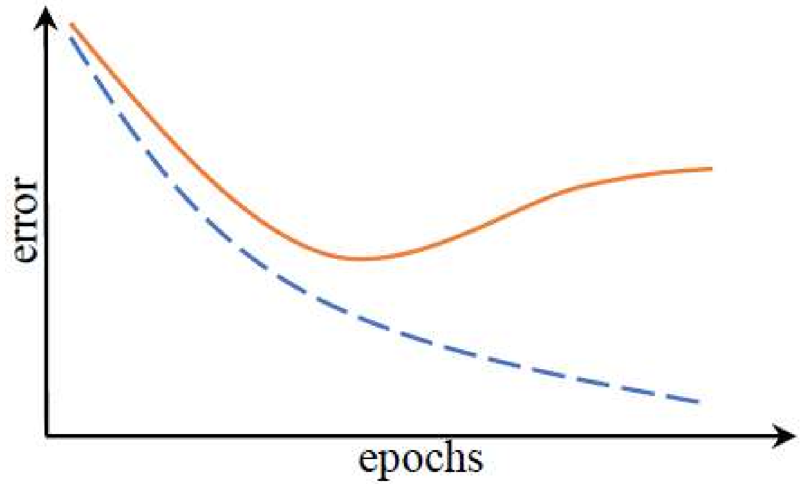
Network Architecture

Reasons for a small network

- Better understandable
- Easier to make adaptations
- Faster training times
- Less prone to overfitting ← main reason; what is overfitting?



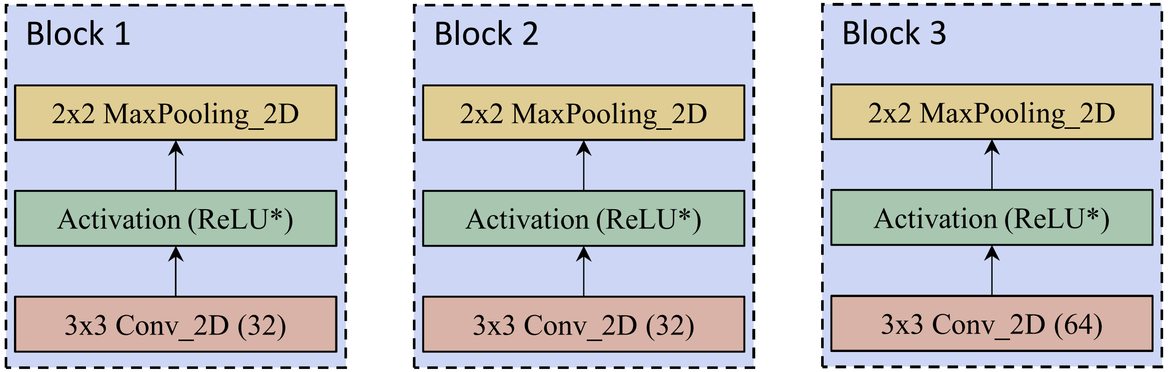
(a) Decision boundaries for an overfitted (dashed blue line) and an regularized model (black line)

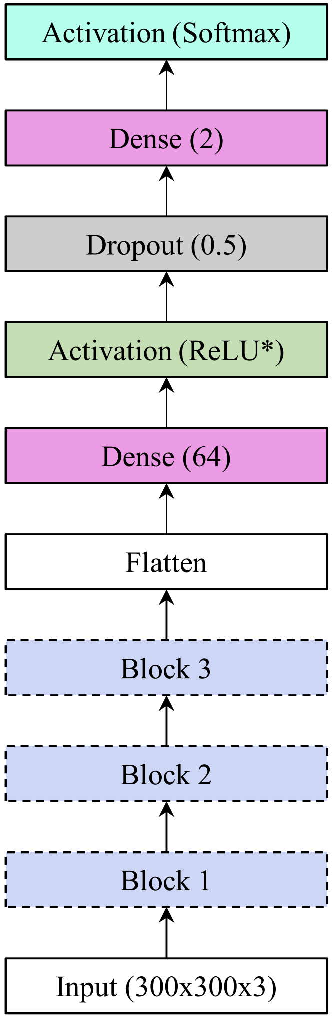


(b) Error rate for training (dashed blue) and test data (orange)

Network Architecture

- Stack of three convolutional layers followed by two fully-connected layers
- Compiled with RMSprop** optimizer and a learning rate of 0.001





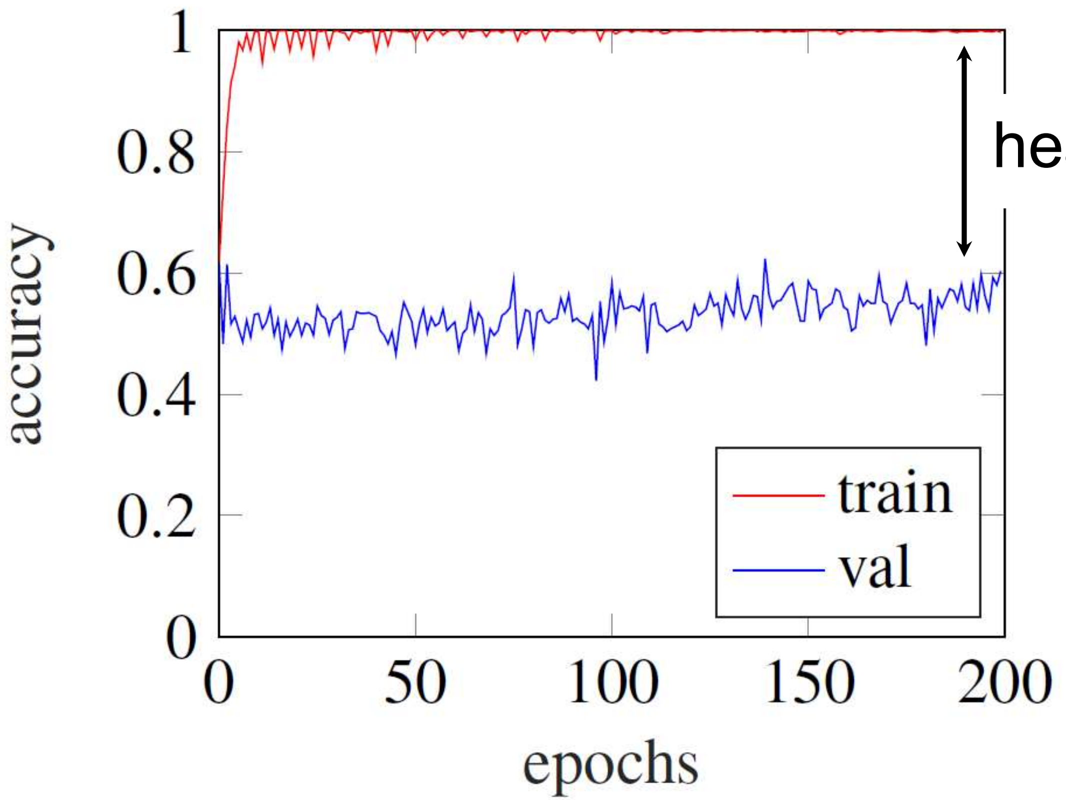
- Input: 300x300 pixel image in range [0, 1)
- Output: probability vector for the classes for each image (e.g. [0.34 0.66])

*Rectifier Linear Unit

**Root Mean Square Propagation

Experiment Setup and Baseline

- default parameter settings
- input images: 300x300 pixel sized crops
- 200 epochs with a batch size of 128
- RMSprop optimizer with a constant learning rate of 0.001

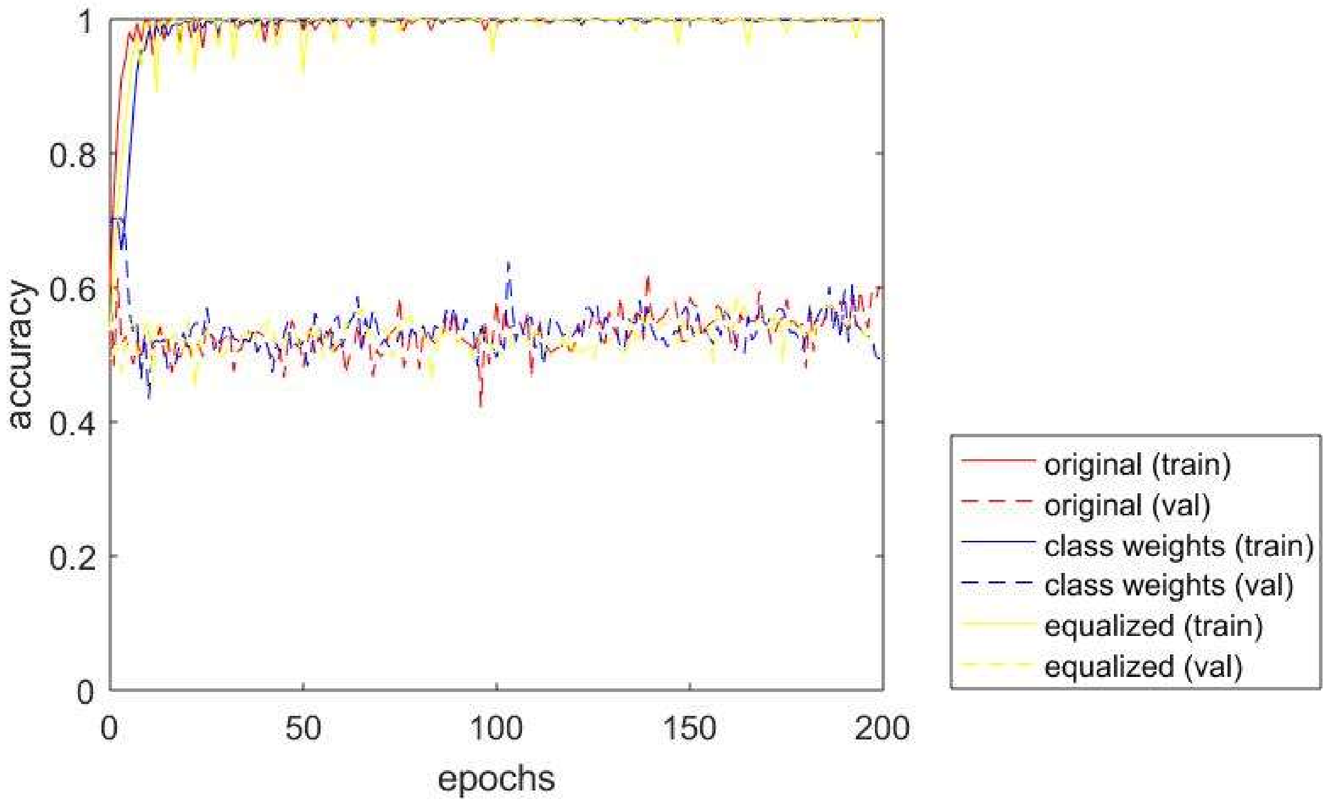


heavy overfitting

Experiment: Class weights

Dataset is not equally split into cancer and healthy cases, there exist three options to deal with it:

- Use original dataset
- Use original dataset, but set class weight dictionary for training
- Reduce dataset to equal size by discarding healthy images



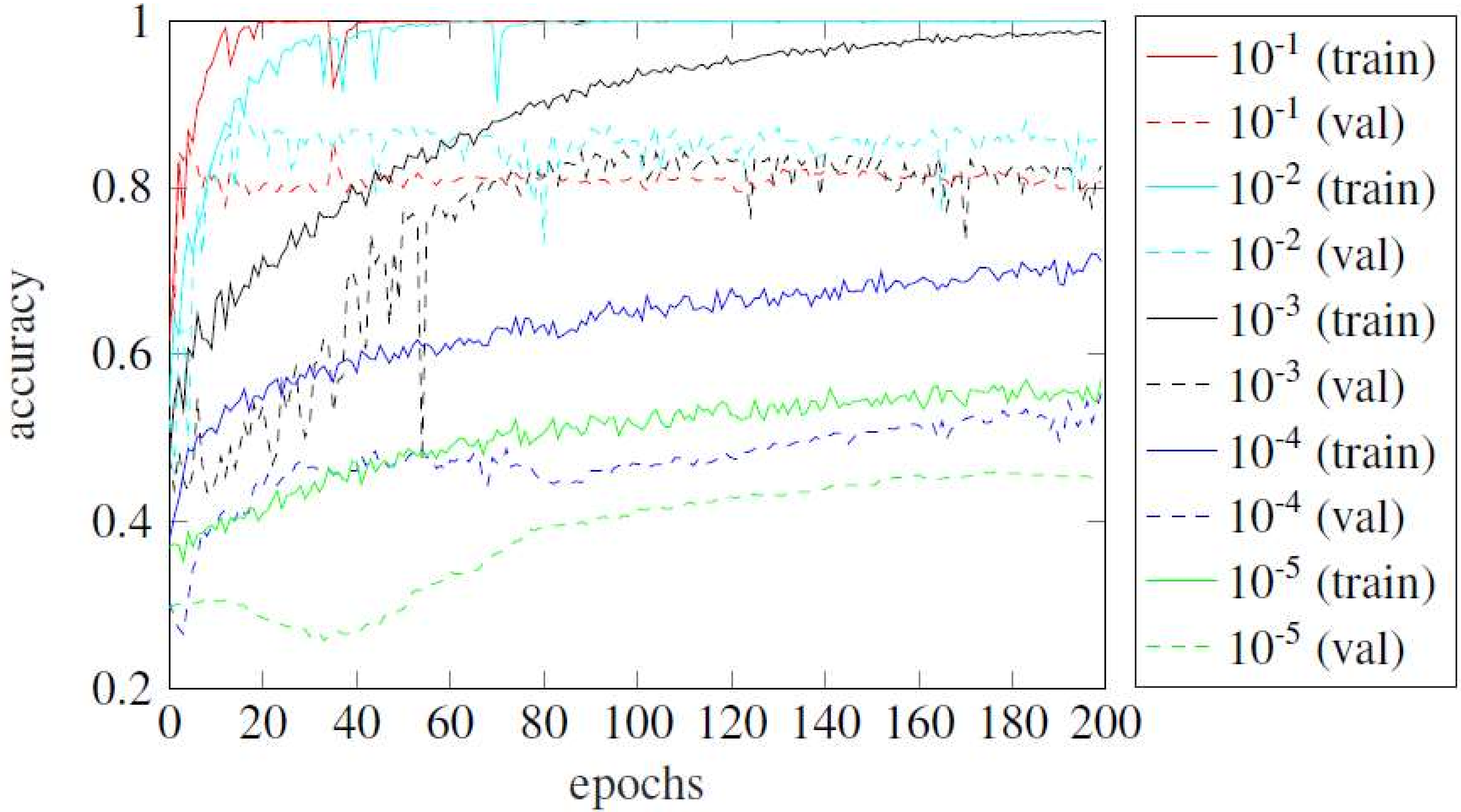
Experiment: Optimizer Type & Learning Rate

Three different optimizer types tested:

- Root Mean Squared Propagation (RMSprop)
- Stochastic Gradient Descent (SGD)
- Adaptive Moment Estimation (Adam)

Experiment: Optimizer Type & Learning Rate

Accuracy curves for training with SGD* optimizer with different learning rates (default: 0.01)



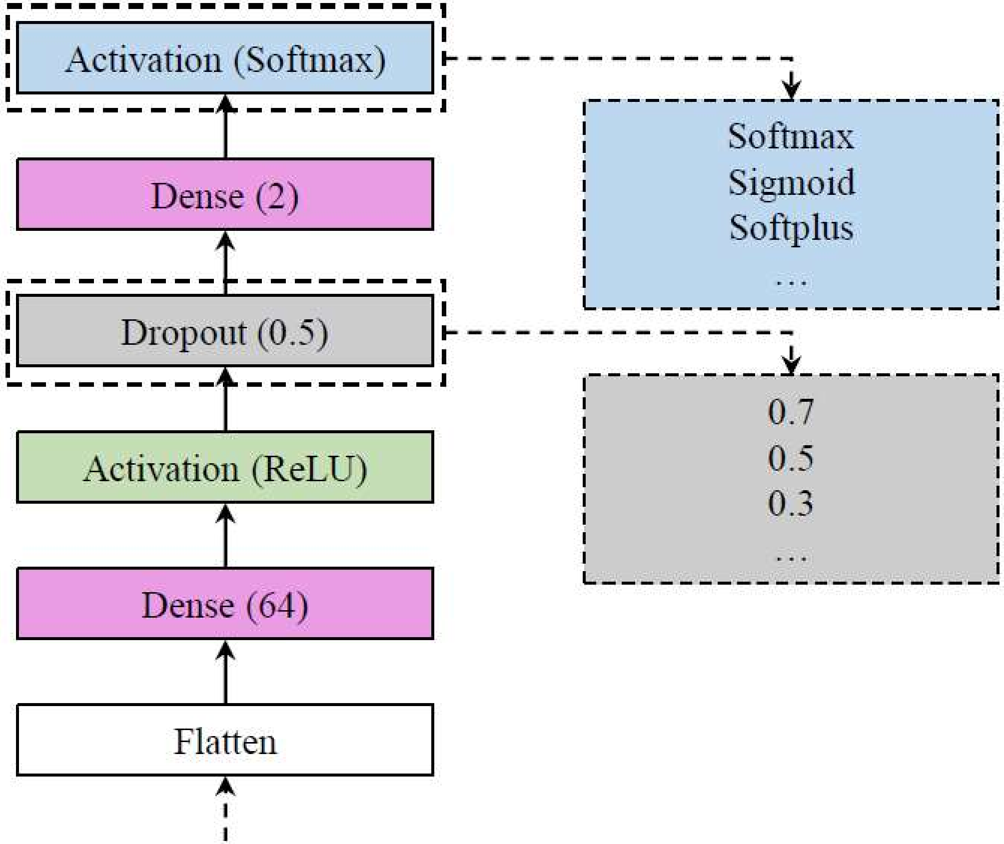
*Stochastic Gradient Descent

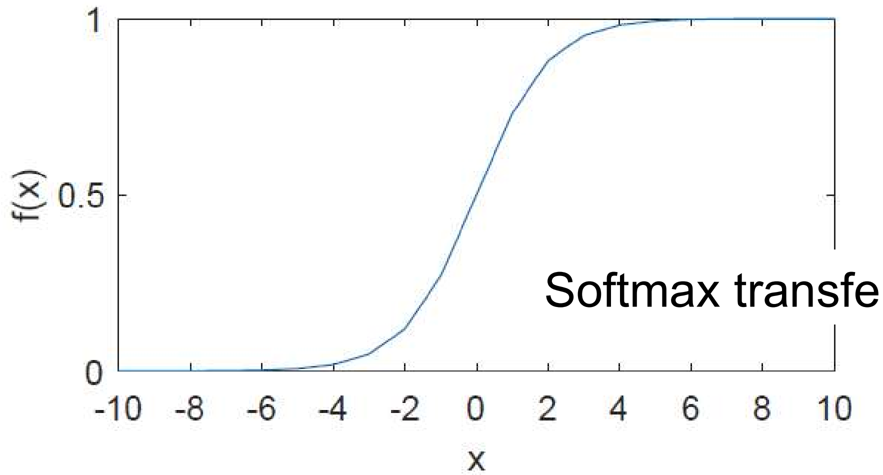
Experiment: Activation Function & Dropout

Which activation function do we change and where is the dropout applied?

Testing activations:

- Softmax
- Sigmoid
- Softplus



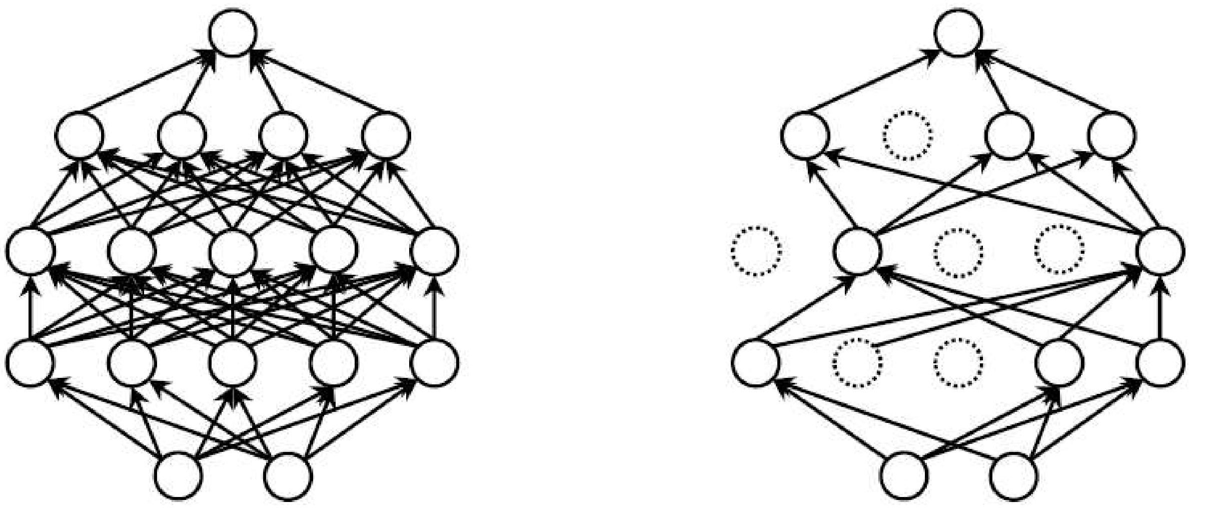


Softmax transfer funtion

Experiment: Activation Function & Dropout

What is dropout?

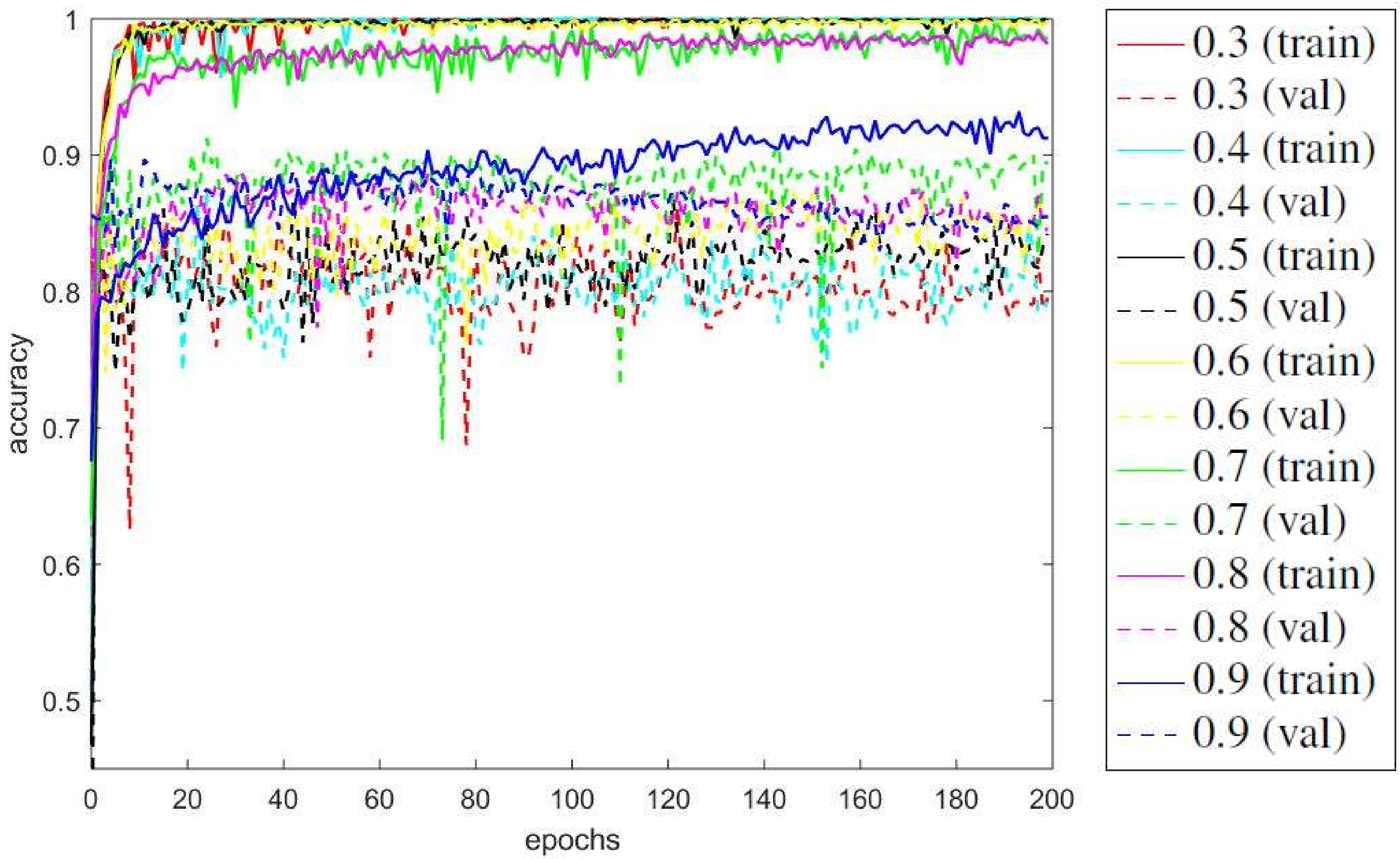
- Technique to reduce overfitting
- At each training stage individual nodes are „dropped out“ (temporarily removed)
- Prevents co-adaption
- Only reduced network is trained, afterwards removed nodes are reinserted with their original weights



(a) Standard neural network (b) Network after applying dropout

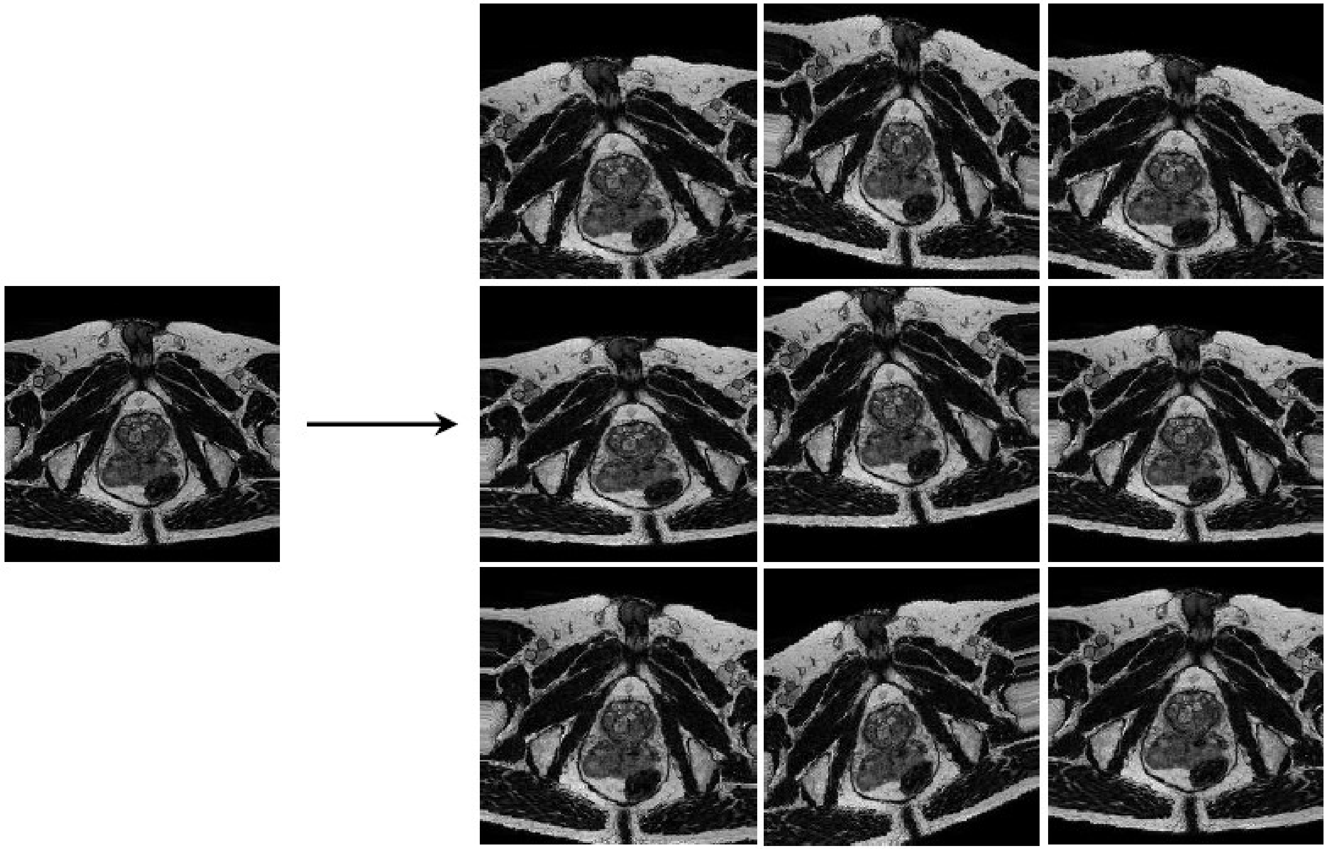
Experiment: Activation Function & Dropout

Training curve with sigmoid activation at the output layer and different dropout values



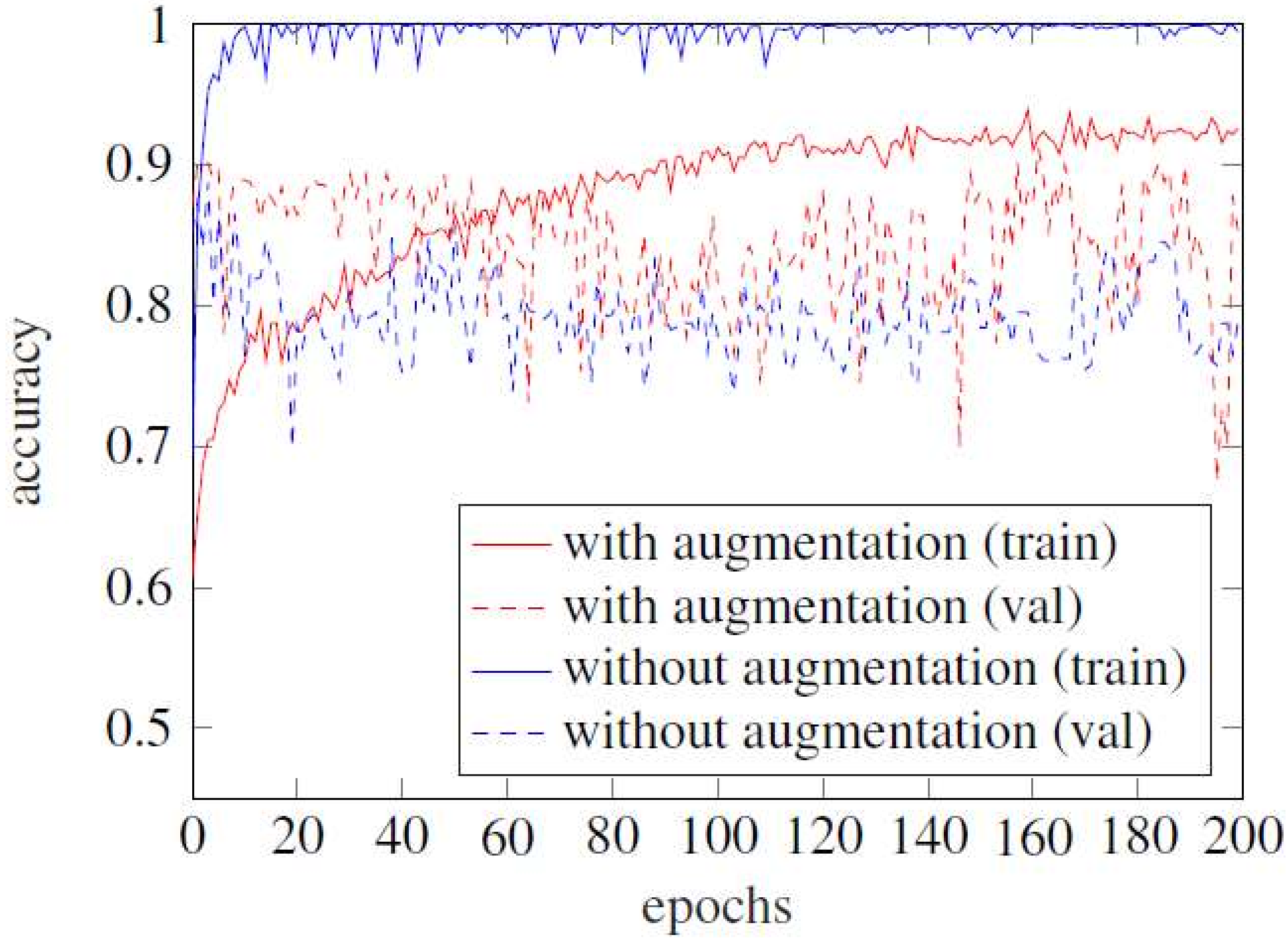
Experiment: Data Augmentation

- Useful tool in many image classification tasks
- Artificial enlargement of the dataset → chance for a better classification result with reduced overfitting
- Difficult for medical applications, as input data is usually more homogeneous than real world images



Experiment: Data Augmentation

- Slower increase of training accuracy with data augmentation
- Higher fluctuations from epoch to epoch with data augmentation
- Higher validation accuracy and reduced overfitting



SUMMARY

Conclusion

The initial network architecture obtained the best classification results with an accuracy of 97%

Even though using a small network with a high dropout, overfitting is quite strong, probably due to the relatively small dataset

Data augmentation can reduce overfitting, but reduces training accuracies

Outlook

- Larger dataset
- Use prostate segmentation as preprocessing step
- Include localization information of the cancer → extend network for cancer localization/segmentation

Thank you!

Questions?