Introduction to Focus Areas – Data Science – WS24/25
Lecturer: Katharina Jahn

Assignment 2

- Deadline Nov 4, 10:00 am
- Upload your notebook (Jupyter or R notebook) and a Pdf created from the notebook to the whiteboard
- State the group number and the names of all group members in the notebook
- If you create additional code for this project (not in notebook), share the files on your university GitLab account and state on your notebook where to find them and describe what the files do

The Data

In this project, we want to classify biomedical (microscopy) images, namely breast cancer histopathology images (malignant vs. benign).

The data is described in this paper: Spanhol, Fabio A., et al. "A dataset for breast cancer histopathological image classification." IEEE Transactions on Biomedical Engineering 63.7 (2015): 1455-1462. (www.inf.ufpr.br/lesoliveira/download/TBME-00608-2015-R2-preprint.pdf) and can be found here: https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/.

The entire dataset contains 7909 images of breast tumor tissue with different magnifying factors. However, you can restrict yourselves to use only one magnifying factor, e.g. 400X. You can also ignore the additional cancer types and focus on the labels "benign" vs "malignant".

The Tasks: Deep Learning for image classification

1. Read the paper: Spanhol, Fabio A., et al. "A dataset for breast cancer histopathological image classification." IEEE Transactions on Biomedical Engineering 63.7 (2015): 1455-1462. (www.inf.ufpr.br/lesoliveira/download/TBME-00608-2015-R2-preprint.pdf)
2. Implement at least three Deep Learning based classifiers that work directly on images and train them to detect malignant tumours. The classifiers should fulfill the following requirement: One needs to be a CNN, one can be a shallow neural network, e.g. a fully connected NN with one hidden layer.
3. Evaluate the performance of your classifiers using appropriate techniques.