

Patient or Tumor Reldentification in Microscopy Images

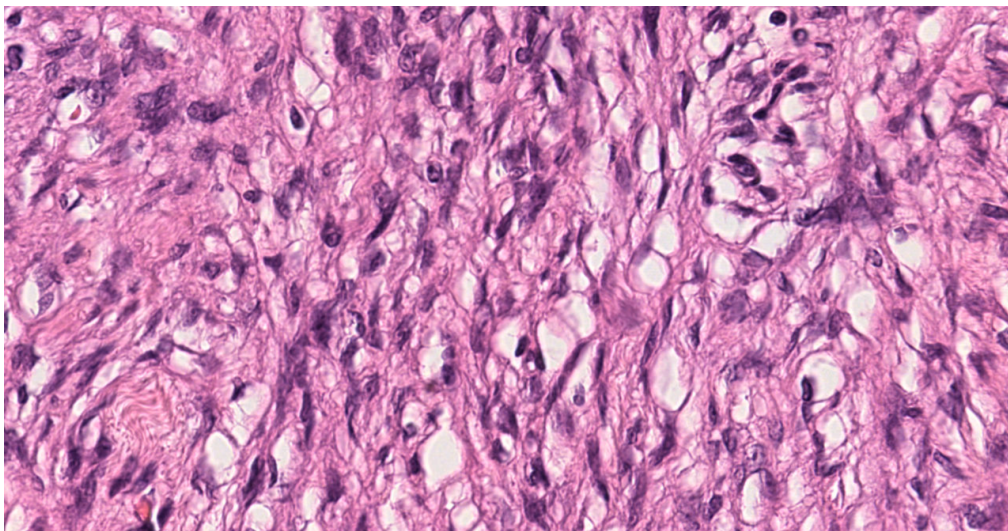
Supervisors: Jonathan Ganz, Marc Aubreville

The rise of deep learning techniques has increased the demand for publicly available large-scale data sets. Also many scientific journals increasingly require making the medical data sets available as a precondition to publishing a paper. However, medical data contains highly sensitive person-related information, which is commonly anonymized prior to publication. The medical image data, however, might still be attributed to a single person, even if the data stems from a large collection, such as recently shown in the field of x-ray images [1].

Another image domain where deep learning is increasingly used and data sets become more and more available is histopathology. For this, a thin slice of tissue, e.g. from a tumor, is cut, stained (colored) and placed under a digital microscope. Stained tissue reveals complex tissue patterns that contain rich information, many of which is unknown to us. It is therefore plausible that individual factors, or certain tissue or tumor signatures might be derived from tumor tissue. This might enable tracing from the tissue back to the patient, if another tissue sample is available.

The thesis shall investigate this question on a data set of more than 1000 microscopy images from around 330 patients that is available from a major German university hospital. It comprises the following tasks:

- Literature review of state-of-the-art deep learning techniques for weakly supervised whole slide image classification
- Design and implementation of a pipeline, preferably using python and torch
- Evaluation of the results
- Documentation of the code and the used methods



[1] K. Packhäuser et al.: Is Medical Chest X-ray Data Anonymous?

<https://arxiv.org/abs/2103.08562>

[2] He et al.: Integrating spatial gene expression and breast tumour morphology via deep learning

<https://www.nature.com/articles/s41551-020-0578-x.pdf>