Disentangled latent representations in computational pathology

Introduction

Histopathologic assessment is the primary mode to establish a diagnosis and to classify cancer. In this process, a tissue section is examined by a pathologist either using a microscope or a high-resolution whole-slide-image (WSI) on a screen. Despite its central role in the diagnostic workflow, histopathologic assessments can be subject to high inter-assessor and inter-laboratory variability, which can e.g. lead to under- and overtreatment. Computational pathology aims to assist and improve histopathologic diagnosis through automating routine tasks and developing tools that can provide novel insights using image analysis and machine learning tools. Comparable to many domains of image analysis, computational pathology has been revolutionized in recent years through the advent of deep learning and the increasing availability of large data sets. First deep-learning-based software products are currently receiving regulatory approval and are expected to have a significant impact on cancer diagnosis in pathology departments. However, differences in sample preparation and image acquisition and the populations in which models are developed and deployed can lead to spurious correlations and confounding. This can reduce the generalizability of model performance and therefore pose a significant obstacle to the clinical application of computational pathology tools.

Contact
Philippe Weitz
philippe.weitz@ki.se

Jun.-Prof. Dr.-Ing Johannes Stegmaier
johannes.stegmaier@lfb.rwth-aachen.de

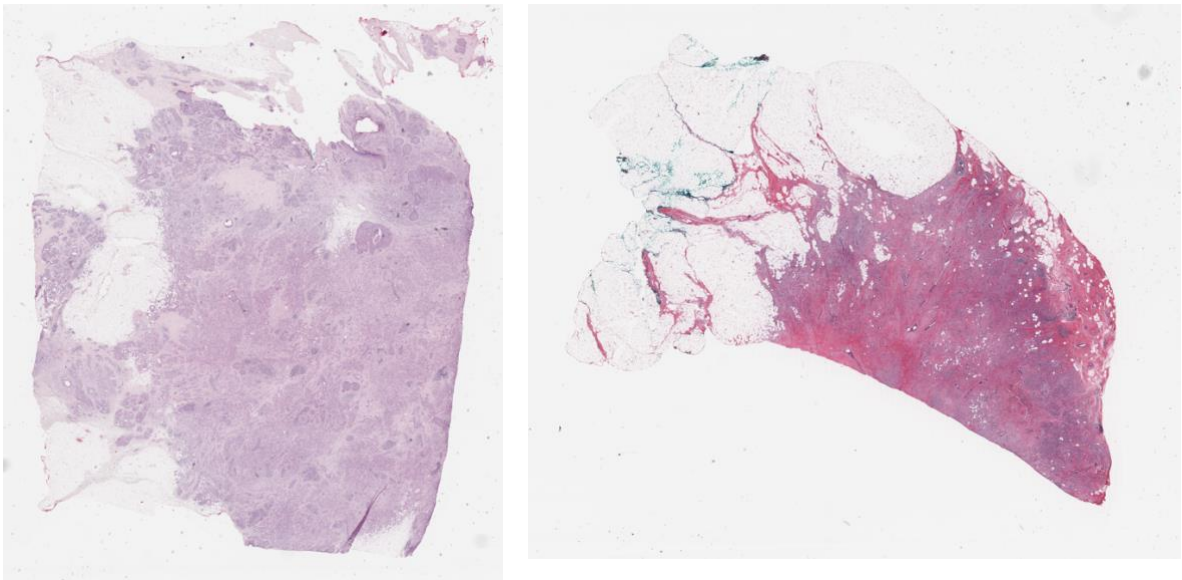

Figure 1: Example of color differences between two breast cancer WSIs from two different pathology departments. Images were obtained through the TCGA research network (https://www.cancer.gov/tcga).

Thesis Aim

A promising approach to these challenges is the application of deep learning models with disentangled latent spaces, in which specific features can be associated with independent (potentially causal) factors, such as stain intensity, luminosity or specific cellular characteristics and morphological patterns. The objective of this thesis project is to optimize existing convolutional neural network (CNN) model architectures and to perform a statistical analysis of the associations of latent space features, e.g. with clinical outcomes and known confounding factors, but also quantitative descriptors of cell morphology. If time allows, the effect of different image normalization techniques could be a further subject of study. We aim to publish results in a relevant technical or medical journal.

Intended Learning Outcomes

The student will become familiar with the theory of contemporary computer vision techniques, particularly deep learning and CNNs, and their application to the domain of histopathology images. However, the main challenges in the development and deployment of machine learning models are almost never exclusively of a technical nature but also related to good experimental/study design and a meaningful analysis and interpretation of results. This thesis project therefore also offers an opportunity to learn about basic methods of statistical inference that are commonly used in medical data analysis.

Requirements

- Interest both in medical image analysis and statistical data analysis
- Knowledge of machine learning theory
- Programming experience in Python
- Good spoken and written English language skills

Desirable:

- Experience with the Python packages PyTorch and Scikit-Learn
- Experience with GitHub

Supervision

The thesis is a joint M.Sc. thesis project between RWTH Aachen University and Karolinska Institute (KI), Stockholm. In Stockholm, the thesis will be supervised by Philippe Weitz, who is currently a PhD student in the Predictive Medicine group at the Department of Medical Epidemiology and Biostatistics (MEB) at KI and his supervisor Prof. Mattias Rantalainen. At RWTH Aachen University, Jun.-Prof. Johannes Stegmaier at the Institute of Imaging and Computer Vision will examine and grade the thesis. It is desirable that the student can spend some time in Stockholm but performing a part of the thesis remotely is possible.