Project Descriptions for Doctoral Research Positions 2022

ID01: Ensembling Experts for Improved Accuracy and Privacy in Predictive Models for Healthcare (Riezler, Durstewitz)

Faculty of Mathematics and Computer Science

Machine learning models for prediction in medicine are mostly based on time series of clinical measurements. Such models can be improved by adding patient-specific personal features like pre-existing medical conditions or personalized features extractable from unstructured data like anamnesis texts. In order to satisfy the data hunger of neural network models, the standard approach is to learn a unified prediction model on data pooled from several patients. Instead, we propose to learn predictive models on data that were labeled by experts for individual patients, and to combine patient-specific models by ensembling techniques. The advantages of ensembles include efficiency due to their parallel nature and improved predictive accuracy, based on the relationship of the generalization error ensembles to the correlation of the component models with each other. Since individual networks are often highly correlated, generalization accuracy can be optimized by searching for orthogonal models. Ensemble methods furthermore permit a simple mechanism to protect the privacy of personalized component models by applying noise perturbation in the ensemble combination process. Clearly, there is a tradeoff between between optimizing ensemble parameters for improved prediction accuracy versus strong protection of privacy. The goal of the proposed project is to find the optimal tradeoff between accuracy and privacy both from a theoretical perspective, and in experimental applications to predictive models for sepsis and diagnosis tasks from psychiatry.

Special requirements for applicants:
• Solid knowledge of machine learning theory and probability theory.
• Knowledge of current deep learning frameworks (e.g., PyTorch) and programming skills (e.g., Python).
• Experience in experimental data science.

ID02: Analysis Pipelines and Data Fusion for Cerebral Organoids (Mikut, Jung-Klawitter)

Faculty of Mechanical Engineering

Based on the definition of Qian, organoids are self-assembled three-dimensional aggregates generated from human pluripotent stem cells (hPSC) with cell types and cytoarchitectures that resemble human organs and tissues. Cerebral organoids are a growing field of interest to understand and model the development of the human brain for health and disease, in particular common and rare inherited diseases. Data of cerebral organoids can include heterogeneous data sources, such as longitudinal Magnetic Resonance Imaging (MRI), immunofluorescence analysis of organoids as well as metabolomic and transcriptomic data at bulk (whole organoids) or single cell level. A main bottleneck of organoid research is the lack of systematic analysis pipelines covering the full range of acquired data. The aim of the proposed project is to establish such a standard analysis pipeline for cerebral organoids and to show the potential of this pipeline for the characterization of human diseases, especially for inborn errors of neurotransmitter metabolism.

Special requirements for applicants:
We expect candidates with a background in machine learning including first experiences in image analysis using deep learning and an interest in biomedical data. Previous experiences with biomedical data (including sequencing data) would be an advantage.

ID03: Interactive Annotation of Volumetric Imaging Data Incorporating Report Information (Stiefelhagen, Kleesiek)

Faculty: Informatics; Field: Computer Vision / Machine Learning

Dense annotation is the underlying factor for success in many deep-learning-based approaches in recent years. The gathering process, however, of medical images is a difficult problem due to the inherent complexity of the data paired with the required expert knowledge. These issues result in a massive time requirement for real world data. Therefore, it is necessary to enhance current labeling tools through the use of active learning approaches to minimize the time and effort required by doctors to annotate complex radiological images. We will develop novel methods for interactive segmentation of medical data using online and active learning from sparse labels. We will also implement methods for speeding up the inference. Building on top of the developed fast interactive labeling approach, we will investigate whether such fast labeling tools will allow us to create personalized segmentation models for individual patients from scratch. We will compare the quality of such models to segmentation results obtained with pre-trained general segmentation models that have been adapted to the individual patients using the available annotated data. Further, the online segmentation methods will be extended to incorporate medical reports to ease and enhance the annotation process. Semantic information regarding disease and anatomical region will facilitate the annotation process and it will be determined how this additional information leads to speed and quality improvements. We will also explore ways to use the information extracted from medical reports to improve segmentation models by using a combination of textual and visual features. The resulting methods and tools will be evaluated with respect to speed versus accuracy as well as regarding the resulting decisions by the doctors and the resulting clinical outcome. The entire thesis project will be carried out in continuous interaction with physicians using whole-body 4D multi-modal PET/CT data in a clinical setting. The data is readily available and is already used for other projects. The code for methods and tools developed in the project will be made available to the scientific community.

Special requirements for applicants

- University degree (M.Sc.) with excellent grades in Computer Science or related fields
- Strong programming skills in at least one programming language (preferably Python and with experience in TensorFlow, PyTorch or similar)
- Good English language skills (your responsibilities include writing publications and giving international presentations)

ID04: Anomaly Detection in Sparse Image Time Series (K. Maier-Hein, Kleesiek)
Faculty: Informatics; Field: Computer Vision/Machine Learning

The project will advance the current state of the art in deep learning methods applicable to modeling entire patient image histories, with a specific focus on advancing change and anomaly detection in image time series. The specific challenge with patient histories is that they are usually irregularly and sparsely sampled on the time axis, often forcing methods into just paying attention to a snapshot of the patient's current condition. If successful, these developments will be an important step towards our ability to analyse and understand processes in healthy aging as well as disease progression. It will provide means to distinguish ordinary patterns from those that require special treatment or further diagnostics. And such models can be employed to extrapolate potential future disease trajectories, or to estimate the effect of measurement points that could be newly acquired.

Special requirements for applicants:
None

ID06: Model-based Artificial Intelligence in Surgical Data Science (L. Maier-Hein, Kenngott)
Faculty of Computer Science

Death within 30 days after surgery has recently been found to be the third-leading cause of death worldwide, with research suggesting that a large proportion of these deaths are due to surgical error. At the same time, recent technological advances in machine learning, especially deep learning, have begun to revolutionize various fields of medicine. As part of this development, the newly established domain of Surgical Data Science aims to improve the quality and outcome of interventional healthcare through the capturing, organizing, analyzing and modelling of data. However, as of yet, for various reasons, clinical translation of data science methods proves difficult. A large international consortium of experts recently attributed the lack of clinical success stories to the lack of large annotated databases. Data sparsity can thus be regarded as the main roadblock in the field of surgical data science. Previous approaches have addressed this roadblock with a diverse set of methods including crowdsourcing, self-supervised learning, active learning and synthetic data generation. In this project, we investigate an entirely complementary approach that is based on integrating existing medical knowledge in neural network-based analysis. Specifically, we propose the integration of prior knowledge encoded in ontologies in a Graph Neural Network (GNN) based-approach to surgical decision support. The method will be validated based on existing data sets from different surgical disciplines.

Special requirements for applicants:
- Well-founded programming skills in Python or C++
- An excellent Master's degree in the field of Informatics, Physics or Mathematics
- Experience in deep learning-based methodology especially Graph Neural Networks
- Excellent ability to work in teams and high intrinsic motivation

ID07: Physics-Based Reconstruction and Visualization for Improved Carbon Ion Radiotherapy (Sadlo, Jäkel)
Faculty of Mathematics and Computer Science

Carbon ion radiotherapy (CRT) provides highly precise and effective cancer treatment at accuracy better than 1 millimeter. However, this comes at the cost of high sensitivity of treatment quality to space-time geometry variations in the patient in the course of CRT, such as changes of tumor size, swelling, or weight gain/loss. Such variations can lead to unplanned irradiation of healthy tissue, or, on the other hand, insufficient dose in the tumor, with direct negative impact on treatment outcome and thus the quality of life of the patient. The group of life science PI Jäkel has been developing a non-invasive method for detecting such internal geometry changes in-vivo. The method is based on exploiting information carried by secondary radiation (nuclear fragments of the therapeutic carbon ions). Such radiation leaves the patient as a by-product. Using this type of radiation for quantitative imaging of the irradiated region would thus avoid additional radiation dose, as would be, e.g., deposited by additional computed tomography (CT) imaging. With exploiting single secondary-iontrajectories, the approach represents an entirely novel kind of medical imaging and is aimed to monitor the irradiation process in the patient. While the technical part of secondary ion detection and path reconstruction is on an advanced level, the extraction of the relevant in-formation on the geometry changes is still at a simple initial level. In particular, it suffers from high uncertainties due to the relatively low amount of recorded data and because only geo-metric information of the ion paths is taken into account. This leads to low signal-to-noise ratio due to spatial averaging, and therefore is limited to detection of relatively large changes. The main goal of this project is to address these weaknesses, with the aim to improve the safety and effectiveness of CRT. This shall be accomplished by (1) reducing uncertainty, (2) spatial reconstruction of the geometry changes, and (3) exploration of the involved (physical) processes. Uncertainty shall be reduced by including physically based (scattering) models and additional ion information, such as energy/velocity, as well as more robust evaluation of the raw sensor data. More reliable spatial information shall be obtained by employing distribution-based models that relate energy deposition to penetration depth with respect to initial energy and location of the primary ion beam. Exploration of the highly interrelated, superimposed, and high-dimensional space spanned by primary and secondary ion trajectories, CT data, energies, and uncertainties is achieved by a visual analytics approach providing interactive exploration. Our novel techniques will be evaluated with regularly taken CT scans of patients (ground truth) in the context of an ongoing observational clinical study on radiotherapy patients (proof-of-concept study), funded by the NCT, Heidelberg.

Special requirements for applicants:

- Master’s degree in computer science (or comparable)
- Advanced experience in software development,
- advanced programming skills in C++ and Python
- Background in calculus and linear algebra
- Preferably experience in visual computing and GPU programming

ID08: Using Anatomical Knowledge to Improve Medical Image Analysis (Stiefelhagen, Kleesiek)

Faculty: Informatics; Field: Computer Vision / Machine Learning

Current biomedical segmentation methods hardly incorporate common sense and established domain knowledge. Especially when reading radiological images, radiologists heavily rely on anatomical knowledge and have an imprinted memory of what is “normal”. In this project we will mimic this process. In prior work we have already created a data set for the 3D segmentation of anatomical structures in the upper body. This data set will be extended to span the entire body as well as additional imaging modalities (CT and MRI). We also will link to existing taxonomies and include anatomical textbook knowledge. Once an anatomical body representation has been established, we will evaluate and exploit this model for different medical image analysis tasks. For this purpose we will utilize two clinical data sets. One data set comprises whole body PET/CT images of prostate cancer patients with metastasis, the other stems from patients with pulmonary embolisms. Both data sets comprise corresponding radiological reports. This will allow us to investigate the performance of anatomical enhanced weakly supervised deep learning models to locate and characterize the pathologies described in the reports. Further, we will evaluate the models on image retrieval tasks, i.e. identify if a given radiological examination contains anatomical structures and where they are located within an image series.

Special requirements for applicants:

- University degree (M.Sc.) with excellent grades in Computer Science or related fields
- Strong programming skills in at least one programming language (preferably Python and with experience in TensorFlow, PyTorch or similar)
- Good English language skills (your responsibilities include writing publications and giving international presentations)