Project Proposals for Doctoral Researcher Positions 2021

ID01: Learning a World Model of Surgical Oncology for Cognitive Assistance Systems in Laparoscopic Surgery (Mathis-Ullrich, Wagner)

Doctorate at the faculty of Informatics

Minimally-invasive surgery is often carried out by tele manipulation of robotic actuators by the surgeon to reduce physical strain and increase dexterity by means of tremor filtration and image magnification. This opens the door to partial or even full automation of simple or repetitive surgical subtasks, with the promise of someday addressing the shortage of highly-trained surgeons and surgical assistants, especially in rural areas. State-of-the-art reinforcement learning (RL) algorithms have been shown to learn complex behaviours from interaction with an environment while maximizing a reward signal. However, they have low sample efficiency, and must be retrained from scratch for each new task and environment. A medical environment is particularly destructible and soft by nature, effectively precluding training an agent for robotic surgery in a real environment. We aim to overcome both limitations (low sample efficiency and difficult transferability) using the method of world models, and successfully transfer an RL policy trained to perform simple surgical tasks in simulation, into the real world surgical application. Model-based RLexhibits significantly higher sample efficiency than other methods as an agent can exploit its understanding of the environment dynamics to plan ahead. World models, a recent class of methods, extend this by learning and planning in a latent space rather than relying on raw pixel observations, allowing for efficient imagining of trajectories many time steps into the future. Because the learned model is not just simply end-to-end, certain elements of the policy can be transferred to the real world with minimal fine-tuning. Our goal is to extend world models to the unique problem of robotic surgery, and develop new methods in transfer learning to bring these benefits into the real world. The proposed project will support our vision of surgery & intervention 4.0 where a single surgical team may simultaneously supervise several remote surgeries, with all but the most critical maneuvers being carried out autonomously by cognitive surgical robots.

ID02: Supervised Machine Learning to Predict Radical Transfer Mechanisms across Collagen Genetic Disorders (SMaRT) (Friederich, Gräter)

Doctorate at the faculty of Informatics

Collagen is the most abundant protein of our body, and plays a critical role in a wide range of diseases caused by >1,000 mutations identified thus far across collagen types. However, mechanistic details on a microscopic level are to date not well understood. In this project, we will develop and harness machine learning enhanced simulation methods to predict the impact of collagen-related genetic disorders, in particular osteogenesis imperfecta, on collagen mechanoradicals. We will systematically uncover the consequences of disease mutants on impaired collagen mechanics and antioxidant defense. To this end, we will develop two complementary methods to predict and analyze hydrogen transfer reactions in collagen tissues on anatomistic level, and validate those by electron-paramagnetic resonance spectroscopy experiments. The first approach is based on a direct prediction of reaction barriers using supervised machine learning regression models based on informative representations

of the local chemical environment in which the reaction takes place. Main challenges lie in the data collection in collaboration with the Gräter lab and in the implementation of problem specific and interpretable machine learning models in the Friederich lab. In the second approach, the reaction is explicitly modeled using machine-learned potentials for molecular dynamics simulations. Such potentials are actively developed world-wide with focus on vastly different application areas in materials science. However, to apply them to reactive biological systems remains a challenge. Understanding hydrogen transfer will allow us to uncover relations between genes and corresponding chemical composition of tissues, radicals in the tissue, and resulting symptoms of collagen-related diseases.

ID03: A Network-based Approach in Support of Comprehensive Data Exploration and Informed Decision Making for Tumor Boards (Gertz, Kleesiek)

Doctorate at the faculty of Mathematics and Computer Science

Tumor Boards (TB) are interdisciplinary meetings in which medical specialists work together to develop clinical decisions in cancer care. Currently, the preparation of these meetings, i.e., gathering medical information regarding a patient's history, as well as the meetings themselves have a high "analogue" character: The information is often gathered in the form of handwritten notes, manually extracted from different clinical systems and there is no coherent visual presentation of the data leading to a decision. Also, the decision itself, the recommendation of the TB, is often only captured as a free text report.

Structured data, e.g., lab values, can be accessed and visualized easily. Information hidden in unstructured data, e.g., free text reports, are much harder to access. Unfortunately, the documentation of therapy regimes and the response of the patients to their therapy are often not available in structured format. On the other hand, the combination of the multi-modal data sources is particularly valuable for making informed decisions, for developing predictive models and for identifying patients with similar disease courses – potentially improving decision making in cancer care.

In this project, we develop and realize a network-based approach to personalized medicine in the context of tumor boards. Heterogeneous and distributed forms of medical data collected for a patient are continuously integrated into an information network that allows users to monitor, visualize, analyze, and explore a patient's up-to-date history recorded as multi-modal data in a context-sensitive manner. Nodes are features automatically extracted from the data sets and edges specify weighted semantic relationships between features. Such networks do not only provide a comprehensive dynamic view of a patient's state but the underlying network model allows for comparing the history of a cohort of patients, determining similar behaviors, e.g., under treatment, and thus allows for more informed decisions by knowledge sharing. A respective framework ranging from data integration components up to a dashboard used by tumor board members will be developed and evaluated in this project.

ID05: INNpin: Invertible Neural Networks for Predicting molecular Interactions (Rother, Köthe, Wade)

Doctorate at the faculty of Computer Science

Molecular dynamics and complexation are essential properties for biomolecular function and can be affected by mutations, such as those due to genetic variations. The high degree of flexibility and many ways in which biomolecules can interact make predicting such interactions a very challenging problem. Molecular dynamics simulation techniques can be used to sample molecular structures and complexation events. However, it is often difficult, or even computationally infeasible, to fully sample the relevant degrees of freedom, and predict the effects of sequence differences or mutations. Furthermore, the models of the interactions, the force fields, may be of insufficient accuracy. Machine learning approaches offer the possibility to overcome some of these problems. Here, we will investigate the use of invertible neural networks (INN) for this purpose. These state-of-the-art networks can, after training, generate new members of an ensemble of interest. In this project, we aim to develop INNs for predicting the structures of molecular complexes and the effects on these complexes of variation in sequence. We will define model systems for developing the methodology and then apply the methods to systems such as protein-peptide complexes of cardiological importance.

ID06: Data-driven Gamification to Improve Quality in Medical Image Annotation Tasks (GaMeIT) (Sunyaev, Wagner)

Doctorate at the Faculty of Economics and Management, Information Systems

Cognitive surgical assistance systems, such as surgical robots require image-based scene understanding to perceive the surgery context, comprehend the surgery procedure, and eventually generate safe trajectories to assist during the surgery. To achieve such scene understanding, recognition and semantic segmentation of different surgery aspects (e.g., shown organs, used surgical tools, different surgery stages) are necessary pre-conditions. Machine Learning (ML) approaches are a promising technology for semantic segmentation of images. To train robots with ML methods, annotated image data (e.g. in the form of videos) is required. Image annotation of surgical images and videos is often manually conducted by healthcare professionals. This is necessary since a certain level of medical expertise is required. The process of manual annotation is prone to human errors since it can be tedious, monotonous, and exhausting. As a consequence, poor label quality is a common problem. However, for surgical robots to improve surgical procedures, sufficient data quality of annotated images is a decisive factor. If ML models for surgical robots are trained based on poorly labeled image data, this may negatively influence patients' health since the robots cannot be utilized to their full potential. In this project, we address the problem of poor label quality of surgical image data by augmenting the annotation process with persuasive technology. In particular, we design, implement, and evaluate a data driven ML-based gamification concept to foster annotators engagement and, thereby, ensure high quality data labeling. By drawing on ML methods, the gamification concept is able to adapt to individual user preferences and overcome the weaknesses of one-size-fits-all gamification approaches.

ID07: Inverse Radiotherapy Treatment Planning using Machine Learning Outcome Prediction Models (Frank, Jäkel)

Doctorate at the Department of Mathematics

Improving the personalization within the planning of a radiotherapy intervention is a primary goal in treatment planning research. One of the major challenges lies in the non-trivial relationship between applied/planned dose and therapeutic outcome, which is bypassed by using surrogate planning objectives depending on empirical dose prescriptions and tolerances. Existing attempts that directly use tumor control probability (TCP) and normal-tissue complication probability (NTCP) in planning show technical feasibility. They rely, however, on simple low-parametric and often univariate models and have been consistently questioned with regards to accuracy and usability in treatment planning. The proposed project aims at superseding these approaches by incorporating state-of-the-art Machine Learning (ML) models for TCP & NTCP into the radiotherapy treatment plan optimization process. Such an integration will face multiple mathematical, computational, and clinical challenges. First, the radiotherapy optimization environment challenges the ML models with respect to stability (i.e., generalization error and respective uncertainty) due to the untypical input given during optimization iterations. Second, fast forward computation and efficient gradient computations/approximations of model-based objective and constraint functions are key to a feasible integration. Finally, one needs to make sure that the resulting treatment plans fulfill clinical acceptance criteria. The project will address these research questions building on substantial previous work on radiotherapy planning software (matRad, open-source) as well as an available Machine Learning outcome prediction framework and respective head-and-neck patient data. The final result will consist of a functioning, open-source software prototype for ML-based treatment planning along with various strategies that facilitate generation of clinically acceptable treatment plans.

ID08: Xeno-Learning for Spectral Image Interpretation in Surgery (L. Maier-Hein, Nickel)

Doctorate at the faculty of Mathematics and Computer Science

Death within 30 days after surgery has recently been found to be the third-leading cause of death worldwide. One of the major challenges faced by the surgeons is the visual discrimination of tissues, for example to distinguish pathologies or critical structures from healthy remaining tissue. To overcome the limitations of visual perception, multispectral imaging (MSI) has been proposed. While conventional medical cameras (e.g. laparoscopic imaging systems) are limited by “imitating” the human eye; multispectral/hyperspectral cameras remove this arbitrary restriction of recording only red, green and blue colors. Instead, they capture multiple specific bands of light that decode relevant information on tissue type and function. The bottleneck related to converting the potential of this novel imaging technique into patient benefit is related to the lack of large annotated data sets needed to train algorithms for clinically relevant tasks. In this project, we aim to address this issue based on a new concept that we refer to as xeno learning. The term has been inspired by the concept of xenotransplantation, which refers to the transplantation of living cells, tissues or organs from one species to another. Here, the shortage of transplantable human organs is addressed by using organs from other species while this project aims to address the shortage of data by transferring knowledge from one species (specifically porcines) to another (here: humans). The core methodological data

science challenge is to develop a data representation that enables the generalization not only to new individuals but also to other species. To this end, we can leverage a huge data set consisting of hyperspectral imaging data comprising several thousand images of animals and humans.

ID09: The Virtual Fridge: Interactive Visualization and Analysis of Biomolecular Dynamics Spanning from Individual Experiments to Large Ensembles (Dachsbacher, Strähle)

Doctorate at the faculty of Computer Sciences

Living organisms rely on many tightly interconnected biomolecular interactions. Deeper understanding of these enables many important applications, e.g. in pharmaceutical and toxicological tests or to understand genetic effects. Using light-sheet and optical fluorescence microscopy, Prof. Strähle studies biomolecular dynamics and their disturbance by genetic as well as physical and chemical influences using the embryonic development of zebrafish (Danio rerio) as a preferred model animal for vertebrates. The goal is to obtain quantitative descriptions of cell death, divisions, movements, interactions, and morphological and molecular changes and physiological states during body homeostasis and in disease models. The data volumes from such experiments are tremendously large: A single experiment produces thousands of gigabyte-sized 3D images (multimodal 3D+t data). Furthermore, experiments are continuously on-going to obtain ensemble data, i.e. data from experiments under similar conditions. While automatic data analysis became an important pillar, visual inspection by domain experts is still important. Not only tasks such as sense making and reasoning, as well as validation and parameter tuning, still require domain expert input. The complexity and variation between individual experiments make guided semi-automation necessary. To date all this is only feasible for data from a single or few experiments. Techniques for a combined interactive exploration of 3D+tdataand derived information (e.g. cell trajectories), ranging from ensemble data summarized with few statistics down to every single piece of experimental data, do not yet exist. Additionally, this requires a suitable visual interface that combines efficient interaction and exploration methods, i.e. optimized for human perception, together with automatic data processing. In this project we will tackle this challenge and develop novel visualization and data analysis techniques.

ID10: Temporal and Global Consistenc Enforcing Segmentation for Real World Radiological Applications (K. Maier-Hein, Kleesiek)

Doctorate at the faculty of Computer Sciences

Within the recent years, especially after the introduction of the U-Net architecture, segmentation accuracies in the biomedical imaging domain have made a great leap forward. Prominent and well-studied applications within the medical field are the segmentation of physiological or pathological structures within radiological images. Given valid and high-quality ground truth segmentations, controlling for data quality, preprocessing of data and incorporating knowledge about the imaging techniques, these networks can be trained leading to impressive dice scores comparable to or even surpassing human inter- and intra-rater variability [1, 2]. Yet, bringing these models into a real clinical environment comes with unexpected headwinds. Frequently the models fail with severe and

unexpected errors that even non-expert humans never would make. As a result, trust in the methods is lost and they are not applied by target group, i.e. radiologist, in real world clinical scenarios.

Even if longitudinal imaging information is available, current methods do only take images of a single time point into account during training and inference. Furthermore, statistics about spatial configuration, volume and signal intensities of lesions, i.e. global information, is usually not considered by algorithms. Thus, incorporating temporal and global information, as also radiologists do when reading imaging studies, will most likely boost the models and lead to better and more robust results.

Within this project we want to address these shortcomings and improve and extend existing models. To obtain robust, consistent and well generalized models the proposed project is embedded in a bi-centric setting comprising the university hospitals of Heidelberg and Essen. As specific use cases two diseases have been identified that usually go along with frequent imaging during their course and for which abundant data is available at both sites: multiple sclerosis (MS) and brain tumors.

ID11: Comparative Visual Analysis of Tissue and Cancer Growth in Simulations and Experimental Data (Dachsbacher, Schug)

Doctorate at the faculty of Computer Sciences

In this project, we focus on the visualization of complex voxel-based time-dependent images ("3D+t") in biomedical contexts where data obtained from both imaging devices and simulations is to be analyzed in a comparative approach. In particular we focus on visualizing and further developing complex high-resolution simulations of tissue and cancer growth by comparing them to data acquired from real samples and experiments. Available standard techniques for visual analysis are not sufficient and multiple challenges need to be overcome. We will develop new visualization techniques for exploring the multivariate 3D+t data, helping the domain experts to reveal and examine, e.g., correlations and outliers, from which we expect a better understanding of the growth processes. We further plan to apply findings from the mathematical theory of persistent homology, which enables us to examine and compare structures on the basis of their coherence or topology. Noise or smaller irregularities can be removed effectively and in a structured manner in order to segment the structures on a suitable level of abstraction. This is particularly important in our target application, as we want to find a suitable level of abstraction for comparing simulations and experimental data. Furthermore, the quantitative comparison of data and their persistence diagrams are subject of research.

ID12: Data science approaches to evaluate and correct bias in single-cell transcriptomics (Brors, Kaster)

Doctorate at the faculty of Computer Sciences/Biosciences

Single-cell genomics (SCG) and transcriptomics (SCT) can provide reliable context for assembled genome fragments and gene expression activity on the level of individual prokaryotic genomes. These methods are rapidly emerging as an essential complement to cultivation-based, metagenomics, metatranscriptomics research approaches by allowing direct access to information

from individual organisms rather than bulk samples. In recent years it became evident that the transcriptomic profile of individual cells can vary even if they are genetically homogenous. However, the current methodology of SCG and SCT requires an amplification step before sequencing which might introduce a bias to the data set and hamper with the actual expression profiles. To date, no systematic approach has been taken to evaluate this bias and correct for it. This projects aims at addressing this issue, thereby making a valuable contribution in the field of bioinformatics omics analyses.

ID13: Leveraging large-scale single-cell datasets for personalized cancer cell-of-origin inference (Stegle, Lutsik)

Doctorate at the faculty of Biosciences, Bioinformatics

Cancer is a highly heterogeneous disease that arises through a series of (epi) genomic mutations. A well-studied dimension of this heterogeneity are somatic mutations in the tumor. Increasingly, it is also clear that the cell-of-origin of the tumor as well as the genetic background of the patient play a key role in shaping cancer heterogeneity, with impact on prognosis and treatment. However, this source of heterogeneity is currently poorly understood. In this project, we will develop machine learning strategies for deciphering cell-of-origin related cancer heterogeneity. Building on variational auto encoders and related dimensionality reduction methods, we will derive generative representations of population-scale single cell data from thousands of individuals. Leveraging these trained models, we will then interpret disease sample sin a reference latent space, thereby enabling patient/cancer specific cell-of-origin identification at unpresented resolution. We will apply these methods to blood cancer sand specifically Acute Myeloid Leukemia (AML) and Chronic Lymphocytic Leukemia (CLL), for which we have access to pertinent data resources. Collectively, the analytics derived here combined with available data, will provide novel avenues for characterizing cancers by accounting for their cell-of-origin. The ultimate goal is the generation of model-driven insights for clinical guidance, enabling early detection, as well as patient-specific strategies for treatment and disease management.

ID14: Visual Data Analysis of Geometry Changes in Carbon Ion Radiotherapy (Sadlo, Jäkel)

Doctorate at the faculty of Mathematics and Computer Sciences

Carbon ion radiotherapy (CRT) provides highly precise cancer treatments, allowing for accuracy better than 1 millimeter. This comes, however, at the cost of high sensitivity of the treatment quality to geometry variations in the treated patient, such as changes of the tumor size in the course of the CRT, swelling, or weight gain/loss. This can lead to unplanned irradiation of healthy tissue, or, on the other hand, insufficient dose in the tumor, with a direct negative impact on the treatment outcome and thus the patient's quality of life. In the department of PI Jäkel, a method has been developed for detecting such internal geometry changes in-vivo in a non-invasive way. It exploits the information carried by secondary radiation (paths of secondary ions caused by the carbon ions), which leaves the irradiated patient as a by-product of the irradiation, and thus avoids additional dose, as would be induced by, e.g., additional computed tomography (CT) scanning. In the current approach, measurement of single ion trajectories, emitted during therapy, are used to monitor the carbon ion irradiation process in the

patient, representing an entirely new kind of medical imaging. However, one of the main challenges in this approach is the low amount of secondary ion data and the resulting high uncertainties. The data analysis, being a crucial component of the method, is up to now based on averaged geometric evaluation of the paths of the secondary ions. It provides a low-signal detection of geometry changes, but due to the spatial averaging, it is not able to provide quantitative nor qualitative insight into the geometry changes in the patient. The central goal of this project is to develop novel data analysis techniques to quantitatively and qualitatively evaluate the ion trajectory information. This shall (i) provide reliable information on the quality of a treatment in terms of sufficient and sufficiently confined irradiation, and (ii) give qualitative spatial information on the geometry changes. The final aim is to achieve an improved detectability of geometry variations in the treated patients (larger significance). Overall, this may allow for reduction of healthy tissue damage and increased tumor control in carbon ion radiotherapy, leading to an increased quality of life for oncologic patients. Our novel technique will be evaluated with CT-images of the patients in the context of a study, taken regularly in the course of the radiotherapy. The data for our project will be obtained as part of an observational study on radiotherapy patients (Proof-of-Concept study), which is funded by the NCT, Heidelberg.