ID01:

Statistical Methods for Infectious Disease Outbreak Detection in a Post-Pandemic World

Data Science supervisor: Prof. Dr. Melanie Schienle

Life Science supervisor: Junior-Prof. Dr. Johannes Bracher

Infectious disease outbreaks can have a considerable impact on public health. Their timely detection helps to contain disease spread and reduce harmful consequences. Real-time monitoring is thus a key challenge for public health agencies, which requires robust algorithms, combined with expertise on epidemiology and data collection. A rich statistical literature and standard procedures like the Farrington algorithm for outbreak detection exist. Nonetheless, exchange with public health experts indicates a substantial gap between current methods and applied requirements. We will tackle three challenges, which we identified with collaborators at RKI and while supporting WHO guideline development on influenza surveillance.

(1) Standard outbreak detection algorithms do not account for delayed reporting, which can hamper sensitivity and timeliness. We will develop an adjustment scheme which is agnostic to the detection algorithm, computationally fast, interpretable and easy to communicate.

(2) Outbreak detection is usually conceived as constructing alarm limits for single time series, accepting a certain rate of false alarms. In practice, however, automated detection across many time series serves to direct limited human resources to patterns worth of investigation. Rather than binary flags, rankings of potential outbreaks by relevance are thus needed. We will conceive and compare ranking methods, and develop methods to assess their quality. (3) The COVID-19 pandemic has demonstrated a need to adjust outbreak detection algorithms after external shocks. We will perform a systematic analysis of changes in seasonal patterns across a comprehensive range of pathogens monitored by RKI. Based on these, we will develop and validate suitable extensions to outbreak detection techniques.

These challenges will be addressed in cooperation between KIT, Uni HD and collaborators from RKI, ensuring a close link between methods development and practical needs.

ID02:

Patient-Specific Physics-Based Virtual Twin for Neurosurgical Simulation Using FEM and Data-Driven Multi-Tissue Modeling

Data Science supervisor: Prof. Dr. Maria Francesca Spadea

Life Science supervisor: Prof. Dr. Uwe Spetzger

Microsurgical dexterity is essential across multiple surgical disciplines, yet reduced training hours and limited operative exposure have created a critical gap in surgical education. While simulation-based training offers a safe alternative to operating room learning, most existing platforms lack haptic feedback, patient-specific anatomical accuracy, and realistic fluid dynamics.

This project proposes a cost- and time-efficient workflow for developing patient-specific augmented-reality surgical simulator that integrates haptic feedback and fluid dynamics, targeting both residency training and preoperative rehearsal of complex microneurosurgical procedures. The work will be conducted over three years in collaboration with the Department of Neurosurgery in Karlsruhe.

The methodology is structured into five sequential phases. First, tissue-specific biomechanical parameters (e.g. Young's modulus, Poisson's ratio, and viscoelastic coefficients) will be collected from the literature and refined through calibration against intraoperative observations. Second, patient-specific anatomical models, with high spatial fidelity, will be generated from CT/MRI data. Third, neurosurgical scenarios of increasing complexity, ranging from standard craniotomy procedures to intracranial aneurysm clipping, will be developed within a hybrid framework integrating physics-based and data driven modeling (to simulate brain tissue deformation and hemodynamic behavior) and real-time haptic rendering (to simulate touch sense). Fourth, Virtual and Augmented Reality environment will be implemented. Fifth, the platform will undergo systematic evaluation of face, content, and construct validity through structured studies involving neurosurgical residents and expert surgeons.

The final outcome will be a reproducible and scalable pipeline for generating patient-specific surgical digital twins, with direct applications in neurosurgical training, surgical rehearsal, and preoperative planning across multiple institutions.

ID03:

Agentic, covariate-aware multi-omics factor analysis for causal biomarker discovery

Data Science supervisor: Prof. Dr. Oliver Stegle

Life Science supervisor: Dr. Junyan Lu

Multi-omics profiling promises a comprehensive view of disease processes across molecular layers, with broad implications for personalised medicine. Despite this potential, the analysis of high-dimensional multi-omic data remains a major theoretical and practical challenge. Clinical multi-omic readouts are noisy, sparsely sampled, and shaped by complex covariate structures, including genetic background, technical variation (e.g. batch effects), temporal and spatial sampling, and experimental conditions such as chemical or genetic perturbations. Together, these factors can preclude the identification of causal and mechanistic signals. Multi-omics factor analysis, pioneered by the partnering PI groups, is among the most promising strategies for distilling biomedical insights from such data. However, existing methods cannot adequately account for the complex covariate structures of clinical cohorts, and they are difficult to specify, adapt, and deploy without substantial methodological expertise, which severely limits their uptake by clinical scientists.

To address these challenges, we propose to develop a flexible, programmatic, covariate-aware multi-omics factorisation framework tailored to heterogeneous clinical cohorts with complex experimental designs. The framework will advance the state of the art through automatic model definition, structured handling of patient and experimental covariates, and strategies that exploit instrumental variables and statistical invariance to enable causal representation learning. In parallel, the framework will be paired with agentic interfaces that make model specification, configuration, and interpretation accessible to clinical scientists without prior expertise in probabilistic modelling. We will apply the resulting solutions to a comprehensive multi-omics cohort of lung cancer patients to identify biomarkers predicting tumor recurrence, the primary cause of treatment failure and cancer-related death.

ID04:
Large-Scale Foundation Models for 3D Radiology

Data Science supervisor: Prof. Dr. Klaus Maier-Hein

Life Science supervisor: Dr. Hannes Kenngott

Large-scale foundation models have transformed natural-image analysis, but multimodal pretraining for 3D radiology remains limited by dataset scale and the difficulty of learning from full volumetric imaging data. This project develops multimodal foundation models for 3D radiology by extending The Human Radiome Project (THRP), a corpus of 2.1 million CT, MRI, and PET volumes, from image-only self-supervised pretraining toward joint learning from imaging, radiology reports, and structured clinical metadata.

The central scientific question is how multimodal supervision can be integrated effectively into large-scale 3D radiology pretraining. The project investigates scalable vision-language pretraining strategies for volumetric imaging, combining contrastive image-text alignment with reconstruction-based objectives and systematically studying supervision granularity from finding-level entities to full reports and volume-level representations.

The resulting models are evaluated across the THRP benchmark on multi-finding classification, retrieval, report generation, visual question answering, regression, segmentation, and zero-shot transfer, with particular focus on robustness under domain shift and low-data settings. Unlike prior work in this area, the required imaging corpus, paired report data, and pretraining infrastructure are already available within the participating groups.

By combining large-scale 3D imaging data with multimodal clinical supervision, the project aims to establish a scalable methodological basis for multimodal radiology foundation models and clinically relevant generative AI applications in medical imaging.

ID 05:
Personalized Modeling and Multimodal Sensing for Imminent Fall Prediction in Geriatric Patients

Data Science supervisor: Prof. Dr. Katja Mombaur

Life Science supervisor: PD Dr. Carl-Philipp Jansen

Falls are a leading cause of injury, hospitalization, and loss of independence in older adults. Falls happen during various Activities of Daily Living (ADL), like walking or standing up from chairs or beds or sitting down, in indoor and outdoor setting, often even while using mobility assistance devices. Despite substantial progress in long-term fall risk assessment and fall detection, current approaches remain fundamentally reactive, either identifying individuals at risk over extended periods of time or detecting falls only after impact. The critical time window immediately preceding a fall, i.e. the second or fraction of a second before, during which intervention could prevent injury, remains largely unexplored.

This project proposes a novel paradigm centered on imminent fall prediction, aiming to identify short-term instability states preceding falls. The central hypothesis is that these states manifest as subtle, individualized deviations in mobility patterns, including gait variability, balance disturbances, and near-falls, which can be captured through multimodal sensing and modeled using advanced data-driven approaches.

A key limitation in the field is the lack of valid datasets, as falls cannot be ethically induced in older populations and continuous monitoring raises concerns regarding measurement burden. This project addresses these challenges by focusing on near-falls, instability events, and (if occurred) actual falls, combined with privacy-preserving sensing strategies and controlled observational studies. Furthermore, the project introduces personalized digital twins, i.e., computational models of individual mobility and balance, to simulate fall scenarios, generate synthetic data, and support the development of novel health technologies for everyday use, ranging from sensor systems to assistive robotic devices.

Methodologically, the project will compare wearable, vision-based and ambient sensing approaches and evaluate the use of the personalized digital twins with different optimization and machine learning paradigms. By integrating real-world data, individualized modeling, and predictive simulations, the project aims to shift fall research from reactive detection to proactive, personalized prevention, contributing to safer aging and advancing data-driven healthcare.

ID06:

Prediction of hypoxia-associated peptides from hematoxylin and eosin histological images

Data Science supervisor: Prof. Dr. Benedikt Brors

Life Science supervisor: Dr. Ina Kurth

Tumor hypoxia is a major driver of poor prognosis and treatment resistance in head and neck squamous cell carcinoma (HNSCC), where it induces extensive transcriptional and proteomic reprogramming. However, comprehensive molecular hypoxia profiling is rarely implemented in routine HNSCC care due to cost and time constraints, whereas hematoxylin and eosin (H&E) staining remains standard for all patients. Deep learning models applied to H&E whole-slide images can predict outcome in HNSCC, but their biological interpretability and explicit linkage to hypoxia biology remain limited. This project follows a two-stage strategy. During development, we combine several molecular and imaging techniques including spatial transcriptomics, hypoxia fluorescence imaging, and, where available, mass spectrometry imaging on the same tumor sections to build a detailed map of where hypoxia occurs and how it looks at the molecular level. This map is then used to teach AI models to recognize the same hypoxia patterns directly on the matched H&E images, and to summarize them in a single, interpretable H&E-based hypoxia score. In application, the score can be calculated from routine H&E slides alone, without any additional molecular tests. The goal is to derive explainable, H&E-based hypoxia scores for biologically informed patient stratification with respect to local control, overall survival, and response to radiotherapy and immunotherapy, and to uncover mechanisms of hypoxia-driven treatment failure as candidates for therapeutic targeting. Multimodal spatial profiling of intratumoral hypoxia thus has the potential to support precision oncology in HNSCC by resolving spatial heterogeneity and enabling risk-adapted, hypoxia-directed treatment strategies.