October 8

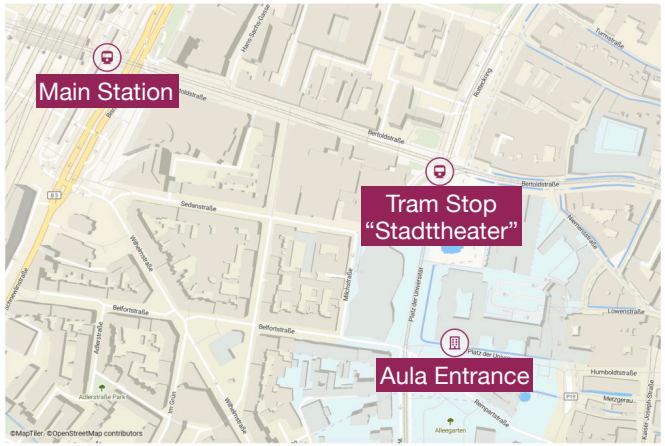
Session 3: The value of mathematical theory for small data real-world applications
Chair: Carsten Dormann

08:30 - 09:00	**Sonja Greven, Berlin** Fusing deep learning and statistics towards understanding structured biomedical data
09:00 - 09:30	**Axel Munk, Göttingen** Statistical optimal transport meets life sciences
09:30 - 09:45	**Carola Heinzel & Lennart Purucker, Freiburg** Improving machine learning for small genetic data using mathematical statistics
09:45 - 10:15	Coffee break
10:15 - 10:45	**Angelika Rohde, Freiburg** Nonparametric maximum likelihood estimation of monotone binary regression models under weak feature impact
10:45 - 11:30	**Panel Discussion** Axel Munk, Angelika Rohde, Sonja Greven, Maren Hackenberg, Holger Dette, Jens Timmer
11:40 - 12:45	**Poster Pitch Tour #3**
12:15 - 13:00	Lunch break

Session 4: Navigating similarity & uncertainty - statistical approaches for robust predictions and inferences in the small data setting
Chair: Anna Köttgen

13:00 - 13:30	**Arnoldo Frigessi, Oslo** Learning the differential equation of the tumour density of one breast cancer patient
13:30 - 14:00	**Jan Gorodkin, Copenhagen** Analysis of CRISPR data and prediction for design of gene editing experiments
14:00 - 14:15	**Nana-Adjoa Kwarteng, Freiburg** Network meta-regression
14:15 - 14:30	Coffee break
14:30 - 15:00	**Harald Binder, Freiburg** The vision of *SmallData*
15:00 - 15:45	**Panel Discussion** Arnoldo Frigessi, Jan Gorodkin, Nadine Binder, Johannes Hertel, Adriani Nikolakopoulou
15:45 - 16:00	**Harald Binder, Freiburg** Wrap up and poster prize

Location



Aula, Kollegiengebäude 1
Platz der Universität 3
79098 Freiburg

universität freiburg

UNIVERSITÄTS KLINIKUM FREIBURG

RUHR UNIVERSITÄT BOCHUM
RUB

Universitätsmedizin GREIFSWALD

Contact

CRC Small Data Office
bemb.smalldataoffice@uniklinik-freiburg.de
Tel.: +49 173 1584176

Website

LinkedIn

small data
MORE FROM LESS
CRC 1597

SmallData Symposium

OCTOBER 7 & 8, 2024

DFG Deutsche Forschungsgemeinschaft

October 7

08:30 - 09:00	Registration
09:00 - 09:30	**Opening Remarks and Welcome Addresses** **Harald Binder**, Speaker of the CRC 1597 *SmallData* **Stefan Rensing**, Vice Rector for Research and Innovation at the University of Freiburg **Lutz Hein**, Dean of the Faculty of Medicine at the University of Freiburg **Frederik Wenz**, Chief Medical Director and Chairman of the Board of the Medical Center - University of Freiburg

Session 1: Hurdles in transferring AI techniques into real-world applications
Chair: Harald Binder

09:30 - 10:00	**Rodolphe Thiébaut, Bordeaux** Integrating gene expression from whole blood into dynamical systems: illustration with a mechanistic model of the antibody response to COVID vaccination
10:00 - 10:30	**Sonja Schimmler, Berlin** A national research data infrastructure for data science and artificial intelligence
10:30 - 10:45	**Masako Kaufmann, Freiburg** Development of CRISPert, a novel deep learning-based tool enabling efficient and safe application of CRISPR-Cas
10:45 - 11:15	Coffee break
11:15 - 11:45	**Anna Köttgen, Freiburg** From population studies to modeling of human metabolism - and back
11:45 - 12:30	**Panel Discussion** Lisa McShane, Matthias Mau, Rodolphe Thiébaut, Sonja Schimmler, Lutz Hein, Eleni Papakonstantinou
12:30 - 14:30	Lunch break
13:00 - 14:15	**Poster Pitch Tour #1**

Session 2: Data driven modelling versus scientific discovery and expert knowledge
Chair: Nadine Binder

14:30 - 15:00	**Jan Hasenauer, Bonn** Sparse clinical data: a call for population-level models
15:00 - 15:30	**Thomas Brox, Freiburg** No free lunch: why small data tasks require big data models
15:30 - 15:45	**Jelena Bratulić, Freiburg** What matters for in-context learning: a balancing act of look-up mechanism and in-weight learning
15:45 - 16:15	Coffee break
16:15 - 16:45	**Frank Hutter, Freiburg** TabPFN v2, a foundation model for small tabular data
16:45 - 17:30	**Panel Discussion** Jan Hasenauer, Thomas Brox, Frank Hutter, Hannah Bast, Noor Awad, Tanja Vogel
17:30 - 19:00	Canapés
17:45 - 19:00	**Poster Pitch Tour #2**

ID	Presenter	Title
Poster Pitch Tour #1		
101	Behrens, M	Identifying data similarity across subgroups and sites
102	Kober, N Bellerino, G	Similarity weights in the nonparametric maximum likelihood estimator Limit theorems for Markov processes
103	Schlosser, A Farhadizadeh, M	Bias-corrected maximum likelihood estimation of parametric competing risk models for small data
104	Lange, Z	Identifying best practice treatment strategies by incorporating information from similar healthcare pathways
105	Schächter, C	Analysis of treatment effects despite switches in measurement instruments by combining variational autoencoders with mixed effects models
106	Tambe-Ndonfack, F	Advanced filtering theory and the Zakai equation for jump-diffusion stochastic processes
108	Secen, E	Dissecting the molecular basis of monogenic neurodevelopmental disorders
109	Jobson Pargeter, W	CRISPert: A transformer-based model for CRISPR-Cas off-target prediction
111	Böhm, S	CoordConformer: Decoding heterogeneous EEG datasets using transformers
112	Hog, J	Meta-learning population-based methods for reinforcement learning
113	Raum, H	Dynamic integration of process models and neural networks to improve predictive performance in ecology
114	Habenicht, H	Similarity evaluation of training vs test data and the potential of process knowledge
115	Karakioulaki, M	A systematic review and metanalysis for inflammation parameters in dystrophic epidermolysis bullosa
116	Yang, H	Calibrating representations of expert knowledge with patient data in latent spaces for synthetic trajectories
118	Bratulić, J	Taxonomy-aware continual semantic segmentation in hyperbolic spaces for open-world perception
119	Walter, S	SPARQL knowledge graph question answering over Wikidata via constrained language modeling
120	Ging, S	Image-text representation learning
121	Arnold, P	Comparing the performance of open and close sourced Large Language Models for automatic CAD-RADS 2.0 classification from cardiac computer tomography radiology reports
122	Fässler, D Huang, C	Methodologies to improve the scope and accuracy of whole-body models of human metabolism
Poster Pitch Tour #2		
123	Scherer, N	Coupling of metabolomics and exome sequencing reveals graded effects of rare damaging heterozygous variants on gene function and human traits and diseases
124	Hoffman, L	Being certain of uncertainty
126	Mesuere, G	Locally stationary hidden Markov models
128	Müller, J	Efficacy of psychotherapy, pharmacotherapy, or their combination in chronic depression: a systematic review and network meta-analysis using aggregated and individual patient data
129	Neubrand, N	1000+ synthetic benchmark problems for parameter estimation in dynamic modelling
130	Kord, Y	Enhancing SNLS optimisation via deep reinforcement learning for adaptive tolerance setting
131	Hasan, M	Generating optimal small datasets for efficient offline reinforcement learning training
132	Zhang, B	Exploration cocktail: automating exploration in reinforcement learning
133	Zabërgja, G	Empirical assessment of paradigms in tabular classification
134	Purucker, L	Applying a foundation model to small tabular data
135	Kabus, F Hackenberg, M	An end-to-end modeling approach for capturing spatiotemporal patterns in two-photon imaging data
e102	Döhler, S	Small data meets high dimensions: some approaches from multiple testing
e103	O'Brien, T	Challenges of small data in biomedical and environmental research
e104	Moringen, A	A meta unit for co-constructing a computational scaffold model to guide human motor learning
e106	Wendland, P	OptAB - an optimal antibiotic selection framework for sepsis patients with artificial intelligence
e107	Dümpelmann, M	Denoising of low dimensional EEG data with deep learning for improved seizure detection
e108	Krutsylo, A	Forward-forward optimization in small data
e109	Brunn, N	Similarity-based refinement of single-cell interactions
e110	Rollin, J	Prediction of cell lineage trajectories by integration of small single-cell RNA datasets into a large reference dataset
Poster Pitch Tour #3		
e112	Brombacher, E	Characterizing the omics landscape based on 10,000+ datasets
e111	Mahendra, M	Convex space learning for tabular synthetic data generation
e116	Umesh, C	Preserving logical and functional dependencies in synthetic tabular data
e115	Archer, L	Uncertainty in clinical risk prediction: perspectives and approaches
e113	Legha, A	Uncertainty-based sequential sample size calculations for developing clinical prediction models using regression or machine learning methods
e117	Pierre Paul, D	Speeding up the clinical studies with biomarker-based enrichment
e119	Schneider, J	Multimodal outcomes in N-of-1 trials: deep-learning based effect estimates in a small data study design
e118	Papakonstantinou, E	Multidimensional investigation of response to treatment with inhaled corticosteroids in COPD patients: insights from the HISTORIC study
e123	Lang, T	AI & statistics in preclinical research and development
e121	Bonetti, M	Two small-sample problems in optimal and exact inference
e114	Eggert, A	When only small data is available in livestock research
e122	Farhadyar, K	Impact of different longitudinal data representations on transformer performance in small data applications
e124	Bodden, D	Allocation bias in group sequential designs
e125	Schoenen, S	Quantifying the impact of allocation bias in randomised clinical trials with multi-component endpoints
e126	Bordoloi, R	Multivariate functional linear discriminant analysis of partially-observed time series