technische universität dortmund

fakultät statistik

Faculty of Statistics
Prof. Dr. Markus Pauly
Joseph- von Fraunhofer Str. 2-4
D-44227 Dortmund

Tel.: 0231/755-90370
Fax: 0231/755-90372
markus.pauly@tu-dortmund.de
www.tu-dortmund.de

Technische Universität Dortmund | D-44221 Dortmund

Secretary
Ulrike Guba
Joseph- von Fraunhofer Str. 2-4
D-44227 Dortmund

Tel.: 0231/755-90371
Fax: 0231/755-90372
gubal@statistik.tu-dortmund.de
www.tu-dortmund.de

Diktatzeichen	Aktenzeichen	Ort	Datum	Dienstgebäude/Raum
	-	Dortmund	29 May 2024	

Potential Theses on
Machine Learning for Rare Diseases

Motivation.
The potential of Machine Learning and AI in life science applications has been discussed extensively. However, this is not true for the special case of rare diseases with inherently rather small sample size which complicates the application. Within this context, we have topics for several theses which may cover one of the following points:

1. Comprehensive Literature Review: Continuing the work [1,2] to review and discuss existing applications and methods' limitations
2. Benchmark Data Sets: Comprehensively collect benchmark data sets for different statistical / ML tasks (associative, predictive, causal, inferential, /interpretability etc.)
3. Benchmark Simulation Models: For an objective comparison is of of particular importance to not only do method's comparisons on benchmark data sets but also in simulations [3]. Here, the aim is to find 'realistic' simulation models.
4. Comparison of methods: Task-specific comparisons of different methods based on benchmark data sets or simulations. The selected methods were chosen based on their relevance and application in current research and the comparisons will be done wrt different criteria.

This master's project is part of an ongoing research project that is currently being worked on intensively. The results will actively contribute to the further development and improvement of the current data science toolbox in biostatistics. We are convinced that this project will not only make a significant contribution to current research, but will also provide an excellent opportunity for the student(s) to immerse themselves in the depths of ML and biostatistics and contribute to a highly topical and socially relevant subject. For further information or if you are interested, please do not hesitate to contact us. We look forward to working with a dedicated master's student on this exciting research project. If you are interested,

please apply with a brief description of your current study situation as well as a transcript of records (from BOSS) to: Markus Pauly (pauly@statistik.tu-dortmund.de).

References

[1] Schaefer, J., Lehne, M., Schepers, J., Prasser, F., & Thun, S. (2020). The use of machine learning in rare diseases: a scoping review. *Orphanet journal of rare diseases*, *15*, 1-10.

[2] Banerjee, J., Taroni, J. N., Allaway, R. J., Prasad, D. V., Guinney, J., & Greene, C. (2023). Machine learning in rare disease. *Nature Methods*, *20*(6), 803-814.

[3] Friedrich, S., & Friede, T. (2024). On the role of benchmarking data sets and simulations in method comparison studies. *Biometrical Journal*, *66*(1), 2200212.