

## **On random forests, omics data and benchmark studies**

After an overview of recent methodological developments from our lab related to the random forest method, I will address two important issues relevant to machine learning algorithms in general and random forest in particular. The first issue is the applicability of prediction rules obtained from machine learning algorithms in practical settings. I will present the results of our recent literature survey on the applicability and reproducibility of random forest prediction rules published in PLOS ONE and discuss lessons learned from this study. The second issue is the impact of the parameters of the method, which can be seen as a source of uncertainty and "researchers degree of freedom". The last part of my talk will be devoted to benchmark studies assessing the performance of random forest in several situations including prognostic modelling with multi-omics data. Special emphasis will be put on the choice of an appropriate benchmark study design.

A.-L. Boulesteix\*, S. Janitza\*, R. Hornung, P. Probst, H. Busen, A. Hapfelmeier, 2019. Making complex prediction rules applicable for readers: current practice in random forest literature and recommendations. *Biometrical Journal* 61:1314-1328. \*equal contribution.

R. Couronné, P. Probst, A.-L. Boulesteix, 2018. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* 19:270.

P. Probst, A.-L. Boulesteix, B. Bischl, 2019. Tunability: importance of hyperparameters of machine learning algorithms. *Journal of Machine Learning Research* 20(53):1-32.

R. Hornung, M. Wright, 2019. Block Forests: random forests for blocks of clinical and omics covariate data. *BMC Bioinformatics* 20:358.