

# Uncovering Interactions in Random Forests

Jacob Michaelson<sup>1,\*</sup>, Marit Ackermann<sup>1</sup>, and Andreas Beyer<sup>1</sup>

1. Cellular Networks and Systems Biology, Biotechnology Center, TU Dresden

\* Contact author: jacob.michaelson@biotec.tu-dresden.de

**Keywords:** Random Forests, interactions, gene regulation

Over the past decade, Random Forests [1] have proven to be a powerful machine learning algorithm in many applications. Its usefulness stems not only from its predictive ability, but also from information it gives about the structure of the model underlying the data. The latter attribute can be particularly appealing in problems where  $N \ll P$  and a majority of the predictor variables do not participate in the underlying true model. Such is the case in many biological problems, where Random Forests can be used to sift through large volumes of data to find meaningful interactions between predictor variables. Measures of variable importance already exist [1, 2] but these generally give information about the "main effects" and do not provide direct insight about relationships between predictors.

In this work we propose a novel approach to detecting variable interactions in Random Forests. Information on predictor co-occurrence in the forest's trees is used as a basis for both a frequentist and a Bayesian approach for uncovering interactions between the variables. To interpret the interdependencies, a graph of the variable interactions is constructed. By depicting this information as a graph, we impose no limit on the order or characteristics of the interactions. We apply the methods to gain new insight on neuronal regulatory pathways in the hippocampus.

## References

- [1] Leo Breiman. Random forests. *Machine Learning*, 45(1):5, 2001.
- [2] Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1):307, 2008.