## Three applications of PARF, a versatile implementation of the Random Forests classifier, to problems in life science

Fran Supek, Rudjer Boskovic Institute

## Abstract

Random Forest (RF) is a classification algorithm which has recently gained popularity in a wide variety of data mining tasks in the biomedical domain. RF is based on an ensemble of decision trees, which usually has top-of-the line predictive performance, on par with the sophisticated Support Vector Machine classifier. Additionally, the computational workload of training a RF classifier can be distributed over a cluster of computers, as implemented in the PARF (Parralel RF) software, freely available from <a href="http://parf.googlecode.com">http://parf.googlecode.com</a> Finally, PARF offers insight into functioning of the trained classifiers via the *attribute importance* and *class prototypes* features. I will demonstrate three applications of RF to datasets from functional genomics, psychiatry and protein biophysics.