A Probabilistic Approach to Classifying Metabolic Stability

A. ter Laak[1]
T Schroeter[2], A. Schwaighofer [2], S. Mika[2], P Lienau[1], A. Reichel[1], Müller, K-R[2], N. Heinrich[1]
[1]*Bayer Schering Pharma AG, Müllerstrasse 178, 13342 Berlin, Germany*
[2]*Fraunhofer FIRST, Intelligent Data Analysis (IDA), Kekuléstraße 7, 12489 Berlin, Germany*

Metabolic stability is an important property for drug candidates. Optimally,it should be taken into account already in the in silico phase of the drug design process. Yet, general purpose predictive tools for this endpoint are inherently difficult to obtain.
We present a machine learning approach to predicting metabolic stability, that is tailored to compounds from the drug development process at Bayer Schering Pharma AG. Our modelling is based on existing measurements of the percentage of each compound remaining after incubation with liver microsomes for 30 minutes. We built independent modelsfor4different species (human, male mouse, female mouse, male rat), with 1000 to 2100 measurements per species. From this data, we develop Bayesian classification models to predict the probability of a compound being metabolically stable.
A particular advantage of the chosen approach is that it implicitly takes the "domain of applicability" into account. For compounds outside the domain of applicability, the model output is 0.5, indicating that it is not possible to tell whether the compound is stable or not.
The developed models were validated on recent project data (200 to 700 compounds, depending on species), showing that the predictions are highly accurate. In particular, we could show that the accuracy of predictions increases when excluding compounds with a predicted probability around 0.5, that is, those that are outside the domain of applicability.

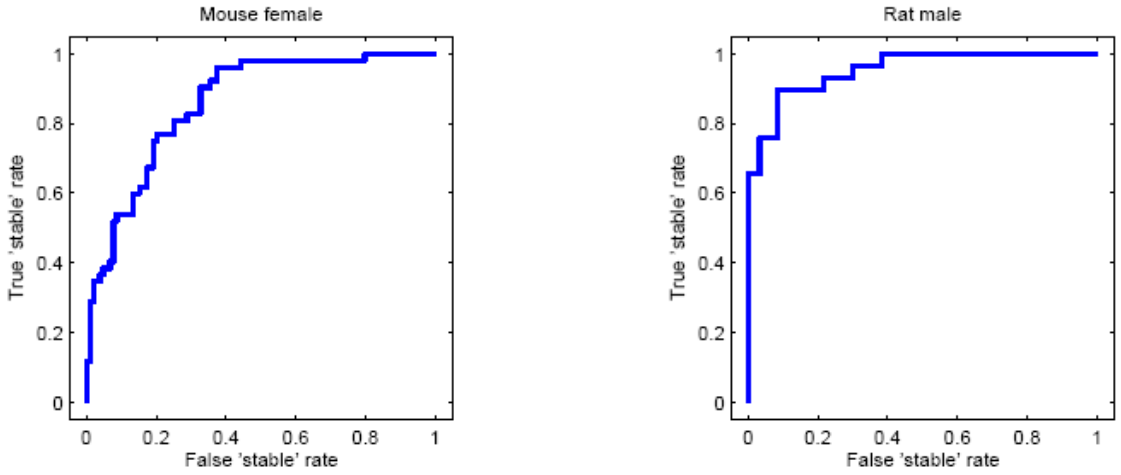

Figure 1. ROC-Curves for validating the developed models on data from recent projects that was not used for model building. Left: Model for female mouse. 258 unambiguous validation measurements, 156 of which are in the domain of applicability. Right: Model for male rat. 183 unambiguous validation measurement, 89 of which are in the domain of applicability.