

# Toxicity Classification from NMR Metabonomic Data Using Artificial Neural Networks

Xin Yu<sup>a</sup>, Dorrit Baunsgaard<sup>b</sup>, Svava Ósk Jónsdóttir<sup>a\*</sup>, Thomas Sicheritz Pontén<sup>a</sup> and Søren Brunak<sup>a</sup>

<sup>a</sup>Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark, DK-2800

Kongens Lyngby, Denmark

<sup>b</sup>Department of Protein Structure and Biophysics, Novo Nordisk A/S, Novo Allé, DK-2880 Bagsværd, Denmark

## ABSTRACT

Early detection of drug toxicity is of great importance in the pharmaceutical industry for reducing the attrition rate of drug candidate failing and providing safer new drug to patients. Metabonomics is one of the approaches that has been rapidly developing for a early drug toxicity screening in the preclinical stage, of which the most effective way is to combine the high resolution NMR spectroscopy of biofluids with the chemometrics methods. However, many of the current studies in this area are based on very limited NMR data sets and the data has been pre-selected in characterized pattern. In addition, most of the prediction tools are based on linear methods.

In this study, based on NMR metabonomic data from urine samples, the possibility of applying artificial neural network (ANN) to whole spectra of descriptors for classifying and prediction drug toxicity has been investigated. The carefully preprocessed data sets were used to set up toxicity classification models, which were further optimized on various time-point and toxicity class combinations together with the neural network parameters. Moreover, a neural network based descriptor analysis has been carried out for finding important pattern for toxicity classification. ANN models with unselected descriptors are robust. They showed high classification performance and prediction accuracy on the toxic, none toxic, liver toxic and mixed toxic compounds. The ANN models with selected descriptors showed very high classification performance on similar data. Larger data sets are needed for further evaluation and developing the toxicity classification models. The results of this study give us very promising starting point for using ANN on drug toxicity classification in the early stage of drug development.

---

\*To whom correspondence should be addressed