A HYBRID COST-OPTIMIZED QSAR MODEL FOR REPRODUCTIVE TOXICITY

F. Lemke[1], A. Roncaglioni[2], E. Benfenati[2]
[1]KnowledgeMiner Software, Berlin, Germany; [2]Istituto Mario Negri, Milan, Italy

Introduction

Alternative testing strategies summarize efforts to replace, reduce and refine animal usage in experimentation. Developing new robust and effective tests has been a priority in the EU for more than 20 years. The challenge is to develop alternative methods, which are validated according to international standards and quality criteria, be recognized by regulatory authorities and be adopted by industry. To build this acceptance and understanding alternative methods have to address these key properties:

- ***Transparency*** of the development and application process;
- ***Reproducibility*** of test results;
- ***Interpretability*** of test results;
- ***Uncertainty*** of results given uncertainty of experimental data;

Within the CAESAR research project we developed a hybrid, cost-optimized Quantitative Structure-Activity Relationship (QSAR) model for reproductive toxicity, following the requirements on alternative testing methods for regulatory purposes, namely for REACH, especially. This model adds two other features to improve accuracy, reliability and applicability of the model:

- ***Hybridization*** by leveraging the concept of model ensembles to increase model accuracy and predictive power, and
- ***Cost-sensitive modeling*** to optimize a model's cost of application, its sensitivity and/or specificity according to a given cost-benefit matrix. This implements elements of risk assessment into the modeling process.

For the implementation of REACH predictive information about reprotoxicity is requested. In fact, every chemical compound has to be characterized for carcinogenicity, mutagenicity and reprotoxicity (CMR). It has been estimated that animal tests for reproductive toxicity would represent one of the major reasons for a highly increased use of animals and for increased costs for implementing REACH. For these reasons alternative testing methods for reprotoxicity are very demanded, but QSAR models for reproductive toxicity have been very scarce yet.

The presented QSAR model will be available free for evaluation and application for both regulatory bodies and industry.

Methods

Toxicity data were taken from Arena et al. (1). Chemical descriptors have been calculated with Dragon and MDL. 13 of them have been selected. The presented QSAR classification model on reprotoxicity was generated by combining three powerful modelling approaches:

- Self-organisation of models from high-dimensional data sets,
- Cost-sensitive modelling to minimize cost and risk of applying a model, and
- Hybrid models.

This unique and innovative combination of modelling approaches is a result of the CEASAR research project and it was driven by the increasing demand for alternative, easy-to-use, reproducible and transparent testing methods, which are especially useful and appropriate for giving decision-support within REACH legislation.

Model Self-organisation

The main scientific foundations of Self-organising Networks of Active Neurons (2) are:

- ***Statistical Learning Network*** theory;
- The ***principle of self-organization*** as an adaptive, automated creation of a net without subjective points given;
- The principle of external complement, which enables the objective selection of a ***model of optimal complexity***,
- The ***principle of regularization of ill-posed tasks***.

An optimal complex model is a model that optimally balances model quality on a given learning data set ("closeness of fit") and its generalisation power on new, previously unseen data with respect to the data's noise level and the task of modelling (prediction, classification, modelling, etc.).

Cost-sensitive modelling

Many real-world applications assign different costs and risks connected to False Positive and False Negative cases. This is an essential and critical fact, which is, however, not yet targeted by known QSAR modelling methods. All QSAR models are usually built using mathematical criteria, which balance positive and negative deviations from the given observed value equally.

In CAESAR, for the first time, a new algorithm has been developed and implemented with the goal to generate QSAR models, which are optimal with respect to both the classification and prediction power of a model and the costs it generates according to an a priori given cost-benefit matrix.

Relative Cost-Benefit Matrix

TP 0	FP 10
FN 100	TN -30

positive values = cost
negative values = benefit

Model Quality Q:
Q = Predictive Power + Classification Accuracy + Benefit of the Model

Best Model M:
$M = \max_{L} Q$
with L = number of all genetated models

This means that in this approach the absolute or relative costs and benefits for True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) cases must be given (and agreed to) in advance of modelling (3).

Hybrid models

All known methods of modelling lead to a single "best" model while the accuracy of the model depends on the variance of the data. A common way for variance reduction is aggregation of similar model results following the idea: Generate many versions of the same predictor/classifier and combine them in a second step. If modelling aims at prediction, it is helpful to use alternative models that estimate alternative forecasts. These forecasts can be combined using several methods to yield a composite forecast of a smaller error variance than any of the models have individually (4). The desire to get a composite forecast is motivated by the pragmatic reason of improving decision-making rather than by the scientific one of seeking better explanatory models. Nonetheless, hybrid models built using Self-organising Networks of Active Neurons are analytically available on the fly by an explicit regression equation.

Results & Discussion

The shown QSAR model was developed following major demands on alternative testing methods:

Accuracy and Generalization
The classification accuracy on the entire data set is shown in Fig. 1. The model was built on a learning data set (234 compounds) with TP = 98%, TN = 62% and was evaluated on an out-of-sample testing data set (58) with TP = 90%, TN = 54%. The performance gap between TPR and TNR is reasonable because the model was built with much higher cost assigned to FN than to FP.

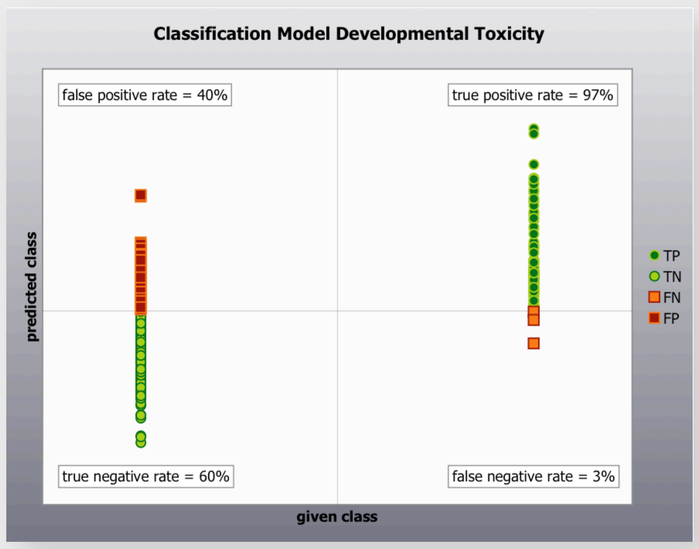

Fig. 1: Accuracy of the hybrid, cost-optimized model

Transparency
The presented model is a composite of three different individual models and is described by this linear equation:

$$\text{Reprotoxicity Class Value} = 0{,}201*\text{Model}_1 + 0{,}584*\text{Model}_2 + 0{,}518*\text{Model}_3 - 0{,}115$$

The steps taken to get this model (and all individual models) are documented and this documentation is part of the final model.

Reproducibility
Taken and following the documentation of the models, it is possible for a third party to reproduce the QSAR models, independently. The results are expected to be the same.

Interpretability
For each self-organized model a linear or non-linear regression equation is provided, automatically, which, for example, looks like this (extract from model 2):

$$Y = +0{,}685X_{238} + 0{,}144X_{393} - 1{,}257X_{236}X_{238} + 1{,}416X_{238}X_{393} + \ldots + \\ - 2{,}602X_{236}X_{238}X_{393} + 0{,}001(X_{12})^2(X_{73})^2 + 0{,}0015X_{12}(X_{73})^3 \\ + 1{,}310E\text{-}6(X_{73})^4 - 5{,}584E\text{-}5(X_{12})^2(X_{73})^3 - 1{,}948E\text{-}6X_{12}(X_{73})^4 \\ + 7{,}242E\text{-}7(X_{12})^2(X_{73})^4 + 0{,}227$$

Although the interpretation capability of non-linear models is limited, it is still very useful information for further model evaluation, analysis and implementation purposes. Also, these QSAR models are predictive models, which not necessarily represent causal relationships. They are approximating the behavior of the complex biochemical process.

Reflection of uncertainty a priori given by experimental data
The output of the model is a most likely prediction along with an interval of prediction, which is computed individually for each compound as shown in figure 2.

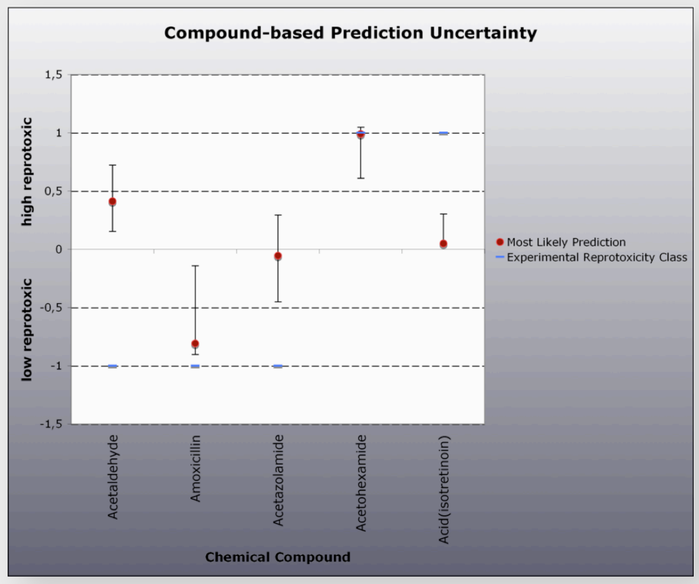

Fig. 2: Most likely prediction and interval of uncertainty

Model cost and risk assessment
A major concern for testing methods for regulatory purposes is minimization of False Negatives. However, different test purposes in different stages of the evaluation process may have different, maybe less restrictive requirements. Furthermore, there is higher cost associated to FN than to FP, usually, and the benefits of TN and TP may also differ. By providing a cost-benefit matrix which reflects actual testing purposes the model optimally adapts to changing requirements and conditions with respect to model benefit, sensitivity and specificity.

Conclusions

We presented a new QSAR model to predict reprotoxicity. The main focus of model development has been in addition to model accuracy and reliability on transparency, reproducibility, interpretability, and model results uncertainty. Furthermore, the model has been optimized to produce a low number of false negatives. Such a model can be used for a screening of chemicals.

In the possible case of the integration of this model with other methods, also experimental ones, the model can be easily adapted according to a given cost-benefit matrix to optimize overall benefit and accuracy (or, if preferred, a higher specificity or sensitivity) of the strategy.

The QSAR model will be available free for evaluation and application for both regulatory bodies and industry.

References

1. V. C. Arena, N.B: Sussman, S. Mazumdar, S. Yu, O:T: Macina. The utility of structure-activity relationship (SAR) models for prediction and covariate selection in developmental toxicity: comparative analysis of logistic regressions and decision tree models. SAR and QSAR Environ. Res. 15, No. 1: 1-18, 2004.
2. J.-A. Mueller, F. Lemke (2000) Self-Organising Data Mining. Extracting Knowledge From Data, BoD, Hamburg.
3. M.H. Zweig, G. Campbell (1993) Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine, Clinical Chemistry, Vol. 39, No.4.
4. J. Elder: The Generalization Paradox of Ensembles, Journal of Computational and Graphical Statistics 12, No. 4: 853-864, 2003

Acknowledgement: We acknowledge the financial support of the EC project CAESAR (SPPI 022674).