Gaining Confidence in Computational Models for Risk Assessment - Combining Approaches and Understanding Uncertainty

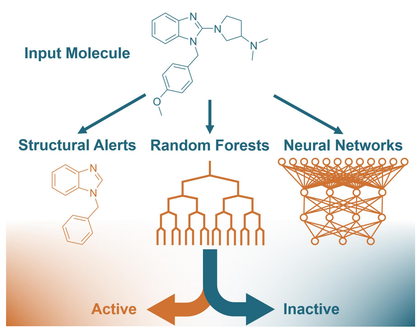
Timothy E H Allen[1,2], Andrew J Wedlake[2], Maria Folia[3], Alistair M Middleton[3], Sam Piechota[3], Elena Gelžinytė[2], Jonathan M Goodman[2], Steve Gutsell[3], Predrag Kukic[3], Paul J Russell.[3]

1. **MRC Toxicology Unit,** University of Cambridge, Gleeson Building, Tennis Court Road, Cambridge, United Kingdom, CB2 1QR
2. **Centre for Molecular Informatics**, Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, United Kingdom CB2 1EW
3. **Unilever Safety and Environmental Assurance Centre**, Colworth Science Park, Sharnbrook, Bedfordshire, United Kingdom, MK44 1LQ

Introduction

The molecular initiating event (MIE) [1,2,3] is the initial chemical-biological interaction that can be thought of as a gateway to the adverse outcome pathway (AOP) [4]. By understanding MIEs we can understand the kinds of interactions molecules make, and hence the kinds of adverse outcomes they might cause. Chemistry is key to understanding the MIE. What is it about these molecules that allow them to do this?

In this study, a variety of computational approaches have been used to try and make activity predictions at human MIEs. Structural alerts have been built automatically using a maximal common substructure algorithm and Bayesian statistics in KNIME [5]. Random forest models were constructed using sklearn and RDKit in Python 3, with 200 physicochemical descriptors as the input [5]. Neural networks were developed with extended connectivity fingerprints as features in Python 3 using TensorFlow [6]. A variety of network architectures, activation functions and hyperparameters were considered and optimised to give the highest level of statistical performance. These binary activity predictions have been compared and combined, to provide the highest model performance and confidence.



Combined Binary Predictions

Different computational models need not always be viewed in competition. Combining such tools increases confidence and allows them to cover one another's weaknesses. Models were trained and evaluated on a consistent dataset extracted from the publicly available databases ChEMBL [7] and ToxCast [8] across 24 human targets from the Bowes list [9]. Statistical performance has been analysed based on model sensitivity (SE), specificity (SP), accuracy (ACC) and Matthews correlation coefficient (MCC). In the consensus approach, predictions were kept where the models agreed and labelled "inconclusive" where they disagreed. [5]

	Test Data			
	SE	**SP**	**ACC**	**MCC**
Structural Alerts	86.6	90.9	90.2	0.782
Random Forests	91.5	86.8	91.3	0.804
Neural Network	90.3	90.1	91.9	0.818
Consensus	92.8	93.8	94.9	0.882

Model predictivity is relatively consistent across the three approaches, with neural networks performing the best overall, followed by random forests. The models were found to agree on 90.4% of all predictions, and a notable increase in predictivity is seen.

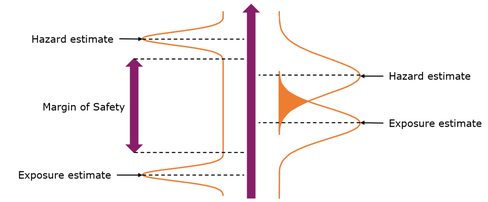
The ICH M7 regulatory guidelines [10] provide an avenue for *in silico* tools to see greater use. The use of complementary modelling procedures is required for these guidelines to be met, and while both our models are statistically derived, they may help bring this conversation forwards.

All binary predictors constructed, model build codes developed and datasets extracted in this work are freely available online through GitHub;

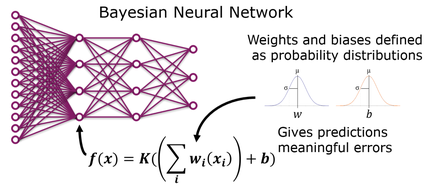
https://github.com/teha2/chemical_toxicology

Quantitative Predictions

Binary predictions allow the rapid evaluation of chemical safety suitable for screening, but for risk assessment information on the quantitative activity and its uncertainty is required for comparison to chemical exposure.

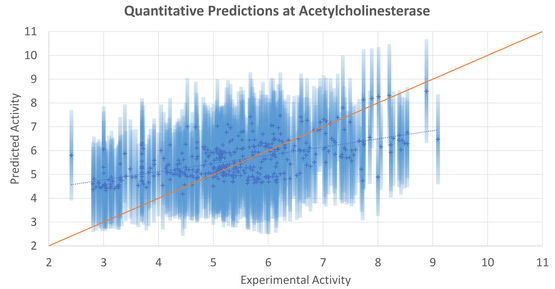


Neural networks have been be trained on quantitative ChEMBL [5] activity data to provide regression models using linear outputs and mean absolute error as the loss function and evaluation statistic. These have then be improved using Bayesian learning in TensorFlow Probability with Dense Variational Layers to replace point value weights and biases throughout the network with probability distributions. This allows the quantitative output to be produced using a Monte Carlo simulation of 500 iterations as a mean and standard deviation, representing the activity estimate and its uncertainty.



Bayesian Learning

Bayesian neural networks have been developed for 21 human MIEs. These models produce high quality quantitative activity estimates (p(IC50), p(EC50), p(Ki), p(Kd)) with average mean absolute errors of 0.6208 ± 0.0505 in test data and 0.9432 ± 0.2229 in external validation data.



Example results for acetylcholinesterase on an external validation set are shown above including 95% confidence intervals. These models have been shown to be able to distinguish between chemicals from the training set, external test set and randomised input strings by producing increasing standard deviations for each of these inputs.

Quantification of uncertainty is considered one of the biggest challenges required for next generation risk assessment (NGRA) to succeed. The Bayesian learning models presented here provide the desired uncertainty values that can be fed into an NGRA procedure or a quantitative adverse outcome pathway for the safety evaluation of a novel chemical.

Conclusions

- Structural alerts, random forest and neural networks have been combined to build high performing models for human MIEs.

- Networks can be updated to provide quantitative predictions more suitable for risk assessment.

- Bayesian learning can further improve these regression models by modelling uncertainty in predictions.

Acknowledgements

Unilever
MRC Toxicology Unit
Centre for Molecular Informatics
St. John's College, Cambridge

References

1. Allen T. E. H. *et. al., Chem. Res. Toxicol.* (2014) 27, 2100-2112.
2. Gutsell S, Russel P. J., *Toxicol. Res.* (2013) 2, 299-307.
3. Allen T. E. H. *et. al., Chem. Res. Toxicol.* (2016) 29, 2060-2070.
4. Ankley G. T. *et al., Environ. Toxicol. Chem.* (2010) 29, 730-741.
5. Wedlake A.J. *et. al., Chem. Res. Toxicol.* (2020) 33, 388-401.
6. Allen T. E. H. *et. al., Chem. Sci.* (2020) 11, 7335-7348.
7. ChEMBL Database (Version 23): *www.ebi.ac.uk/chembl*
8. ToxCast: *https://www.epa.gov/chemical-research/toxicity-forecasting*
9. J. Bowes. *et al, Nature Reviews* (2012) 11, 909-922.
10. European Medicines Agency (2014) ICH M7 Guidelines for Mutagenic Impurities.