Gaining Confidence in Combined Methods - Uniting Structural Alerts, Random Forests and Neural Networks for Use in Risk Assessment

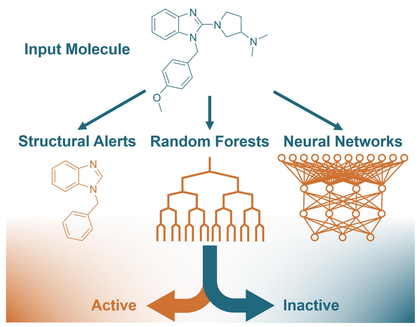
Timothy E H Allen[1,2], Andrew J Wedlake[2], Maria Folia[3], Sam Piechota[3], Elena Gelžinytė[2], Jonathan M Goodman[2], Steve Gutsell[3], Predrag Kukic[3], Paul J Russell.[3]

1. MRC Toxicology Unit, Hodgkin Building, Lancaster Road, Leicester, LE1 7HB, United Kingdom **2. Centre for Molecular Informatics, Department of Chemistry,** Lensfield Road, Cambridge, CB2 1EW, United Kingdom **3. Unilever Safety and Environmental Assurance Centre,** Colworth Science Park, Sharnbrook, Bedfordshire, MK44 1LQ, United Kingdom

Introduction

The molecular initiating event (MIE) [1,2,3] can be thought of as a gateway to the adverse outcome pathway (AOP) [4] - the initial chemical interaction. 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. Random forest models were constructed using sklearn and RDKit in Python 3, with 200 physicochemical descriptors as the input. Neural networks were developed with extended connectivity fingerprints as features in Python 3 using TensorFlow. A variety of network architectures, activation functions and hyperparameters were considered and optimised to give the highest level of statistical performance.



Binary Prediction Performance

Models were trained and evaluated on a consistent dataset extracted from the publicly available databases ChEMBL [5] and ToxCast [6] across 79 human targets, including some from the Bowes [7] and Sipes [8] lists. Statistical performance has been analysed based on model sensitivity (SE), specificity (SP), accuracy (ACC) and Matthews correlation coefficient (MCC). Average model performance is shown below.

	Training Data			
	SE	SP	ACC	MCC
Structural Alerts	91.0	95.8	95.0	0.882
Random Forests	94.9	94.7	96.4	0.915
Neural Networks	92.3	96.8	95.9	0.904

	Test Data			
	SE	SP	ACC	MCC
Structural Alerts	84.1	93.5	91.1	0.790
Random Forests	89.0	90.4	92.2	0.815
Neural Networks	87.9	93.6	92.8	0.832

Model predictivity is relatively consistent across the three approaches, with neural networks performing the best overall, followed by random forests. Random forests appear to perform better at predicting experimental negatives, while structural alerts and neural networks perform better on experimental positives.

A Consensus Approach

Making comparisons between different modelling approaches allows us to identify their strengths and weaknesses. However, different models need not always be viewed in competition. By combining the structure-based structural alerts and physicochemical-based random forests a consensus model can be produced. This was done across 24 biological targets. Predictions were kept where the models agreed and labelled “inconclusive” where they disagreed. [9]

	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
Consensus	92.2	92.6	94.1	0.865

The models were found to agree on 92.1% of all predictions, and a notable increase in predictivity is seen. This combination approach increases confidence in computational predictions and allows predictions derived from different types of molecular features.

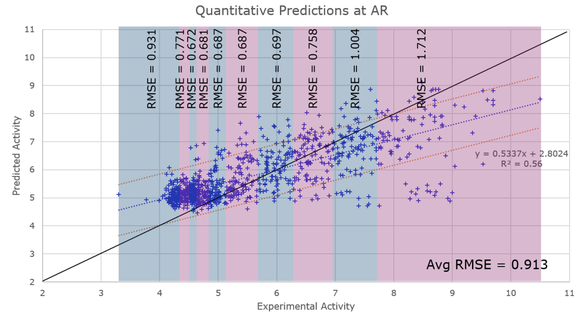
The ICH M7 regulatory guidelines [10] provide an avenue for *in silico* tools to see greater use. The use of complementary modelling procedures is required for these guidelines to be met, and while both our models are statistically derived, they may help bring this conversation forwards.

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

https://github.com/teha2/chemical_toxicology

Quantitative Predictions

Neural networks for binary activity prediction have been adapted for quantitative activity predictions based on data extracted from ChEMBL [5]. These predictions are more suitable for tasks in risk assessment where quantitative values are required for comparison to chemical exposures.



Here predicted activity at the androgen receptor (AR) is plotted against experimental activity for test set data. The graph is broken down into coloured blocks for each 10% of the test set. Root mean squared error values are shown for each section and in total. The data is skewed to p(Activity) vales between 4 and 5, making this a challenging prediction task. Despite this the network performs well, averaging predictions within one Log unit.

Conclusions

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

- These models have been combined to provide an increase in model performance.

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

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. ChEMBL Database: *www.ebi.ac.uk/chembl*
6. ToxCast: *https://www.epa.gov/chemical-research/toxicity-forecasting*
7. J. Bowes. *et al, Nature Reviews* (2012) 11, 909-922.
8. N. Sipes. *et al, Chem. Res. Toxicol.* (2013) 26, 878-895.
9. Wedlake A.J. *et. al., Chem. Res. Toxicol.* (2020) 33, 388-401.
10. European Medicines Agency (2014) ICH M7 Guidelines for Mutagenic Impurities.