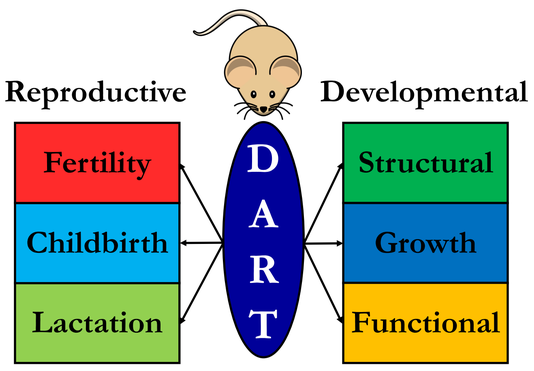
USING AUTOMATED MACHINE LEARNING FOR THE PREDICTION OF DEVELOPMENTAL AND REPRODUCTIVE TOXICITY

MARCUS W. H., WANG[A]; JONATHAN M., GOODMAN[A]; KATARZYNA R., PRZYBYLAK[B]

INTRODUCTION

- Developmental and reproductive toxicity (DART) is an important regulatory endpoint in health hazard assessment.
- Need to comply with the regulatory ban on animal testing for finished cosmetic products and cosmetic ingredients
- Important to develop non-animal testing alternative approaches such as *in silico* methods to assess the potential DART toxicities of chemicals.
- Limited data in the literature, coupled with the complex nature of the DART endpoint makes it a challenging endpoint to study.
- Machine learning models in an automated machine learning (AutoML) process have been used to investigate DART in this work.



METHODOLOGY

- Data was compiled from 11 different sources (includes existing data in the literature such as from datasets used to build models) covering a range of endpoints in DART. The toxicity value for each compound was recorded for each source.
- This gives the largest known database with 3622 compounds (1900 positives, 1722 negatives) for the general prediction of DART. This dataset will henceforth be referred to as the DART dataset.
- An AutoML process (AutoGluon package) was used to develop ML models on this database.
- Morgan fingerprints with 2048 bits and radius 2 were used as the features for model input.

BENCHMARK RESULTS

Model	SE (%)	SP (%)	Accuracy (%)	MCC
Jiang et al. 2019	78.5	88.1	83.6	-
Feng et al. 2021	77.3	90.7	84.4	-
ExtraTreesGini_BAG_L1	80.6 ± 2.0	88.8 ± 1.2	84.8 ± 1.5	0.697 ± 0.030

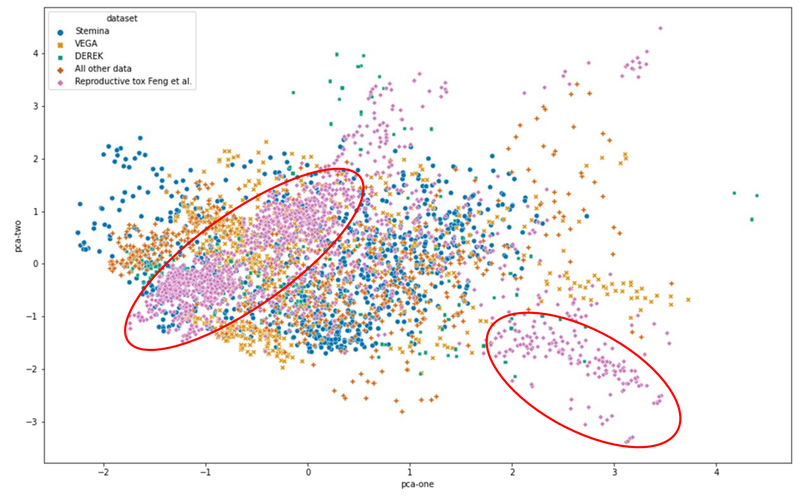
- AutoML models were benchmarked against literature results.
- Better accuracy than all results so far on the benchmark dataset.
- AutoML process can subsequently be used for the DART dataset.

VARYING TOXICITY THRESHOLDS

Threshold (for positives) for overall toxicity	SE (%)	SP (%)	Accuracy (%)	MCC
At least 1 positive	76.7 ± 1.8	70.6 ± 3.0	73.9 ± 1.8	0.475 ± 0.035
Positives more than negatives	68.0 ± 2.7	80.0 ± 1.6	74.5 ± 0.8	0.485 ± 0.019

- A screening tool requires high sensitivity (SE).
- A confirmation tool requires high specificity (SP).
- Model performance is stable across both thresholds.
- Further improvements unlikely given consistent performance of all models so far.
- Model performance ultimately limited by current data quantity and quality.

DATA VISUALISATION



- A Principal Component analysis (PCA) plot was constructed for the DART database.
- Easier to find similar chemicals for reproductive toxicity (main clusters circled in red) (Feng et al.) as compared to developmental toxicity.
- Plot matches expectations on complexity of developmental and reproductive aspects of DART.
- Areas with low number of data points in the plot highlights the need for more experimental data.

MODEL PERFORMANCE ON DART

Model	SE (%)	SP (%)	Accuracy (%)	MCC
ExtraTreesGini_BAG_L1	76.7 ± 1.8	70.6 ± 3.0	73.9 ± 1.8	0.475 ± 0.035
ExtraTreesGini_BAG_L2	75.1 ± 2.2	72.4 ± 2.7	73.8 ± 1.7	0.474 ± 0.034
RandomForest-Gini_BAG_L2	74.8 ± 2.2	72.5 ± 3.2	73.7 ± 1.8	0.473 ± 0.035

- 24 ML models with a variety of algorithms were trained using the DART dataset compiled in this study.
- Top three models for predicting DART are shown.
- Results are reported as an average across five runs.
- Consistent results with low standard deviations across all five runs for all 24 models.
- Reasonable results given complexity of DART and lack of quality data.

CONCLUSION

1. About 3600 compounds for the general prediction of DART has been compiled.

Largest known database

2. The best-performing machine learning model has been developed for predicting the general DART endpoint.

Best accuracy of 74%

REFERENCES

1. Rajagopal, R., Baltazar, M.T., Carmichael, P.L., Dent, M.P., Head, J., Li, H., Muller, I., Reynolds, J., Sadh, K., Simpson, W. and Spriggs, S., **2022**. Beyond AOPs: A Mechanistic Evaluation of NAMs in DART Testing. Frontiers in toxicology, 4.
2. Feng, H., Zhang, L., Li, S., Liu, L., Yang, T., Yang, P., Zhao, J., Arkin, I.T. and Liu, H., **2021**. Predicting the reproductive toxicity of chemicals using ensemble learning methods and molecular fingerprints. *Toxicology Letters*, 340, pp.4-14.
3. Fentem, J., Malcomber, I., Maxwell, G. and Westmoreland, C., **2021**. Upholding the EU's Commitment to 'Animal Testing as a Last Resort'Under REACH Requires a Paradigm Shift in How We Assess Chemical Safety to Close the Gap Between Regulatory Testing and Modern Safety Science. Alternatives to Laboratory Animals, 49(4), pp.122-132.
4. Baltazar, M.T., Cable, S., Carmichael, P.L., Cubberley, R., Cull, T., Delagrange, M., Dent, M.P., Hatherell, S., Houghton, J., Kukic, P. and Li, H., **2020**. A next-generation risk assessment case study for coumarin in cosmetic products. Toxicological Sciences, 176(1), pp.236-252.
5. Jiang, C., Yang, H., Di, P., Li, W., Tang, Y. and Liu, G., **2019**. In silico prediction of chemical reproductive toxicity using machine learning. *Journal of Applied Toxicology*, 39(6), pp.844-854.
6. Dent, M., Amaral, R.T., Da Silva, P.A., Ansell, J., Boisleve, F., Hatao, M., Hirose, A., Kasai, Y., Kern, P., Kreiling, R. and Milstein, S., **2018**. Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. Computational Toxicology, 7, pp.20-26.
7. Wu, S., Fisher, J., Naciff, J., Laufersweiler, M., Lester, C., Daston, G. and Blackburn, K., **2013**. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chemical research in toxicology, 26(12), pp.1840-1861.
8. Hewitt, M., Ellison, C.M., Enoch, S.J., Madden, J.C. and Cronin, M.T.D., **2010**. Integrating (Q) SAR models, expert systems and read-across approaches for the prediction of developmental toxicity. *Reproductive Toxicology*, 30(1), pp.147-160.

CONTACT DETAILS

Marcus Wang (mwhw3@cam.ac.uk)

[A] Centre for Molecular Informatics, Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

[B] Unilever Safety and Environmental Assurance Centre, Colworth Science Park, Sharnbrook, Bedfordshire MK44 1LQ, United Kingdom

ACKNOWLEDGEMENTS

We thank Unilever for funding this study.