# Demographic Factors in Drug-Induced Cardiotoxicity Prediction Models

M. Iwan[1,2,*], A. Roncaglioni[2], F. Grisoni[1]
E-mail: m.iwan@tue.nl

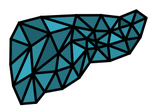
[1] Eindhoven University of Technology
[2] Istituto di Ricerche Farmacologiche Mario Negri IRCCS
* Current affiliations: TU/e, Bayer AG

## Introduction

Drug-induced cardiotoxicity (DICT) represents a significant challenge in drug development, accounting for approximately 14% of all post-market drug withdrawals.[1] The associated risk is modulated by demographic factors that are not yet fully characterised. Despite this, existing *in silico* DICT prediction models implicitly assume that all patients respond similarly to medications, regardless of sex, age, or body mass.

Women face higher risk of arrhythmia at equivalent doses.[2]

Elderly patients show reduced renal clearance and elevated drug plasma concentrations.[3]



Obesity alters hepatic drug distribution and increases comorbidity risk.[4]

## Motivation and Aims

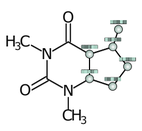
The FDA Adverse Event Reporting System (FAERS) contains millions of adverse event reports—with demographic information—yet has never been used to develop demographically-aware DICT models. Using FAERS, we developed CARBIDE, a collection of 27 dataset variants, to evaluate if useful structure-demographic interactions can be learned from pharmacovigilance data.

## Data processing

I) Data source: FAERS database[5]

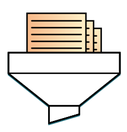
- 2012–2024
- 7.2 million patients
- Information on sex, age, and weight

II) Drug names processing



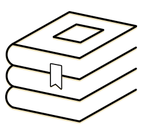
- 311,451 unique drug descriptions
- 8,260 unique drug combinations
- 4,333 unique compounds

III) FAERS filtering



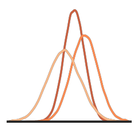
- Monotherapy
- Most likely drug
- Up to 10 drugs

IV) Adverse Events classification



- Focused cardiac terms
- Expanded cardiac terms
- Cardiovascular terms

V) Disproportionality Analysis

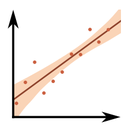


- Proportional Reporting Ratio
- Reporting Odds Ratio
- Information Component

VI) 27 CARBIDE variants

- 15–45 thousand entries
- 950–2,000 unique compounds
- 54–62% cardiotoxic entries

VII) ML models training and ablation studies



- 3 model types
- 6 molecular descriptors
- Feature contribution analysis

VIII) Results and analysis

## Results

Results shown for the optimal CARBIDE variant using an XGBoost model with ECFP molecular fingerprints.

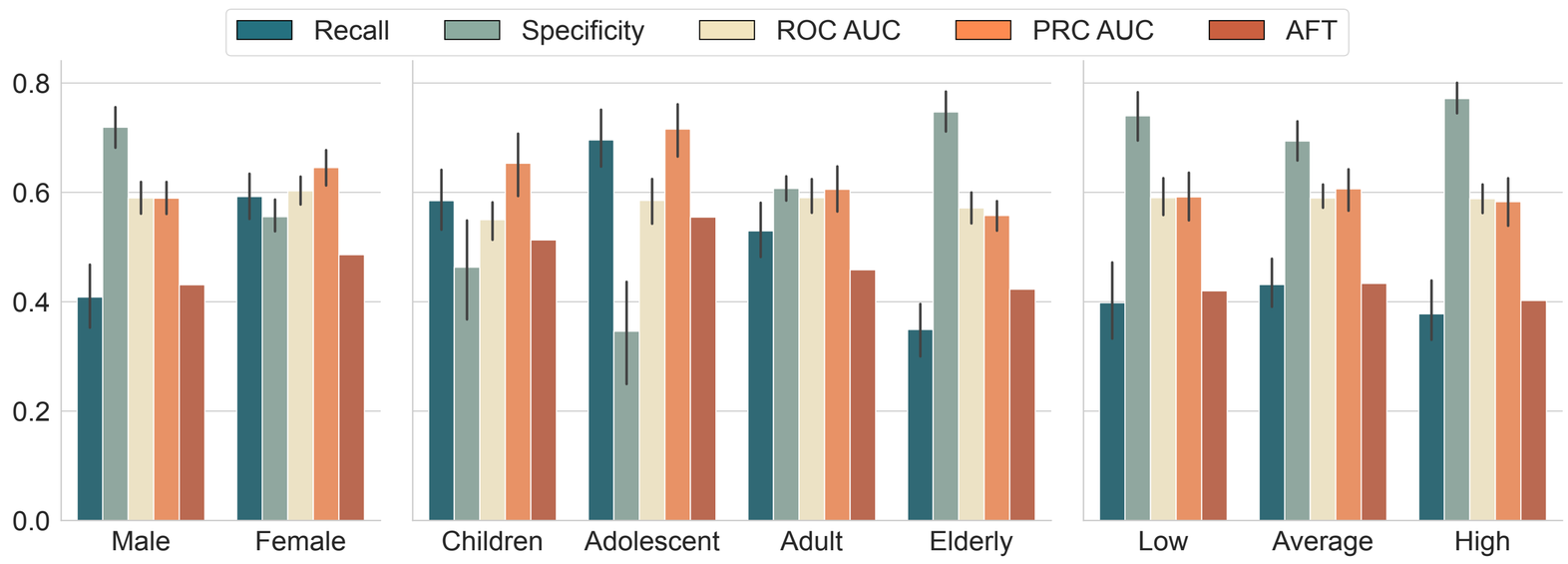



Figure 1. Performance of trained ML models in subpopulations. ROC AUC and PRC AUC are generally comparable across groups; Recall–Specificity ratio (RS ratio) varies substantially. AFT—Adjusted Fraction of Toxic compounds.

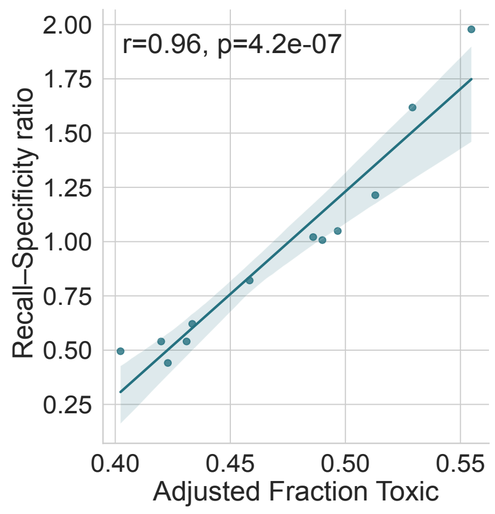



Figure 2. RS ratio and Adjusted Fraction of Toxic compounds show a near-perfect linear relationship.

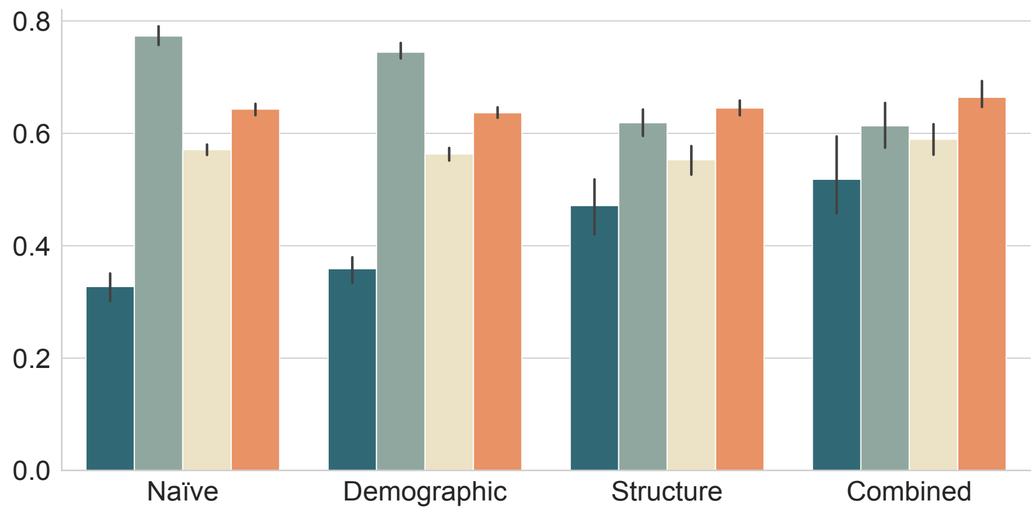
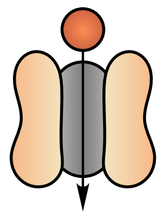



Figure 3. Performance of Naïve model predicting the per-subpopulation majority class vs machine learning models trained on only the Demographic, Structural or on Combined descriptors.

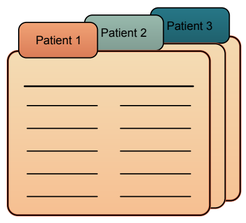
## Conclusions

- Molecular structure alone captures most predictive signal: structure-only models match the combined model, with no substantial benefit from adding demographic features.
- Models exploit demographic features as a prior, learning population-level cardiotoxicity statistics rather than meaningful drug–patient interactions.
- Our findings align with known limitations of pharmacovigilance data—reporting biases, duplicates, and noise—and suggest that DPA alone is insufficient to overcome them.

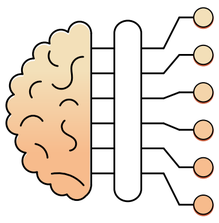
## Future directions



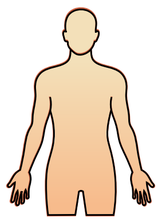
Biological features




Health records



Deep learning



PBPK models

## Funding

This study was funded by the Horizon Europe funding programme, under the Marie Skłodowska-Curie Actions Doctoral Networks grant agreement "Explainable AI for Molecules - AiChemist" no. 101120466.

## References

[1] I. J. Onakpoya et al., *BMC Med.*, vol. 14 (**2016**)
[2] M. Holmes et al., *J. Mol. Cell. Cardiol. Plus*, vol. 13 (**2025**)
[3] N. N. Ngcobo et al., *Clin. Pharmacokinet.*, vol. 64 (**2025**)
[4] Z. Almuwaqqat et al., *JAMA Netw. Open*, vol. 7 (**2024**)
[5] U.S. Food and Drug Administration, *FAERS*

Full paper