Initial Insights into Cost-Efficient AI Toxicity Profiling: Cell Painting + Chemical Structures

Krzysztof Rataj, Aleksandra Nowak, Magdalena Otrocka, Adriana Borowa, Nasim Jamali, Michał Warchoł
Ardigen, Kraków, Poland

ABSTRACT

Toxicity is one of the most common reasons for failure in late stage drug development, accounting for 30% of non-successful projects[1]. While multiple assays for various toxicity types are available, they usually are quite costly and unfit to be run in a high-throughput setting. This makes them only feasible at the late stage drug discovery, where failure results in a significant time and resource penalty to the project. Quantitative Structure-Activity Relationship (QSAR) toxicity prediction methods could be applied at the earlier stage of the projects, however, their use is limited to the chemical space of the training data and the method often fails for novel chemical structures.

Here we present an endeavor to create a dataset of small molecule treatments visualized with the target-agnostic Cell Painting assay[2], that captures the early onset of various types of toxicity, such as hepato-, cardio-, and genotoxicity. This data, screened in relevant cell lines and analyzed with multimodal Artificial Intelligence algorithms allows us to build models capable of assessing the treatment toxicity with high accuracy and throughput.

For the proof of concept project we selected a single hepatocyte HUH7 cell line and tested an initial small set of compounds with known toxicities. The results show that using multiple modalities, here image and molecular structure, improves the predictive power over single modality models. Additionally, we are able to detect and predict cell-line independent toxicities, such as cardiotoxicity signals in a hepatocyte cell line.

DATASET

We generated a data set of Cell Painting images for approximately 200 compounds with public toxicity annotations from various databases, including DILIst[3] and hERGCentral[4]. The compounds were tested in HUH7 hepatocyte cell line at 5 concentrations in 2 technical and 2 biological replicates.

Data acquisition

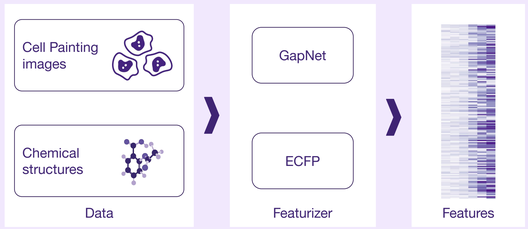
CONCLUSIONS

- **A multimodal approach improves the efficiency of predictive models:** Combining Cell Painting and chemical structures consistently improved predictive toxicity models, yielding better ROC AUC values across nearly all tests.
- **A bigger dataset is needed to train better models:** Though well-performing toxicity prediction classifiers could be trained even using a small dataset, larger data set is essential to achieve significant improvement in model performance and predictive accuracy.
- **Compound concentration range should be optimised:** The low concentrations used to avoid simple cell death failed to induce morphological changes in most compounds. Adjusting the concentration range is needed to capture relevant toxicity signals.

METHODS

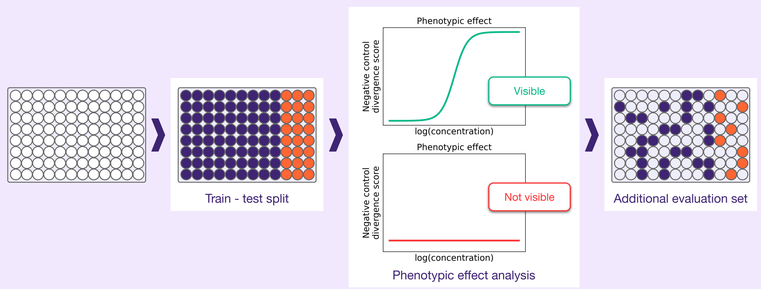
Feature generation

Image features were generated using convolutional neural network GapNet and ECFP was applied to featurise chemical structures.



Evaluation scheme

The data was split into train - test split where for each of the considered toxicity types 80% of compounds known to have that toxicity was used to train the AI model, while the remaining 20% were used to test it. This split was performed 20 times with different random seeds to ensure that the effect of compound selection would be accounted for. Additionally, a separate subset of compounds with visible effect (approximately 25% of the compounds) on cell morphology was selected for additional evaluation.

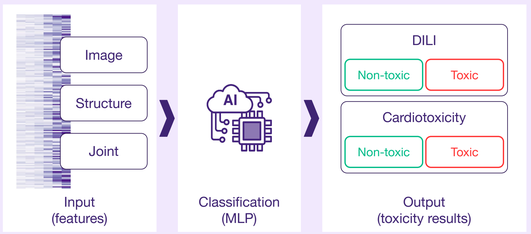


Predictions

We tested multiple training setups where Multiple Layer Perceptron (MLP) is trained to predict known toxicity. Training setups differ in: modality (image, structure, or joint), toxicity type (DILI - Drug-Induced Liver Injury and Cardiotoxicity), and the definition of the non-toxic class. We highlight 3 different possible sets of non-toxic classes.

Non-toxic

- **DMSO** vehicle treatment,
- **Negcon** non-toxic compounds,
- **Other** compounds toxic in different type of toxicity than the tested one.



RESULTS

- Comparison of results across various data setups indicates that multimodal models that combine image and chemical structure representations achieve better average performance for both DILI and cardiotoxicity prediction, outperforming models using only a single data modality.
- The classifiers trained to distinguish toxic compounds from the DMSO control display predictably high performance. For models using imaging features, performance is high (ROC AUC > 0.7 in almost all cases), and it reaches the maximum (ROC AUC = ~1) for models combining structure and imaging features. However, performance significantly drops for models trained to distinguish between toxic and non-toxic compounds.
- Training the models exclusively on compounds showing visible phenotypic effects yielded high average performance, even for image-based models. Critically, this training subset also introduced the highest variability in the resulting ROC AUC values, depending on the specific train-test split used for the experiment.
- Interestingly, despite the experiments being conducted in a hepatocyte cell line, the cardiotoxicity prediction models showed good results—in some cases, even better than the DILI models. This counterintuitive finding may indicate that cardiotoxicity prediction based on general cellular morphology changes may be cell line-agnostic.

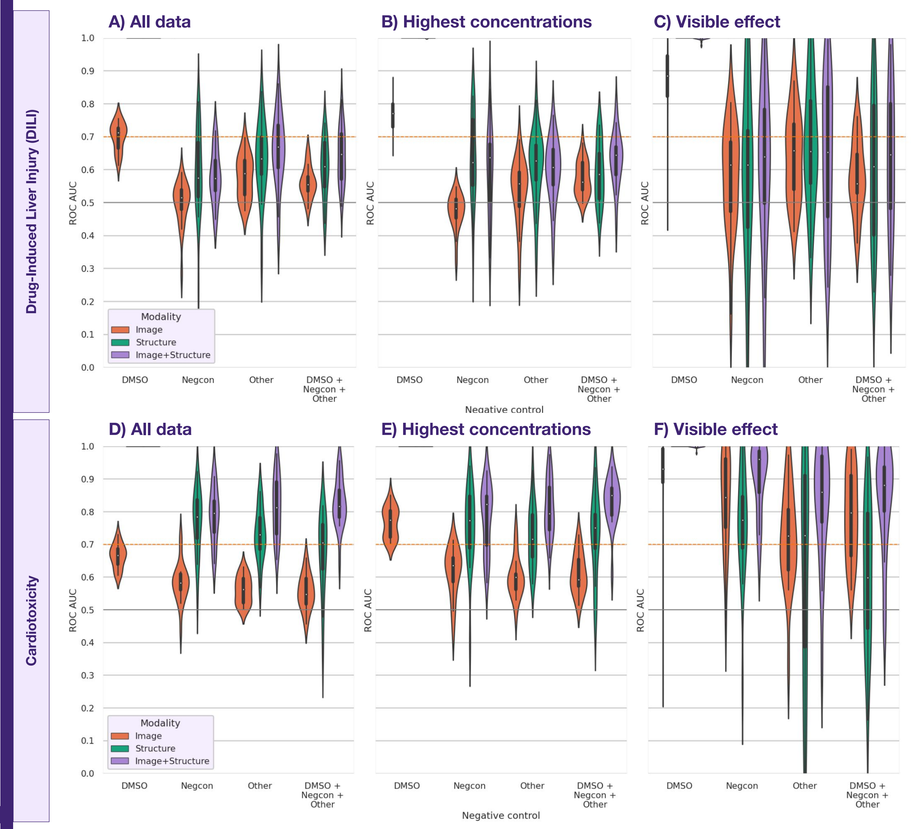
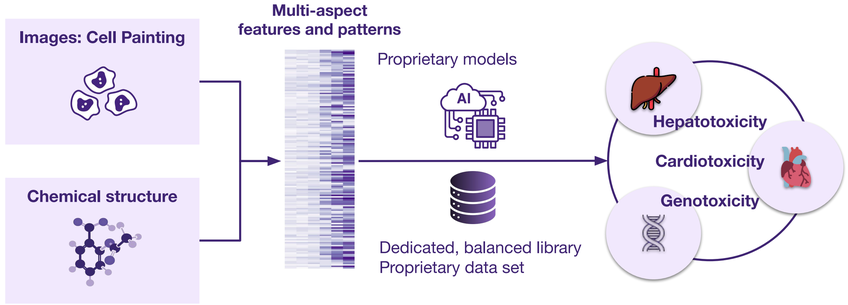


Fig. 1 Performance of MLPs trained to predict **Drug-Induced Liver Injury (DILI)** and **Cardiotoxicity** using:
A, D) All data: Model trained on all compounds across all tested concentrations,
B, E) Highest concentrations: Model trained only on data points from all compounds at the two highest tested concentrations,
C, F) Visible effect: Model trained only on data points corresponding to compounds that showed a visible phenotypic effect at the highest concentrations.
Violin plots illustrate the distribution of ROC AUC metric across all train-test splits. The models were trained to distinguish either DILI or Cardiotoxic compounds from DMSO, Negcon, Other (see Methods, Predictions) compounds, and combined controls. The plots are color-coded based on the data modalities used for training of each respective model (Image, Structure, joint: Image+Structure).

FUTURE PERSPECTIVE

Ardigen is building a toxicity prediction tool based on a dedicated Cell Painting dataset.



If you want to know more or collaborate on the project come to the booth #22

Toxicity tool advantages

- **Toxicity profiles:** multiple types of toxicity in a single experimental screen.
- **Time and cost-efficiency:** screening hundreds of compounds for multiple toxicity types in a single experiment within 4 weeks.
- **Higher chance of identification** of toxic compounds at an early stage of drug discovery project.

REFERENCES

1. Sun, Duxin, et al., 2022. “Why 90% of Clinical Drug Development Fails and How to Improve It?” Acta Pharmaceutica Sinica. B 12, no. 7 : 3049–62. https://doi.org/10.1016/j.apsb.2022.02.002.
2. Bray et al., 2016: "Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes". Nature Protocols. 11 (9): 1757–1774.
3. Thakkar et al., 2020: “Drug-induced liver injury severity and toxicity (DILIst): binary classification of 1279 drugs by human hepatotoxicity” Drug Discovery Today. 25 (1), 201-208.
4. Du et al. 2011: “hERGCentral: a large database to store, retrieve, and analyze compound-human Ether-à-go-go related gene channel interactions to facilitate cardiotoxicity assessment in drug development.” Assay Drug Dev Technol. 9 (6):580-8.

European Funds Smart Growth
Republic of Poland
NCBR National Centre for Research and Development
European Union European Regional Development Fund

This work was supported by Smart Growth Operational Programme 2014-2020/European Regional Development Fund (Project no. POIR.01.02.00-00-0017/17-00, HiScAI - Development of cell-based phenotypic platform based on high content imaging system integrated with artificial intelligence.