Integrating Cheminformatics and Phenotypic Screening: A Novel Approach for Enhanced Mode of Action Prediction

Michał Koziarski[1,2,], Piotr Gaiński[1,3,*], Krzysztof Rataj[1], Adriana Borowa[1,3], Konrad Wójtowicz[1,3], Jakub Gwóźdź[1], Magdalena Otrocka[1], Dawid Rymarczyk[1,3], Michał Warchoł[1]*

1. Ardigen, Kraków, Poland
2. Department of Systems and Computer Networks, Wrocław University of Science and Technology, Wrocław, Poland
3. Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland
** denotes equal contribution*

contact: **michal.warchol@ardigen.com**

ABSTRACT

The Cell Painting protocol, recognized as a key tool for drug discovery phenotypic screening, generates rich, single-cell resolution data. While traditional analysis methods have focused on image data, integrating chemical structure information of compounds offers a more comprehensive insight.
We combined both open-source descriptors for images and molecules as well as deep learning ones to enhance mode of action predictions. This multimodal approach marks a significant leap in HCS analysis, promising more accurate drug discovery outcomes.

DATASET

We used **a dataset released by Bray et al. [1]** to implement and test our approach. It is a publicly available dataset of High Content Screening (HCS) images and morphological profiles of 30,000 small-molecule treatments.

The dataset was generated by applying Cell Painting assay protocol. For **each compound, 6-48 fields of view were acquired** with five fluorescent channels. Figure 1 presents some examples.

Using data from the ChEMBL repository [4], we assigned **19 Modes of Action (MoAs) to 2221 compounds** from the dataset (Figure 4). One compound may have more than one MoA assigned.

To ensure reliability of the presented results, we use a **structural split** based on the hierarchical clustering of ECFP representations. As a result, the data was split into 1740 training and 481 test compounds.

METHODS

The goal of our work is to obtain the model accurately **predicting MoA** of a compound using HCS images and chemical structures as input. To address polypharmacology and the possibility of simultaneous modes of action, we defined our problem as a multi-label classification task. We compared a variety of data representations (human-defined and AI-based), as well as uni- and multimodal approaches.

Phenotypic representations. We compared two types of phenotypic representations: human-defined features obtained with Cell Profiler (CP) and AI-based features extracted from the penultimate layer of a deep convolutional neural network (GapNet-PL [2]). To obtain a phenotypic representation of a well, we used maximum aggregation over fields of view.

Structural representations. Along with the commonly used human-defined Extended-Connectivity Fingerprints (ECFP), we used deep feature representations from a proprietary graph transformer model: Relative Molecule Attention Transformer (R-MAT) [3].

Combining modalities. To fuse the visual and chemical modalities, we combined phenotypic and structural representations via concatenation. For the deep learning-based representations, we first trained the individual models in a unimodal setting, and used extracted features for concatenation.

Classification. Random Forest is used to obtain a final prediction.

RESULTS AND DISCUSSION

To analyze how meaningful the representation types are, we visualized the latent space using the UMAP algorithm. The representation obtained from deep learning-based method clusters compound of the same MoA together, thus creating more meaningful representations that increases the classifier accuracy (Figure 3).

The effectiveness of the MoA prediction models is measured using the ROC AUC metric. One can observe that the deep learning-based models exploiting both types of data achieve the best performance and obtain the highest ROC AUC score for each of the MoAs (Figure 4).

Using an averaged ROC AUC score (Figure 5), we conclude that **information fusion from both modalities (structural and phenotypic) is the most effective** due to their synergy (Figure 5).

CONCLUSIONS

We demonstrated a multimodal approach for the task of Mode of Action prediction. **Our analysis shows that phenotypic and structural modalities are complementary in case of both human-defined and deep learning representations.** Moreover, use of deep learning improves the performance and reduces the computation time of MoA prediction.

REFERENCES

[1] Bray, Mark-Anthony, et al. "*A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay.*" Gigascience 6.12 (2017).
[2] Rumetshofer, Elisabeth, et al. "*Human-level protein localization with convolutional neural networks.*" International conference on learning representations (2018).
[3] Maziarka, Łukasz, et al. "*Relative Molecule Self-Attention Transformer.*" arXiv preprint arXiv:2110.05841 (2021).
[4] Mendez, David, et al. "ChEMBL: towards direct deposition of bioassay data." Nucleic acids research 47.D1 (2019): D930-D940.
[5] McInnes, Leland, John Healy, and James Melville. "UMAP: Uniform manifold approximation and projection for dimension reduction." arXiv preprint arXiv:1802.03426 (2018).

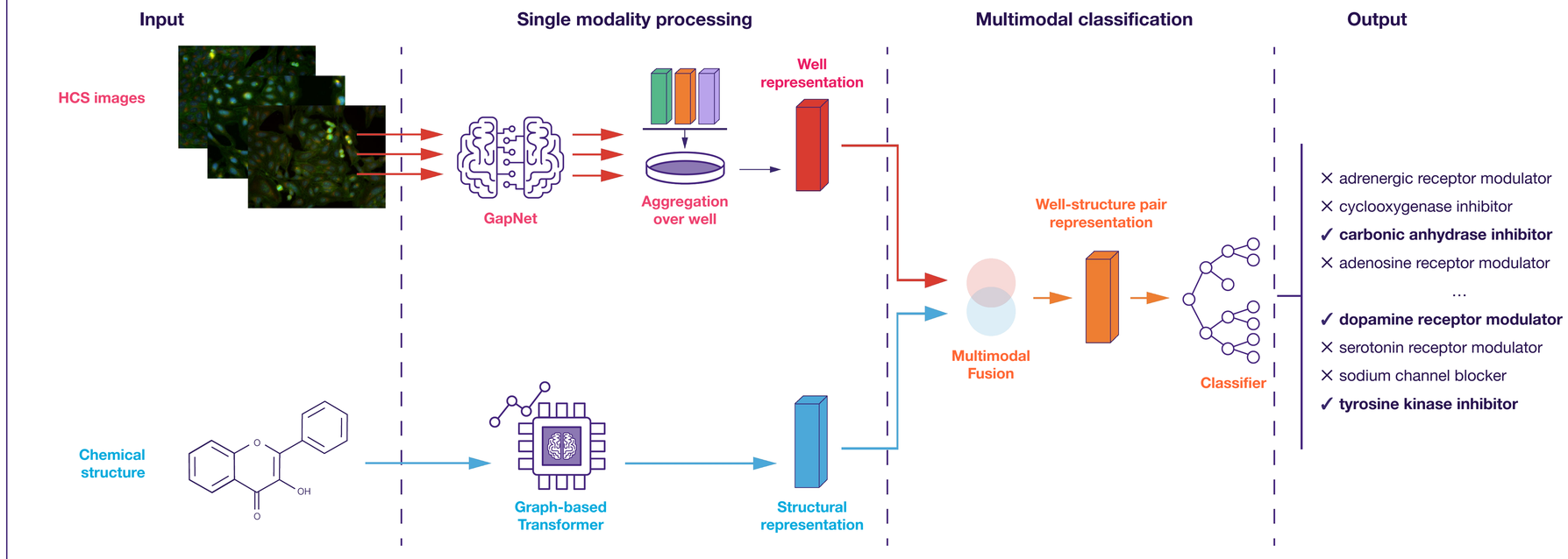

Figure 2. Architecture for multimodal MoA classification. Firstly, HCS images and compound structures are passed through deep learning architectures (GapNet and graph-based transformer, respectively) to obtain their multidimensional representations. Then, they are fused to create a multimodal feature vector passed to the MoA classifier. ✓ indicates MoA presence, × indicates its absence.

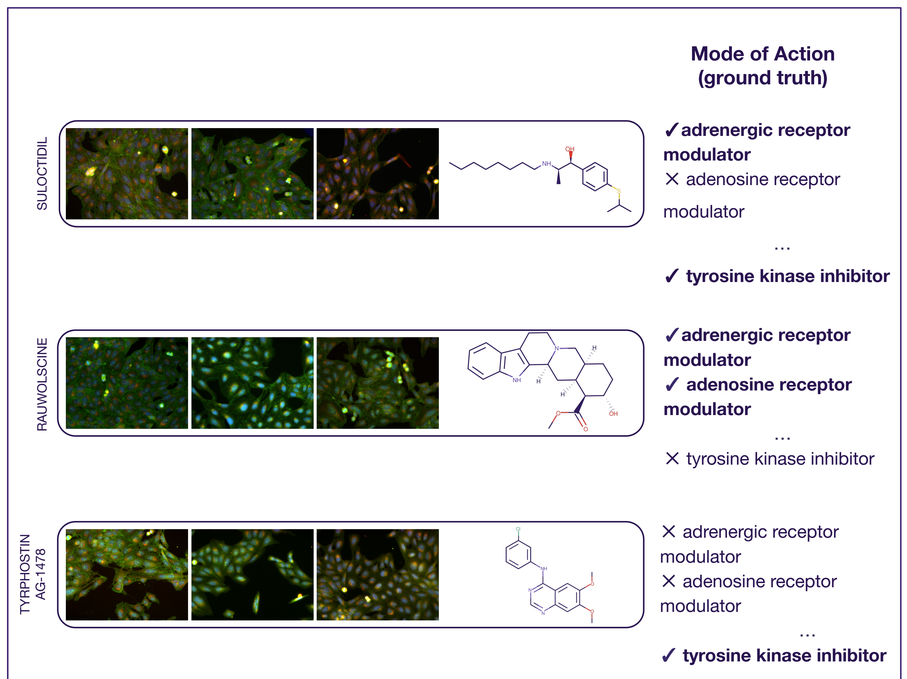

Figure 1. Randomly selected fields of view and associated chemical structures of compounds with diverse known MoAs. ✓ indicates assigned MoA, × indicates not assigned MoA (ground truth).

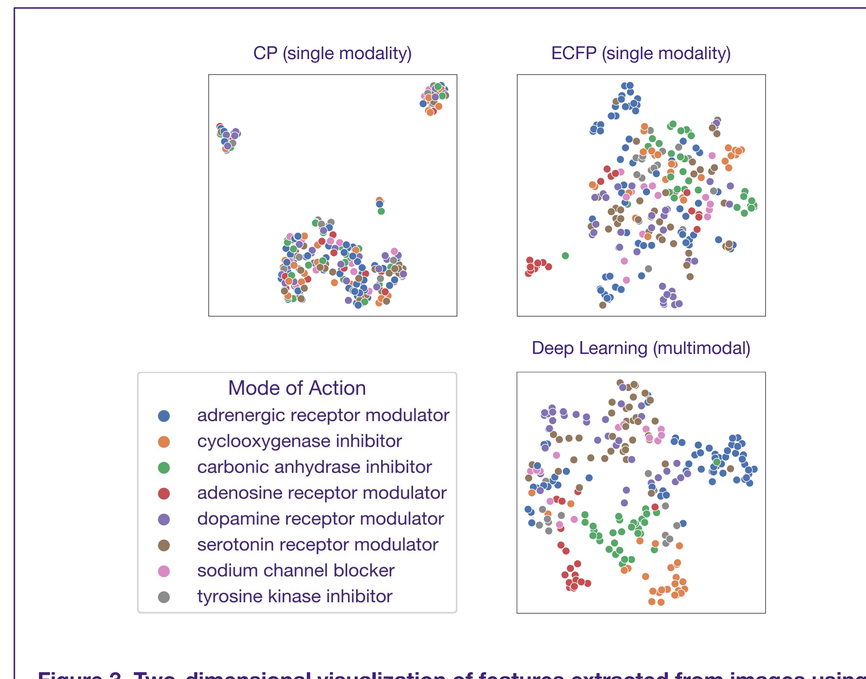

Figure 3. Two-dimensional visualization of features extracted from images using CellProfiler (CP), chemical structures (ECFP), and our multimodal approach based on deep learning embeddings. One point corresponded to a single compound and points were generated using the UMAP algorithm [5]. One can observe that the multimodal representations are superior in separating different MoAs to the human-defined features such as CP and ECFP.

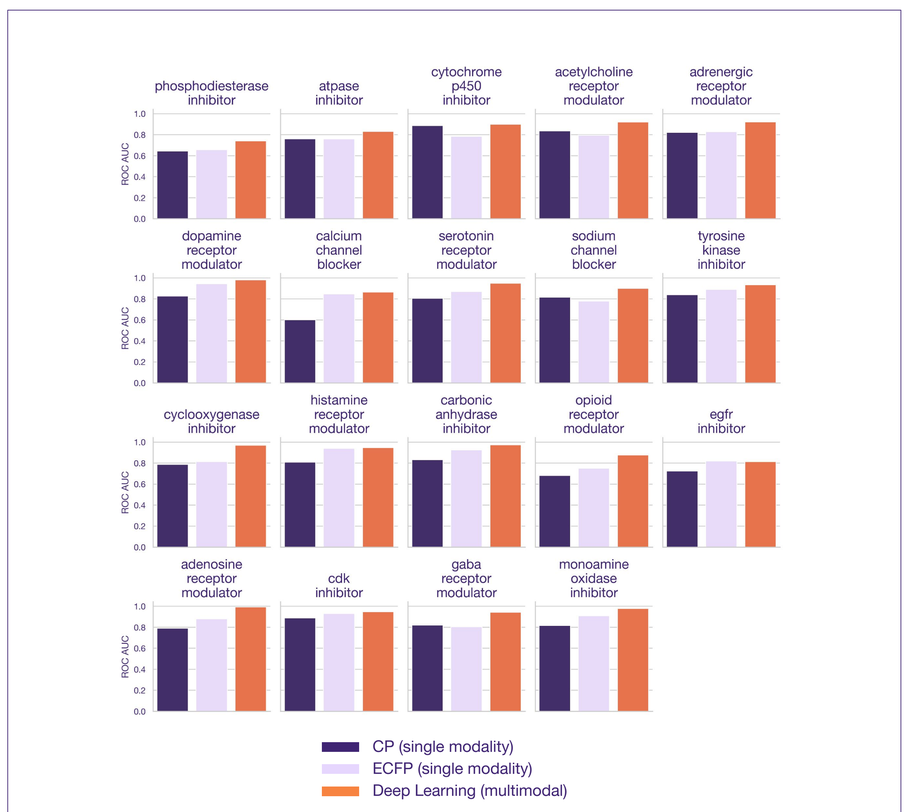

Figure 4. Performance of the models trained on different representations for each MoA measured by the ROC AUC score. Here, we compare a classifier trained on multimodal features, generated by the proposed model, with classifiers trained on human-defined image descriptors (CP) and chemical structure descriptors (ECFP). We observe that models trained on multimodal representations are superior to the ones trained on human-defined features.

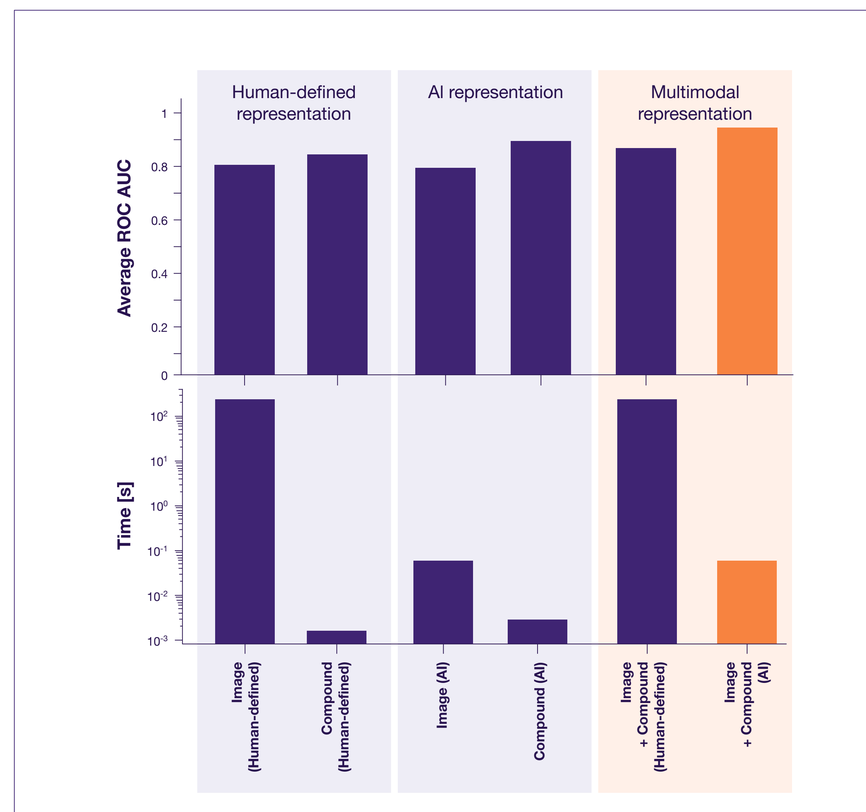

Figure 5. ROC AUC averaged over all of the considered MoAs (top) and inference time comparison (bottom). Deep feature representations outperform traditional chemical structure descriptors. Deep image representations, while achieving a comparable performance to CellProfiler features, are much faster to compute and synergize better in a multimodal model. Combining modalities significantly improves the performance over models trained on individual modalities.