Artificial Intelligence boosts Mode of Action and bioactivity prediction through the integration of Cell Painting images and small molecule structures

Maurycy Chronowski[1], Szymon Adamski[1], Sakshi Garg[2], Adriana Borowa[1, 3], Dawid Rymarczyk[1, 3], Magdalena Otrocka[1], Bolek Zapiec[2], Michał Warchoł[1]

1. Ardigen, Kraków, Poland
2. Merck Healthcare KGaA, Darmstadt, Germany
3. Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland

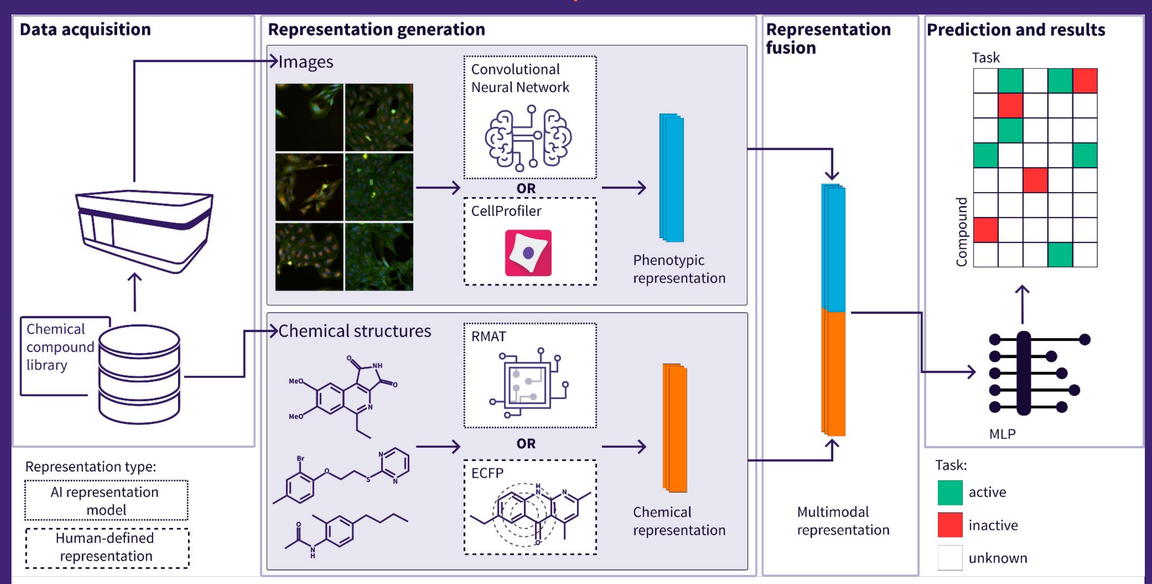

Figure 1. Method overview: Images and structures are processed using either AI model or human-defined features and passed through MLP to obtain predictions.

ABSTRACT

Advancements in the field of Artificial Intelligence opened paths to leveraging the potential of computational screening methods on an unprecedented scale. Vast datasets of High Content Screening (HCS) images can be used to train AI models and subsequently apply them to effectively support the drug discovery process; with the ambition of boosting speed and probability of success. Understanding how to make the best use of such datasets becomes more crucial the bigger our screening grows.

In our experiments, we demonstrate the effectiveness of predicting the Mode of Action (MoA) and biological activity of small molecules using Cell Painting, an HCS Assay which visualises cell phenotype in a target-agnostic manner. We further explore the approach of combining phenotypic features extracted from the images with encodings of chemical structures, which has been shown to improve the accuracy of MoA prediction. Our proposed setup compares how human-defined features score against AI representations, with the latter being also significantly more cost- and time-efficient to extract.

We benchmarked our method using a dataset consisting of thousands of compounds. For each input modality, we trained a neural network for compound activity prediction. We present a quantitative summary of the activity prediction performance and a qualitative comparison of how well different data modalities can predict different MoAs and activity against the targets of interest. We display the benefit of combining the phenotypic and chemical features, allowing to fully utilise the predictive power of Cell Painting datasets.

DATASET

We used **CellPainting images of U2OS cells** generated at Merck. The cells were treated with a library containing a wide range of small molecules from different sources, including proprietary compounds and a bioactive reference set. The latter is composed of **2505 molecules with DrugBank and Therapeutic Target Database (TTD) annotations, 153 SGC/KCGS probes and 147 Merck Legacy compounds from internal Merck projects**. Compounds were selected based on their highest clinical stage and selectivity (annotated targets per compound) and then prioritised in order to balance between the best possible target coverage and number of compounds per target (mainly below 10).

From the entire dataset, **we sampled 3 hold-out test sets** grouped by chemical scaffold similarity to avoid a structural information leak. For each hold-out, we generated 5 randomised train/validation splits repeated until achieving sufficient label coverage of at least 5 positive and 5 negative training examples (i.e. unique compounds, not replicates) per task. We scored **147 Modes of Action (MoAs) and 556 Bioactivity Properties** that passed the filtering. On average, we had 15 positive and 102 negative examples per task in the whole dataset. In each fold, the train set consisted of 8k-9k compounds, and the test set 1k-1.5k.

METHODS

We applied a method presented in Fig. 1 to analyse the performance of MoA and biological activity prediction models for small molecules using images and chemical structures. We start by generating a vector representation of both images and chemical structures, which is then concatenated to create a multimodal representation of a molecule. Then, this representation is passed through a Multi-Layer Perceptron (MLP) to obtain a prediction for each property. Each label describes either MoA or property and is assigned an active (1) or inactive (-1) value. Additionally, we compare a uni-modal approach where only one representation is used, either phenotypic or chemical.

Artificial Intelligence representations Phenotypic representation is generated using a Deep Convolutional Neural Network (GapNet [1]) pre-trained using a compound-matching task (DL). To generate representation of chemical structures, we use a proprietary model: Relative Molecule Attention Transformer (RMAT [2]).

Human-defined representations Morphological features of cells are obtained using CellProfiler (CP [3]) and structures are described using Extended-Connectivity Fingerprints (ECFP [4]).

RESULTS AND DISCUSSION

The combination of CP and RMAT provides the best results as depicted in Fig. 2 which presents a summary of results of all methods with respect to prediction tasks. Although, it is noteworthy that the difference between top performing methods is not statistically significant. On the other hand, the addition of DL image features enables more tasks to achieve the highest ROC AUC brackets. We observe performance gain both with RMAT and ECFP representations, compared to using only phenotypes as presented in Fig. 3 showing distributions of tasks for all modalities.

We observed that both, using a multimodal approach (Fig. 4A) as well as applying AI features (Fig. 4B), lead to the improvement of ROC AUC prediction score for most of the tested kinases from all families.

The improvement by multimodality is especially profound for CMGC kinases, where the activity was predicted with higher confidence for 11 out of 14 kinases and improvement was seen for all DYRK and GSK3Beta enzymes. Interestingly, for enzymes such as CDK2A, CDKE, CLK4 and DYRK1A, ROC AUC above 0.7 was achieved only with multimodal tasks.

By contrast, multimodal approach was not improving the prediction of the CAMK family but applying AI approach increased ROC AUC for 5 out of 6 kinases, whereas for CMGC improvement of activity prediction by AI methods was seen only for 6 out of 17 enzymes.

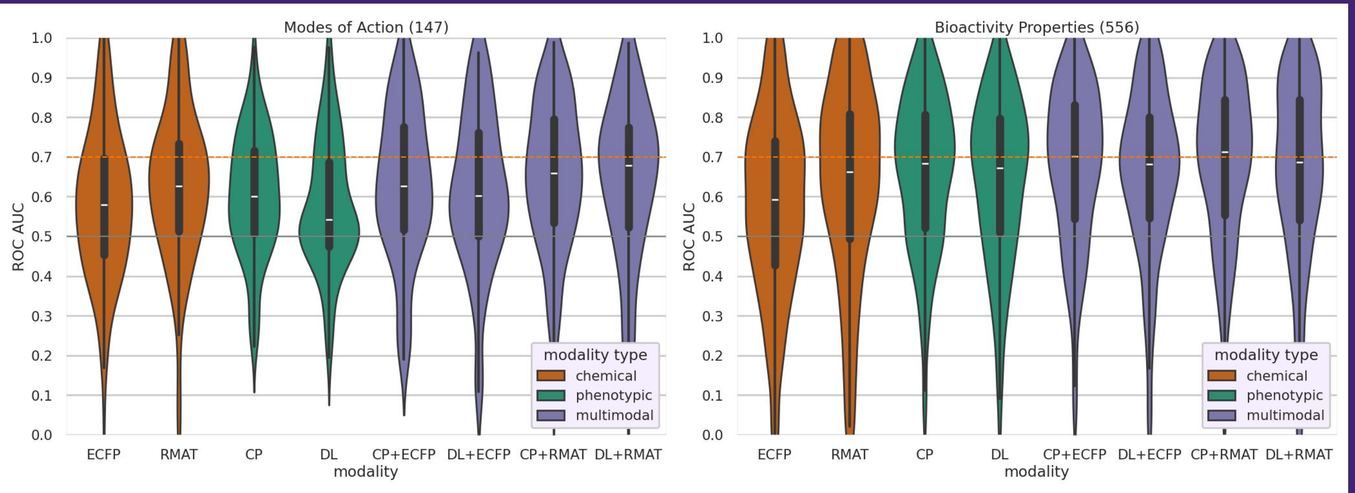

Figure 3. Comparison of distributions of scores obtained with each modality. On average, multimodality improves scores when compared to using only phenotypic representations.

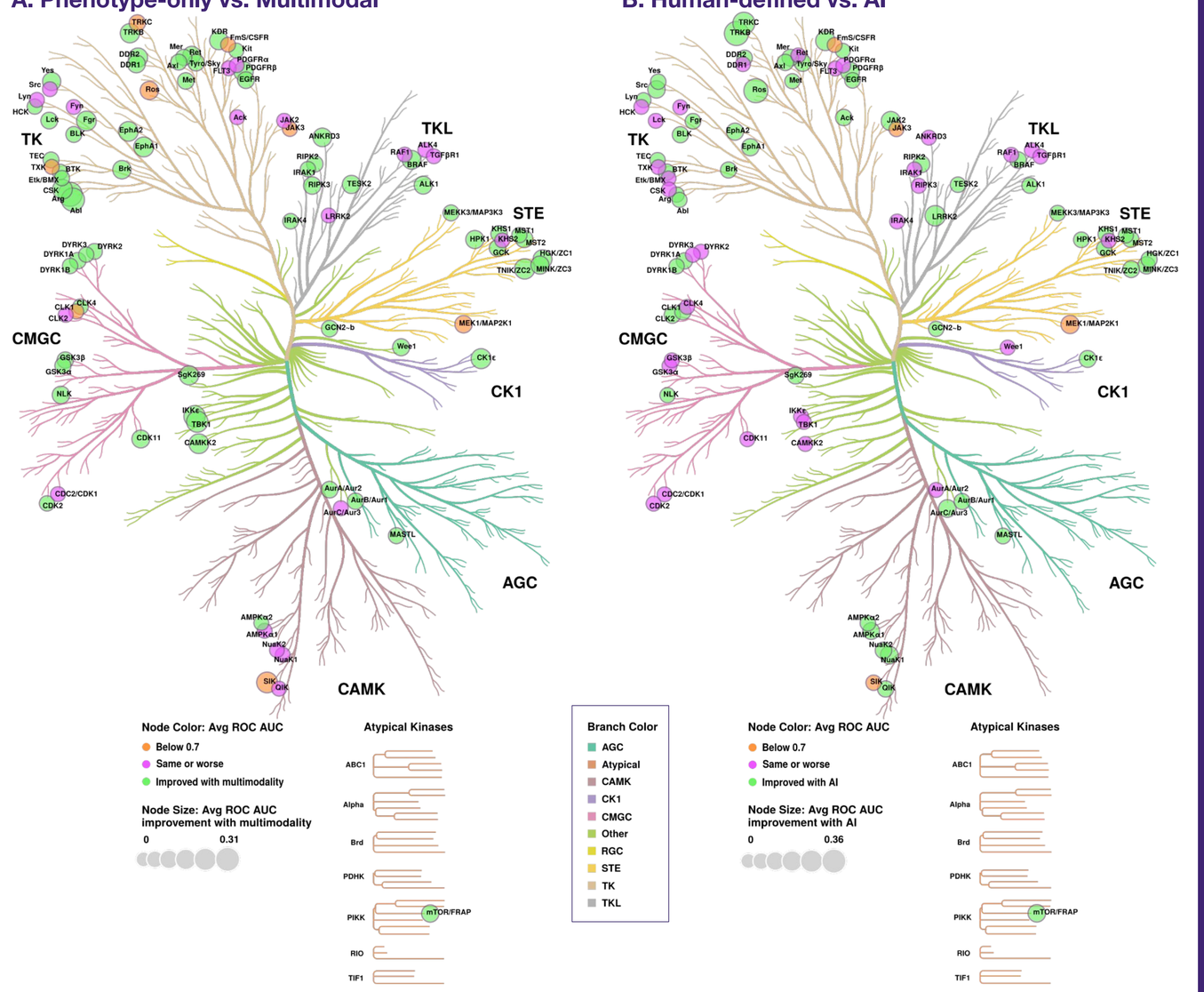

Figure 4. Kinome trees. Node color represents improvement of Avg ROC AUC while using A) multimodal approach or B) AI methods for particular kinases. Node size represents the values of this improvement. Utilisation of multimodality and AI-based representations leads to performance improvements in each family of the kinome. Tested kinases within the CAMK group appear to be exceptionally phenotype-driven with only one target being improved by chemical features.

	Modes of Action (147 tasks)				Bioactivity Properties (556 tasks)			
	Average ROC AUC	# tasks >= 0.7	# tasks >= 0.8	# tasks >= 0.9	Average ROC AUC	# tasks >= 0.7	# tasks >= 0.8	# tasks >= 0.9
ECFP	0.58 ± 0.12	36	19	9	0.58 ± 0.16	184	94	45
RMAT*	0.63 ± 0.11	52	23	12	0.63 ± 0.16	244	162	76
DL*	0.58 ± 0.09	35	14	4	0.64 ± 0.14	248	134	47
CP	0.60 ± 0.10	40	14	3	0.66 ± 0.14	257	148	55
DL+ECFP*	0.61 ± 0.12	47	27	**13**	0.66 ± 0.13	249	140	65
CP+ECFP	0.63 ± 0.11	53	33	12	0.68 ± 0.14	279	169	72
DL+RMAT*	0.64 ± 0.11	**65**	29	10	0.66 ± 0.16	262	**181**	**92**
CP+RMAT*	**0.65 ± 0.13**	62	**36**	11	**0.68 ± 0.14**	**287**	176	86

Figure 2. ROC AUC scores - numerical comparison between modalities. Asterisks denote modalities that utilise AI-based representations, as opposed to human-defined features and bold values represent the best result.

CONCLUSIONS

- **Multimodal representation consistently improves prediction scores** across all tested setups as showcased by our experiments.
- Cell phenotypes described with either **CellProfiler or Deep Learning features perform comparably** when supported by chemical representations. Therefore, other advantages shall be considered, like the straightforward interpretability of CP features, or the easiness of extraction and shorter computation time of DL representations.
- **AI-driven methods improve the quality of predictions on a biologically diverse collection of tasks** spanning over the entirety of the kinome.

REFERENCES

1. Rumetshofer, Elisabeth, et al. "Human-level protein localization with convolutional neural networks." International conference on learning representations (2018).
2. Maziarka, Łukasz, et al. Relative molecule self-attention transformer. J Cheminform 16, 3 (2024). DOI: 10.1186/s13321-023-00789-7
3. Stirling, D.R., et al. CellProfiler 4: improvements in speed, utility and usability. BMC Bioinformatics 22, 433 (2021). DOI: 10.1186/s12859-021-04344-9
4. Rogers, David, and Hahn, Mathew. Extended-Connectivity Fingerprints J. Chem. Inf. Model. (2010), 50, 5, 742–754 DOI: 10.1021/ci100050t
5. Manning, G., et al. The Protein Kinase Complement of the Human Genome (2002). Science 298:1912-1934