

DeepCIM: Prediction of myelosuppressive anti-cancer molecules using drug-target interaction datasets

Lee, Taeyeub¹, Thomas, Philipp², Serra Traynor, Carlos³

¹Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK ² Department of Mathematics, Imperial College London, London, UK ³Clinical Pharmacology and Quantitative Pharmacology, Clinical Pharmacology and Safety Science, R&D, AstraZeneca, Cambridge, UK

Abstract

DeepCIM is a new deep neural network that predicts chemotherapy-induced myelosuppression for small-molecule drugs. It integrates SMILES, protein sequences, and pIC50 values and was pre-trained on large datasets like DrugBank and BindingDB. External validation on the Davis dataset indicates the potential of such techniques for improving preclinical drug development strategies.

In an extension of our analysis, we compared NTP Gr3/4 rates between drugs and drug combinations selected by DeepCIM against those not selected, utilizing Flatiron real-world evidence (RWE).

Introduction

- Study Focus:** Introduce DeepCIM, engineered to assess CIM leveraging cSMILES, protein sequence and pIC50.
- Model Workflow and Performance:** Triple input system, multi-layer architecture, adds dropout to prevent overfitting—head-to-head comparison with SVM.
- External Validation:** The accuracy and precision of DeepCIM were externally validated using the Davis dataset.
- Applications on RWE:** The study extends its findings through an RWE analysis, comparing NTP rates between drugs selected and not selected by DeepCIM.

Methods

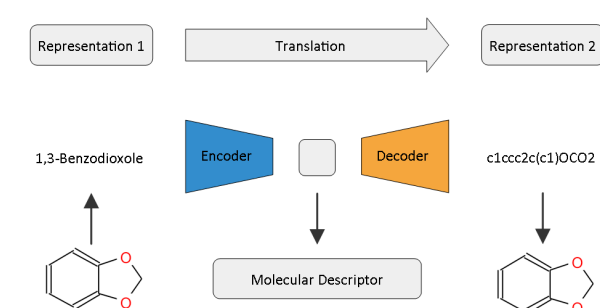
Data Collection and Integration:

Using DrugBank [2](16,575 drugs) and BindingDB [3] (2.8 million data points), identifying 890 drug-target interactions involving 288 protein targets and 112 myelosuppressive molecules.

Utilizing the molecular descriptors for encoding the molecular structures (Fig 1). External validation using Davis dataset [4], comprising 72 small molecules and 442 protein kinases.

Utilizing Flatiron RWE (56,479 outcomes) to assess impact of DeepCIM's NTP prediction on post-marketing first-line of therapy data.

Figure 1 Pretrained Autoencoder for DNN classification: Distinguishing Myelosuppressive and Non-Myelosuppressive Molecules from DrugBank

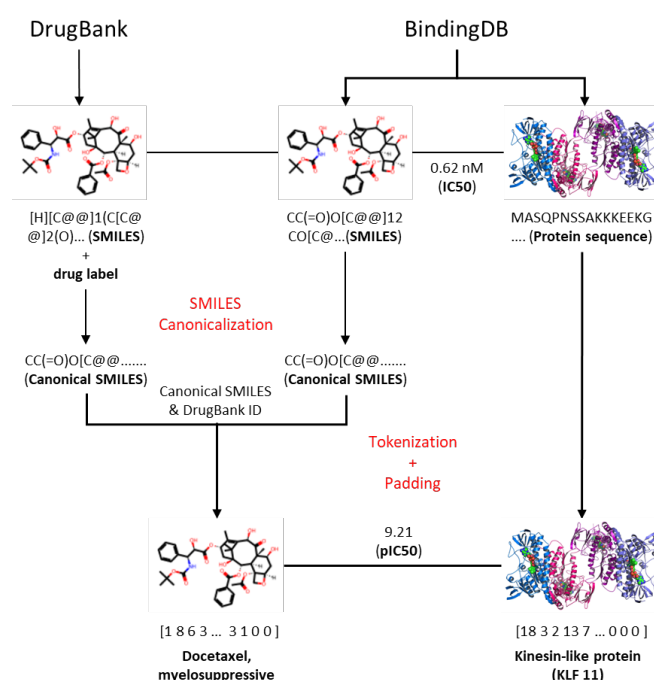


Analysis roadmap

An Overview of Key Components:

- Model Development and Architecture**
 - DeepCIM architecture overview (Fig 2)
- Training and Validation on Davis dataset**
 - DrugBank and BindingDB datasets for training
 - External validation using the Davis dataset
- Performance and Results**
 - Accuracy, precision, and FPR of DeepCIM and SVM
 - Integration of real-world evidence (RWE)
- Applications and Future Directions**
 - Potential applications in novel modalities and other therapeutic areas.

Figure 2 DeepCIM overview of data integration workflow: SMILES canonization, target tokenization.

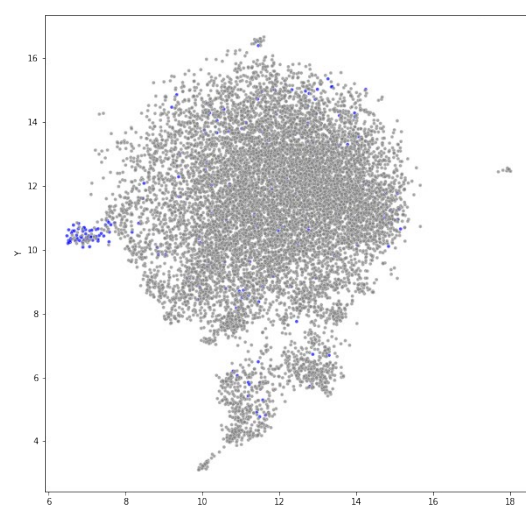


Results

Molecular Representation and Decoding:

Prior iteration of DeepCIM (Auto-Encoder DNN) learns molecular descriptors. These descriptors are employed in a PCA to visualize the first 2 components, with UMAP color coding indicating myelosuppressive versus non-myelosuppressive grouping Fig. 3.

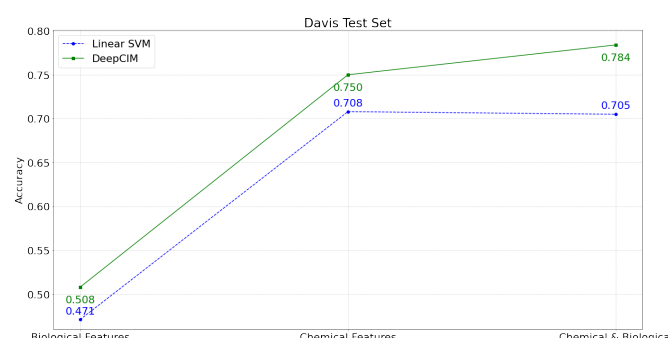
Figure 3 Visualizing Molecular Patterns with PCA: Grouping Myelosuppressive (Blue) and Non-Myelosuppressive (Grey) Molecules



Predicting CIM and Drug-Target Interaction

Improved accuracy was attained by incorporating protein sequences in DeepCIM's final iteration, reaching 0.79 compared to 0.75. This outperformed SVM's accuracy of 0.70, achieving an F-1 score of 0.81 in predicting CIM during external validation (Fig 4).

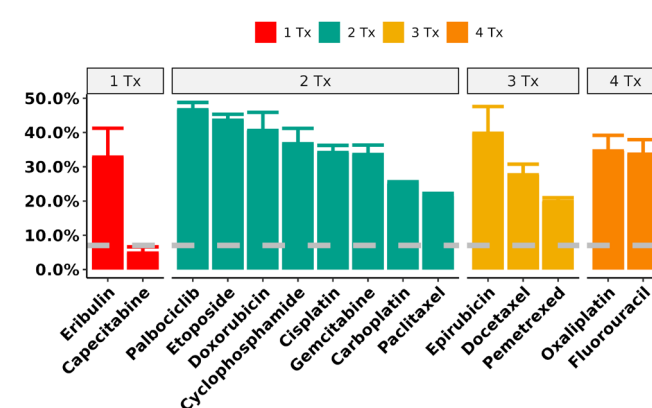
Figure 4 External Validation Results : Head-to-head comparison with SVM on Davis dataset



DeepCIM-selected vs RoTx using RWE

Integrating RWE from Flatiron to compare NTP Gr3/4 rates. DeepCIM-selected Tx showed a NTP GR3/4 rates of 25% (N = 38,673) compared to 7% (N = 17,806) for the remaining treatments (Fig 5).

Figure 5 Prevalence of NTP GR3/4 at 1st LoT by No. Tx. DeepCIM-selected vs RoTx (grey dashed)



Conclusions

- Used a comprehensive range of structural data to train the novel DeepCIM model.
- Showed DeepCIM's ability to predict a compound's myelosuppressive effects based on ligand structure and target sequence binding affinity values.
- RWE revealed substantial differences in NTP rates, indicating clinical impact.

References

- Lee, Taeyeub. Imperial College London, 2024
- Wishart, D.S., et al., *Nucleic Acids Res.* 46 (2018): 1074-1082.
- Gilson, M.K., et al., *Nucleic Acids Res.* 44 (2016): 1045-53.
- Davis, M.I., et al., *Nat Viotechnol.* 29 (2011): 1180-1192.
- Traynor, Carlos, et al. *Journal of Personalized Medicine* 11.12 (2021): 1356.

Acknowledgements

We would like to thank Flatiron, the provider of RWE data. We extend special thanks to our colleagues from AZ-Imperial for their instrumental role in making this collaboration possible.

