MACHINE LEARNING-GUIDED DISCOVERY OF SMALL MOLECULE INHIBITORS TARGETING PD-L1 FOR CANCER IMMUNOTHERAPY

Hunzallah Usmani[1], Wim Dehaen[1][2]

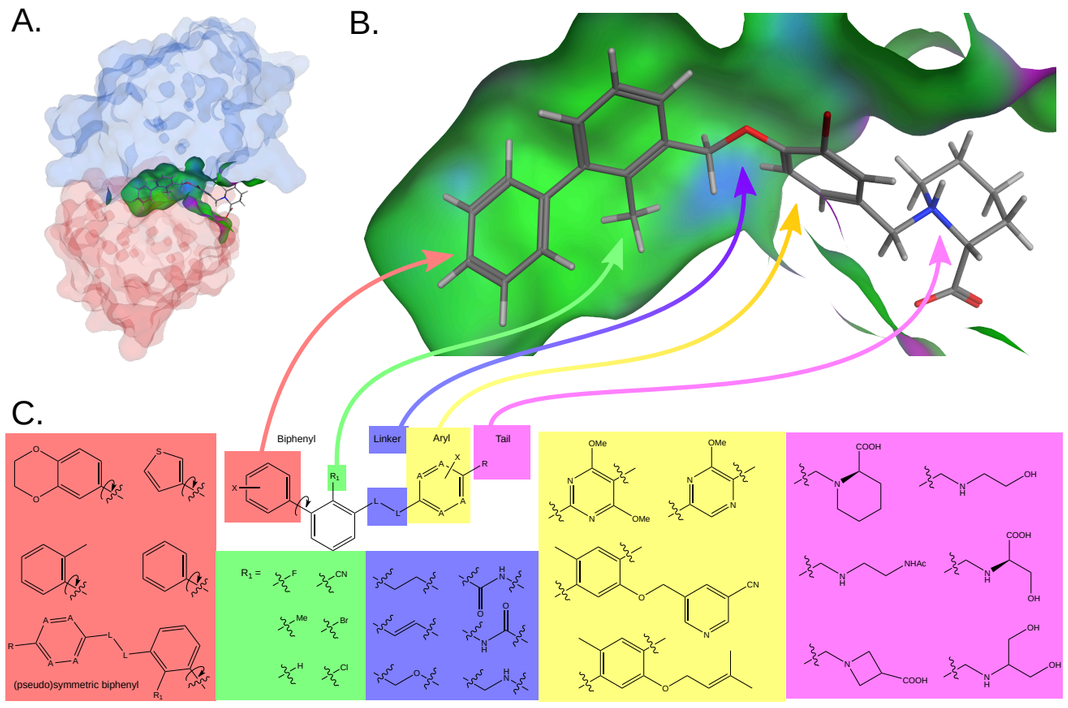
Affiliations

[1] CZ-OPENSCREEN, Department of Informatics and Chemistry, Faculty of Chemical Technology, University of Chemistry and Technology Prague, Technická 5, 16 628 Prague 6, Czech Republic;
[2] Department of Organic Chemistry, Faculty of Chemical Technology, University of Chemistry and Technology Prague, Technická 5, 16 628 Prague 6, Czech Republic;

UCT PRAGUE

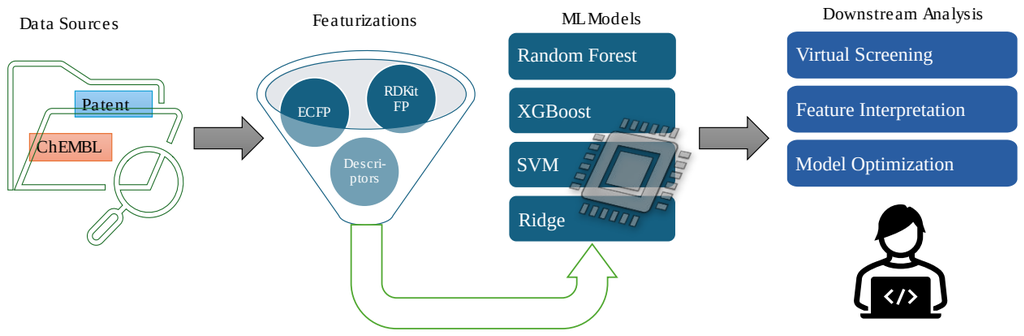
INTRODUCTION

Cancer cells often evade immune detection by utilizing Programmed Cell Death Ligand 1 (PD-L1), a protein that disarms the body's immune response. While commercially available antibodies target this pathway, they have certain practical issues, such as bioavailability [1]. Our project developed a machine learning-guided pipeline to accelerate the discovery of novel small-molecule PD-L1 inhibitors. Trained on diverse bioactivity data we created an end-to-end pipeline that accelerates inhibitor discovery.



A. Crystal structure of PD-L1/PD-L1 complex: PDB 5J8O **B.** Enhanced view of binding interface of PD-L1/PD-L1 complex (5J8O) **C.** Known structural motifs for PD-L1 inhibitors, illustrating the general pharmacophore, consisting of a biphenyl head group, a linker, an aryl group, and a tail.

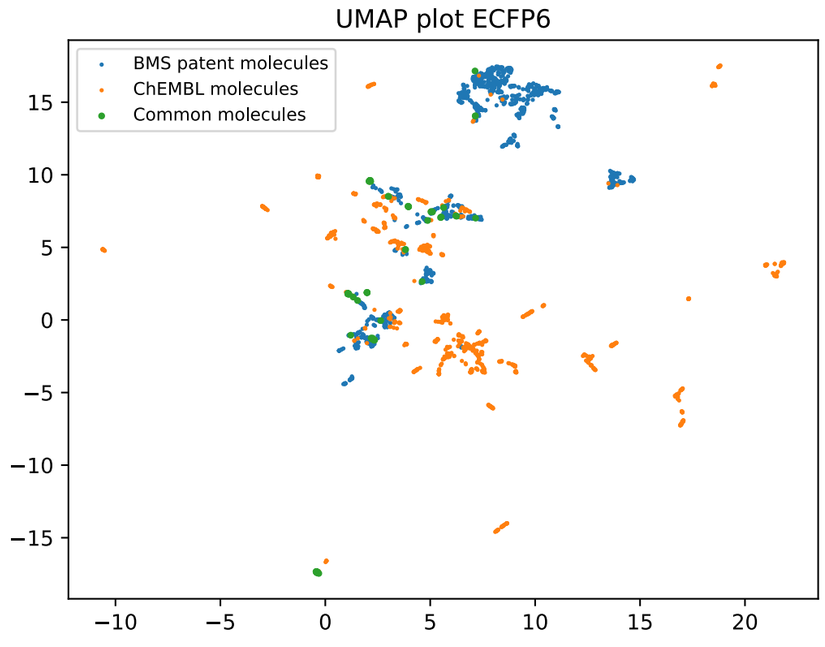
WORKFLOW



ML pipeline from data to discovery: molecular structures featurized as molecular fingerprints (FP) and physicochemical descriptors are used to train bioactivity models that guide virtual screening of potential PD-L1 inhibitors.

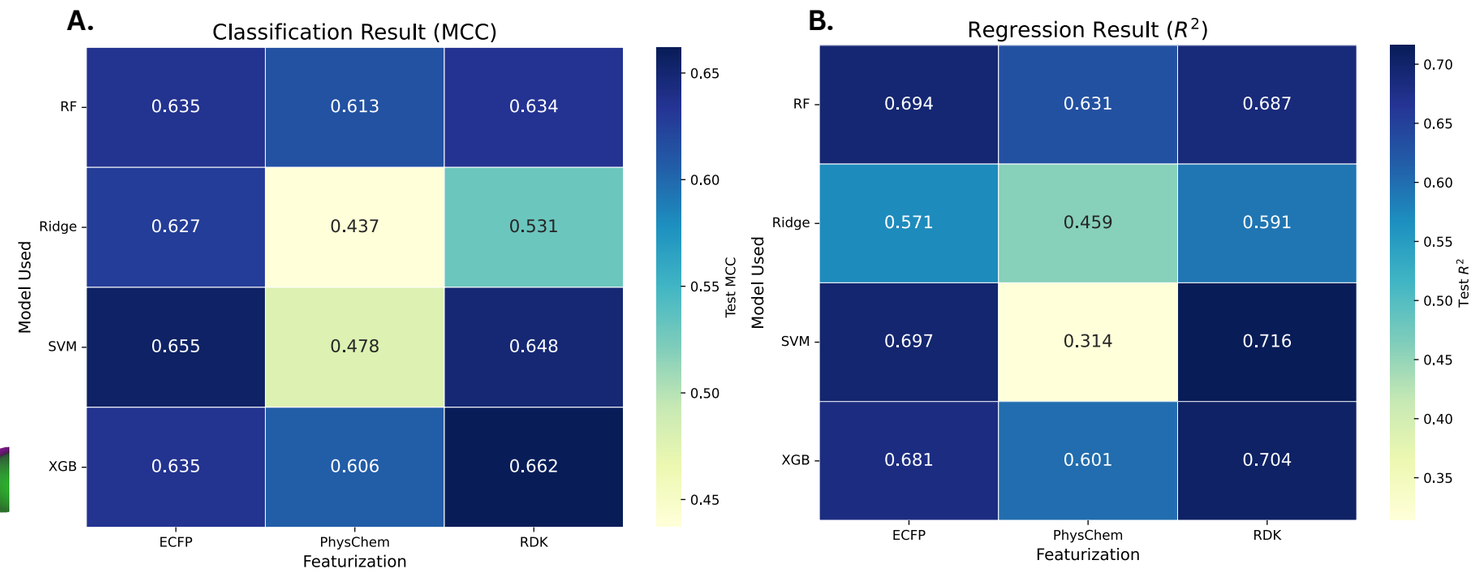
RESULTS

Chemical Space Visualization



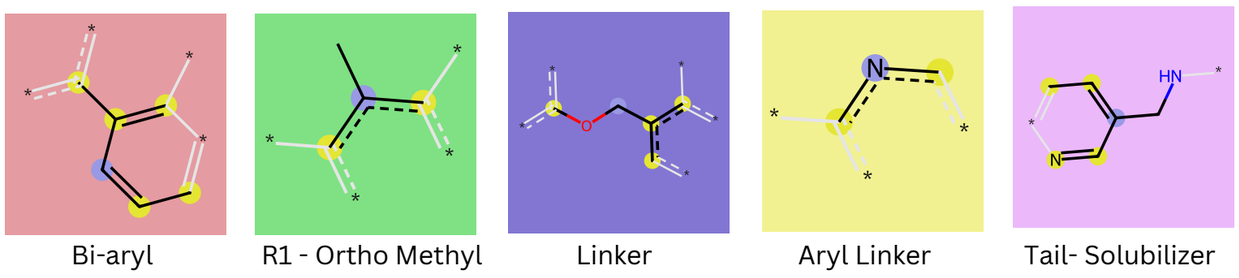
A UMAP plot showing the distribution of PD-L1 inhibitors from different data sources. The plot visualizes molecules from the BMS patent dataset, the ChEMBL database, and the compounds found in both [2].

Model Metrics



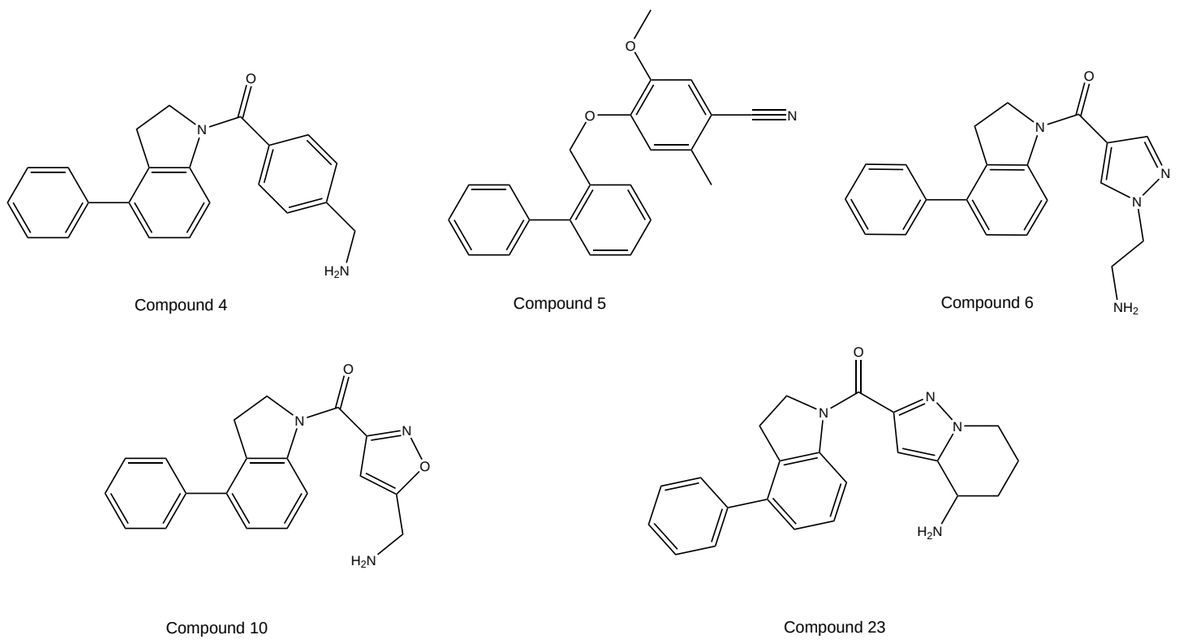
A. Test MCC (Matthews Correlation Coefficient) score for different combinations of classification models and featurizer.
B. Test R^2 (Coefficient of Determination) reported for each combination of regressor model and featurizer.

Model Interpretation



Top-ranked features correspond to common PD-L1 inhibitor pharmacophores, including aryl linkers and solubilizing tails. Features collected from built Random Forest Model with ECFP Fingerprints

Virtual Screening



Enamine in-stock collection (4.5 million molecules) was screened using the built model, Random Forest with ECFP Fingerprints. Top virtual hits predicted to bind PD-L1 are shown based on model score and structural diversity.

CONCLUSION

Our models, particularly those leveraging ECFP features, demonstrated high predictive accuracy in both classification and regression tasks. Crucially, the model's feature importance analysis highlighted key pharmacophoric motifs that align with known structural requirements for PD-L1 binding, providing a strong validation of the pipeline's chemical relevance. This pipeline offers a fast and cost-effective way to help identify new drug candidates for cancer immunotherapy.

REFERENCES AND ACKNOWLEDGMENTS

CZ OPENSCREEN

Grant number LM2023052

[1] Jiang, Y., Chen, M., Nie, H., & Yuan, Y. (2019). PD-1 and PD-L1 in cancer immunotherapy: clinical implications and future considerations. Human Vaccines & Immunotherapeutics, 15(5), 1111–1122. https://doi.org/10.1080/21645515.2019.1571892
[2] Patil, S. P., Fattakhova, E., Hofer, J., Oravic, M., Bender, A., Brearey, J., Parker, D., Radnoff, M., & Smith, Z. (2022). Machine-Learning guided discovery of bioactive inhibitors of PD1-PDL1 interaction. Pharmaceuticals, 15(5), 613. https://doi.org/10.3390/ph15050613